Fourteenth
Biennial Scientific Report

March 2017 – February 2019
# Contents

I Overview – The Institute 1

1 Overview 3
  1.1 Organization and Staff . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
  1.2 Scientific Vision and Strategic Goals . . . . . . . . . . . . . . . . . . . 4
  1.3 Long-Term Achievements and Impact . . . . . . . . . . . . . . . . . . . 5
  1.4 Highlights 2017–2019 and New Research Directions . . . . . . . . . . . 6
  1.5 Career Mentoring . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
  1.6 Collaborations and Strategic Partnerships . . . . . . . . . . . . . . . . . 9
  1.7 Results 2017–2019 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

II Overview – The Research Units & Senior Researchers 13

2 D1: Algorithms and Complexity 15

3 D2: Computer Vision and Multimodal Computing 21

4 D3: Internet Architecture 29

5 D4: Computer Graphics 33

6 D5: Databases and Information Systems 41

7 RG1: Automation of Logic 49

8 Zeynep Akata: Multimodal Deep Learning 53

9 Bjoern Andres: Combinatorial Image Analysis 55

10 Vahid Babaei: Computation, Appearance and Manufacturing 57

11 Klaus Berberich: Text+Time Search & Analytics 59

12 Jasmin Christian Blanchette: Interactive Theorem Proving 61

13 Karl Bringmann: Fine-Grained Complexity and Algorithms 63

14 Andreas Bulling: Perceptual User Interfaces 65
## Contents

36.4.2 Human Action Recognition ........................................ 253
36.4.3 Human Pose Estimation ........................................... 254
36.4.4 Human Body Models ............................................... 258
36.4.5 Multi-Person Tracking ............................................ 259

36.5 Object Detection and Recognition – *Shanshan Zhang and Bernt Schiele* . 260
36.5.1 3D Object Recognition ........................................... 260
36.5.2 Non-Maximum Suppression ........................................ 263
36.5.3 People Detection .................................................. 263
36.5.4 Person Recognition ............................................... 265

36.6 Learning – *Zeynep Akata* ........................................... 266
36.6.1 Few-shot Learning ................................................ 267
36.6.2 Zero-shot Learning ............................................... 270
36.6.3 Multilabel and Weakly Supervised Learning ................. 273

36.7 Privacy and Security – *Mario Fritz and Bernt Schiele* ............ 274
36.7.1 Adversarial Perturbations ....................................... 275
36.7.2 Attacks and Defenses for Machine Learning .................. 279
36.7.3 Enforcing Privacy ............................................... 280
36.7.4 Understanding Privacy ......................................... 282

36.8 Scalable Learning and Perception – *Mario Fritz* .................. 284
36.8.1 Deep Learning for Graphics and Physics ...................... 284
36.8.2 Uncertainty in Deep Learning .................................. 289

36.9 Virtual Humans – *Gerard Pons-Moll* ................................ 294
36.9.1 Human Pose Estimation from Video and IMU ................ 294
36.9.2 Perceiving 3D People from Visual Data ....................... 296

36.10 Combinatorial Image Analysis – *Evgeny Levinkov and Bjoern Andres* . 301
36.10.1 Correlation Clustering ......................................... 301
36.10.2 Biomedical Image and Data Analysis ......................... 304
36.10.3 Semi-supervised Learning ..................................... 306

36.11 Vision and Language – *Zeynep Akata* ................................ 307
36.11.1 Datasets ......................................................... 307
36.11.2 Explainable AI .................................................. 309
36.11.3 Generative Models with Vision and Language .............. 314
36.11.4 Visual Description and Question Answering ................. 316

36.12 Perceptual User Interfaces – *Andreas Bulling* ...................... 319
36.12.1 Calibration Methods ............................................ 319
36.12.2 Eye Contact Detection ....................................... 321
36.12.3 Gaze Estimation ............................................... 322
36.12.4 Gaze Interaction .............................................. 325
36.12.5 Interaction Behaviour Understanding ......................... 327
36.12.6 Interaction with Mobile Devices ............................... 329
36.12.7 Mobile Eye Tracking .......................................... 331
36.12.8 Non-Verbal Behaviour Analysis ................................ 334
36.12.9 Public Display Interaction .................................... 336
36.12.10 Symbiotic Human-Machine Vision .......................... 338
36.12.11 Usable Security ............................................... 341
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>36.13</td>
<td>Image and Video Segmentation – Anna Khoreva and Bernt Schiele</td>
<td>343</td>
</tr>
<tr>
<td>36.13.1</td>
<td>Image Segmentation with Weak Labels</td>
<td>343</td>
</tr>
<tr>
<td>36.13.2</td>
<td>Video Object Segmentation with Different Levels of Supervision</td>
<td>345</td>
</tr>
<tr>
<td>36.13.3</td>
<td>Semantic Segmentation</td>
<td>348</td>
</tr>
<tr>
<td>36.14</td>
<td>Combinatorial Computer Vision – Paul Swoboda</td>
<td>350</td>
</tr>
<tr>
<td>36.14.1</td>
<td>Graph Matching</td>
<td>351</td>
</tr>
<tr>
<td>36.14.2</td>
<td>Convex Optimization</td>
<td>353</td>
</tr>
<tr>
<td>36.15</td>
<td>Academic Activities</td>
<td>353</td>
</tr>
<tr>
<td>36.15.1</td>
<td>Journal Positions</td>
<td>353</td>
</tr>
<tr>
<td>36.15.2</td>
<td>Conference and Workshop Positions</td>
<td>354</td>
</tr>
<tr>
<td>36.15.3</td>
<td>Invited Talks and Tutorials</td>
<td>359</td>
</tr>
<tr>
<td>36.15.4</td>
<td>Other Academic Activities</td>
<td>362</td>
</tr>
<tr>
<td>36.16</td>
<td>Teaching Activities</td>
<td>363</td>
</tr>
<tr>
<td>36.17</td>
<td>Dissertations, Habilitations, Awards</td>
<td>364</td>
</tr>
<tr>
<td>36.17.1</td>
<td>Dissertations</td>
<td>364</td>
</tr>
<tr>
<td>36.17.2</td>
<td>Offers for Faculty Positions</td>
<td>365</td>
</tr>
<tr>
<td>36.17.3</td>
<td>Awards</td>
<td>365</td>
</tr>
<tr>
<td>36.18</td>
<td>Grants and Cooperations</td>
<td>366</td>
</tr>
<tr>
<td>36.19</td>
<td>Publications</td>
<td>368</td>
</tr>
<tr>
<td>37</td>
<td>D3: Internet Architecture</td>
<td>381</td>
</tr>
<tr>
<td>37.1</td>
<td>Personnel</td>
<td>381</td>
</tr>
<tr>
<td>37.2</td>
<td>Visitors</td>
<td>381</td>
</tr>
<tr>
<td>37.3</td>
<td>Group Organization</td>
<td>382</td>
</tr>
<tr>
<td>37.4</td>
<td>BGP Communities</td>
<td>382</td>
</tr>
<tr>
<td>37.4.1</td>
<td>BGP Community-based Attacks</td>
<td>383</td>
</tr>
<tr>
<td>37.4.2</td>
<td>Mitigating DDoS Attacks Using Advanced Blackholing</td>
<td>384</td>
</tr>
<tr>
<td>37.5</td>
<td>On Traffic Volume Asymmetry</td>
<td>386</td>
</tr>
<tr>
<td>37.6</td>
<td>Optimal Transport for Video Streaming</td>
<td>387</td>
</tr>
<tr>
<td>37.7</td>
<td>Socket Intents</td>
<td>389</td>
</tr>
<tr>
<td>37.8</td>
<td>Distributed and Timely Flow Summarization at Scale</td>
<td>390</td>
</tr>
<tr>
<td>37.9</td>
<td>The Need for New Computing Primitives</td>
<td>391</td>
</tr>
<tr>
<td>37.10</td>
<td>Academic Activities</td>
<td>392</td>
</tr>
<tr>
<td>37.10.1</td>
<td>Conference and Workshop Positions</td>
<td>392</td>
</tr>
<tr>
<td>37.10.2</td>
<td>Invited Talks and Tutorials</td>
<td>393</td>
</tr>
<tr>
<td>37.11</td>
<td>Teaching Activities</td>
<td>394</td>
</tr>
<tr>
<td>37.12</td>
<td>Dissertations, Habilitations, Awards</td>
<td>394</td>
</tr>
<tr>
<td>37.12.1</td>
<td>Dissertations</td>
<td>394</td>
</tr>
<tr>
<td>37.12.2</td>
<td>Awards</td>
<td>394</td>
</tr>
<tr>
<td>37.13</td>
<td>Publications</td>
<td>395</td>
</tr>
<tr>
<td>38</td>
<td>D4: Computer Graphics</td>
<td>397</td>
</tr>
<tr>
<td>38.1</td>
<td>Personnel</td>
<td>397</td>
</tr>
<tr>
<td>38.2</td>
<td>Visitors</td>
<td>398</td>
</tr>
<tr>
<td>38.3</td>
<td>Group Organization</td>
<td>401</td>
</tr>
</tbody>
</table>
### Contents

38.4 Digital Geometry Processing – Renjie Chen, Rhaleb Zayer, and Hans-Peter Seidel

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.4.1 Shape Deformation and Mapping Optimization</td>
<td>402</td>
</tr>
<tr>
<td>38.4.2 Path Planning with Divergence-Based Distance Functions</td>
<td>404</td>
</tr>
<tr>
<td>38.4.3 Design and Optimization of Space Structures</td>
<td>405</td>
</tr>
<tr>
<td>38.4.4 High Performance Surface Tessellations</td>
<td>407</td>
</tr>
<tr>
<td>38.4.5 High Performance Mesh Subdivision</td>
<td>408</td>
</tr>
<tr>
<td>38.4.6 Linear Algebra Kernels and Graph Algorithms</td>
<td>410</td>
</tr>
</tbody>
</table>

38.5 Computational Videography – Christian Theobalt

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.5.1 Inverse and Video-based Rendering</td>
<td>414</td>
</tr>
<tr>
<td>38.5.2 Advanced Synthesis and Editing of Human Actor Videos</td>
<td>415</td>
</tr>
</tbody>
</table>

38.6 Reconstructing the Static and Dynamic Real World – Christian Theobalt

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.6.1 High-quality Static Scene Reconstruction with Lightweight Sensors</td>
<td>419</td>
</tr>
<tr>
<td>38.6.2 Deformable Surface and Performance Capture in Controlled and Uncontrolled Scenes</td>
<td>420</td>
</tr>
<tr>
<td>38.6.3 Marker-less 3D Human Pose Estimation and Skeletal Motion Capture in General Scenes and with Lightweight Sensors</td>
<td>422</td>
</tr>
<tr>
<td>38.6.4 High-quality Lightweight Facial Performance Capture</td>
<td>424</td>
</tr>
<tr>
<td>38.6.5 Marker-Less Motion Capture of Hands in Interaction</td>
<td>427</td>
</tr>
</tbody>
</table>

38.7 Realistic and Real-time Rendering – Gurprit Singh

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.7.1 Real-time Distribution Effects</td>
<td>435</td>
</tr>
<tr>
<td>38.7.2 Vector Graphics, Appearance Synthesis and Simulations</td>
<td>436</td>
</tr>
<tr>
<td>38.7.3 Analysis and Synthesis of Sample Correlations</td>
<td>437</td>
</tr>
</tbody>
</table>

38.8 Perception and Advanced Displays – Karol Myszkowski

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.8.1 Mono2Stereo and Stereo2Multiview Image Conversion</td>
<td>439</td>
</tr>
<tr>
<td>38.8.2 Light-Fields and Advanced Displays</td>
<td>440</td>
</tr>
<tr>
<td>38.8.3 Image and Light-Field Quality Evaluation</td>
<td>441</td>
</tr>
<tr>
<td>38.8.4 Eye Tracking</td>
<td>444</td>
</tr>
</tbody>
</table>

38.9 Computational Fabrication – Vahid Babaei

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.9.1 Appearance Fabrication</td>
<td>447</td>
</tr>
<tr>
<td>38.9.2 Fabrication of Digital Drawing Tools</td>
<td>448</td>
</tr>
<tr>
<td>38.9.3 Display Design and Fabrication</td>
<td>450</td>
</tr>
</tbody>
</table>

38.10 Software and Datasets

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.10.1 TheCaptury GmbH – A Spin-off Company Commercializing</td>
<td>451</td>
</tr>
<tr>
<td>38.10.2 GVVPerfCapEva – A Repository of Human Shape and Performance</td>
<td>451</td>
</tr>
<tr>
<td>38.10.3 PFSTOOLS for Processing High Dynamic Range Images and Video</td>
<td>454</td>
</tr>
<tr>
<td>38.10.4 LocVis – Local Visibility Maps of Artifacts and Distortions in Images</td>
<td>454</td>
</tr>
<tr>
<td>38.10.5 MPI Light-Field Archive</td>
<td>456</td>
</tr>
</tbody>
</table>

38.11 Academic Activities

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>457</td>
</tr>
<tr>
<td></td>
<td>459</td>
</tr>
</tbody>
</table>
38.11.1 Journal Positions .................................................. 459
38.11.2 Conference and Workshop Positions ........................ 459
38.11.3 Invited Talks and Tutorials ...................................... 461
38.11.4 Other Academic Activities ..................................... 463
38.12 Teaching Activities .................................................. 464
38.13 Dissertations, Habilitations, Awards ............................ 465
38.13.1 Dissertations ........................................................ 465
38.13.2 Offers for Faculty Positions ...................................... 465
38.13.3 Awards .............................................................. 466
38.14 Grants and Cooperations .......................................... 467
38.15 Publications .......................................................... 469

39 D5: Databases and Information Systems .......................... 479
39.1 Personnel ............................................................... 479
39.2 Visitors ................................................................. 480
39.3 Group Organization .................................................. 482
39.4 Knowledge Base Construction and Curation – Simon Razniewski and Daria Stepanova ........................................ 482
39.4.1 YAGO Knowledge Base ........................................... 485
39.4.2 Commonsense Knowledge ....................................... 486
39.4.3 Fictional Domain Knowledge .................................... 488
39.4.4 On-the-fly Knowledge Base Construction ..................... 490
39.4.5 Knowledge Base Completeness ................................ 492
39.4.6 Personal Knowledge Bases ....................................... 494
39.4.7 Rules from Knowledge Bases and Text ......................... 497
39.4.8 Explainable Fact Checking ....................................... 499
39.4.9 Knowledge for Visual Content Tagging ....................... 501
39.5 Data Mining and Exploratory Data Analysis – Jilles Vreeken and Pauli Miettinen ............................................. 504
39.5.1 Dependency Discovery ............................................ 505
39.5.2 Causality ............................................................ 508
39.5.3 Subgroup Discovery in Multivariate Data ..................... 513
39.5.4 Redescription Mining ............................................. 518
39.5.5 Patterns, Anomalies, and Interestingness ...................... 521
39.5.6 Reductions and ASP Solvers for Frequency-Based Pattern Mining ......................................................... 526
39.5.7 Matrix Factorization .............................................. 528
39.5.8 Large Entries of Matrix Products ............................... 532
39.5.9 Community Models for Graphs ................................ 534
39.5.10 Effective Parallelization for Machine Learning ............. 536
39.6 Information Retrieval, Information Extraction, and Text Analysis – Andrew Yates and Klaus Berberich ......................... 537
39.6.1 Neural Matching and Ranking ................................... 538
39.6.2 Event-centric Search and Exploration ......................... 540
39.6.3 Querying Text with Semantic Annotations ................. 544
39.6.4 Analyzing Health Forums and Clinical Texts ................ 547
# Contents

39.6.5 Information Extraction from Dialogues .......................... 548
39.6.6 Information Extraction for Higher-Arity Relations ............... 551
39.6.7 Search on Health Corpora ...................................... 552
39.6.8 Ranked Query Processing on Knowledge Graphs ................. 553
39.6.9 Quantities in Text and Tables ................................... 555

39.7 Natural Language Understanding – Paramita Mirza, Rishiraj Saha Roy, and Jannik Strötgen .......................................................... 558
39.7.1 AmbiverseNLU ....................................................... 559
39.7.2 Named Entity Recognition and Disambiguation .................. 562
39.7.3 Temporal and Spatial Expressions in Text ......................... 564
39.7.4 Counting Quantifiers in Text ..................................... 569
39.7.5 Emotion Extraction ............................................... 572
39.7.6 Dialogue Generation .............................................. 575
39.7.7 Template-based Question Answering .............................. 577
39.7.8 Temporal Question Answering ..................................... 580
39.7.9 A Benchmark for Question Answering ............................. 582
39.7.10 Graph-based Question Answering ................................ 584

39.8 Privacy, Trust and Fairness – Gerhard Weikum ................. 586
39.8.1 Trust Assessment for Online Forums .............................. 587
39.8.2 Credibility Assessment of Claims ................................ 590
39.8.3 Detecting and Classifying Fake News ............................. 592
39.8.4 Reconciling Privacy and Utility .................................. 595
39.8.5 Fairness in Rankings .............................................. 597
39.8.6 Learning Fair Representations ................................... 598
39.8.7 Explaining Social Feed Recommendations ....................... 600
39.8.8 Dense Subgraphs in Social Media ................................ 602
39.8.9 Understanding Controversies in Online Forums ................ 604

39.9 Academic Activities .................................................... 607
39.9.1 Journal Positions .................................................... 607
39.9.2 Book Positions ....................................................... 607
39.9.3 Conference and Workshop Positions ............................. 607
39.9.4 Invited Talks and Tutorials ....................................... 615
39.9.5 Other Academic Activities ........................................ 617

39.10 Teaching Activities ..................................................... 617
39.11 Dissertations, Habilitations, Awards ................................. 619
39.11.1 Dissertations ....................................................... 619
39.11.2 Offers for Faculty Positions ...................................... 620
39.11.3 Awards ............................................................ 620

39.12 Grants and Cooperations ............................................. 621
39.13 Publications ........................................................... 622

40 RG1: Automation of Logic .................................................. 639
40.1 Personnel ............................................................... 639
40.2 Visitors ................................................................. 639
40.3 Foundations of Automated Reasoning ................................. 640
## Contents

40.3.1 Novel Decidable Fragments of First-Order Logic 640
40.3.2 The Approximation Refinement Approach to First-Order Theorem Proving 644
40.3.3 Clause Learning from Simple Model Representations for the Bernays–Schönfinkel fragment 645
40.3.4 Representation of Inherently Infinite Models 646
40.3.5 Prime Implicate Generation in Equational Logic 647
40.3.6 Scalable, Fine-Grained Proofs of Preprocessing for SMT Solving 648
40.3.7 Decidable and Undecidable Fragments of First-Order Logic Modulo Linear Arithmetic 649

40.4 Arithmetic 650
40.4.1 Real Quantifier Elimination by Virtual Substitution 650
40.4.2 Subtropical Methods for Real Constraints 651
40.4.3 New Techniques for Linear Arithmetic: Cubes and Equalities 652
40.4.4 A Reduction from Unbounded Linear Mixed Arithmetic Problems into Bounded Problems 654
40.4.5 Multistationarity of Biological Reaction Networks 655

40.5 Towards Higher-Order Automated Reasoning 657
40.5.1 Extension of the Superposition Calculus with Lambda-Free Higher-Order Terms and (Co)datatypes 657
40.5.2 Extension of Term Orders to Lambda-Free Higher-Order 658

40.6 Formalizing Logic 659
40.6.1 (Co)datatypes, (Co)recursion, and Generalized Nominal-Logic-Style Binders for Isabelle/HOL 659
40.6.2 Verifying CDCL 660
40.6.3 Verifying Bachmair and Ganzinger’s Ordered Resolution Prover 661
40.6.4 Superposition with Types and Datatypes 661

40.7 Logic for Machine Learning 662
40.7.1 Formalization of the Expressiveness of Deep Learning in Isabelle/HOL 663
40.7.2 Logic Fragments Reduction for Inductive Logic Programming 663

40.8 Software and Applications 664
40.8.1 Zipperposition 664
40.8.2 Reduce and Redlog 665
40.8.3 The SPASS Workbench 666
40.8.4 SPASS-IQ: An LA Theory Solver for SPASS 667
40.8.5 SPASS-SATT: A Ground Linear Arithmetic Solver 668

40.9 Academic Activities 669
40.9.1 Journal Positions 669
40.9.2 Conference and Workshop Positions 670
40.9.3 Invited Talks and Tutorials 672
40.9.4 Other Academic Activities 673

40.10 Teaching Activities 673
40.11 Dissertations, Awards 674
40.11.1 Dissertations 674
Part I

Overview – The Institute
1 Overview

1.1 Organization and Staff

Directors and Departments: The Max Planck Institute for Informatics was established in 1990, with Kurt Mehlhorn as founding director. Harald Ganzinger was appointed shortly afterwards; he passed away in 2004. Currently, the institute has five directors (with appointment periods given in parentheses): Kurt Mehlhorn (1990–2019) heading the Department on Algorithms and Complexity (D1), Bernt Schiele (2010–2035) heading the Department on Computer Vision and Multimodal Computing (D2), Anja Feldmann (2018–2033) heading the Department on Internet Architecture, Hans-Peter Seidel (1999–2026) heading the Department on Computer Graphics (D4), Gerhard Weikum (2003-2023) heading the Department on Databases and Information Systems. In addition to these five departments, the institute has one permanent independent Research Group on Automation of Logic, headed by Christoph Weidenbach. In total, the institute currently has 153 scientists, out of which 96 are doctoral students and 57 have a doctoral degree.

Senior Researchers: The institute has four scientific ranks: directors, senior researchers with tenure, senior researchers, and researchers. Senior researchers with tenure and senior researchers roughly correspond to tenured associate professor and non-tenured assistant professor in the North-American system. The process for appointing senior researchers is similar to faculty appointment procedures, and involves reference letters from international top researchers. The appointment committee comprises the directors of the institute and a faculty member of Saarland University. Currently the institute has, in addition to its directors, 27 senior researchers.

Senior researcher positions are not tenure-track, in general. We expect our senior researchers to become professors, ultimately full professors, or leading researchers in industry. A strong indicator for the success of the model is the high number of faculty positions that our alumni reached at universities all over the world (see Section 1.5).

Figure 1.1 shows the organizational structure and the research areas of the institute. Each department pursues a number of research areas and each area has its coordinator(s). The coordinating scientists are senior researchers or postdoctoral researchers with strong potential for becoming senior researchers.

―Zeynep Akata, Bjoern Andres, Klaus Berberich, Jasmin Blanchette, Karl Bringmann, Andreas Bulling, Renjie Chen, Mario Fritz, Olga Kalinina, Andreas Karrenbauer, Christoph Lenzen, Tobias Marschall, Pauli Miettinen, Karol Myszkowski (tenured), Nico Pfeifer, Gerard Pons-Moll, Simon Razniewski, Michael Sagraloff, Rishiraj Saha Roy, Marcel Schulz, Daria Stepanova, Jannik Strötgen, Thomas Sturm, Christian Theobalt (tenured), Jilles Vreeken, Christoph Weidenbach (tenured), Andrew Yates
Doctoral Students: As of March 2019, there are 96 doctoral students who are supervised by members of the institute, including 21 women (about 22%). 59 of the 96 students are non-German. The institute does not grant degrees. In this regard, we closely collaborate with Saarland University. Members of the institute teach courses at the University and supervise students at all levels. Upon appointment to Senior Researcher, the Computer Science Department of Saarland University decides whether to grant the Senior Researcher the right to supervise doctoral students. Currently, this right has been granted to all senior researchers of the institute.

Joint Administration and Technical Support: The institute shares the IT support group, administration, library, and facility management with the Max Planck Institute for Software Systems. The current head counts for both institutes are 54 full-time employees including 20 IT support people.

1.2 Scientific Vision and Strategic Goals

Our central research theme is algorithms for multimodal computing. Algorithms have always been the main focus of the institute. They are the core of what makes computer systems useful and productive. They influence every aspect of our daily lives and are the basis for industrial change. Throughout the last decade, major parts of our research effort have focused on multimodal computing. The grand challenge is to understand, search, and organize large, distributed, noisy, incomplete, and diverse information in a robust, efficient, and intelligent manner. Our research ranges from foundations (algorithms and complexity, automation of logic) to a variety of multimodal domains (computer graphics and vision, geometric...
computation, intelligent information systems, adaptive networks). The overarching mission of the institute is to be one of the world’s top players and strategic trendsetters on this theme of multimodal computing.

Most of the major advances in computer science have come through the combination of new theoretical insights and application-oriented experimental validation, driven by outstanding researchers. Our goal is, thus, to have impact through i) research and publications, ii) software and services enabled by our research, and iii) people alike. In the following, Section 1.3 presents our achievements regarding the first two dimensions, and Section 1.5 discusses our track record with respect to the third dimension.

1.3 Long-Term Achievements and Impact

Over the last twenty years, the institute has pursued a number of high-risk high-gain endeavors, starting with foundational science and ultimately making great practical impact. In the following we outline the highlights of the institute’s scientific achievements.

- Kurt Mehlhorn and his group initiated and advanced the field of Algorithm Engineering: setting the trend and making ground-breaking contributions towards reconciling the rigorous design and complexity analysis of algorithms with the development of practically viable software libraries. Highlight results with huge impact are the software libraries LEDA and CGAL, where advanced algorithms have been coded with guarantees on their correctness and run-time properties. These libraries are widely used all over the world; LEDA was part of the software that Celera used for sequencing the human genome in the early 2000s.

- Bernt Schiele’s group has been working on tracking multiple people for several years now. In joint work with Bjoern Andres published at CVPR’15 and CVPR’17 multi person tracking has been formulated as a graph decomposition problem leveraging strong person detectors as unary costs and re-identification classifiers as pairwise costs. This work as won the Multi-Object Tracking Challenge twice (at ECCV’16 and at CVPR’17).

- Anja Feldmann’s research vision is to obtain insights from Internet measurements as foundation for shaping the evolution of Internet by proposing optimizations and investigating alternative designs. As such her research group has and continue to address many challenges in Internet measurement, e.g., the use/abuse of Interdomain routing, as well as cross-layer network optimization, e.g., by highlighting the benefits of joint optimization of ISPs/CDNs or applications/network.

- Hans-Peter Seidel and his group stand out for their integrated view of 3D image analysis and synthesis. The group developed groundbreaking results on multiresolution modeling with special emphasis on new metaphors for editing and shape deformation, and novel data structures for high performance geometry processing. Recent research on markerless performance capture has pioneered methods for the reconstruction of detailed dynamic 3D models of humans from both multiview and monocular video and has led to entirely new ways of fusing and deeply integrating model-based and deep learning-based scene reconstruction. The group has been and continues to be one of the pioneers and a driving force in perception-based graphics.
Overview

Christoph Weidenbach and his group have been investigating foundations of automated reasoning since the early 90's. Starting from the development of sound and complete first-order calculi and their implementations, the group is now driving towards calculi for more expressive logics that also automatically discover and exploit structures.

Gerhard Weikum and his team pioneered the theme of Knowledge Harvesting: automatically building comprehensive knowledge bases from Internet contents. This work provided the blueprint for industrial-strength knowledge graphs that are key assets for search engines, question answering and text analytics (at Google, Microsoft, etc.). The knowledge base YAGO was used by IBM Watson when it won the Jeopardy quiz show in 2011.

High-Risk Research: When the above long-term projects started, they were far from the mainstream in their scientific communities; most were considered elusive and some even characterized as useless. Needless to say, not all of our bold endeavors worked out. The sections on the five departments give examples. Nevertheless, most of this work resulted in novel insights (sometimes about what is, fundamentally or practically, non-viable) and often in influential publications.

1.4 Highlights 2017–2019 and New Research Directions

Two important developments for the institute are the retirement of Thomas Lengauer (who became an Emeritus in 2018) and the upcoming retirement of the institute’s founding director, Kurt Mehlhorn, later this year. Kurt Mehlhorn will move to Emeritus status at the end of August 2019. Similar to Thomas Lengauer he will continue as a researcher in the institute. In order to have some guidance for the upcoming search for directors we decided early on to (i) keep the algorithms and complexity slot and (ii) to make an effort to attract at least one outstanding woman as a director.

In the meantime we have been able to identify and attract Anja Feldmann. Anja joined the institute as a new director in 2018 and is heading the Internet Architecture Department. The appointment process for a new director in Algorithms and Complexity is on-going.

Thomas Lengauer retired on May 31, 2018, after more than 15 years as Director of the Bioinformatics Department and dedicating his life to science. With a symposium on bioinformatics and beyond the institute and more than 200 participants celebrated the work and career of Thomas on May 25 2018. The list of scientific speakers ranged from collaborators to former students to colleagues from the institute: Gene Myers, Nico Pfeiffer, Yves Moreau, Kasia Bozek, Robert Tarjan, Jörg Rahmenführer, Bastian Beggel, and Anja Feldmann.

The following are selected highlights from the scientific results that the institute has achieved in the last two years.

– During the reporting period, Kurt Mehlhorn’s group worked mainly on fine-grained complexity, distributed computing, reliable computing, optimization, and algorithmic game theory.
Bernt Schiele’s group has been working on variety of essential problems that will be also followed in the future. An important example is how to learn without or very little supervision. In joint work with Zeynep Akata we have been able to push the state-of-the-art in zero-shot as well as few-shot learning substantially e.g. by combining the strength of VAE and GANs within a joined framework. In joint work with Mario Fritz we have worked on various aspects at the intersection of privacy and security on the one side and computer vision and machine learning on the other side. E.g. we have contributed to the better understanding of privacy implications of visual data dissemination through and also showed, that compute vision model stealing is possible under weak assumptions, thereby raising concerns about the security of current machine learning techniques.

Anja Feldmann’s group has been working on understanding the capabilities of today’s Internet Inter-Domain routing. In recent work they have been highlighting the good (CoNext 18), the bad (IMC 17), and the evil (IMC 18) aspects of today’s collaborative system. This work has also been presented at RIPE as well as at IETF (via an IETF Applied Networking Research award).

Hans-Peter Seidel’s group developed a novel sparse matrix representation for unstructured grids and demonstrated its power in different geometry processing applications (EG’17, ICS’17, SC’18, SIGAsia’18) (joint with R. Zayer and M. Steinberger). C. Theobalt developed VNect (SIG’17), the first method for accurate and real time 3D skeletal motion capture from monocular RGB and MoFA (ICCV’17, PAMI’18, CVPR’18), the first unsupervised approach for the reconstruction of high-quality 3D faces from a single color image. Deep Video Portraits enables realistic re-animation of the entire posture and face expression in a portrait video using only an input color video (SIG’18, CommACM’19). With T. Ritschel and G. Singh we obtained far reaching results on the simulation of distribution effects and the generation of well-behaved sampling patterns in rendering (SIG’17, SIG’18, CGF’19). K. Myszkowski introduced an award winning advanced near-eye display design (TVCG’17), and we developed novel fundamental algorithms for perception-aware gaze contingent rendering (SIG’17), 2D-to-3D conversion (TVCG’18), and light field quality metrics (CVPR’17, TOG’18).

Christoph Weidenbach’s group has continued the discovery of new (un)decidable first-order fragments (CADE 2017, FROCOS 2017) and has established new techniques in solving linear arithmetic constraints over the integers (CADE 2017, IJCAR 2018, SMTCOMP 2018).

In the context of the ERC Synergy Grant imPACT, we address the issue of trust in online information, particularly, aiming to analyze and assess the credibility of textual statements in news, discussion forums and other kinds of online communities. This research led to the doctoral thesis of Subhabrata Mukherjee which won an honorable mention for the KDD dissertation award 2018. More recent work (with publications in WWW 2017, WWW 2018 and EMNLP 2018) focused on incorporating web evidence for estimating the credibility of claims made in news or by influential people (politicians, bloggers, partisan media etc.). We devised a suite of methods, including pipelined classifiers, probabilistic graphical models for joint inference on source trustworthiness.
and content credibility, all the way to deep-learning networks with tailored attention mechanisms.

**New Research Directions:** The overriding theme of multimodal computing will be continued. However, we expect a leap in research directions by the ongoing blending of virtual worlds with the real world. This trend involves several grand challenges on which we will embark. As digital contents and output from all kinds of sensors keeps exploding, we need to move from data to understanding situations and anticipating user behavior. We need to support immersive interactions across all modalities, considering visual signals like facial expressions, gestures and body language, in conjunction with language, contextual knowledge and social interactions. Finally, with the rapid advance of machine learning and data-driven algorithmic decision making, we need to better understand how to make computer behavior comprehensible. These challenges are the motivation for our foundational research on algorithms, visual computing, internet architecture and knowledge discovery in the coming years. Moreover, we will continue our collaboration with the Bioinformatics Center at Saarland University even after Thomas Lengauer’s departure, and we remain committed to a strong research environment in this area.

### 1.5 Career Mentoring

The institute has a strong track record on educating and mentoring young researchers, at both the doctoral student level and the postdoctoral level.

**Doctoral Student Training:** Since the institute was established in 1990, a total number of 313 doctoral students graduated. These include 56 women, and 130 non-Germans. Among these students, 13 have been awarded the Otto Hahn Medal of the Max Planck Society, Andrey Rybalchenko (2007) and Fabian Suchanek (2012) have received the Otto Hahn Award for the best dissertations in the Chemistry, Physics, Technology Section of the Max Planck Society. This award includes a 5-year scholarship for an independent research group, and only one award per year is given (since 2007). Maximilian Dylla won the DBIS Dissertation Award in 2014 Marcus Rohrbach won the DAGM MVTec Dissertation Award in 2015. Karl Bringmann won the EATCS Distinguished Dissertation Award in 2015. Pablo Garrido and Petr Kellnhofer both won Eurographics Dissertation Awards in 2018.

**Young Scientist Career Advancement:** A unique strength of the institute is its successful fostering of young scientists. These mentoring and career support efforts are most pronounced at the level of our senior researchers. Since the notion of senior researchers was explicitly introduced in 2007, a total of 76 young scientists held such positions. 52 of them have meanwhile left the institute. Out of these, 51 have accepted a tenured or tenure-track position offer at universities or university-like research organizations. Several alumni (i.e., people who graduated here or spent at least two years at the institute and left) and current
1 Overview

senior researchers of the institute have won prestigious awards: Leibniz Prizes\textsuperscript{2}, ERC Grants\textsuperscript{3}, and other honors.

Support for Women: The fraction of women in our institute currently is 22\% for doctoral students and 16\% at the postdoctoral level (including senior researchers). To increase the representation of women in our field, we have established the Lise Meitner Fellowship for outstanding female scientists at the postdoctoral level. So far, these two-year fellowships have been awarded to 10 women.\textsuperscript{4} A large number of female alumni of the institute have become professors.\textsuperscript{5}

1.6 Collaborations and Strategic Partnerships

An overriding goal that our institute has been contributing to since its beginning in 1990, is to establish Saarbrücken as one of the world’s premier sites in computer science. Indeed, over the last three decades the site as a whole has a unique track record, as exemplified by 4 ACM Fellows, 6 Leibniz Prizes (the highest scientific honor in Germany) and a total of 11 ERC Grants (at all levels).

Saarland Informatics Campus: Our most important partner is the Computer Science Department of Saarland University. We have a long-standing tradition of teamwork and joint engagement in both research and teaching. The are numerous collaborations with faculty members of the university, and with researchers of various institutes on campus, including colleagues from the Department for Computational Linguistics, the Department for Biology and the Medical School of the university.

\textsuperscript{2}Susanne Albers 2008, Leif Kobbelt 2014, Peter Sanders 2012
\textsuperscript{5}Anna Adamaszek (Univ. Copenhagen, Denmark), Zeynep Akata (Univ. Amsterdam, Netherlands), Susanne Albers (TU Munich), Iris Antes (TU Munich), Hannah Bast (Univ. Freiburg), Carola Dör (CNRS, France), Panagiota Fatourou (Univ. Ionnina, Greece), Lilia Georgieva (Heriot-Watt Univ., UK), Katja Hose (Aalborg Univ., Denmark), Georgiana Ifrim (UC Dublin, Ireland), Mouna Kacimi (Univ. Bozen-Bolzano, Italy), Ruxandra Lasowski (Univ. of Applied Sciences Furtwangen), Petra Mutzel (Univ. Dortmund), Alice McHardy (Univ. Düsseldorf), Nicole Megow (TU Munich), Nada Nakashole (UC San Diego, USA), Marina Papatriantafilou (Univ. Gothenburg, Sweden), Ruzica Piskac (Yale Univ., USA), Nicoletta Preda (Univ. Versailles, France), Maya Ramanath (IIT Delhi, India), Ina Schäfer (TU Braunschweig), Renate Schmidt (Univ. Manchester, UK), Viorica Sofronie-Stokkermans (Univ. Koblenz-Landau), Kavitha Telikepalli (Tata Institute, India), Yafang Wang (Shandong Univ., China), Nicola Wolpert (Univ. of Applied Sciences Stuttgart), Shanshan Zhang (Nanjing U of Science and Technology, China), Hang Zhou (École Polytechnique Paris, France), Anke van Zuylen (College of William & Mary, USA)
Cluster of Excellence on Multimodal Computing and Interaction (MMCI): The MMCI cluster was established by the German Science Foundation (DFG) in the framework of the German Excellence Initiative in 2007 and successfully renewed in 2012, with funding until 2019. The institute is a key contributor to the success of the cluster through its scientific contributions. All directors of the institute are principal investigators of the cluster, and the scientific coordinator of the cluster is Hans-Peter Seidel.

Max Planck Center for Visual Computing and Communication (MPC-VCC): The center was established jointly by MPG and Stanford University with support from the German Ministry of Research and Education (BMBF) in 2003. Following a successful evaluation in 2012, MPC-VCC has been extended until 2020. Selected young scientists have the opportunity to work two years at Stanford as visiting assistant professor, and then continue as leaders of independent research groups at our institute. The center is jointly directed by Hans-Peter Seidel (MPI-INF) and Bernd Girod and Leo Guibas (Stanford).

Indo Max Planck Center for Computer Science (IMPECS): This center fosters collaboration between top Indian universities, MPI-INF and MPI-SWS, and the universities in Kaiserslautern and Saarbrücken. It is funded by the BMBF, the Indian Ministry of Science and Technology and the MPG. The center is directed by Kurt Mehlhorn (MPI-INF) and Naveen Garg (IIT Delhi).

The Research Centre on Interactive Media, Smart Systems and Emerging Technologies (RISE): This centre is located in Cyprus and was founded in 2017. It is funded within the framework of Horizon 2020. MPI is a strategic international partner of RISE and represented on the Board by Hans-Peter Seidel.

ERC Synergy Grant imPACT: The ERC Synergy Grant has been awarded to Michael Backes (Helmholtz Center for Information Security), Peter Druschel (MPI for Software Systems), Rupak Majumdar (MPI for Software Systems) and Gerhard Weikum (MPI for Informatics) for joint research on the strategic research theme of “Privacy, Accountability, Compliance, and Trust for the Internet of Tomorrow”. The project is funded with a total of 10 Million Euros for the timeframe 2015–2020.

IMPRS for Computer Science: The International Max Planck Research School for Computer Science (IMPRS-CS) was established in 2001, as a joint program of the MPI for Informatics and Saarland University. Currently, IMPRS-CS supports 96 doctoral students, including 21 women (about 22%) and 59 non-Germans. In addition, IMPRS-CS currently supports 20 students with fellowships towards Master’s degrees.

The Max Planck Graduate Center for Computer and Information Science: The center is a highly selective doctoral program that grants admitted students full financial support to pursue doctoral research in the broad area of computer and information science, with faculty at the Max Planck Institute for Informatics, the Max Planck Institute for Software Systems, the Max Planck Institute for Intelligent Systems, and some of the best German
universities. Students normally start their graduate studies at the Saarland Informatics Campus in Saarbrücken. While taking courses, they have the opportunity to explore research in different areas as part of immersion labs at different MPIs and universities. The center started in 2018 and the first batch of admitted students started beginning of 2019.

**Saarbrücken Graduate School for Computer Science:** The school was established in 2007 and encompasses all doctoral training in computer science on campus. The school was largely modeled after the IMPRS-CS and adopted many of its elements. IMPRS-CS provides fellowships for doctoral students, within the structural framework of the Graduate School.

### 1.7 Results 2017–2019

**Publications, Software, Startups:** In the two-year timeframe 2017–2019, the institute published more than 600 papers in peer-refereed conferences and journals. Many of these appeared in top-tier venues, with competitive conferences typically accepting only 10 to 20 percent of their submissions. Several publications won best paper awards or best student paper awards (at CVPR 17, EMNLP 17, Eurographics Symposium 17, IEEE VR 17, IMWUT 17, ISWC 17, MUM 17, UIST 17, 3DV 18, ACM CHI 18, ETRA 18, Eurographics 18, IEEE ICDM 18, IMC 18, ISS 18, MobileHCI 18, SOS 18, WWW 18).

Two startups that spun off from our research in 2014 and 2012, Captury and Logic4Business, respectively, are gaining traction in their respective markets. A new startup, Ambiverse, was founded in 2016 with seed-funding from the EXIST program of the German Ministry for Economy (BMWi). It aims to market the scalable software tools on entity linking and deep competence on knowledge-based language understanding for text analytics.

**People:** In the two-year timeframe 2017–2019, 39 of our doctoral students graduated. These include 10 women. In the same time period, 16 of our researchers left the institute to take a tenured or tenure-track faculty position. These include 1 women.

**Awards:** Members of the institute won prestigious awards. The following are the most prominent examples; full lists are in the respective sections of the departments.

- Anja Feldmann received the Vodafone Innovation award 2018 and the Shelling award of the bavarian academy of sciences 2018; Bernt Schiele has been elected IEEE Fellow as well as IAPR Fellow; Hans-Peter Seidel received the Eurographics Medal 2017; Christoph Weidenbach won the Thoralf Skolem Award 2017; Gerhard Weikum’s team received the Seoul WWW Test of Time Award at the Web Conference 2018 for the impact of the original

*Appendix:*

- Bastian Beggel (U. of Applied Sciences Kaiserslautern), Mario Boley (Monash U. Melbourne, Australia), Andreas Bulling (U. Stuttgart), Dan Casas (U. Madrid, Spain), Avishek Chatterjee (IIT Kharagpur, India), Renjie Chen (U. Science and Technology of China (USTC)), Piotr Didyk (USI Lugano, Switzerland), Mario Fritz (CISPA), Christian Ikenmeyer (University of Liverpool), Caigui Jiang (Xi’an Jiatong U.), Moti Medina (Ben-Gurion U., Israel), Pauli Miettinen (U. of Eastern Finland, Finland), Emanuele Natale (U. Côte d’Azur, France), Tim Oosterwijk (Maastricht U., Netherlands), Helge Rhodin (Univ. British Columbia, Vancouver, Canada), Michael Sagraloff (U. of Applied Sciences Landshut), Karteek Sreenivasiah (IIT Hyderabad, India), Jilles Vreeken (CISPA), Hang Zhou (École Polytechnique Paris, France); Björn Andres Adjunct Professor (U. of Tübingen)
Yago paper at WWW 2007, co-authored by Fabian Suchanek, Gjergji Kasneci and Gerhard Weikum.

At the level of Senior Researchers, Christian Theobalt won the Karl Heinz Beckurts-Prize 2017 and received the ERC Consolidator Grant 4DRepLy; Piotr Didyk received the ERC Starting Grant PERDY; Gerhard Pons-Moll received an Emmy-Noether grant; Andreas Bulling received the ERC Starting Grant ANTICIPATE; Karol Myszkowski was recognized with the IEEE TVCG Best Associate Editor Award in 2017; Jilles Vreeken won the IEEE ICDM Tao Li Early Career Award 2018; Karl Bringmann won the EATCS Presburger Award 2019 and the Heinz Maier-Leibnitz-Prize 2019. Gerard Pons-Moll received a Google Faculty Research Award 2019.

At the student level, Franziska Müller received the Google PhD Fellowship in Machine Perception 2017 and the Women STEM Award 2017. Marvin Künnemann won the Otto-Hahn-Medal 2017, Anna Rohrbach won the Otto-Hahn-Medal 2018, Petr Kellnhofer received the Dr.-Eduard-Martin Prize 2017, Mateusz Malinowski received the Dr.-Eduard-Martin Prize 2018. Mateusz Malinowski won the DAGM MVTec Dissertation Award. The Guenter-Hotz Medal 2017 was awarded to Marc Habermann. David Stutz received the STEM Award IT 2018; Pablo Garrido and Petr Kellnhofer won Eurographics PhD Dissertation Awards 2018.

Further honors with considerable visibility include the following. Kurt Mehlhorn received a Honorary Doctorate Degree from the University of Patras and was appointed Honorary Member of the “Gesellschaft zur Förderung des Forschungstransfers”. Bernt Schiele was elected Fellow of the IAPR 2018.

**Gender Proportion and Diversity:** We have been successful in attracting an outstanding woman as a director. At the level of senior researchers, we will expand and intensify our efforts to attract more women. At the postdoc level, the institute offers a distinguished fellowship for women, named Lise Meitner Scholarship, since 2013. The interest in this program has been strongly increasing in terms of both quantity and quality of applicants. In the last two years, we selected three young women as recipients of this fellowship. The best postdocs are often candidates for becoming senior researchers after two years.

**Outreach and Visibility:** To increase the international visibility of Saarbrücken as a world-class CS hub, we have established an agreement for joint branding: all CS players on campus now use the label “Saarland Informatics Campus” as part of their official addresses. The label will also be used in Google Scholar profiles, academic rating sites, and other PR efforts (see http://sic.saarland/).

Within the Max Planck Society, several institutes have set up a joint web site to increase visibility towards young talents (see http://www.cis.mpg.de/).

The institute continues its role as a provider of for the BWINF, “Bundesweit Informatik Fördern”, promoting young computer science talent in Germany.
Part II

Overview – The Research Units & Senior Researchers
2  D1: Algorithms and Complexity

History, Group Organization, and Development

The Algorithms and Complexity group (D1) was established in 1990 as one of the two initial groups of the institute. Kurt Mehlhorn leads the group since its foundation. Kurt Mehlhorn will retire as director of the department on August 31st, 2019. He will continue as a scientist after this date. The appointment process for a new director is on-going.

As of March 1st, 2019, the senior scientists and subgroup coordinators are Alkimini Sgouritsa, Karl Bringmann, Andreas Karrenbauer, Christoph Lenzen, and Kurt Mehlhorn. Section 35.1 lists the names of current and recent group members and the current positions of the group members that left during the report period. Our alumni continue to get very good positions, see page 228. Nine group members completed their PhD, see page 228. Group members received prestigious awards, see page 229, e.g., Karl Bringmann received the EATCS Presburger Award 2019 and the Heinz Maier-Leibnitz-Prize 2019, Marvin Künnemann received an Otto Hahn Medal of the Max Planck Society in 2017, Emanuele Natale received an Award for Best Italian PhD Thesis in Theoretical Computer Science by the Italian Chapter of EATCS in 2017, Eunjiin Oh received the Lise Meitner Award from the MPI-INF in 2018, and Kevin Schewior received a Dissertation Award of the German Operations Research Society (GOR) in 2017. Some group members hold their own grants, see page 229.

We have published extensively and in excellent venues, e.g., 12 papers in SODA 2018 and 2019, 4 papers in FOCS 2017 and 2018, 11 papers in ICALP 2017 and 2018, 5 papers in ESA 2017 and 2018, and 5 papers in SoCG 2017 and 2018.

Vision and Research Areas

The vision for D1 is to be a first class algorithm group and a trendsetter in algorithmics, and to have impact on the research community and society through research results, people, software, and scientific leadership.

About 80% of our effort is theoretical work, and about 20% is experimental and software construction. Our research is organized into five areas. The area coordinators are shown in brackets (Coordinators as of March 1st, 2019, in italics).

- Algorithmic Game Theory and Online Algorithms (Kurt Mehlhorn and Alkimini Sgouritsa)
- Combinatorics, Computing, and Randomness (Karl Bringmann).
- Combinatorial Optimization (Andreas Karrenbauer)
– Geometry and Algebra (Christian Ikenmeyer and Michael Sagraloff)
– Theory of Distributed and Embedded Systems (Christoph Lenzen).

Together, we span a large part of algorithmic research. The emphasis changes over the years as group members come and go. We hire postdocs mainly based on quality and less on thematic fit. There is considerable interaction and collaboration between the areas. The entire group meets twice a week to discuss science (Tuesday and Thursday noon seminar) and biweekly to discuss administrative matters.

We teach at all levels; see page 226 for details. Frequently group members pair for lectures.

We are involved in the Indo Max Planck Center for Computer Science (IMPECS).

Main Results

I discuss some of the main results obtained in the reporting period under the headings fine-grained complexity, algorithm engineering, geodesic Voronoi diagrams in simple polygons, dynamical systems, geometric complexity, distributed algorithms for network flow, and theory of distributed and embedding systems. For all but the third heading, I cover groups of results.

Fine-Grained Complexity

The classic theory of NP-hardness provides a coarse-grained classification of problems into efficient (= polynomial-time solvable) and intractable (= NP-hard). Modern fine-grained complexity theory yields quantitative running time lower bounds, and thus provides a more fine-grained classification of problems according to their time complexity. Naturally, such stronger lower bounds rely on stronger hypotheses than \( P \neq NP \), e.g., the Strong Exponential Time Hypothesis for Satisfiability or the Orthogonal Vectors Conjecture. Karl Bringmann, Marvin Künneman, Bhaskar Chaudhury, Andre Nusser, Philip Wellnitz and their network of co-workers develop fine-grained complexity theory by designing proof techniques for tight fine-grained lower bounds and applying them to a variety of problems. They also contributed to the foundations of fine-grained complexity theory, e.g., by studying the hypotheses that this theory relies on.

They obtained a multitude of results and so I mention only three referring to the Section of fine-grained complexity on page 131 and to Bringmann’s section (see page 63) for details.

– An algorithm for subset sum than runs in time \( \tilde{O}(n + t) \), where \( n \) is the number of input integers and \( t \) is the target and the corresponding conditional lower bound. The previous best algorithm dating back to the sixties ran in time \( O(nt) \).

– Strong additional evidence for the orthogonal vectors conjecture is given, e.g., if the conjecture is false then there is a randomized \( O(n^{(1-\epsilon)k}) \)-time algorithm for the Min-Weight-k-Clique problem on \( d \)-hypergraphs with \( m \) vertices. The conjecture is a corner stone of fine-grained complexity. It states that given two sets \( A \) and \( B \) of \( n \) vectors from \( \{0,1\}^d \), to decide whether there are vectors \( a \in A \) and \( b \in B \) such that \( a \) and \( b \) are orthogonal, requires essentially time \( \Omega(n^2) \) provided that \( d \) is sufficiently large.

– An algorithm for the word-break problem that runs in \( O(nm^{1/3} + m) \). Given a set \( D \) of strings of total length \( m \) and a text \( s \) of length \( n \), is it possible to partition \( s \)
into substrings that all belong to $D$? Surprisingly, a matching conditional lower bound holds for “combinatorial algorithms” (essentially algorithms that are not allowed to use fast matrix multiplication).

**Algorithm Engineering** The algorithm $\text{ANewDsc}$ developed by Kobel, Roullier (INRIA), and Sagraloff for isolating the roots of a univariate polynomial will become the default root finder in the next release of MAPLE; it is based on theoretical work by Mehlhorn and Sagraloff. On ill-conditioned instances, the implementation is considerably faster than previous methods, and on small or well-conditioned instances, it can compete with the fastest existing methods. Moreover, it applies not only to polynomials with integer coefficients but to polynomials with arbitrary computable coefficients. See page 193 for more details.

In 2015, the French Ministry of Culture published a document criticizing the lack of an official standard for a French keyboard layout. It pointed out that all layouts in use were lacking some special characters. The Ministry was concerned that this curbs proper use and development of the French language. AFNOR, the French national organization for standardization, was tasked to produce a standard that would support correct French and include the missing characters. AFNOR set up a committee to design the new standard. A. M. Feit (Aalto), M. Nancel (Inria), M. John, A. Karrenbauer, D. Weir (Aalto), A. Oulasvirta (Aalto) provided computational support to the committee. They modelled the problem as a quadratic assignment problem and optimized for performance, ergonomics, intuitiveness, and familiarity. Intermediate results were presented to the committee which would then typically further modified the objectives. The new standard, see Figure 2.1, was approved recently. The quadratic assignment problem is notoriously difficult. John and Karrenbauer contributed a novel column-generation method that exploits the sparsity of typical instances to allow for quicker assessments of solutions. I refer to Karrenbauer’s section for more details, see page 75.

**Geodesic Voronoi Diagrams in Simple Polygons** The *geodesic distance* between two points inside a simple polygon is the length of the shortest path connecting the two points and contained in the simple polygon. The *geodesic Voronoi diagram* of a set of $m$ point sites in...
a simple $n$-gon is defined as the subdivision of the simple polygon into cells, exactly one cell per site, such that every point in a cell has the same nearest site under the geodesic distance.

Clearly, $\Omega(n + m \log m)$ time is required for computing the geodesic Voronoi diagram and it was asked already 30 years ago, whether this time bound can be achieved. Since ’89, progress was made in a sequence of papers. Eunjin Oh finally settled the problem and presented an algorithm with optimal running time, see page 186 for more details.

**Dynamical Systems** Systems that evolve over time play according to simple rules are ubiquitous and are a common theme for several group members.

An important question in Computational Economics is to provide algorithmic explanation why market equilibrium can be achieved in settings, where price updates on different goods in a market are made in a distributed and non-coordinated manner. In other words, price updates are **asynchronous**. Yun Kuen Cheung and Richard Cole (NYU) showed that CES Fisher markets converge toward market equilibrium amid general asynchronous tatonnement price updates.

Tatonnement price updates, already introduced by Léon Walras in his 1874 book “Elements of Pure Economics, or the theory of social wealth”, can be briefly described as follows: if a good is over-demanded (i.e., its demand exceeds its supply), raise the price of the good; and vice versa. Synchronous tatonnement in a CES Fisher market was known to converge; each round of the process reduces a (carefully chosen) potential function by a certain fraction.

When tatonnement price updates are asynchronous, each update might be using outdated information, so the progresses will be countered by errors. The key question to ask is whether such errors can accumulate to nullify the progresses, and to bar the price updates from converging to equilibrium. Cheung and Cole show that this cannot happen, see page 118 for more details.

In a series of wet-lab experiments it was shown that the slime mold Physarum polycephalum possesses the ability to solve shortest path problems. The slime’s adaption process can be modeled by a system of differential equations. In earlier work, we showed that this dynamical system converges for all graphs and all initial conditions.

Nature cannot be expected to follow the differential equation exactly. In this period, we (Karrenbauer, Kolev, Mehlhorn) showed convergence for a considerably more general dynamics. Let $G$ be a graph with two designated nodes $s_0$ and $s_1$. Each edge $e$ has a positive cost $c_e$ and a time-varying positive capacity $x_e$. The capacity vector evolves according to the dynamics

$$\dot{x}_e = a_e(x, t) (|q_e| - x_e),$$

where $q_e$ is the electrical flow through edge $e$ in a network where edge $e$ has resistance $c_e/x_e$ and one unit of current is sent from $s_0$ to $s_1$, and $a_e(x, t)$ is any Lipschitz-continuous function that is bounded away from zero and from infinity, i.e., there is a positive constant $C$ such that $1/C \leq a_e(x, t) \leq C$ for all $x$ and $t$. In the original result, $a_e(x, t) = 1$ for all $x$ and $t$. We refer to page 179 for more details.

Emanuele Natale and co-workers study consensus dynamics on the complete graph of $n$ nodes. Initially, each node supports one from up to $n$ opinions. Nodes randomly and in
parallel sample the opinions of constantly many nodes. Based on these samples, they use an update rule to change their own opinion. The goal is to reach consensus, a configuration where all nodes support the same opinion.

They compare two well-known dynamics: 2-Choices and 3-Majority dynamics. In the former, each node samples two nodes and adopts their opinion if they agree. In the latter, each node samples three nodes: If an opinion is supported by at least two samples the node adopts it, otherwise it randomly adopts one of the sampled opinions.

They show that the 3-Majority dynamics reaches consensus with high probability in $\tilde{O}(n^{3/4})$ rounds, while the 2-Choices dynamics can need $\Omega(n/\log n)$ rounds. They thus get the first unconditional sublinear bound for 3-Majority dynamics and the first result separating the consensus time of these processes. Along the way, they develop a framework that allows a fine-grained comparison between different consensus processes, see page 217 for more details.

Geometric Complexity Theory Geometric Complexity Theory was initiated by Mulmuley and Sohoni in order to resolve Valiant’s famous determinant vs permanent conjecture (the permanent cannot be computed by a polynomial size algebraic circuit). Mulmuley and Sohoni laid out a line of attack by formulating a sequence of conjectures. Christian Ikenmeyer, Karl Bringmann and Gorav Jindal in cooperation with Markus Bläser, Vladimir Lysikov (Saarland University), Peter Bürgisser (TU Berlin), Fulvio Gesmundo (University of Copenhagen), Stefan Mengel (CRIL), Ketan Mulmuley (University of Chicago), Greta Panova (University of Southern California), Michael Walter (University of Amsterdam), and Jeroen Zuiddam (IAS Princeton) refuted some of these conjectures\(^1\); this is a severe blow to the program.

They also investigate the fundamental capabilities and limitations of algebraic proofs for complexity lower bounds. We provide a setting in which superpolynomial algebraic lower bounds proofs require superpolynomial circuit size under a reasonable Boolean complexity assumption. Moreover, we show how geometric complexity theory can in principle break this barrier by encoding hard functions by concise representation-theoretic labels. As a side result from our techniques we prove that there is a constant multiplicative error to which tensor rank is NP-hard to approximate.

We refer the reader to page 198 for details.

Distributed Algorithms for Networks Flows Shortest path problems and maximum flow problems are notoriously hard for distributed algorithms. The standard combinatorial algorithms for these problems are inherently sequential and do not seem to be a good starting point for distributed algorithms. This is different for newer algorithms based on interior point methods.

M. Ghaffari (ETH), A. Karrenbauer, F. Kuhn (Freiburg), C. Lenzen, and B. Patt-Shamir (Tel Aviv) designed a near-optimal distributed algorithm computing a $(1 + o(1))$-approximation for the maximum flow in undirected weighted networks that runs in $O((D + n)n^{o(1)})$ rounds of communication in the Congest model; here $D$ is the diameter of the network. This is the first improvement over the trivial bound of $O(n^2)$ and nearly matches

\[^1\]This paragraph is more vague that the other paragraphs as KM has only superficial knowledge of these results.
the $\Omega(D+n)$ round complexity lower bound. The development of their algorithm entails two sub-results of independent interest: (1) An $O((D+n)n^{o(1)})$-round distributed construction of a spanning tree of average stretch $o(1)$ and (2) an $O((D+n)n^{o(1)})$-round distributed construction of an $o(1)$-congestion approximator consisting of the cuts induced by $O(\log n)$ virtual trees.

Moreover, R. Becker, A. Karrenbauer, S. Krinninger, and C. Lenzen studied undirected transshipment and single-source shortest paths (SSSP) problems in distributed and streaming models of computation. They presented a tailored gradient descent algorithm that computes near-optimal solutions (up to a multiplicative error of $1+\varepsilon$) within $\varepsilon^{-O(1)} \log^{O(1)} n$ iterations, each comprising of a polylog-approximation of an auxiliary instance of a transshipment problem. Such solutions of poly-logarithmic error can be obtained by computing a solution on a sparse spanner of logarithmic stretch, which leads to interesting results in various distributed and streaming models of computation:

1. **Broadcast Congest model:** $(1+\varepsilon)$-approximate SSSP using $\tilde{O}((\sqrt{n}+D)\varepsilon^{-O(1)})$ rounds.
2. **Broadcast Congested clique model:** $(1+\varepsilon)$-approximate shortest transshipment and SSSP using $\tilde{O}(\varepsilon^{-O(1)})$ rounds.
3. **Multipass streaming model:** $(1+\varepsilon)$-approximate shortest transshipment and SSSP using $\tilde{O}(n)$ space and $\tilde{O}(\varepsilon^{-O(1)})$ passes.

I refer to page 173 for details.

**Theory of Distributed and Embedded Systems** The work on Design of Fault-tolerant Hardware (see page 203) and Hazard-freedom and Metastability-containment (see page 207) is to a large extent financed by Christoph Lenzen’s ERC grant and so I give Christoph the priority to report about them; see page 77.

The other topics in this research area are distributed algorithms and computational dynamics. I already reported about them in earlier paragraphs.
3 D2: Computer Vision and Multimodal Computing

Group Overview

The Computer Vision and Multimodal Computing group (D2) was established in 2010 with the appointment of Bernt Schiele. At the time of writing, the group encompasses three research group leaders, three associated research group leaders, three postdocs and 19 PhD students. Among those 22% are female (one research group leader, three postdocs, three PhD students).

The research group leaders and senior scientists each have their own PhD-students to conduct research in their respective area. The current three research group leaders are Zeynep Akata (appointed 2014), Gerard Pons-Moll (appointed 2017), and Paul Swoboda (appointed 2018). The three associated research group leaders have been appointed as senior researchers at the institute before and still supervise PhD students at the institute, but have accepted offers for faculty positions elsewhere in the meantime. These are: Andreas Bulling (faculty at Uni Stuttgart since November 2018), Mario Fritz (faculty at CISPA since July 2018), and Bjoern Andres (faculty at Uni Tübingen and group leader at Bosch BCAI Renningen since October 2017).

Nine group members completed their PhD during the reporting period and one more will defend his PhD in early April. Three of them obtained prizes for their dissertation. Our researchers get very good offers for faculty positions in academia, postdoc positions in academia, and research positions in industry.

Vision and Research Areas

Understanding sensor information is a fundamental problem in computer science. Scientific challenges cover the entire pipeline from single-sensor processing, over spatial and temporal fusion of multiple and divergent sensor modalities to the complete description of large-scale multimodal sensor streams. At the same time we observe a tremendous increase in both the quantity as well as the diversity of sensor information due to the increasing number of sensors.

\footnote{Anna Rohrbach obtained the Otto-Hahn Medal of the Max-Planck Society, Siyu Tang and Mateusz Malinowski both received the DAGM MVTec Dissertation Award 2018, and Mateusz Malinowski also the Dr. Eduard-Martin Award}

\footnote{Björn Andres (U Tübingen), Zeynep Akata (U Würzburg), Andreas Bulling (U Stuttgart, U Freiburg), Mario Fritz (CISPA, U Marburg)}

\footnote{Anna Rohrbach (UC Berkeley), Siyu Tang (MPI Intelligent Systems), Xucong Zhang (ETH Zurich)}

\footnote{Jan Hosang (Google, Zurich), Anna Khoreva (Bosch, BCAI Renningen), Maksim Lapin (Amazon, Berlin), Wenbin Li (Continental, Regensburg), Mateusz Malinowski (Google Deepmind, UK), Seong Joon Oh (LINE Plus Clova ML, Korea)}
such as cameras, GPS, or inertial sensors) embedded in a wide variety of digital devices and environments as well as due to the increasing storage of multimodal sensor data (such as surveillance data, personal storage of digital information, multimedia databases, or simply the Internet). While storing and indexing large amounts of sensor data has made tremendous progress, understanding of this multimodal sensor data still lacks far behind. Therefore the long-term goal of D2 is to make progress on how to process, structure, access, and truly understand multi-sensory data both for online use as well as for large-scale databases.

The group currently focuses on two areas, namely computer vision and multimodal sensor processing. In the area of computer vision we address some of the most fundamental problems of image and video understanding such as object class recognition, people detection and tracking, and scene understanding. In the area of multimodal computing we currently focus on perceptual user interfaces. As a cross-cutting theme for both areas we also work in the area of machine learning. It is clear that only advanced machine learning techniques will enable to infer higher-level information from noisy sensor data and enable to deal with the large-scale nature of current and future multimodal databases and sensor-streams.

**Highlights of Current Research**

In the last report we discussed long-term research results e.g. in the areas of human pose estimation and tracking as well as object recognition. In this report we mainly focus on recent research highlights from the current reporting period.

**Zero and Few-Shot Learning**  
**Investigators: Zeynep Akata and Bernt Schiele**

Many approaches in generalized zero- and few-shot learning rely on cross-modal mapping between the image feature space and the class embedding space. As labeled images are rare, one direction is to augment the dataset by generating either images or image features. However, the former misses fine-grained details and the latter requires learning a mapping associated with class embeddings.

An important achievement in this area during the reporting period are CVPR 2017 and TPAMI 2018 papers that establish a benchmark on zero-shot learning which has since then become one of the major resource for benchmarking (their cumulative citation count is over 200 since June 2017). In addition, we proposed connections between zero-shot learning and human gaze (CVPR 2017, spotlight). Furthermore, we improved zero-shot learning by generating images of unseen classes (ICML and NIPS 2016) and image features of unseen classes (CVPR 2018) which established the new state of the art on six benchmark datasets.

Furthermore, in a recently accepted CVPR’19 paper, we propose to develop a conditional generative model that combines the strength of VAE and GANs and in addition, via an unconditional discriminator, and learns the marginal feature distribution of unlabeled images to learn highly discriminative CNN features generalizable across datasets. We have also demonstrated that our learned features are interpretable: we visualize them by inverting them back to the pixel space and we explain them by generating textual arguments of why they are associated with a certain label. This way we aim to connect low-shot learning with explainability.
**Privacy and Security**  *Investigators: Mario Fritz and Bernt Schiele*

While recent advances in computer vision and machine learning bring many benefits to society, we equally have to be aware of emerging security and privacy risks.

*Adversarial Perturbation* are small changes to the input data, that can mislead state of the art classifiers. We make significant contributions to the understanding of this phenomena and investigate the impact on more complex vision/language tasks and deep learning policies of agents, as well as, investigate the interplay between attacks and defenses in a game theoretic framework. This line of work was published in top tier venues: 1 ICCV, 2 CVPR.

*Attacks and Defenses for Machine Learning* have gained relevance as more and more such approaches are deployed in various application scenarios. We have shown for the first time that details such as architecture parameters can be reverse engineered by observing only the output of a neural network. Beyond this, we show that modeling stealing is possible under weak assumptions, thereby raising new concerns about the security of these techniques and how intellectual property can be protected. We complement these new attacks with new defenses to counter such strategies. The work is published in top venues: 1 ICLR, 1 CVPR.

*Enforcing Privacy* has received increasing attention, as the latest machine learning techniques can extract private information from multi-modal data with high accuracy. We have pioneered work on visual privacy where we take a systematic approach to categorize and sanitize a broad range of private information in visual data. Our work extends to natural language and also decentralized learning. The work is published in top venues: 2 CVPR, 1 ECCV, 1 USENIX Security, 1 AISTATS.

*Understanding Privacy* is important in order to assess the amount of private information in the first place. We focus on visual data in social media type data, where we evaluate and advance person recognition as well as demonstrate recognition of social relations in a domain based approach. We have presented first steps towards the overall vision of a Visual Privacy Advisor that can support users to enforce their privacy policies on social media. The work is published in top venues: 1 ICCV, 1 CVPR.

**Scalable Learning and Perception**  *Investigator: Mario Fritz*

With the advances of deep learning techniques, recent success has spurred the hope to address more holistic challenges that encompass not only scene understanding but also include light, surfaces, physics and multiple agents. We have made significant progress towards connecting the success of Deep Learning in vision to computer graphics, simple physics in robotics as well as in learning of agent behavior. In particular, we have proposed advanced approaches that use machine learning to extract surface and light information, synthesize new appearance or automatically edit visual content. We have proposed a first, fully data-driven approach (“intuitive physics”) to learn simple stability events and demonstrated for the first time how such models can be leveraged for robotic manipulation. This line of work is published in top tier venues: 1 ICCV, 2 TPAMI, 1 ICRA, 1 NIPS, 1 CVPR.

The second focus, in cooperation with Bernt Schiele, is on modeling uncertainty in Deep Learning approaches. While we have seen a steady increase in accuracy – in many application scenarios this is insufficient. In particular, we focus on anticipating on future events in a scene, which is of high relevance to autonomous driving. We have advanced modeling of uncertainty in Deep Learning approaches in particular when multi-modal posteriors are
concerned and also demonstrated such techniques in relevant application scenarios. Also this line of work is published in top tier venues: 2 CVPR, 1 ICLR, 1 AAAI.

**Real Virtual Humans**  
*Investigator: Gerard Pons-Moll*

The faithful digitization of human beings has the potential to redefine the way we think and communicate (with other humans and with machines), and it is necessary for many applications. While digital human models entail many components of artificial intelligence, we are focusing on modelling and capturing the 3D human appearance and motion of people.

Over the last years, we have made significant progress, and have produced statistical body models of pose, shape, soft-tissue and clothing (Dyna-Siggraph’15, SMPL-SiggraphAsia’15, CVPR’17, ClothCap-Siggraph’17) which are widely used in industry and academia, and are regarded as state of the art. While useful, to train these models we required high quality 3D scans which makes it difficult to scale. In addition, no easy to use models of people in clothing exist.

Our main challenge for the future is to faithfully reconstruct and model 3D people in clothing from single images and video, for which data is abundant. Our solutions will have to effectively incorporate 3D model knowledge within deep learning frameworks. In essence, we want to train machines to learn about the 3D world by looking at images and video. We have done first steps in this direction and introduced algorithms to reconstruct people in clothing from video for the first time (CVPR’18), and to reconstruct pose and shape from a single image (3DV’18 best paper). We also recently made publicly available the first dataset, recorded with a moving camera, of people in natural scenes with ground truth 3D poses for which we devised a fusion algorithm that combines Inertial Sensors and video data (ECCV’18). In the future, we plan to learn models of clothing geometry, topology and appearance, as well as algorithms to reconstruct and model people interactions.

In summary, our research aims at answering the following two inter-related research questions: *How do we efficiently digitize humans without losing the detail that make us real?* and *How can we train machines to perceive such rich representations from visual data?*

The work has been published at the major international computer vision and computer graphics conferences and journals including (10 papers at CVPR, ICCV, ECCV), (6 papers at Siggraph, Siggraph Asia, TOG), BMVC(Best Paper Award 2013), Eurographics(Best Paper Award 2017), 3 papers at 3DV(2018 Best Student Paper Award), IJCV and TPAMI. Gerard Pons-Moll has received two starting grants, the Emmy Noether Programme 1.6Mi in December 2018, and a grant from Zentrum Digitalisierung Bayern 1.25Mi in January 2017 (declined), a Google Faculty Research Award (2019), a Facebook research cooperation agreement (2018), as well as 3 best paper awards at vision and graphics conferences.

**Combinatorial Image Analysis**  
*Investigator: Bjoern Andres*

Algorithms for image analysis are beginning to have major impact on our society, e.g., in the area of computer vision for autonomous driving, and in the areas of image and data analysis for biomedical research and applications. The research group Combinatorial Image Analysis studies mathematical abstractions of image analysis tasks in the form of combinatorial optimization problems. It develops algorithms for solving these problems exactly, partially or locally. It analyzes these algorithms and the solutions they output in
terms of application-specific metrics and benchmarks. The goal of this research is to enable practical applications of expressive combinatorial optimization problems in the fields of computer vision and biomedical image and data analysis.

Our main achievement during the reporting period consists in the understanding that seemingly unrelated computer vision tasks, including instance-separating semantic image segmentation, multiple object tracking and human body pose estimation, are fundamentally linked by a single combinatorial optimization problem, a generalization of the well-known correlation clustering problem. A publication at CVPR 2017 can be seen as the culmination of this work that spans a series of papers published over a period of three years at ICML, CVPR, ICCV and ECCV. Also during the reporting period, Bjoern Andres has accepted an offer of an Honorary Professorship by the University of Tübingen.

**Vision and Language**  
**Investigator:** Zeynep Akata

Deep models are frequently seen as opaque and are unable to explain their decisions. In contrast, humans can justify their decisions with natural language and point to the evidence in the visual world which led to their decisions. However, existing models which generate textual explanations enforce task relevance through a discriminative term loss function, but such mechanisms only weakly constrain mentioned object parts to actually be present in the image.

Our major achievements in this field during the reporting period are CVPR 2018 (spotlight) paper that proposed two datasets for both activity and VQA pointing and explanation tasks. In addition, we proposed improvements in ECCV 2018 over our generated textual explanations proposed in ECCV 2016 and extended it to videos in the context of self-driving cars in ECCV 2018. These works were important in establishing a benchmarks and state of the art in various settings through well curated datasets, and well designed models which are generalizable across domains. Our efforts have been awarded with a DARPA grant with the title Deeply Explainable Artificial Intelligence to a team composed of UC Berkeley, University of Boston and University of Amsterdam. At the University of Amsterdam this funding supports 30% of an assistant professor and a PhD student.

In the short term, we propose to generate explanations by utilizing localized grounding of constituent phrases in generated explanations to ensure image relevance. From the perceptual input point of view, this grounding can have different forms, e.g. attention maps, bounding boxes, etc. where the important regions of the image are determined in a post-hoc manner.

A complementary direction is that, instead of producing a classification in a single step, we propose to iteratively make binary sub-decisions which, when combined as a whole, aim to produce the same classification result while revealing a decision tree as thought process. In the short term, we aim to incorporate attribute information at the class-level to give the binary decisions a semantic meaning and allowing to trade-off interpretability and classification accuracy. We anticipate that the decision tree resulting from the sequence of binary decisions would reveal a hierarchical clustering of the data and can be treated as attributes learned from scratch.

In the long term, we would like to utilize deep and explainable frameworks in more mainstream tasks. For instance, deep neural perception and control networks have become key components of self-driving vehicles. User acceptance is likely to benefit from easy-to-interpret visual and textual driving rationales which allow end-users to understand what
triggered a particular behavior. As a starting point, we plan to use visual (spatial) attention model to train a convolutional network end-to-end from images to steering angle commands. The attention model identifies image regions that potentially influence the network’s output. We would then apply a causal filtering step to determine which input regions causally influence the vehicle’s control signal. As for the human-machine communication, we would use a video-to-text language model to produce textual rationales that justify the model’s decision. The explanation generator would then use a spatiotemporal attention mechanism, which is encouraged to match the controller’s attention.

**Perceptual User Interfaces  Investigator: Andreas Bulling**

Developing human-computer interfaces that fully exploit the information content available in natural human behavior is challenging, particularly in unconstrained daily life settings. The group works at the interface of human-computer interaction, computer vision, as well as wearable and ubiquitous computing. We develop novel ambient and on-body sensing systems as well as computational methods to analyze human behavior automatically. We specifically focus on visual and physical behavior as these modalities are most promising for developing interfaces that offer human-like interactive and social capabilities. We study these systems and methods in the context of specific application domains, most importantly pervasive eye-based human-computer interfaces and computational human behavior analysis.

Major achievements in the reporting period include novel computational methods for everyday eye contact detection using unsupervised gaze target discovery (Best Paper Honourable Mention Award at UIST 2017), for fully articulated eye gaze redirection in video (Best Paper Honourable Mention Award at Eurographics 2018), for forecasting user attention during everyday mobile interactions using device-integrated and wearable sensors (Best Paper Award at MobileHCI 2018), as well as a ground-breaking new approach to mobile eye tracking using multiple tiny, low-resolution cameras and learning-based gaze estimation (Distinguished Paper Award in the Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies). We have also released OpenGaze (www.opengaze.org), the first open source toolkit for camera-based gaze estimation and interaction that aims to democratize their use in HCI and therefore designed the toolkit specifically for gaze interface designers. In the area of human-computer interaction, we have contributed a detailed analysis of the past, present, and future of gaze-enabled handheld mobile devices (Best Paper Honourable Mention Award at MobileHCI 2018), a study on the identifiability of different visual representations of oneself on public displays (Best Paper Honourable Mention Award at ACM CHI 2018), an investigation into the viability of a threat model that involves multiple shoulder surfers (Best Paper Honourable Mention Award at MUM 2017), and we have introduced, prototypically implemented, and empirically evaluated the novel concept of error-aware gaze-based interfaces for robust mobile gaze interaction (Best Paper Award at ETRA 2018).

**Image and Video Segmentation  Investigators: Anna Khoreva and Bernt Schiele**

Image and video segmentation is a very challenging task and one of the most crucial steps toward scene understanding. Much progress has been made in this field over the last years. To a large extent, the success can be attributed to the strong appearance models completely learned from data, in particular using deep learning methods. However, to
perform best these methods require large representative datasets for training with expensive pixel-level annotations, which for many applications are prohibitive to obtain in practice. This considerably restricts the potential to transfer these models to approach different domains or various object categories. Therefore, there is a need to relax the constraint of exploiting expensive pixel-level annotations for training and to consider alternative forms of supervision, which are easier and cheaper to collect.

In the last few years we focused on developing approaches for training segmentation networks with weaker forms of supervision, such as leveraging bounding boxes (CVPR’17) or image labels (CVPR’17) as supervision, making use of unlabeled and synthetic data (ArXiv’18) for semantic and instance labelling tasks, exploiting static images (CVPR’17) and sophisticated data augmentation techniques (IJCV’19) as well as language supervision (ACCV’18) for video object segmentation. Our contributions in this area allowed us to achieve high-quality image and video segmentation results while relying on less expensive annotations, advancing the state of the art on multiple challenging benchmarks. Major achievements in the reporting period have been our spotlight presentation at CVPR’17 (acceptance rate 9%), oral presentations at GCPR’17, workshops at CVPR’17, ECCV’18 and ACCV’18 and the second place in the 2017 DAVIS Challenge on Video Object Segmentation.

In the long term, we would like to continue working on weakly supervised approaches for segmentation task. We are investigating using conditional generative adversarial networks (GANs) for data and feature synthesis, exploiting temporal coherence and dynamics in videos as supervision as well as leveraging multi-modal data for segmentation annotation generation to resolve ambiguities caused by partial observations of a specific modality. We believe that the proposed strategies can further boost the results and help to relax the data dependency constraint.

Publications, Cooperations and Awards

Chapter 36 contains a detailed report of the publications, cooperations and awards of the reporting period. From the journal publications 16 have been published or accepted at either IJCV or IEEE PAMI. From the conference publications, 55 have been published or accepted at one of the major computer vision conferences (CVPR, ICCV, ECCV). An additional 9 have been published at NIPS, ICML, ICLR or AAAI. Another 14 papers have been published at CHI, UIST, IMWUT, and ETRA. In the reporting period (2017-2019) the group has been cooperating with a wide range of research groups worldwide. Cooperations that have led to joint publications during the reporting period include: UC Berkeley, Stanford U, Oxford U, ETH Zurich, EPFL, KU Leuven, Lancaster University, TU Munich, U Tübingen, U Freiburg, MPI for Intelligent Systems, as well as Researchers from Deepmind, Google, Facebook, Intel, and Amazon.

Mateusz Malinowski received the DAGM MV Tec and Dr. Eduard Martin Dissertation Awards for his PhD. Siyu Tang also received the DAGM MV Tec Dissertation Award. Anna Rohrbach received the Otto Hahn Medal for her PhD. Several best paper awards have been awarded to members of the group including at 3DV (Mohamed Omran, Gerard Pons-Moll and Bernt Schiele), MobileHCI (Andreas Bulling), ETRA (Andreas Bulling), and IMWUT (Andreas Bulling). Bernt Schiele has been elected IEEE Fellow as well as IAPR Fellow.
4 D3: Internet Architecture

History, Group Organization and Development

The Internet Architecture department (D3) was established in January 2018, and group’s organization structure and details of members are provided in §37.3.

Vision and Research Strategy

The Internet is a immensely successful human-made artifact that has fundamentally changed the society. In becoming such an immensely successful infrastructure the use of the Internet and, consequently, the Internet itself has changed and continues to change, as highlighted by some of this group’s research efforts. These changes are in part driven by the user/eye-ball interests as well as by how content, including user generated content is made available. The AS-level topology of the Internet has also experienced significant changes (or “flattening”) over time, i.e., it has evolved from a highly hierarchical topology to a flatter (non-hierarchical, simpler) topology. In contrast, content providers and content delivery networks are relying on sophisticated back-office infrastructures that include crawlers, caching hierarchies, as well as infrastructure to deliver advertising. The future challenges in this context are (i) continual observation of the usage as well as the underlying infrastructure, (ii) locating current performance bottlenecks within the infrastructure, (iii) understanding how novel applications interact or should interact with the infrastructure, (iv) designing network management mechanisms to minimize the need for manual configuration within the infrastructure as well as the across infrastructures and automating security mechanisms for protecting our infrastructure, and lastly (v) incentivizing efficient network usage and network upgrades.

We will be generating staggering volumes of data in a few years, if not already doing so, almost everywhere. Our analytical processing capabilities will also have further advanced, e.g., offer intelligent machine learning mechanisms, in this time frame. The increasing demands of users to have a ubiquitous access to information from anywhere at any time, in a context where data is generated in a highly distributed manner (i.e., in diverse geographical locations and networks) poses unique, extremely hard challenges to solve. A feasible solution will have to enable data streams to be processed and distributed in a coordinated manner in real time. Such an approach requires a distributed processing platform where both computations and data can move around freely as well as securely in an optimal fashion providing fast reaction times and minimal resource usage. In this context data provenance, quality criteria, and time constraints, both varying per customer, will have to be taken into account, necessitating the integration of information processing and networking into a single paradigm.

Part of the success of the early Internet is that it relied on “working code and rough consensus.” The Internet infrastructure today is, unfortunately, not quite as simple as
before, since the performance requirements—100 Gbps to 10 Tbps—of today’s networks, particularly, in the core of the Internet require custom hardware solutions. This requirement has lead to the use of specialized hardware and software and closed-box solutions for the main components of the Internet—the routers and switches—thereby resulting in the ossification of the Internet. Software defined networking is one way in which to tackle this problem and some of the challenges in this are include (i) taking advantage of the software capabilities, (ii) supporting wide-area data analytics, and (iii) integrating network resources, namely storage, CPU, and data, into the concept of SDN.

Our approach to tackle these challenges is to follow a data-driven systems research agenda: collect data from operational networks, analyze them using big-data analytics to identify invariants, revisit assumptions, and detect where current performance bottlenecks in the Internet are. We also employ simulation environments to validate our analysis results and to enable “what-if” studies. The insights obtained from the measurements form the foundation for shaping the future Internet via optimizations and alternative designs. Hereby, our output includes protocol enhancements, novel network management tools, and concepts for software-defined networking.

Research Areas and Achievements

Our research work spans three broad areas, (1) Internet measurement, (2) future-proofing the Internet, and (3) wide-area data analytics.

Internet Measurement

The Internet is massively heterogeneous and also continually evolving ecosystem. Naturally, no one vantage point can accurately capture the breadth of these changes. This limitation notwithstanding, there is a dire need to monitor and analyze the use of the Internet infrastructure as well as the characteristics of the infrastructure itself, especially given the constantly evolving nature of the ecosystem. Meticulously gathering measurements from diverse vantage points, and systematically analyzing these measurements to characterize the performance and operation of the Internet ecosystem, identify vulnerabilities and issues in the different components of this ecosystem, e.g., network protocols and devices, characterize the use of the infrastructure, e.g., through analyses of network traffic volume and dynamics, and understand how to evolve and upgrade the networks for the future constitutes a core research focus of our department. Analyzing the use of BGP communities in the wild and the routing vulnerabilities they introduce, leveraging BGP communities to mitigate large-scale distributed denial-of-service (DDoS) attacks, and analyzing the traffic volume asymmetry at scale are some of the ongoing research efforts in this space. In addition to providing crucial insights into the current state of the Internet, our measurement-driven systems design approaches reveal how to also upgrade and evolve the Internet infrastructure in anticipation of novel applications and changing usage patterns.
Future-proofing the Internet

The continually evolving nature of the Internet ecosystem routinely introduces new, unforeseen challenges. With the increasing adoption of sensors and Internet of Things (IoT) devices, we are generating an unprecedented volume of data. If the generated data volumes are not challenging enough, that such data now comes from diverse endpoints, widely distributed throughout the network at its edge, makes the collection and analysis of this data in real time poses a grand challenge for the networking community. Addressing the challenge necessitates the design and development of new tools and techniques, or even better computing primitives that can make scalable, accurate, real-time analyses of such data feasible in practice. The changes in the Internet ecosystem also provide unprecedented opportunities to revisit some long standing networking problems and design practical, scalable solutions that exploit or leverage these recent changes. Redesigning protocols or applications by exploiting new opportunities facilitates the applications to better adapt to the changing network; the infrastructure in turn benefits from having well-designed protocols and applications that make more efficient use of the infrastructure.

Wide-area data analytics

We are generating an astounding volume of data in the Internet, and virtually all forecasts indicate that the amount of data we will generate, gather, store, and analyze will only continue to increase at a staggering pace. The tools and techniques to analyze this data in a systematic, scalable, and practical manner, whether for security audits or for mundane operations such as infrastructure planning, however, has not kept pace with the growth in data generation. Besides, data sources today are often widely distributed across different networks and geographies, making the already hard problem even harder. Aggregating data from widely distributed vantage points necessitate might necessitate techniques to compress or consolidate the data to minimize the volume of data transferred from different locations to a centralized location for analyses. It is also often hard to know a priori the type and nature of queries that will be run against these data sets and this limitation highlights the need for generic tools and techniques to represent data in a succinct form while still remaining amenable to be queried in a scalable manner. Alternatively, we need to also explore about how to decompose a query to run in a truly distributed manner or reorganize, reshape, and relocate data in such a way to reduce the query time without sacrificing much accuracy.

Prizes and Awards

Prof. Anja Feldmann, Ph.D.:

- Vodafone Stiftung: Vodafone Innovationspreis, 2018
- Bayerische Akademie der Wissenschaften: Schelling Prize, 2018
- TU-Berlin: Honorarprofessor, 2018

Press and Media Coverage

Prof. Anja Feldmann, Ph.D.:
– 3SAT: Scobel on Blockchain, http://www.3sat.de/mediathek/?obj=73309

Teaching

Data Networks (A. Feldmann and B. Chandrasekaran)
5 D4: Computer Graphics

Group Overview

The computer graphics group (D4) was established in 1999 with the appointment of Hans-Peter Seidel. Karol Myszkowski joined in 2000, Christian Theobalt joined in 2009. Over the last two decades the group graduated more than 60 PhD students, and more than 40 former group members got offers for tenured faculty position. During this period our academic offspring received a variety of prestigious grants and awards, including two ERC Consolidator Grants, eight ERC Starting Grants, four DFG Emmy Noether Fellowships, seven EG Young Researcher Awards, and three German Pattern Recognition Awards.

At the time of writing, the group encompasses two senior researchers with tenure (K. Myszkowski and C. Theobalt), four group leaders and senior researchers without tenure (R. Chen (GL since 2015, SR since 2018), V. Babaei (GL since 2018), G. Singh (GL since 2019), and R. Zayer (GL since 2018)), six postdocs, and 18 PhD students. Six group members completed and handed in their PhD thesis during the reporting period, and several of our young researchers and former PhD students got offers for faculty appointments or postdoc positions. Full details on current and recent group members are provided in Section 38.1.

Vision and Research Strategy

During the last few decades computer graphics firmly established itself as a core discipline within computer science. New and emerging technologies such as digital media, social networks, digital television, digital photography and the rapid development of new sensing devices, telecommunication and telepresence, virtual and augmented reality further indicate its potential and pose new challenges in the years to come.

To address these challenges, and in particular to seamlessly blend real and synthetic footage, we have adopted a new and more integrated scientific view of computer graphics as 3D Image Analysis and Synthesis that takes into account the whole image processing

---

6 D. Casas (Univ. Madrid, Spain, 2018), A. Chatterjee (IIT Kharagpur, India, 2017), R. Chen (University of Science and Technology of China (USTC), 2019), P. Didyk (USI Lugano, Switzerland, 2018), C. Jiang (Xi’an Jiaotong University, 2019) H. Rhodin (Univ. British Columbia, Vancouver, Canada, 2019)
7 P. Garrido (Technicolor), P. Kellnhofer (MIT), T. Leimkühler (INRIA), S. Sridhar (Stanford), M. Zollhöfer (Stanford)
pipeline from scene acquisition to scene reconstruction to scene editing to scene rendering. We also take into account human perception on all levels of the pipeline, and we exploit the abundance of digital visual data and novel concepts from machine learning to extract powerful priors that can assist us during the acquisition, reconstruction, editing, and image formation processes.

Our vision and long term goal are completely immersive, interactive, and visually rich environments with sophisticated scene representations and the highest visual quality, fused seamlessly with the real world. Standard 2D screens are being replaced with high dynamic range displays, stereo and automultiscopic screens, portable and wearable displays. Imaging algorithms with embedded perceptual models ensure that the perceived quality and viewing comfort is maximized. Interaction is intuitive and light weight.

In order to make progress along the lines above, our work is both theoretical and practical with a focus on first-class research and new methods and algorithms, as well as on the integration of new algorithms into functioning software systems, and the experimental validation of systems in specific application scenarios that are of practical relevance. We also try to provide a stimulating environment for junior researchers that allows them to develop and build their own research programs and groups.

**Research Areas and Achievements**

Our research is currently organized into the following six research areas (coordinators in brackets):

- Digital Geometry Processing (R. Chen, R. Zayer, and H.-P. Seidel)
- Computational Videography (C. Theobalt)
- Reconstructing the Static and Dynamic Real World (C. Theobalt)
- Realistic and Real-time Rendering (G. Singh)
- Perception and Advanced Displays (K. Myszkowski)
- Computational Fabrication (V. Babaei)

However, we are not organized into disjoint subgroups, and there is little hierarchy. While each of the areas has its specific focus, some of them also have significant overlaps. Likewise, the students and researchers working in each area are dynamically formed teams rather than specifically dedicated staff. The senior researchers and group leaders together with Hans-Peter Seidel serve as an internal steering committee for the group. They also act as advisors or co-advisors of doctoral students.

**Digital Geometry Processing**

Digital geometry processing is concerned with the representation, analysis, manipulation, and optimization of digital shapes. During the reporting period we developed a novel sparse matrix representation for unstructured grids and demonstrated its power in different geometry processing applications, and we substantially improved the performance of the underlying linear algebra kernels. We also continued our successful work on shape deformation.
A key advantage of working with structured grids is the ability to directly tap into the powerful machinery of linear algebra. This is not much so for unstructured grids where intermediate bookkeeping data structures stand in the way. On modern high performance computing hardware, the conventional wisdom behind these intermediate structures is further challenged by costly memory access, and by prohibitive memory resources on environments such as graphics hardware. We developed a novel sparse matrix representation for unstructured grids and demonstrated its power in different geometry processing applications such as high performance surface tessellation and high performance mesh subdivision. Our representation reduces the memory storage requirements but also cuts down on the build of data movement from global storage to the compute units. Our framework translates geometric computations and topological modifications into concise linear algebra operations, and we capitalize on the nature of sparse matrix-vector multiplication which allows avoiding explicit transpose computation and storage (EG’17, SIGAsia’18, CoRR’18).

Linear algebra kernels and graph algorithms are the back bone of modern high performance computing, and our work above demonstrates the power of these kernels for high performance geometry processing. In a sequence of papers we have specifically addressed sparse matrix-vector product (SpMV), general sparse matrix-matrix multiplication (SpGEMM), sparse matrix assembly, and dynamic graph data structures, and substantially improved the performance of these key operations (2xHPEC’17 (Best Paper Award), 2xICS’17, SC’18, ACM PPoPP’19).

We also continued our successful work on shape deformation and mapping optimization and proposed an interactive deformation system that is driven by user specified positional constraints. Our method is based on an unconstrained minimization of isometric energies, and is guaranteed to produce smooth locally injective maps by operating within a reduced dimensional subspace of harmonic maps. A key result is a simple and fast analytic modification of the Hessian of the energy such that it is positive definite, which is crucial for the successful operation of a Newton solver. We also extended the Newton method to triangle mesh discretizations on surfaces (SGP’17, SIGAsia’17, CGF’18).

In other works, we investigated the design and optimization of space structures and path planning with divergence-based distance functions (SIG’17, SGP’18, CAGD’18).

**Computational Videography**

Advanced computational videography aims to provide algorithms and tools for the realistic editing of live videos. We developed *Deep Video Portraits*, the first approach that enables photo-realistic re-animation of the entire posture and face expression of a human in a portrait video using only an input color video. The core of the approach is a generative adversarial neural network with a space-time architecture that converts computer graphics renderings of a (reconstructed) face model into photo-realistic and temporally-coherent video. We also addressed the gaze-aware reenactment of stereo video content for VR telepresence applications (*Face VR*). The developed approach makes it possible to computationally remove the VR headset from a stereo face video of a person, while reproducing the full face expression and eye gaze “behind” the display. We also developed a novel learning-based approach for the animation and reenactment of human actor videos (SIG’18, TOG’18, TOG’19).

In a second line of research we developed new advanced approaches for inverse rendering,
i.e., estimation of lighting and reflectance from monocular color and depth video. Examples are the first real-time method for BRDF estimation from monocular RGB, as well as the first real-time approach for live user-guided intrinsic video (TVCG’17, CVPR’18).

Reconstructing the Static and Dynamic World

We advanced the state-of-the-art of both static and dynamic scene reconstruction methods along several important dimensions: the generality of scenes that can be handled, the accuracy and quality of the reconstruction, the efficiency and robustness of reconstruction, and the simplicity of sensors needed for capturing. Notably, we researched entirely new ways of fusing and deeply integrating model-based and deep learning-based scene reconstruction. This paves the way for a new generation of future real world reconstruction, perception and understanding methods that can learn and continuously refine their internal algorithmic concepts and model representations on even sparsely labeled or unlabeled real world examples.

For static scene reconstruction we developed BundleFusion, the first approach for globally-consistent online reconstruction of large scale scenes with an RGB-D camera. Global consistency in the captured geometry is achieved by real-time bundle adjustment on a truncated signed distance field that is dynamically updated and effectively stored in space and time. We also developed CurveFusion to scan very thin and tubular structures, and we proposed a new algorithm for point cloud denoising (EG STAR’18, TOG’17, SigAsia’18, 3DV’18).

We greatly advanced the capabilities of marker-less human performance capture methods, as well as methods capturing general deformable objects with cameras. Our MonoPerfcap algorithm is the first marker-less approach for temporally coherent 3D performance capture of a human with general clothing from monocular RGB video. We tackle this challenging problem by using a novel approach that employs sparse 2D and 3D human pose detections from a convolutional neural network using a batch-based pose estimation strategy and then refining the surface geometry based on fully automatically extracted silhouettes. While MonoPerfcap is an off-line method, LiveCap is the first template-based real-time human performance capture approach that reconstructs dense, space-time coherent deforming geometry of entire humans in general clothing from just a single RGB video. We also further advanced performance capture from multi-view video (3DV’18, CVPR’18, TOG’18, TOG’19, IJCV’17).

We proposed new methods for substantially improved marker-less 3D skeletal motion capture, without surface reconstruction, from video. We developed VNect, the first method for accurate and real time 3D skeletal motion capture from monocular RGB video. It combines a learning based approach with generative fitting of a kinematic skeleton model. The system uses a novel fully-convolutional 3D pose estimation approach which strongly couples 3D pose inference of each body joint to its image evidence. We also proposed the first real-time approach for the egocentric estimation of 3D human body pose in a wide range of unconstrained everyday activities (PAMI’17, 3DV’17, 3DV’18, SIG’17, IEEE VR’19).

Reconstruction of high-quality 3D models of moving human faces is essential for the creation of digital human actors. We proposed MoFA, the first unsupervised approach for monocular reconstruction of the identity geometry, albedo texture, face expression, and scene illumination from a single color image. This was made possible by a new model-based face autoencoder which tightly integrates in a new way a 3D parametric face model, differentiable image formation and layers of a CNN. The algorithm can be trained on unlabeled community
image data. In later follow-up work we substantially extended and improved the approach. A different approach to monocular reconstruction of 3D faces is presented in our InverseFaceNet algorithm (ICCV’17, PAMI’18, CVPR’18, SIG’18, EGSTAR’18, CommACM’19).

We have developed state-of-the-art techniques to track hands, in real-time without markers or gloves, and using only a single RGB-D or color camera. The deep methodological integration of model-based and learning-based reconstruction also enable some of the first methods that handle non-trivial hand-object interactions in cluttered scenes (ICCV’17, CVPR’18).

We also investigated foundational algorithmic questions that are of cross-cutting relevance to key aspects of both expert designed and learned image-based reconstruction methods in particular, and visual computing or machine learning algorithms in general, including scalable multi-matching (TOG’17, CVPR’18, CoRR’18).

Realistic and Real-time Rendering

During the reporting period we have focused on the simulation of distribution effects and on the analysis and synthesis of sample correlations.

Distribution effects like depth-of-field and motion blur are an important factor to cinematic quality in synthetic images. Drawing point spread functions (PSF) for every pixel is a general algorithm for achieving high-quality distribution effects. While this is usually an order of magnitude faster than classical Monte Carlo (MC) methods, it is still not suitable for applications requiring interactive feedback. We proposed a novel approach, based on a sparse representation of the Laplacian of the PSFs, that significantly accelerates splitting of PSFs, and we suggested a novel efficient framework to produce complex distribution effects by exploiting coherency among RGB-D images rendered via pin-hole cameras (CGF’17, SIG’18).

Real-time renderers are paving their way towards full global illumination pipelines (e.g., in the game industry). However, much is left to understand regarding sample correlations that directly affect the quality of the images (with aliasing or noise) irrespective of the underlying application. We have achieved significant progress on MC based sampling techniques that are critical in approximating high dimensional light integrals representing radiant light energy. These works establish sound theoretical formulations that can represent error in closed-form and propose the first deep learning framework that can synthesize samples with good properties (low discrepancy, blue noise spectra with anisotropy and progressivity) without any special hand-crafting or mathematical intricacies (SIG’17, CGF’19, CGF’19, CoRR’18).

Other works address novel view synthesis (PAMI’17), expressive single scattering for light shaft stylization (TVCG’17), deep shading (EGSR’17), and CPU vector graphics rendering (Eurographics’19). For our previous work on stackless KD-tree traversal for GPU ray tracing we received the High Performance Graphics (HPG) Impact Award 2017.

Perception and Advanced Displays

Modeling the dynamics of perceptual mechanisms in the human visual system (HVS) offers great promise for designing novel and advanced display systems which can attain much higher visual quality and viewing comfort. We introduced an award-winning advanced near-eye
display (NED) design, and we developed novel fundamental algorithms for perception-aware gaze-contingent rendering and for the automatic generation of high-quality 3D content. Finally, we developed a novel CNN-based metric for light field quality evaluation which outperforms existing solutions.

We improved the state-of-the-art in displaying high-quality visuals by introducing an advanced NED design, which addresses the lack of accommodative cues in conventional displays. Enabling accommodative cues is especially critical in AR applications where computer-generated overlays must be properly combined with real world stimuli. Our display prototype, which has been designed in collaboration with UNC, essentially solves this problem with a novel deformable membrane half-mirror setup. The membranes are installed on air-tight chambers and the optical principle behind our design is approximating spherical concave surface whose curvature is controlled by changing the air pressure inside the chamber (changing the virtual depth level). The design provides a wide field of view, and the approach successfully resolves the well-known mergence-accommodation conflict. Our design received the Best Paper Award at IEEE VR’17, and the Emerging Technologies DCEXPO Special Prize at SIG’17.

Gaze-contingent rendering in combination with the use of eye tracking focuses the rendering budget on those areas where it is most needed. The main bottleneck is the limited sampling and refresh rates of affordable eye trackers and displays, resulting in an increased latency, visible as disturbing temporal artifacts. We developed a model for predicting the landing positions of saccade eye movements that solves this problem (CGF’17, SIG’17, PACMCGIT’18, TVCG’19). Moreover, we developed novel fundamental algorithms for the automatic generation of high-quality 3D content, based on Mono2Stereo and Stereo2Multiview Image conversion (SIG’17, TMM’18, TVCG’18).

As light fields are gaining in popularity, there is a need to measure and quantify light field quality in the processing pipeline. We collected an extensive dataset of reference and distorted image pairs together with user markings indicating whether distortions are visible or not. Using this dataset, we developed a novel CNN-based metric for light field quality evaluation which outperforms existing solutions (TOG’18, CVPR’19).

**Computational Fabrication**

Our interest is in the visual appearance of objects. Similar to 2D printers, multimaterial 3D printers are also binary devices where only a single material with a fixed concentration can be deposited in every volume element. Our “contoning” approach eliminates previous halftoning artifacts by printing layers of transparent inks around an opaque body made of white, diffuse material and mixing the colors within the volume of the object. Capturing and fabricating the object’s spectral reflectance in the visible range of light is called spectral reproduction. We focused on the faithful reproduction of paintings and used a deep neural network to both predict the spectrum of a printed ink stack and optimize for the stack layout that best matches a target spectrum. Current printing materials for inkjet 3D printing suffer from unwanted translucency that results in volumetric light transport (cross-talk) between surface points and severe blurring of details. We employed a general Monte-Carlo simulation of heterogeneous scattering that predicts the color at each location on the surface (SCF’17, SIG’17, SIGAsia’17, SIGAsia’18).
In other works we also addressed the perception-aware modeling and fabrication of drawing tools and the design and analysis of directional front projection screens (SCF’17, SIG’18).

Software and Datasets

As part of our research we have developed several libraries, tools, and large corpora of reference data sets that are being made available to the research community at large. These include the GVVPfPerfCapEva repository of human shape and performance capture datasets, the PFSTOOLS for processing high dynamic range images and video, the LocVis dataset of locally annotated images, and the MPI light field archive. My former PhD students N. Hasler, C. Stoll and M. Richter, together with C. Theobalt, also formed the startup The Captury, a spin-off company commercializing marker-less motion and performance capture.

Some Performance Indicators

The group continues to be a leader in computer graphics, with a specific focus on the integrated view of 3D image analysis and synthesis. Our work is being referenced widely, and we are among the trend-setters in this direction.

During the reporting period we have again made significant progress in our research along the lines above, and members of the group have actively published in the top conferences and journals in the field. As a syntactic indicator, within the two-year reporting period from spring 2017 to spring 2019, members of the group published 18 papers at SIGGRAPH/SIGGRAPH Asia/ACM TOG, 13 papers at Eurographics/CGF/TVCG, 19 papers at ICCV/ECCV/CVPR/3DV/ICCV/PAMI, and 4 papers at ICS/SC (see Section 38.15 for details). We have actively participated in the program committees of major conferences (SIGGRAPH, SIGGRAPH Asia, Eurographics, Pacific Graphics, SGP, EGSER, SCA, HPG, Graphics Interface, CAD/Graphics, VMV, ICCV, ECCV, CVPR, 3DV, GCPR), and we hold editorial board seats with journals such as ACM TOG, ACM TAP, IEEE TVCG, IEEE CG&A, CAGD, GMOD, Int. J. Shape Modeling, and Visual Informatics. We have given numerous invited talks and tutorial presentations at major national and international events (see Section 38.11.3 for details). Our software and datasets have been successfully used in a variety of projects (see Section 38.10), and many young researchers from the group have spread out to other institutions.

Projects and Cooperations

The group has been cooperating with a wide range of research groups worldwide. Cooperations that have led to joint publications during the reporting period include Stanford, MIT, Princeton, UNC Chapel Hill, U. Hong Kong, U. Tokyo, Aalto U., Bar-Ilana U., U. Bath, Cambridge U., CNRS, TU Delft, U. Edinburgh, U. Erlangen, Fraunhofer IIS, IST Austria, TU Graz, Technion Haifa, UCL London, EPFL Lausanne, USI Lugano, Charles U. Prague, U. Zaragoza, ETH Zürich. In addition, we also collaborated with some leading industrial research labs, including Google, Microsoft, Oculus, Intel, Nvidia, Disney, Adobe and Technicolor. Formal cooperations exist with Stanford (Max Planck Center for Visual
Computing and Communication (MPC-VCC)), Fraunhofer IIS (Perceptually-aware light field capture, processing and display), and the RISE Research Centre of Excellence in Cyprus. We also participate in the Horizon 2020 Training Networks DISTRO and RealVision. There are also several collaborations within the institute and with other groups on campus. For brevity we just mention the Cluster of Excellence MMCI and the Saarbrücken Graduate School of Computer Science.

**Awards**

Hans-Peter Seidel received the Eurographics Gold Medal (2017), and the High-Performance Graphics (HPG) Impact Award (2017). Christian Theobalt received an ERC Consolidator Grant (2017), and he was awarded the prestigious Karl Heinz Beckurts Prize (2017). Piotr Didyk received an ERC Starting Grant (2018). Karol Myszkowski was awarded the 2016 IEEE TVCG Best Associate Editor Award (announced Dec. 2017).

Pablo Garrido (2018) and Petr Kellnhofer (2018) both won a Eurographics Dissertation Award. Petr also won the Eduard Martin Dissertation Prize of Saarland University (2017). Franziska Müller received a Google PhD Fellowship (2017), and the national Women STEM Award (2017). Marc Habermann received the Günther Hotz Medal of Saarland University for the best Master students in computer science (2017). Petr Kellnhofer, Piotr Didyk, and Karol Myszkowski (IEEE VR’17 and SIGGRAPH’17 Emerging Technologies DCEXPO Special Prize), Thomas Leimkühler (EGSR’17), Franziska Müller and Christian Theobalt (ACM ISS’18), and Rhaleb Zayer (IEEE HPEC’17) all won best paper awards at major international conferences.
6 D5: Databases and Information Systems

History and Group Organization

D5 was established in 2003. It is headed by Gerhard Weikum and, as of May 2019, consists of 21 doctoral students, 5 senior researchers (Klaus Berberich, Simon Razniewski, Rishiraj Saha Roy, Jilles Vreeken and Andrew Yates) and 6 other post-doctoral researchers. During the reporting period 2017–2019, three senior researchers (Pauli Miettinen, Daria Stepanova, Jannik Strötgen) left the group for faculty(-like) positions at universities or research labs.

Scientific Vision and Research Areas

The group’s general objective is to develop methods for knowledge discovery in a broad sense: extracting, organizing, searching and exploring various kinds of knowledge from structured, semistructured, textual and multimodal information sources. Our approach towards this goal combines concepts, models, and algorithms from several fields, including database systems, information retrieval, natural language processing, web science and data mining.

Our research has been driven by the overarching vision of automatically constructing, growing and curating large-scale and high-quality knowledge bases from Internet sources. We have spearheaded this research avenue and refer to it as knowledge harvesting. To boost search, data analysis and language understanding, machines need to be equipped with comprehensive knowledge about the world’s entities, their semantic properties and their relationships among each other. In addition to such encyclopedic facts, machines should also have commonsense knowledge about properties of everyday objects (e.g., coffee being black, liquid and hot) and human activities (e.g., pouring milk in coffee or the visit of an evaluation panel), and should even capture socio-cultural contexts of propositions (e.g., varying beliefs about who invented the computer).

For illustration, envision a next-generation social network with “cyber-human” agents as additional participants and augmented-reality facilities for human users. The agents should behave similarly to their human peers. They will understand discourse context, situated language, facial expressions, gestures, emotions and actions of users, and harness rich world knowledge to infer the users’ intentions and anticipate their behavior. With these abilities, a software agent could be the host in a talk show with human-like behavior, or join in on a chat about movies. The agent will be able to answer questions about murders, suspects and motives in a crime movie, or discuss public protests and their underlying political controversies when watching a news clip.

Our research is currently organized into five technical areas:

– Knowledge Base Construction and Curation
– Data Mining and Exploratory Data Analysis
Enhancing computers with “machine knowledge” that can power intelligent applications has been a long-standing goal of computer science (going back to AI pioneers like Feigenbaum and Lenat). Major advances in knowledge harvesting, with our group as a trendsetter, have made this formerly elusive vision practically viable today.

Our work on knowledge harvesting was motivated by the objective of semantic search, starting in 2004. Later it became the Yago-Naga project, with the first release of the Yago knowledge base (yago-knowledge.org) in February 2007. The unique strength of Yago is its high-quality type system with hundred thousands of classes. When IBM Watson won the Jeopardy quiz show, it leveraged Yago’s knowledge of fine-grained entity types for semantic type checking.

Impact: Knowledge harvesting has been adopted at big industrial stakeholders, and knowledge bases have become a key asset in semantic search (for queries about entities), question answering, analytics (e.g., aggregating by entities), recommendations and data integration (i.e., to combine heterogeneous datasets). Examples are the knowledge graphs for search engines (e.g., Google, Bing, Baidu) and social networks (e.g., Facebook) as well as domain-specific knowledge bases (e.g., Amazon, Bloomberg, Mayo Clinic, Walmart). In addition, knowledge bases have found wide use as a distant supervision source for a variety of tasks in natural language processing. Our Yago-Naga project has served as a blueprint for many of these follow-up endeavors. The original Yago paper at WWW 2007 has nearly 3000 citations, and the Yago2 paper from 2013 already has more than 900 citations. The influence and value of Yago has been recognized by the research community through the AIJ Influential Paper Award 2017\(^1\) (for the 2013 Yago2 paper in the Artificial Intelligence Journal) and the W3C Seoul Test of Time Award 2018\(^2\) (for the original WWW 2007 paper on Yago).

High-Risk Research: The Yago-Naga project has been a high-risk (and high-gain) endeavor. In the first few years, hardly any of our colleagues believed that large knowledge bases would become viable and make impact. Some dimensions of this research theme did not work out, though. We started with the goal of developing a search engine that understands semantic concepts for computing precise and concise answers to sophisticated queries. This branch of our research led to insights and novel methods, influential publications and advanced prototypes like Bingo!, TopX and Naga, but did not succeed in building a full-scale system that could be deployed for Internet search. We also had the ambition to build the envisioned search engine in a completely decentralized manner as a peer-to-peer system. On this theme

\(^1\)http://aij.ijcai.org/aij-awards-list-of-previous-winners
\(^2\)https://www.iw3c2.org/updates/ToT/
as well, we were very successful in terms of insight and publications, but our advanced prototype system, Minerva, did not make the practical impact that we had aimed for.

**Further Highlights:** The Yago-Naga theme also spun off side projects which have been very influential. One of these is the RDF-3X database engine for efficient storage and querying of subject-predicate-object triples, primarily developed by Thomas Neumann who is now a professor at TU Munich and has received the VLDB Early Career Award 2014 for this work. The four main papers on this research together have more than 1200 citations, and the RDF-3X system is widely used in the Semantic Web community. Another contribution with high impact is the AIDA method and software for Named Entity Recognition and Disambiguation. Here, the knowledge base is leveraged as a background asset for better language understanding and text analytics. The EMNLP 2011 paper has about 700 citations. This work has spawned a startup called Ambiverse and the development of the ambiverseNLU software suite, available as open source code.

**Contributions and Impact: Major Results 2017–2019**

We publish our results in top-tier conferences in several communities: Web research (Web), data mining (DM), database systems (DB), information retrieval (IR) and natural language processing (NLP). In the two-year timeframe 2017–2019, the group had 17 full papers in first-rate Web venues (WWW, WSDM, ISWC, ICWSM, TWEB, IJCAI, AAAI), 12 in DM (KDD, ICDM, SDM, TKDD, NIPS), 5 in DB (VLDB, ICDE, EDBT, TODS), 6 in IR (SIGIR, CIKM, ECIR), and 9 in NLP (ACL, EMNLP, COLING, NAACL). In addition, we are successful in building prototype systems and publishing demo papers at top venues like ACL, VLDB, SIGIR, WWW, etc. Three of these publications won best paper awards at top-tier conferences (EMNLP 2017, WWW 2018, ICDM 2018).

**Rule Mining from Knowledge Bases and Web Text:** We successfully continued our research on learning rules for intensional world knowledge and applying them to the task of knowledge base curation (with publications in ISWC 2017, ISWC 2018, IJCAI 2018, WSDM 2019, WWW 2019). Most notably, we introduced a notion of (in-)completeness awareness to rule mining algorithms, and we incorporated latent embeddings to overcome gaps in the underlying knowledge base. The latter involved tapping into text and web sources, and we further enhanced this theme to compute evidence and generate explanations for doubtful statements which are either confirmed or refuted using rule-based inference.

**Discovering Dependencies in High-Dimensional Data:** We developed new methods for discovering approximate dependencies between attributes in multivariate datasets such as census data, economic statistics or materials properties. In contrast to traditional approaches of analyzing correlations, our methods can produce robust and interpretable outputs, this way explaining which attributes depend on which other attributes. Experimental results include studies with material scientists, and these colleagues (at the Fritz Haber Institute in Berlin) got truly excited about the findings and their interpretability. Publications on these
results include papers at KDD 2017 and ICDM 2018 (winning the best paper award) as well as an article in the New Journal of Physics 2017.

**Quantities in Tables, Text and Knowledge Bases:** We are among the first to address the issue of extracting and semantically organizing quantities that appear in ad-hoc tables and text sources on the Web. These are numeric expressions that denote measures with units and a reference frame, such as annual revenue of companies, battery-only range of hybrid cars, carbon footprint of data centers, or conductivity of thermoelectric materials. We have developed methods for linking noisy expressions of quantities in text and tables to a knowledge base (published in CIKM 2016), and for aligning approximate and aggregate quantities between text and tables in rich-content reports to support analysts in exploring details and spotting explanations of numeric facts (ICDE 2019 and WWW 2019). As quantities are often part of higher-arity relations (e.g., dosages of drugs for specific patient groups with specific disorders), we devised new information extraction methods that go beyond the traditional binary-relation model (WWW 2018 best paper). We also developed specific methods for mining counting quantifiers from natural language sources, in order to assess the coverage of relations in knowledge bases and to aid knowledge completion (ACL 2017 and ISWC 2018).

**Question Answering over Knowledge Bases and Web Text:** We revived our earlier research on answering fact-oriented questions posed in natural language (incl. telegraphic queries), with both knowledge bases and text corpora as underlying data. To tackle the vocabulary mismatch between phrases in questions and the entities, types, predicates and concepts in a knowledge base, we enhanced template-based methods for translating questions into structured queries with “never-ending learning” from user feedback (incl. pseudo-relevance) and semantic similarities among questions. Recently, we also devised a novel approach to question answering based on dynamically building quasi knowledge graphs from Web content and using advanced graph algorithms (Group Steiner Trees) to compute answers. This method is completely unsupervised, can handle a wide variety of user inputs (incl. ungrammatical utterances), and can harness both text and structured data and knowledge in a unified manner. Publications on this research theme include papers at WWW 2017, EMNLP 2017, VLDB 2018, WWW 2018, NAACL 2019 and SIGIR 2019.

**Credibility and Trust of Claims:** In the context of the ERC Synergy Grant imPACT, we address the issue of trust in online information, particularly, aiming to analyze and assess the credibility of textual statements in news, discussion forums and other kinds of online communities. This research led to the doctoral thesis of Subhabrata Mukherjee which won the runner-up position for the KDD dissertation award 2018. More recent work (with publications in WWW 2017, WWW 2018 and EMNLP 2018) focused on incorporating web evidence for estimating the credibility of claims made in news or by influential people (politicians, bloggers, partisan media etc.). We devised a suite of methods, including pipelined classifiers, probabilistic graphical models for joint inference on source trustworthiness and content credibility, all the way to deep-learning networks with tailored attention mechanisms.
Young Researchers

Our group has a strong track record in promoting young researchers in their careers. The academic offspring of D5 includes internationally visible scientists such as Mario Boley (at Monash University, Australia), Gerard de Melo (at Rutgers U, USA), Rainer Gemulla (at U Mannheim, Germany), Katja Hose (at U Aalborg, Denmark), Georgiana Ifrim (at UC Dublin, Ireland), Mouna Kacimi (at U Bolzano, Italy), Sebastian Michel (at U Kaiserslautern, Germany), Pauli Miettinen (at U Eastern Finland), Ndapa Nakashole (at UC San Diego, USA), Thomas Neumann (at TU Munich, Germany), Maya Ramanath (at IIT Delhi, India), Ralf Schenkel (at U Trier, Germany), Fabian Suchanek (at U Telecom ParisTech, France), Martin Theobald (at U Luxembourg), and further alumni who have obtained faculty positions. A good fraction of our graduates joined industrial research labs, for example, Ralitsa Angelova (Google), Asia Biega (Microsoft), Maximilian Dylla (Google), Subhabrata Mukherjee (Amazon), Josiane Parreira (Siemens), Daria Stepanova (Bosch), Niket Tandon (Allen Institute for Artificial Intelligence) and Mohamed Yahya (Bloomberg). In the two-year timeframe 2017–2019, we had 10 doctoral students graduating.

Awards

Jilles Vreeken received the IEEE ICDM Tao Li Early Career Award at ICDM 2018, the data mining community’s first award for recognizing the excellence and impact of young researchers at most ten years after graduation. Subhabrata Mukherjee received the ACM KDD Dissertation Runner-Up Award 2018.

Patrick Ernst (and his co-authors Amy Siu and Gerhard Weikum) won the Best Paper Award at the Web Conference (WWW) 2018. Panagiotis Mandros (and his co-authors Mario Boley and Jilles Vreeken) won the Best Paper Award at the ICDM 2018 Conference. Andrew Yates (and his co-authors Arman Cohan and Nazli Goharian) won the Best Paper Award at the EMNLP 2017 Conference. Dhruv Gupta and Thomas Tanon (and their respective co-authors) obtained honorable mentions for best demo and best student paper awards at WWW 2018 and ISWC 2017, respectively. Francesca Lisi and Daria Stepanova won the best poster award at the International Conference on Rules and Reasoning 2017.

The long-term impact of Yago has been recognized by the AIJ Influential Paper Award 2017 (for the 2013 Yago2 paper in the Artificial Intelligence Journal) and the W3C Seoul Test of Time Award 2018 (for the original WWW 2007 paper on Yago).

Teaching

The department is also committed to a fair amount of teaching at Saarland University, including the regularly offered core lecture on Information Retrieval and Data Mining. In addition, a total of 26 Bachelor’s and Master’s theses in the 2017–2019 timeframe were completed under the supervision of the group’s senior researchers.
Cooperations

**Institute and Max Planck Society:** Within the institute, we collaborate with D1 on efficient algorithms for large graphs and matrices, and with D2 on background knowledge for understanding visual contents. The collaboration with D2 has led to jointly authored papers in CVPR and AAAI. Within the Max Planck Society, we collaborate with Peter Druschel, Krishna Gummadi and Rupak Majumdar, all at the Max Planck Institute for Software Systems in Saarbrücken and Kaiserslautern. This has resulted in joint papers at KDD and SIGIR. In addition, our group is a member of the MaxNet research network on Big Data Driven Materials Science, which involves 9 Max Planck Institutes. In this context, members of our group published joint work with the Fritz Haber Institute in the New Journal of Physics.

**External Partners and Competitive Grants:** D5 participates in the DFG-funded Excellence Cluster on Multimodal Computing and Interaction. Gerhard Weikum is one of the principal investigators and responsible for the research area on knowledge management. The Excellence Cluster is a framework for extensive cooperation on the Saarland University campus. We have collaborated with other groups in computational linguistics and in human-computer interfaces, with joint publications in premier venues. We also participate in the DFG-funded Collaborative Research Center (Sonderforschungsbereich) on Methods and Tools for Understanding and Controlling Privacy, where we collaborate with Michael Backes from the CISPA Helmholtz Center and Jens Dittrich from Saarland University. There are various individual collaborations with researchers at other universities and research labs across the world (including U Aberdeen, Allen Institute for AI, Amazon Research, VU Amsterdam, NU Athens, Bosch Center for AI, Northeastern U Boston, U Cyprus, IIIT Delhi, FORTH Heraklion, Georgetown U, U Melbourne, U ParisTech, Southwest Jiaotong U, Volkswagen Research, Warwick U, U Washington, etc.).

The single most important collaboration is the ERC Synergy Grant 610150 on Privacy, Accountability, Compliance and Trust for Tomorrow’s Internet (imPACT), with Michael Backes, Peter Druschel, Rupak Majumdar and Gerhard Weikum as principal investigators. This ERC grant is one of the highest reputed scientific prizes in Europe; it provides the four PIs with a total budget of 10 Million Euros for the timeframe 2015–2020. Our research agenda aims at fundamental insight on reconciling the tensions between the four PACT properties (privacy, accountability, compliance, trust). This includes the goal of developing game-changing methods and tools that assist users in analyzing, understanding and managing their privacy risks in social media and other kinds of digital traces.

Future Research Focus

**New Directions in Machine Knowledge:** Current knowledge bases focus on facts about entities like prominent people, places and products, and still lack other knowledge dimensions like properties of everyday objects, human activities and socio-cultural contexts. For advanced search, conversational bots, visual understanding and next-generation AI, computers need to be equipped with broader kinds of knowledge: quantitative modifiers of facts, plausibility
invariants, properties of everyday objects, human activities and spatio-temporal as well as socio-cultural context.

**Responsible Data Science:** With the amazing advances of AI and its strong role in data science across all fields, managing this powerful technology in a responsible manner often seems to be a mere afterthought. We plan to give this neglected theme its due prominence by research on various aspects of trustworthy and responsible computing. AI systems in human-centric applications must prevent obscured, biased and unfair consequences in real-life situations. Examples where this is of utmost importance are when data-driven machine learning is used for algorithmic decision making, such as automation of loan requests, visa approvals, insurance claims, job interviews, all the way to AI components for e-government and political decisions. It is crucial that such systems be trustworthy and transparent and have explainable behavior, and that they deal with data about people in an unbiased and fair manner.
7 RG1: Automation of Logic

History

The Automation of Logic Group has been established in September 2005 and is headed by Christoph Weidenbach. The group covers the complete pipeline from basic research on (new) logics and their automation up to applications in research and industry.

There are currently 5 researchers and 4 PhD students in the group. Ching Hoo Tang, Andreas Teucke and Daniel Wand have finished their PhD theses during the reporting period and left the group. Ching Hoo Tang and Andreas Teucke joined the group’s spin off L4B, while Daniel Wand accepted an offer by Bosch. Marco Voigt has submitted his PhD thesis and Martin Bromberger and Mathias Fleury will submit by the end of 2019. Thomas Sturm and Jasmin Blanchette have appointments at CNRS and VU Amsterdam, respectively, but remain associated with the group.

Thomas Sturm, working in computer algebra, Jasmin Blanchette, working in interactive theorem proving and its mechanization are, together with Christoph Weidenbach, the senior researchers of the group. Sophie Tourret and Uwe Waldmann work on higher-order logic automation.

Vision and Research Strategy and Achievements

The vision of the group is to increase the productivity of formal analysis/verification/problem solving technology through a higher degree of automation of the underlying logics. The following challenges motivate main parts of our work: (i) drive the development of first-order, and beyond first-order reasoning calculi, (ii) understand the mechanics of reasoning with respect to concrete models, (iii) build specific reasoning procedures for arithmetic theories, (iv) show applicability of our methods to reasoning challenges from other areas, (v) formally verify our own results, and (vi) scale the applicability of our methods to the size of real world industrial applications.

About half of our work is of a theoretical nature and the other half is experimental, in particular on the basis of developed tools. Implementation of our methods is important to check their automation potential in practice, to increase the impact of our theoretical results by providing our software to other people and eventually to detect further challenges in theory development. For example, the results on sufficient tests for the satisfiability LIA (Linear Integer Arithmetic) constraints, Section 40.4.3, have been implemented in SPASS-SATT, Section 40.8.5, winning the QF_LIA category of SMT-COMP 2018; the development of subtropical methods for real constraints, Section 40.4.2, and its implementation in the computer algebra system Redlog yielded new insights in bioinformatics. In 2012 we founded the company “Logic for Business” (L4B) providing consulting and software for the overall
life-cycle management of complex products.

We have structured our research along the following topics: (i) Foundations of Automated Reasoning, (ii) (First-Order) Arithmetic, (iii) Towards Higher-Order Automated Reasoning, (iv) Formalizing Logic, (v) Logic for Machine Learning, (vi) Software. The structure does not impose a structure on the group. In fact, most of us contribute to several areas and almost all of us to the development of software.

A selection of scientific results for topics (i), (ii), (iii), and (iv), respectively, are:

**New Decidable Fragments of First-Order Logic (JLC 2017, LICS 2017, CADE 2017).** The idea of separated variables, already reported during the previous reporting period, has shown further potential. It yields strict, decidable extensions of the monadic fragment, the Bernays-Schönfinkel fragment, the two-variable fragment, the Ackermann fragment, the Gödel-Kalmár-Schütte fragment, the fluted fragment, the guarded fragment, the loosely guarded fragment, and the guarded negation fragment. We have also extended decidability results for two clause fragments towards recursive predicates, and non-Horn clauses, respectively.

**Fast Decision Procedures for LIA Constraint Solving (FMSD 2017, IJCAR 2018).** We have continued our research on fast decision procedures for linear arithmetic in the context of automated reasoning. Through the so-called *Double-Bounded-Reduction* and afterwards applied *Mixed-Echelon-Hermite* transformation, any unbounded linear arithmetic problem can be transformed in polynomial time into a fully bounded problem. Solutions of the resulting problem can be translated in polynomial time into solutions of the original problem.

**Towards Higher-Order Logic Automated Theorem Proving (CADE 2017, FSCD 2017, IJCAR 2018).** Lifting superposition-based first-order reasoning to higher-order logics while keeping the syntactic restrictions of the calculus poses several challenges. We have defined suitable orderings and lifted the calculus to a lambda-free higher-order logic.

**Subtropical Methods for Real Constraints (FROCOS 2017, CASC 2018).** We had developed *subtropical methods* for heuristically checking real constraints in the first orthant during the previous reporting period. We generalized this method from a single to finitely many ordering constraints. The resulting method performs surprisingly well both in the automated reasoning (SMT) and the bioinformatics context.

**Formalizing Automated Reasoning Calculi in Logic (IJCAI 2017, FROCOS 2017, CPP 2018, JAR 2018, POPL 2019, IJCAR 2018, CPP 2019).** We continue our “eat your own dog food” line of research on the formalization of (our developed) logical calculi in Isabelle and by this way testing our enhancements to Isabelle. In the current reporting period we formalized further variants of CDCL (Conflict-Driven Clause Learning) for SAT and the ordered resolution prover from Bachmair and Ganzinger’s article in the Handbook of Automated Reasoning. Furthermore, we added to (Co)datatypes, (Co)recursion and Binders of the Isabelle/HOL implementation.
Projects and Cooperations

Together with Stephan Merz’s group (Inria Nancy) we constitute the Inria project VeriDis (http://www.inria.fr/en/teams/veridis), where we investigate automated reasoning support for the verification of distributed algorithms. The project is accompanied by the EU FET Open action SC-Square on bringing together automated reasoning and computer algebra research, led by Thomas Sturm. In the transregional collaborative research center TRR 248, established in January 2019, we investigate automated reasoning in the context of perspicuous software systems in close cooperation with researchers from TU Dresden.
8 Zeynep Akata: Multimodal Deep Learning

Group Development

The group \textit{Multimodal Deep Learning} was established in the department of Computer Vision and Multimodal Computing in 03/2017. In addition to the group head, Zeynep Akata, it currently consists of one PhD student Yongqin Xian and one more PhD student Moritz Böhle will join in January 2018.

Vision and Research Strategy

The group’s goal is to develop deep learning methods that observe and process multimodal input coming from the environment, make further connections through inference and communicate the system output to the user. The group’s current research focus is exploring the interplay between vision and language for several tasks. The group’s research is typically based on a three step approach. The first step is to identify and define the task. The main questions to answer in this step are: What is the task trying to accomplish, has this task been defined before and why is this task interesting and necessary? The second step is to designing the framework to accomplish the defined task. The main questions to answer are: What would be the components of the framework, what kind of tools exist previously that we can use to build over and what would be the steps of our proposed model? The third step is to implement the designed framework, experiment it and make improvements over time. The main questions to answer are: Is there code online that is useful for us available, is there a dataset that is useful for us or do we need to collect our own data and are the results expected, interesting, do they improve the previous state of the art, if not what can we be done to improve it?

Research Areas and Achievements

The research of this group is mainly at the intersection of vision and language for them being complementary sources of information. It is broadly divided into two subfields: \textit{low-shot learning} and \textit{explainable artificial intelligence}.

\textbf{Low-Shot Learning.} Low-shot learning can be further divided into zero-shot learning and few-shot learning. Zero-shot learning for image classification involves learning a model about the environment given a set of observations that belongs to a certain set of classes. The set of classes at training and at test time are disjoint. As, the classic supervised learning algorithms that rely on the full set of class labels can not be employed for the zero-shot learning task, the group uses language as side information to build a structure in the label
space. In few-shot learning the task is to recognize classes for which only a few labeled samples are provided. In both low-shot learning tasks, the group develops generative models using side information as a data augmentation mechanism. Many approaches in generalized zero-shot learning rely on cross-modal mapping between the image feature space and the class embedding space. As labeled images are rare, one direction is to augment the dataset by generating either images or image features. However, the former misses fine-grained details and the latter requires learning a mapping associated with class embeddings. The group develops models that combine multiple approaches to improve this idea. This work has been published in:

- 2 x CVPR 2017 (full), CVPR 2018 (full), TPAMI 2018 (full)

**Explainable Artificial Intelligence.** Deep models are frequently seen as opaque and are unable to explain their decisions. In contrast, humans can justify their decisions with natural language and point to the evidence in the visual world which led to their decisions. However, existing models which generate textual explanations enforce task relevance through a discriminative term loss function, but such mechanisms only weakly constrain mentioned object parts to actually be present in the image. As deep learning methods got deeper and more expressive, several attempts have been made to visualize and describe the internal decision process of a network. One aspect of explainable AI that has been gaining momentum is generating visual and textual explanations of a classification decision. Image explanation through text differs from the traditional sense of captioning as the generated sentences are not only image specific but also they are class discriminative. The group focuses on generating visual explanations and pointing to the evidence for a classification decision of a deep multimodal learning framework.

- CVPR 2018 (full), 2 x ECCV 2018 (full)

**Projects and Cooperations**

Within the campus the group closely collaborates with Prof. Bernt Schiele’s group *Computer Vision and Multimodal Computing*, Mario Fritz’s group *Multimodal Learning and Interaction*, and Andreas Bulling’s group *Perceptual User Interfaces*.

Externally the group collaborates with Prof. Trevor Darrell’s group (University of California Berkeley), Prof. Max Welling’s group (University of Amsterdam), Prof. Honglak Lee’s group (University of Michigan, Ann Arbor).

**Prizes and Awards**

Zeynep Akata started work as a tenure-track assistant professor position at the University of Amsterdam in 2017. Yongqin Xian was awarded the European Finalist of Qualcomm Innovation Fellowship in 2017.
9 Bjoern Andres: Combinatorial Image Analysis

Group Development

The research group Combinatorial Image Analysis was established in September 2013. Its current members are Bjoern Andres and the doctoral students Jan-Hendrik Lange and Evgeny Levinkov. During the reporting period, Bjoern Andres has accepted an offer of an Honorary Professorship by the University of Tübingen where he is building a research group for the company Bosch. Bjoern Andres continues to supervise Evgeny Levinkov and Jan-Hendrik Lange at the Max Planck Institute for Informatics, with support from Paul Swoboda. Both doctoral students are successful in their research and on track to completing their doctorates. Evgeny Levinkov is scheduled to defend his dissertation in April this year. Jan-Hendrik Lange has published four conference articles during the reporting period and submitted one already on joint work with Bjoern Andres and Paul Swoboda. Andrea Hornakova, previously a doctoral student in the Combinatorial Image Analysis group, has moved to the Combinatorial Computer Vision group.

Vision and Research Strategy

Algorithms for image analysis are beginning to have major impact on our society, e.g., in the area of computer vision for autonomous driving, and in the areas of image and data analysis for biomedical research and applications. One path to further improving existing algorithms is to advance the mathematical abstractions of real-world tasks solved by these algorithms. The research group Combinatorial Image Analysis studies mathematical abstractions of image analysis tasks in the form of combinatorial optimization problems. It develops algorithms for solving these problems exactly, partially or locally. It analyzes these algorithms and the solutions they output in terms of application-specific metrics and benchmarks. The goal of this research is to enable practical applications of expressive combinatorial optimization problems in the fields of computer vision and biomedical image and data analysis.

Research Areas and Achievements

Correlation Clustering During the reporting period, the group’s research has been driven by the realization that seemingly unrelated computer vision tasks, including instance-separating semantic image segmentation, multiple object tracking and human body pose estimation, are fundamentally linked by a single combinatorial optimization problem, a generalization of the well-known correlation clustering problem. Feasible solutions of the generalized problem...
define both a decomposition and a node labeling of a graph. The analysis of this problem, a generalization of the well-known correlation clustering problem, as well as the design of practical algorithms for computing feasible solutions have been the primary focus of the group. Firstly, we have studied the problem theoretically, establishing inefficiency results in the research area of clustering. Secondly, we have demonstrated the impact the problem empirically in its application to instance-separating semantic image segmentation, multiple object tracking and human body pose estimation. Thirdly, we have contributed algorithms for solving parts of the problem exactly and efficiently in cases where a total solution is impractical. Our work has been published in ICML (2x) and CVPR (3x).

**Biomedical Image and Data Analysis** A second focus of our research has been in the field of biomedical image and data analysis. In particular, we have improved our algorithms for lineage tracing, the tracking of living cells as they move and divide in a sequence of microscopy images. Lineage tracing is a powerful tool in the field of developmental biology, e.g., for determining how the structure of living tissue emerges, and for quantifying the dynamics of cancer cells. During the reporting period, we have improved our integer programming algorithm for the moral lineage tracing problem such that it finds accurate solutions on published instances and scales to larger instances, leveraging moral lineage tracing to practical significance. We have demonstrated the efficacy of the algorithm in an application to sequences of images acquired by the technique of lens-free microscopy. Our work has been published in Medical Image Analysis, Bioinformatics, ICCV and MICCAI.

**Semi-supervised Learning** A third and minor focus of the group has been on semi-supervised learning. In collaboration with TU Munich, we have contributed algorithms for learning the coefficients of combinatorial optimization problems in a semi-supervised fashion. Our work has been published in CVPR.

**Projects and Cooperations**

Within the department of Computer Vision and Multimodal Computing, we have worked closely with Bernt Schiele’s group on the problems of multiple object tracking and human body pose estimation.

Across departments of the Max Planck Institute for Informatics, we have worked with Andreas Karrenabuer on partial optimality conditions for the weighted correlation clustering problem.

Across institutes, we have worked with Daniel Cremers’ group at TU Munich on a problem in the field of semi-supervised machine learning.

**Prizes and Awards**

- Best Reviewer Award of the NIPS Foundation (2017)
- Winner of the CVPR Multiple Object Tracking Challenge (2017)
Group Development

The Computation, Appearance and Manufacturing group started in September 2018 funded by the Max Planck Center for Visual Computing and Communication (MPC-VCC). Currently, the team is led by Vahid Babaei and employs two master students who work as part time research assistants (Sebastian Cucerca and Navid Ansari).

Vision and Research Strategy

We are witnessing the day-to-day improvement of manufacturing devices in the era of advanced manufacturing. Advanced manufacturing bridges the gap between R&D activities and product development, boosts the economic competitiveness and creates high-quality jobs. In Computation, Appearance and Manufacturing group, we focus on inventing new computational tools that release the full potential of advanced manufacturing processes, such as additive manufacturing (also known as 3D printing). With the immense growth of the manufacturing hardware in resolution, scale and speed, the algorithm complexity increases even more dramatically. Our group therefore aims at developing hardware-aware, scalable algorithms for advanced manufacturing.

We have a particular interest in visual appearance of objects and strive for better algorithms that help manufacturing products with novel and useful appearance characteristics. Design for manufacturing of objects with high-fidelity appearance specifications is a key engineering task. Thus, there remains a huge opportunity for revolutionizing the computational appearance design and manufacturing, and significantly improving the appearance quality of many products. The results of this research will immediately enable numerous applications in rapid prototyping and manufacturing of end-use products. This spans several application domains from medical devices and surgical training, to cultural heritage preservation and anti-counterfeiting.

Research Areas and Achievements

Multi-material 3D Appearance Printing. The hardware capable of multi-material 3D printing is very new: 3D printers capable of “full-color” printing are becoming available only now. It is, however, challenging to exploit this capability to its fullest, to perfectly reproduce a certain appearance, since this requires finding the ideal arrangement of basis materials. Our research in this axis aims at overcoming several technical challenges: (a)
The addition of a third dimension along which we can arrange materials introduces both opportunities and challenges. For example, V. Babaei recently showed how to take advantage of this property by introducing a novel ink combination strategy without the need for the status-quo halftoning methods (SIGGRAPH 2017). (b) It is crucial that we can predict the appearance of a given printed structure accurately and quickly. Prediction accuracy is the paramount requirement for many high-end applications. The simulation should also be rapid since the iterative optimization procedures for computing the multi-material structures resulting in a certain appearance rely on the forward simulation at each optimization step. (c) Multi-material 3D printing of appearance is an unexplored research area, with almost no existing standards. Today, we do not have a clear insight about the necessary “primary” materials in order to span a reasonable gamut of appearance (unlike the CMYK or RGB systems in color reproduction).

Computational Fine Art Reproduction. Fine art objects are instruments for aesthetic contemplation as well as social scientific studies. Fine art artifacts are exposed to different dangers, such as aging and destruction even though they already incur huge costs to museums for conserving them. It is essential that we learn to preserve this heritage for future generations and ourselves. We take fine art as a primary case study for our research on appearance reproduction. Beside the importance for cultural heritage preservation, work of art is an ideal case study since all elements of appearance are present: 3D texture, spectral color, gloss and translucency. We have just published a system for multispectral reproduction of oil paintings, which leverages 3D printing capabilities resulting in unprecedented spectral fidelity (SIGGRAPH ASIA 2018). In the context of fine art facsimile printing, we are currently interested in the selection of the type and concentration of printing inks, which is a challenging optimization problem.

Novel Computational Appearance Manufacturing Methods. Additive manufacturing brings many advantages with its ability to produce almost any shape and its unique multi-material capabilities. Yet some objects are impossible (wood, stone), or complicated (metal), or extremely slow (any large object) to make with additive manufacturing. We are developing computational tools for appearance fabrication using other manufacturing technologies. Particularly, we are interested in physical editing of existing products using, for example, direct ink writing, subtractive milling or laser marking.

Projects and Cooperations

We have an ongoing collaboration on volumetric prediction of multi-material 3D prints with Karol Myszkowski (MPI-INF). Externally, we collaborate with Wojciech Matusik (MIT), Szymon Rusinkiewicz (Princeton University), Piotr Didyk (USI), Bernd Bickel (IST), Tim Weyrich (UCL), and Jaroslav Krivanek and Alexander Wilkie (Charles University).
11 Klaus Berberich: Text+Time Search & Analytics

Group Development

The research group Text+Time Search & Analytics was established in May 2013. It currently consists of Klaus Berberich (head) and Dhruv Gupta (Ph.D. student). During the reporting period, Kai Hui and Arunav Mishra successfully completed their Ph.D. work, and Supratim Das successfully completed his M.Sc. thesis.

Vision and Research Strategy

Our objective is to develop efficient and effective methods to search and analyze natural language texts that come with semantic annotations. These include temporal expressions, which convey time periods a text refers to, as well as named entities, which connect the text to other data sources such as a knowledge graph (e.g., Wikidata or YAGO).

Search, as one direction, targets situations when the user’s information need is precise and can be satisfied by some piece of our data, typically a document. Analytics, as another direction, targets situations when the user’s information is best satisfied by some derived data such as aggregated statistics or a summary of multiple documents.

Research Areas and Achievements

Temporal Summarization. In this line of work, we look into how a collection of documents describing a specific real-world event can be summarized effectively. To this end, we have proposed extractive multi-document summarization methods that take into account the semantic annotations available in documents. More precisely, our methods consider coverage and diversity not only of textual terms, but also of temporal expressions, geographic locations, and other named entities. We published our work in:

- ECIR 2017 (full), ECIR 2018 (short)

Focus Time Estimation. We also look into how the time period, that a snippet of text or a fact from a knowledge graph refers to, can be identified, even if the text or fact does not include temporal information. The key idea here is to leverage background knowledge that can be obtained from semantically annotated document collections. Thus, one of our methods relies on trained embeddings for words and temporal expressions, which can then be used to identify temporal expressions that are semantically similar to the words in a text snippet. We published our work in:
Low-Cost Evaluation for Information Retrieval. When evaluating information retrieval systems, typically significant human effort is required to assess the relevance of retrieved documents. Existing work has looked into reducing the required human effort by judiciously choosing documents to assess and predicting missing relevance assessments. We address this problem with a focus on novelty & diversity where one wants to retrieve relevant documents that cover diverse query aspects. More recently, we have started to apply similar ideas to preference judgments, as an alternative way of evaluating retrieval systems. Here, one key idea is to exploit the transitivity of preferences to reduce the overall number of judgments required. We published our work in:

- WWW 2017 (poster), ECIR 2017 (short+full), ICTIR 2017 (full)

Neural Information Retrieval. Making use of deep neural networks for re-ranking documents in ad-hoc retrieval has been another active area of research during the reporting period. Our models have been designed to pick up information about the order, proximity, and context of term occurrences in retrieved documents. Previously, such signals had to be manually crafted and be integrated into retrieval models. The developed models were able to achieve substantial improvements over established retrieval models (e.g., unigram language models and OKAPI BM25) on different TREC testbeds. We published our work in:

- WWW 2017 (poster), EMNLP 2017 (full), WSDM 2018 (full)

Structured Search in Semantically Annotated Document Collections. While it is nowadays feasible to adorn even large document collections with semantic annotations, it remains an open issue how those can be leveraged for efficient and effective retrieval. One line of our research in the reporting period looked into how so-called semantic aspects, as interesting combinations of semantic annotations, can be identified automatically to help users explore the result of a query. In another line of research, we looked into how semantically annotated document collections can be queried effectively and efficiently. This included the design of a query language, inspired by regular expressions, as well as an indexing infrastructure, which relies on indexing carefully selected combinations of words and semantic annotations. We published our work in:

- JCDL 2018 (poster), CIKM 2018 (full), WSDM 2019 (demo), ESWC 2019 (full)

Projects and Cooperations

The research group maintains loose cooperations with L3S Hannover, NTNU Trondheim, and the University of Stavanger. In the past it has contributed to the European Union project Longitudinal Analytics of Web Archive data (LAWA).

Prizes and Awards

Our poster Identifying Time Intervals for Knowledge Graph Facts received an honorable mention at WWW 2018.
12 Jasmin Christian Blanchette: Interactive Theorem Proving

Group Development

The Interactive Theorem Proving subgroup of Automation of Logic was established in 2015. In addition to the group head, Jasmin Blanchette, it currently consists of one PhD student, Mathias Fleury, who works under Christoph Weidenbach’s cosupervision. One PhD student, Daniel Wand, graduated from the group during the reporting period. The group closely collaborates with Sophie Tourret and Uwe Waldmann from Automation of Logic. It also hosted Anders Schlichtkrull (DTU Copenhagen) for four months. A former MSc student from the group, Alexander Bentkamp, is employed as a PhD student at the Vrije Universiteit Amsterdam under Blanchette’s and Waldmann’s joint supervision.

Vision and Research Strategy

The group’s main goal is to bring the benefits of automatic theorem proving to interactive theorem proving and vice versa. Being embedded in the Automation of Logic team is ideal to achieve this double objective.

Proof assistants, also called interactive theorem provers, are increasingly used to verify hardware and software systems. Despite the success stories, theorem proving remains laborious and requires specialized experts. Our goal is to make proof assistants easier to use. In an ideal world, whenever the user entered a conjecture, the system would quickly produce a proof or a counterexample. In practice, we can achieve considerable success by developing suitable logical calculi and heuristics, notably for higher-order reasoning.

Conversely, we contend that the time is ripe to use proof assistants to formalize the rich metatheory of automatic theorem provers. In 2015, we instigated the IsaFoL (Isabelle Formalization of Logic) project to start formalizing classical and modern results about automated reasoning and logic in the Isabelle/HOL proof assistant.

Research Areas and Achievements

Formalization of Logic. Researchers in automated reasoning spend a significant portion of their work time specifying logical calculi and proving metatheorems about them. These proofs are typically carried out with pen and paper, which is error-prone and can be tedious. As proof assistants are becoming easier to use, it makes sense to employ them. In this spirit, we started the IsaFoL effort that aims at developing libraries and methodology for formalizing modern research in the field, using the Isabelle/HOL proof assistant. The
project has grown to include contributions from researchers and students in Amsterdam, Copenhagen, Gothenburg, Grenoble, Munich, Saarbrücken, and Zurich. Blanchette gave an invited talk at Certified Proofs and Programs (CPP) 2019 about this work.

Our initial emphasis is on established results about propositional and first-order logic. In particular, Fleury has formalized large parts of Weidenbach’s forthcoming textbook, tentatively called *Automated Reasoning: The Art of Generic Problem Solving*, notably the sections dedicated to CDCL, the proof calculus underlying modern SAT solvers. His verified SAT solver is, by far, the most efficient tool of its kind, showcasing theorem proving technology as a way to develop optimized yet trustworthy code.

In addition, Blanchette, Schlichtkrull, Tourret, and Waldmann are working on building a framework for reasoning about saturation-based provers, including superposition provers. We started with Chapter 2 of *Handbook of Automated Reasoning* but are now generalizing and simplifying this framework. We are gradually moving from carrying out case studies, where the proof assistant aspect is predominant, to developing new metatheory.

**Higher-Order Proof Automation.** To make interactive verification more cost-effective, we propose to deliver very high levels of automation to users of proof assistants by fusing and extending two lines of research: automatic and interactive theorem proving. Our starting point is that first-order automatic provers are the best tools available for performing most of the logical work. Our approach is to enrich the superposition calculus and SMT techniques—the leading proof methods in the first-order world—with higher-order reasoning in a careful manner, to preserve their desirable properties. We are designing proof rules and strategies, guided by representative benchmarks from interactive verification.

At the same time as we develop higher-order superposition and higher-order SMT on the theoretical front, are are developing prototype and real implementations of our ideas in three systems: the superposition-based provers Zipperposition (developed in Amsterdam) and E (developed in Stuttgart and Amsterdam) and the SMT solvers veriT (developed in Nancy) and CVC4 (developed in the US). In future work, we plan to integrate these provers in proof assistants, including Isabelle/HOL, Lean, and the TLA+ Proof System.

**Projects and Cooperations**

The group is closely involved in Blanchette’s ERC Starting Grant 2016 *Matryoshka: Strong Higher-Order Automation for Fast Interactive Verification*, which takes place primarily at the Vrije Universiteit Amsterdam, where Blanchette is assistant lecturer, and Inria Nancy. Pascal Fontaine and Stephan Merz at Inria Nancy provide expertise with SMT solving and the TLA+ Proof System. In addition, we have a close collaborator at DHBW Stuttgart, Stephan Schulz, whose automatic prover is one of the key software tools for Matryoshka.

**Prizes and Awards**

Blanchette was awarded an NWO Vidi personal grant in 2017 (EUR 800k) and an ERC Starting Grant in 2016 (EUR 1500k). For his PhD thesis, he was awarded in 2012 the Heinz Schwärtzel Prize for Foundations of Computer Science, jointly by three Munich universities.
13 Karl Bringmann: Fine-Grained Complexity and Algorithms

Group Development

Karl Bringmann joined MPI-INF as a post-doctoral researcher in January 2016 and was appointed senior researcher in June 2017, establishing the research group Fine-Grained Complexity and Algorithms. The group consists of the PhD students André Nusser, Bhaskar Ray Chaudhury, and Philip Wellnitz and the post-doc Marvin Künnemann. The group supervised two Bachelor theses and three research immersion labs in the reporting period.

Vision and Research Strategy

The traditional way of establishing a problem as intractable is to prove it to be NP-hard. This makes a polynomial-time algorithm unlikely, so even for medium-size instances we cannot expect to solve the problem in reasonable time, at least on worst-case instances.

However, in a modern big data world with inputs such as DNA sequences, social network graphs, or sensor network measurings, even a quadratic-time algorithm can be too slow. This shift necessitates changes in intractability theory. In order to avoid searching forever for a faster algorithm that does not exist, we need intractability tools that establish far-from-linear lower bounds. In other words, we need an analogue of NP-hardness for polynomial-time problems. Unfortunately, P vs. NP is too coarse and no techniques for proving unconditional lower bounds of this form are known.

Therefore, the modern approach is to prove conditional lower bounds. To this end, we start from a widely believed conjecture about the time complexity of a key problem, and transfer the conjectured intractability to another problem via a fine-grained reduction, yielding a conditional lower bound on how fast the other problem can be solved. The resulting area of fine-grained complexity theory had initial results in the early 90s and started to mature in the last 10 years. Karl Bringmann made significant early contributions to the field during his PhD studies, by proving the first conditional lower bound for any dynamic programming problem, namely the Fréchet distance, which paved the way for many further lower bounds for related problems, e.g., edit distance and dynamic time warping.

The group’s main agenda is to design conditionally best-possible algorithms, i.e., to design an efficient algorithm and to complement it by a conditional lower bound ruling out any significantly faster algorithms, for a variety of different problems. This uncovers the inherent time complexity of a problem (conditional on a plausible complexity-theoretic assumption) and provides strong indication to stop searching for a faster algorithm.
Research Areas and Achievements

Knapsack-type Problems  Subset Sum and Knapsack are fundamental problems in theoretical computer science and mathematical optimization. We aim at a precise understanding of the time complexity of these problems with respect to various parameters. In the reporting period, we have found a conditionally best-possible algorithm for Subset Sum (with respect to the target number as parameter). The ultimate goal is to generalize these insights to a precise understanding of various classes of integer programming problems.

Computational Geometry  Analysis of trajectory data as produced by GPS trackers is a ubiquitous task, giving rise to a variety of problems in computational geometry such as curve simplification or pattern recognition. We study these problems from the viewpoint of fine-grained complexity theory, aiming at conditionally best-possible algorithms. Additionally, we also implement our algorithms and perform algorithm engineering in order to transfer our insights to practice, in particular, in the reporting period we developed an implementation for computing the Fréchet distance which is now the state of the art.

Foundations of Fine-Grained Complexity Theory  We also study the theoretical foundations of fine-grained complexity theory. Since several conjectures have been used to prove conditional lower bounds, a driving question of the area is whether there exists a unifying conjecture from which we can prove all our lower bounds. Our group contributed to this question by proving that the weighted $k$-Clique conjecture unifies both the All Pairs Shortest Path conjecture and the Orthogonal Vectors conjecture.

Achievements.  In the reporting period the group published over 25 conference publications, most notably 5 papers at STOC/FOCS and 7 papers at SODA.

Projects and Cooperations

Within MPI-INF, Karl Bringmann is collaborating with Dr. Christian Ikenmeyer (D1, now at MPI-SWS) on the topic of algebraic complexity theory. On campus, Karl Bringmann is collaborating with Dr. Holger Dell (UdS and MMCI) in research on foundations of fine-grained complexity theory and in teaching the course “Multivariate Algorithmics” (Winter 2018/19). Off campus, the group has many widespread collaborators including Dr. Amir Abboud (IBM Almaden), Prof. Kasper Green Larsen (Aarhus University), Prof. David Woodruff (CMU), and Prof. Mikkel Thorup (University of Copenhagen).

Prizes and Awards

Karl Bringmann received an EATCS Presburger Award for Young Scientists in 2019 for outstanding contributions in theoretical computer science, and he was awarded a Heinz Maier-Leibnitz Prize in 2019. Marvin Künnemann received an Otto Hahn Medal of the Max Planck Society in 2017. André Nusser received the “Publikationspreis der Universität Stuttgart” in 2018 (together with Stefan Funke and Sabine Storandt).
14 Andreas Bulling: Perceptual User Interfaces

Group Development

Andreas Bulling joined MPI-INF in February 2013. In June 2013 he was granted an Independent Research Group Leader (W2) position and his group also became affiliated with the Cluster of Excellence on Multimodal Computing and Interaction (MMCI). In 2018, he received offers for full professorships (W3) at the University of Freiburg and the University of Stuttgart, Germany, the latter of which he accepted (started August 1, 2018). Xucong Zhang graduated (summa cum laude) in September 2018 and is now a PostDoc at ETH Zurich. Mohamed Khamis (jointly supervised with Prof. F. Alt at Bundeswehr University Munich) graduated in September 2018 and is now a Lecturer at the University of Glasgow, UK. Michael Xuelin Huang joined Google as a Research Scientist in December 2018. Julian Steil (PhD student) and Philipp Müller (PhD student) remain in Saarbrücken until graduation.

Vision and Research Strategy

Developing human-computer interfaces that fully exploit the information content available in natural human behaviour is challenging, particularly in unconstrained daily life settings. The group works at the interface of human-computer interaction, computer vision, as well as wearable and ubiquitous computing. We develop novel ambient and on-body sensing systems as well as computational methods to analyse human behaviour automatically. We specifically focus on visual and physical behaviour as these modalities are most promising for developing interfaces that offer human-like interactive and social capabilities. We study these systems and methods in the context of specific application domains, most importantly pervasive eye-based human-computer interfaces and computational human behaviour analysis.

Research Areas and Achievements

Pervasive Eye-Based Human-Computer Interfaces. Previous work on computational methods for eye tracking and eye-based human-computer interfaces mainly focused on single user, single device and WIMP-style interactions. Previous work also did not explore the full range of information available in users’ everyday visual behaviour. We aim to analyse and exploit the information contained in visual behaviour in all explicit and implicit interactions that users perform with computing systems throughout the day. We envision a future in which our eyes universally enable, enhance and support such interactions, and in which eye-based interfaces fully exploit the wealth of information contained in visual behaviour.
Computational Human Behaviour Analysis. A large body of previous work focused on the development of computational methods to robustly recognise activities that users perform. We aim to investigate and develop computational methods for multimodal sensing and modelling that allow us to better capture, analyse, and understand qualitative and social aspects of human behaviour. We specifically focus on automatic analysis and information extraction from non-verbal bodily cues. We combine this with information derived from gaze behaviour as one of the most important cues in non-verbal communication. We aim to study physical and visual behavioural cues during natural interactions in unconstrained real-world environments. These studies will provide new insights into the fundamentals of interactions between humans and machines and into the methods required to capture them.

Achievements. In the reporting period, the group has published 40 papers at conferences and in journals, most notably 9 papers at CHI/UIST/IUI, 2 papers in IMWUT, 4 papers in TPAMI/CVPR/ICCVW, and 6 papers at ETRA.

Projects and Cooperations
At MPI-INF, Andreas Bulling has been collaborating closely with Dr. M. Fritz on symbiotic human-machine vision. Locally, he has been collaborating with Prof. A. Krüger and Dr. Sonntag at DFKI. In Germany the main collaborator is Prof. F. Alt at the Bundeswehr University Munich. International collaborations exist with Prof. L.P. Morency (CMU, US), Prof. P. Robinson (University of Cambridge, UK), Prof. H. Gellersen (Lancaster University, UK), and Dr. T. Loetscher (University of South Australia, AU).

Prizes and Awards
  – ERC Starting Grant (2018)
  – Best Paper Award, ACM MobileHCI (2018)
  – Best Paper Honourable Mention Award, ACM MobileHCI (2018)
  – Best Paper Award, ACM ETRA (2018)
  – Best Presentation Award, ACM ETRA (2018)
  – Best Paper Honourable Mention Award, Eurographics (2018)
  – Best Paper Honourable Mention Award, ACM CHI (2018)
  – Best Paper Honourable Mention Award, ACM UIST (2017)
  – Best Paper Honourable Mention Award, MUM (2017)
15 Balakrishnan Chandrasekaran: The Server-to-Server Landscape

Vision and Research Strategy

The majority of the Internet traffic today is delivered via content delivery networks (CDNs). The ever increasing end-user demands for bandwidth and latency seem to assure that the volume of content delivered through CDNs will only increase going forward. Typically, the CDN hauls data from origins (i.e., content providers) to its back-end servers, moves this data (over a sophisticated overlay network) to its front-end servers, and serves the data from there to the end users. If we focus on the path taken by content from origins to end users via a CDN a simple fact becomes apparent: a significant fraction of this path traverses the CDN infrastructure, between the back-end and front-end servers. The server-to-server “landscape” formed by these paths is increasingly becoming “longer,” as front-end servers are being deployed closer to end users to minimize the last mile latency. Since CDNs have been expanding their infrastructure by deploying more servers in diverse networks and geographic locations, the landscape is also becoming “wider.” This server-to-server landscape provides a few unique perspectives as well as opportunities for solving some long-standing networking problems in the Internet. Our research revolves around characterizing this server-to-server landscape, leveraging the unique perspectives offered by this landscape to provide insights into the Internet’s structure and performance of protocols deployed on the Internet, and, lastly, exploiting the landscape to improve the Internet’s performance and security.

Research Areas and Achievements

Selectively reliable transport for video streaming

We are currently exploring how to leverage the server-to-server landscape to question the status quo in video streaming. Today, TCP is the dominant transport protocol for video streaming, due to the widespread use of dynamic adaptive streaming over HTTP (DASH). The rich body of prior work on optimizing TCP, adaptive bitrate selection algorithms, or TCP variants, however, highlights TCP’s shortcomings. Our preliminary investigation reveals that even at a loss rate of 0.16%—lower than that typically observed in the Internet—the video player spends 20% of the total video time in stalls (i.e., in waiting for the lost packets to arrive at the playback buffer) when streaming using TCP. But with CDNs (e.g., Akamai Technologies) and popular Web browsers (e.g., Google Chrome) already supporting QUIC, it is worth revisiting this status quo in streaming.

A simple observation should, nevertheless, highlight that reliable transports are ill-suited for video streaming: video data consists of different types of frames, some types of which
do not require reliable delivery. The loss of some types of frames has minimal or no impact (since such losses can be recovered) on the end-user quality of experience (QoE). Therefore, by adding support for unreliable streams in QUIC and offering a selectively reliable transport, wherein not all video frames are delivered reliably, we can optimize video streaming and improve end-user experiences. This approach has several advantages: (a) it builds atop QUIC that is rapidly gaining adoption; and (b) it involves only a simple, backward compatible, incrementally deployable extension—support for unreliable streams in QUIC. We are currently working on a thorough examination of the use of unreliable streams for video streaming and the interplay between such a partially reliable transport with the application-layer adaptive bitrate schemes is in the pipeline.

Towards a Speed of Light Internet

Reducing latency across the Internet is of immense value and content providers have clearly demonstrated the impact latency has on their bottom lines. Microsoft’s Bing, for instance, found that a two second slowdown translated to a drop in revenue per user of 4.3%. Latency is also crucial to facilitate the idea of running software in the cloud. A low-latency network allows more computations to be offloaded to the cloud, while giving the end users an illusion that they are running their computations locally (on their own machines). The Internet, however, is shockingly slow today! Prior work showed that the median time to fetch just the HTML documents of popular Web sites was 37-times slower than the round-trip speed-of-light latency (between the corresponding clients and servers), and that infrastructural improvements alone could reduce latency by at least a factor of three.

While there are already efforts underway on designing a low-latency speed-of-light (or “cSpeed”) Internet that at least connects the city centers of the most populous locations, our efforts are in an orthogonal dimension: How can a CDN make the best use of a speed-of-light network? The objective of a CDN to deliver content as quickly as possible combined with their ability to identify the latency-sensitive content among the overall traffic makes CDNs an ideal user of a cSpeed network. How would the design of a CDN change, if it had access to a cSpeed (but bandwidth-limited compared to fiber) backbone, supplementing the CDN’s existing connectivity? Addressing these and other related questions will continue to guide our research efforts in this space.

Projects and Cooperations

We are currently collaborating with researchers at the University of Massachusetts Amherst and Akamai technologies to evaluate the design and deployment of a selectively reliable transport using real video traffic and understand the suitability and effectiveness of the transport at Akamai for video streaming. We also have ongoing collaborations with Duke University, ETH Zürich, University of Illinois Urbana-Champaign, University of California Santa Cruz, and Yale University in the latency-related efforts.
16 Renjie Chen: Images and Geometry

Group Development

The research group Images and Geometry was established in July 2015 and is supported by the Max Planck Center for Visual Computing and Communication. Currently, the group consists of Renjie Chen and PhD student Björn Golla. Asif Omer finished his Master theses with the group.

Vision and Research Strategy

As its core mission, computer graphics endeavors to deliver natural-looking and convincing graphic contents, such as images, videos and 3D models for various applications, including design, entertainment, education, simulation, etc. In many cases, “natural-looking” can be interpreted as low distortion with respect to some reference. Depending on the application, the distortion can be measured as the amount of feature stretching, non-feature noise, change of scale, self-overlapping, and so on. As distortions can largely affect human perception of the contents, we want to generate images and shapes with no distortion or controlled amount of distortions, while satisfying the user-defined constraints for various applications. My research has been mainly focusing on designing theoretically sound and computationally efficient algorithms for geometry processing and its applications in image editing.

Research Areas and Achievements

Shape Modeling. Harmonic maps are extensively used in geometry processing applications, as they possess various nice properties such as smoothness. We established a sufficient and necessary condition for a harmonic mapping of any (possibly multiply-connected) domain to have bounded conformal and isometric distortions, solely based on the boundary behavior of the mapping. We construct a harmonic linear subspace which allows us to cast the locally injective mapping problem as a boundary value problem. Within this low dimensional subspace, the induced map and its differentials can be expressed in closed-form leading to simple, efficient and accurate evaluation. Moreover, we provide a closed-form recipe for a positive projection of the Hessian of the isometric energies, and this leads to effective-and-efficient Newton iterations. Our algorithm is specifically designed for modern parallel graphics hardware architectures, making it the first GPU-based locally injective deformation method.

Divergence-Based Distance Functions. Distance functions can be used to automatically plan a gradient-descent path towards a given target point in the domain, avoiding obstacles
that may be present. A key requirement from such distance functions is the absence of spurious local minima, and this has led to the common use of harmonic potential functions. This choice guarantees the absence of spurious minima, but is slow to compute. To alleviate this problem, we propose a family of novel divergence distances based on $f$-divergence of the Poisson kernel. We show that divergence distances generate paths identical to those generated by the potential function. However using divergence distances has a significant computational advantage, as, following a pre-processing stage, they may be computed online up to an order of magnitude faster than the others. Additionally, the computation is “embarrassingly parallel”, so may be implemented on a GPU with up to three orders of magnitude speedup.

We prove that the discrete version of the $f$-divergence distance also delivers, based on reduced coordinate vectors, which are the harmonic measures of a partition of the boundary into segments. Since in practice, the path is often restricted to follow the edges of a discrete network defined on a finite set of sites, any method that works well in the continuous setting must be discretized appropriately to preserve the important properties of the continuous case. We show how to define a network connecting a finite set of sites, such that a greedy routing algorithm based on our reduced coordinates is guaranteed to generate a path in the network between any two sites.

**Fastest-Path Computations.** In the age of real-time online traffic information and GPS-enabled devices, fastest-path computations in a road network graph, where each edge is weighted by a “travel time” value, are becoming a standard feature of many navigation-related applications. To support this, efficient computation of these paths in very large road networks is critical. Fastest paths may be computed as minimal-cost paths in a weighted graph, but traditional minimal-cost path algorithms based on variants of the classic Dijkstra algorithm do not scale well, as in the worst case they may traverse the entire graph. A common improvement, which can dramatically reduce the number of traversed graph vertices, is the A* algorithm, which requires a good heuristic lower bound on the minimal cost. We introduce a simple, but very effective and highly scalable, heuristic function based on a small number of values assigned to each graph vertex. The values are based on graph separators and computed efficiently in a preprocessing stage. Experimental results show that our heuristic provides estimates of the minimal cost which are superior to those of other heuristics, and when used in the A* algorithm, this heuristic can reduce the number of vertices traversed by an order of magnitude compared to other heuristics.

**Projects and Cooperations**

Within the department we collaborate with the High Performance Digital Geometry Processing group of Rhaleb Zayer. Externally, we collaborate with Prof. Dr. Craig Gotsman (New Jersey Institute of Technology), Prof. Dr. Kai Hormann (Università della Svizzera italiana), Prof. Markus Steinberger (TU Graz), Prof. Ofir Weber (Bar-Ilan University) and Dr. Edward Chien (Bar-Ilan University).
Group Development

Mario Fritz joined MPI-INF in 2011 and became Senior Research in 2013. In the reporting period, Mateusz Malinowski and Wenbin Li graduated. Mateusz received the DAGM MVTec dissertation award as well as the Dr.-Eduard-Martin award for his PhD thesis. Mario Fritz has received an offer for a W3 faculty position, but decided in June 2018 to fully move to the newly founded CISPA Helmholtz Center for Information Security as a permanent faculty member.

Vision and Research Strategy

With the advance of new sensor technology and abundant data resources, machines can get a detailed “picture” of the real-world – unlike ever possible before. The previously wide gap between these raw data sources and the semantic understanding of humans is starting to close. Driven by big data, increased compute power and advances in machine learning, we see a new generation of systems emerging that achieve new levels of performance on a range of competences such as visual scene understanding and natural language comprehension as well as robotic control. Such success rekindles the hope of building artificial intelligence that rivals human capacity. These developments are likely to have a profound impact on society and are hypothesized to lead to another industrial revolution. My research aims at understanding the limits of such technology by pushing the boundaries of what is possible today. Equally, I am addressing issues of privacy, safety and trust in order to make these emerging technologies compatible and compliant with society. In particular, I focus on advancing and analyzing Deep Learning techniques as they have recently transformed the academic and industrial landscape and are a driving force behind the recent performance leaps across a wide range of domains and tasks.

Research Areas and Achievements

Deep Learning for Graphics and Physics I have broadened my research thread on Scalable Learning and Perception and further explored intersections with Graphics and Physics. On the one hand, we explore learning-based techniques that analyze the interaction between light and surfaces and thereby overcome limitation of traditional – e.g. model-based and optimization-based approaches. E.g. we show to recover surface properties and real-world
environment light conditions in challenging real-world scenarios as well as estimate stability in real-world scenes which we in turn use for robotic manipulation.

**Modeling Uncertainty** While Deep Learning approaches have focused to a large extent on point estimates – our research focuses on how to equip such approaches with models of uncertainty. E.g. we cast deep learning approaches into Bayesian formulation and achieve accurate predictions over multi-modal distributions. We demonstrate our methods in several real-world scenarios – including the extrapolation of pedestrian trajectories into the future for the purpose of safe autonomous driving.

**Privacy and Security in Vision and Machine Learning** As computer vision techniques become more advanced, the scalable analysis of personal visual data becomes an increasing privacy risk. We have to the understanding, analysis and awareness of such problems. Furthermore, we have presented a first approach that helps the users to enforce privacy policies – which we call the Visual Privacy Advisor. Our techniques on automatic content sanitization provide methods and tools to the user in order to remove privacy threatening content from visual data. Beyond the privacy of the data, we also focus on the privacy and security of machine learning models. In our latest work, we raise and address important question on how to protect intellectual properties in Machine Learning service.

**Projects and Cooperations**

Within MPI there were tight collaborations with Bernt Schiele, Andreas Bulling, Christian Theobalt due to overlapping interests in the broad area of machine learning, computer vision, HCI as well as privacy. On campus, we collaborated with Prof. Muecklich on classification of steel microstructures. International, we collaborated with Prof. Ales Leonardis from University of Birmingham on the topic of intuitive physics for robotics, with Prof. Tinne Tuytelaars, Prof. Tobias Ritschel and Prof Luc Van Gool on the topics of deep learning for graphics and latent representations.
Group Development

The group Geometric Complexity Theory was founded in March 2017 with the promotion of Christian Ikenmeyer (group head) to senior researcher. The postdoc Nitin Saurabh joined in January 2018 and the postdoc Reuven Hodges joined in September 2019. In Saarbrücken the group collaborates with Professor Markus Bläser (Saarland University, computational complexity chair) and his working group, and with Professor Frank-Olaf Schreyer (Saarland University, algebraic geometry chair).

The student Julian Dörfler successfully completed a Research Immersion Lab. The student Umangathan Kandasamy is currently working on his master’s project. In the summer 2018 the research intern Farnoosh Hashemi visited the group for 3 months.

Vision and Research Strategy

Computational complexity theory is concerned with the study of the inherent complexity of computational problems. Its flagship conjecture is the famous $P \neq NP$ conjecture, which is one of the seven Millenium Problems of the Clay Mathematics Institute, ranking this conjecture as the most important one at the intersection of mathematics and theoretical computer science. In order to attack the $P \neq NP$ conjecture with algebraic methods, in the 1970s Leslie Valiant introduced an intriguing algebraization of the classical boolean complexity theory called algebraic complexity theory. His flagship conjecture states that the determinantal complexity of the permanent polynomial grows superpolynomially fast, or succinctly put $VBP \neq VNP$. Towards answering Valiant’s conjecture in 2001 Mulmuley and Sohoni proposed an approach based on algebraic geometry and representation theory for which they coined the name geometric complexity theory.

The group’s main agenda is to prove computational complexity lower bounds and to understand the capabilities and limitations of the geometric complexity theory approach. More precisely, we strive to use tools from representation theory and algebraic geometry to prove complexity lower bounds in Valiant’s algebraic complexity theory, and we analyze potential limitations of the methods proposed by Mulmuley and Sohoni, such as occurrences of irreducible representations in coordinate rings of group varieties and a potential algebraic natural proofs barrier.
Research Areas and Achievements

In the following we highlight two of the group’s projects on geometric complexity theory.

Variety Membership Testing and Algebraic Natural Proofs

Algebraic natural proofs were introduced in 2017 by Forbes, Shpilka and Volk and independently by Grochow, Kumar, Saks and Saraf as an attempt to transfer Razborov and Rudich’s famous 1997 barrier result for Boolean circuit complexity to algebraic complexity theory. Razborov and Rudich’s barrier result relies on a widely believed assumption, namely, the existence of pseudo-random generators. Unfortunately, there is no known analogous theory of pseudo-randomness in the algebraic setting.

As a testbed for geometric complexity theory, we introduce the geometric complexity measure of border completion rank of a matrix with affine linear forms as entries and construct a small family of such matrices and a bound $b$, such that at least one of these matrices does not have an algebraic natural proof of polynomial size against all matrices of border completion rank $b$, provided $\text{coNP} \not\subseteq \text{MA}$. This is an algebraic barrier result that is based on a well-established and widely believed Boolean conjecture.

Moreover, in the same paper we establish in a toy setting how representation theoretic methods from geometric complexity theory could break this barrier.

Representation Theoretic Multiplicities

The geometric complexity theory approach proposed by Mulmuley and Sohoni in 2001 studies the $\text{GL}_n^2$ orbit closures $\Delta_n$ of the $n \times n$ determinant polynomial and $\Gamma^m_n$ of the padded $m \times m$ permanent polynomial. The flagship conjecture in geometric complexity theory states that if the function $n = n(m)$ is polynomially bounded in $m$, then there exists a large $m$ with $\Gamma^m_n \not\subseteq \Delta_n$. This is a natural geometric strengthening of Valiant’s classical determinant vs permanent conjecture. If a type of an irreducible representation occurs in the coordinate ring $\mathbb{C}[\Gamma^m_n]$ with a higher multiplicity than in $\mathbb{C}[\Delta_n]$, then this proves $\Gamma^m_n \not\subseteq \Delta_n$. Mulmuley and Sohoni conjectured that one could find types that would occur in $\mathbb{C}[\Gamma^m_n]$ but not in $\mathbb{C}[\Delta_n]$ and prove the determinant vs permanent conjecture in this way. These types are called occurrence obstructions and in 2016 we showed that not enough of these exist to prove the determinant vs permanent conjecture. But in a recent preprint we show that multiplicities could be an actual way of proving it: we present the first situation in which multiplicities have provably stronger separation capabilities than just occurrences.

Projects and Cooperations

Besides local collaborations, the group collaborates with Mateusz Michalek (MPI-MiS, non-linear algebra) and his working group. The group maintains international collaborations with the working groups of Joseph Landsberg (Texas A&M University), Greta Panova (University of Southern California), and Takeshi Tokuyama (Tohoku University). Greta Panova is a DFG Mercator Fellow on Christian Ikenmeyer’s Sachbeihilfe grant.
Group Development

The group has been established in January 2013 with the support of the Max Planck Center for Visual Computing and Communication. In 2018, two of its members (Eugenia Holm and Ruben Becker) have successfully defended their PhD under the supervision of Andreas Karrenbauer. Currently, Eugenia is on maternity leave, and Ruben has joined the Gran Sasso Science Institute in L'Aquila, Italy, as a postdoc. The group further consists of PhD students Davis Issac and Maximilian John.

Vision and Research Strategy

Optimization problems are ubiquitous. Whenever there is a choice between several possibilities, there is an inherent optimization problem. Our core competence in this area is to recognize such optimization problems, to establish mathematical models, to tackle them by state-of-the-art methods, and to invent new techniques to obtain satisfactory results. To this end, we combine methods from discrete and continuous optimization. Our research is application-driven. This also includes inter-disciplinary research with collaborators from areas such as engineering, physics, chemistry, biology, and economics. Furthermore, connections and synergies with the field of machine learning are explored.

Research Areas and Achievements

In this report period, one focus of our research was in the area of computational interactivity. That is, the use of computational methods in the field of Human-Computer-Interaction. Clearly, there are multiple alternative designs for a user-interface for a given task, so that we are faced with an optimization problem. However, in contrast to classical optimization where constraints and objectives are mathematically well-defined, the big challenge here is that the notion of a good (better/worse) design is highly subjective. Typically, many stakeholders are involved having competing opinions, which may even evolve over time.

Designing interactive technology entails several objectives, one of which is identifying and selecting appropriate functionality. Given candidate functionalities such as “print,” “bookmark,” and “share,” a designer has to choose which functionalities to include and which to leave out. Such choices critically affect the acceptability, productivity, usability, and experience of the design. However, designers may overlook reasonable designs because there is an exponential number of functionality sets and multiple factors to consider. We are the first to formally define this problem and propose an algorithmic method to support designers to explore alternative functionality sets in early stage design. Based on interviews
of professional designers, we mathematically define the task of identifying functionality sets that strike the best balance among four objectives: usefulness, satisfaction, ease of use, and profitability. We develop an integer linear programming solution that can efficiently solve very large instances (set size over 1,300) on a regular computer. We present realistic special cases that can be solved by combinatorial algorithms in polynomial time. Further, we build on techniques of robust optimization to search for diverse and surprising functionality designs. Empirical results from a controlled study and field deployment are encouraging. Most designers rated computationally created sets to be of comparable or superior quality than their own. Designers reported gaining better understanding of available functionalities and the design space.

One of our research highlights in this area is our contribution to the new French keyboard standard (NF Z71-300). Keyboard layouts have traditionally been designed manually and have evolved incrementally. They have turned out to be difficult to improve even when a better solution is known, because of human factors (learnability etc.) or socio-technical ones such as standards. Designing a new keyboard is a hard problem from a computational perspective, an instance of the NP-complete quadratic assignment problem. In 2019, France will become the first nation to adopt a keyboard standard whose design was aided by computational methods, developed by members of this group. The new layout supports many more characters, optimizes access to frequent special characters to facilitate typing correct French, reduces risks of repetitive strain injury, and nevertheless achieves positions similar to the original design's. It transpired that off-the-shelf optimization methods were not ideal for such a large multi-stakeholder effort with differing, varying constraints and conflicting opinions. The work highlights a need to focus on participatory methods for optimization, wherein computational methods are used to ask the right questions, define objectives collaboratively, expose assumptions and tradeoffs, and involve expert stakeholders in numerous fields in every step.

Projects and Cooperations

Our work on computational interactivity was done in collaboration with Antti Oulasvirta, Daryl Weir, Perttu Lähteenlahti (Aalto University, Finland), Anna Maria Feit (Aalto University, Finland; ETH Zurich, Switzerland), Mathieu Nancel (Inria, University of Lille, France).

Within MPI-INF, we collaborated with researchers from D2 (Jan-Hendrik Lange and Bjoern Andres) on optimization problems in machine learning.
20 Christoph Lenzen: Theory of Distributed and Embedded Systems

Group Development

Ph.D. Students: Stephan Friedrichs graduated summa cum laude in late 2017 and works now at TomTom. Attila Kinali, who joined the group in January 2015, plans to finish his thesis in the course of the current year. Ben Wiederhake and Johannes Bund joined the group in January and August 2018, respectively, both after completing their master thesis under the supervision of Christoph Lenzen.

Postdocs: Reut Levi and Moti Medina left the group in August 2017; Reut is currently postdoc at the Weizmann Institute of Science and Moti is professor at the Ben-Gurion University of the Negev. Emanuele Natale continues his stay that started in November 2016. Saeed Amiri and Will Rosenbaum joined in December 2017 and August 2018, respectively.

Vision and Research Strategy

Broadly speaking, distributed computing concerns systems that consist of multiple agents that act based on local information. The main challenge is typically how agents gain access to sufficient information to collaboratively solve a task quickly, despite limits on communication, faults, or inaccurate data. One of the key differences to many other models of computing is the possibility that the system may tolerate faults of individual agents. Our group applies this paradigm to hardware, viewing, e.g., a multi-core processor as a distributed system.

Research Areas

Parallel and Distributed Distance Problems. Naturally, solving well-known computational problems involving distances such as All Pairs Shortest Paths, Single Source Shortest Paths, Steiner Trees and Forests, or flow type problems is of great utility in distributed systems, too. We tackle these problems with all available means: randomization, approximation, decomposition, etc. Key differences to centralized algorithms are that each node can perform its own computations, is initially aware of its neighborhood in the graph only, and needs to learn its local share of the output only (e.g. its distances from other nodes). To this end, nodes communicate over the edges of the graph; this communication is the main cost factor. For parallel algorithms, there is more flexibility: all nodes have access to a shared memory.

Design of Fault-tolerant Hardware. Today’s Very Large Scale Integrated (VLSI) circuits operate on such small length and time scales that they become distributed systems in
their own right. Small subcircuits take the role of nodes that communicate via wires – the “edges” of the communication graph – and billions of individual components render fault-tolerance mission-critical. Hence, insights from the area of distributed computing are useful for designing reliable hardware. However, VLSI circuits add unique challenges. In low-level hardware, even simple computations require substantial effort, and violation of timing constraints may induce so-called metastability, a state outside of the abstraction of binary logic. We seek to adapt algorithmic ideas from distributed computing and other areas to this environment, with the goal of devising hardware with strong, provable fault-tolerance properties. A key challenge in this process is the design of clock generation and propagation schemes that are simultaneously highly resilient, accurate and precise, yet cost-efficient.

**Hazard-freedom and Metastability-containment.** One specific challenge when dealing with low-level hardware (and clocking in particular) is metastability, an unstable third equilibrium state of bistable elements such as latches and registers. The standard solution is to use synchronizers whenever signals cross clock domain boundaries or performing analog-to-digital conversions. However, this takes a couple of clock cycles, which has negative impact when time is critical, e.g., in control loops operating at very high frequencies. Instead, one can contain metastability, i.e., accept it as normal part of the operation and study circuit behavior in three-valued Kleene logic. The goal is to develop small and low-latency circuits in this setting, posing plenty of circuit design and complexity questions.

**Sublinear Algorithms and Property Testing.** When dealing with very large graphs, no single machine can process or even hold the entire graph in its memory. This necessitates a different model of computing, where the graph is probed locally. Nonetheless, queries like “is node v part of the maximal independent set?” assume an underlying consistent graph structure. Concretely, when performing multiple independent queries of the above type, the yes/no answers should be correct with respect to a single maximal independent set. The goal here is to minimize the query complexity, i.e., the number of queries of the form “what are the neighbors of node v?” or “what is neighbor number 3 of v?” etc. is to be minimized.

Conceptually closely related is property testing in the Congest model. Here, the goal is to, e.g., separate the cases of an input graph being triangle-free or “far from” triangle-free (i.e., many edges would have to be removed to make it triangle-free), where any answer is feasible for graphs in between. Communication between nodes is limited by an $O(\log n)$ message size restriction and the optimization target of small running time, implying that each node can only learn about a small part of the graph.

**Computational Dynamics.** Dynamics are simple stochastic processes on networks, in which agents update their own state according to a symmetric function of the state of their neighbors and of their current state, with no dependency on time or on the topology of the network. In previous decades, this kind of systems has been investigated from a computability point of view, attracting the interest of mathematicians and physicists. Recently it has been subject to a renewed interest from computer scientists, as new techniques for analyzing this class of processes have made it possible to answer questions regarding their efficiency and capability as distributed algorithms.
21 Pauli Miettinen: Data Mining

Group Development

The group Data Mining was established in early 2013. Its head, Pauli Miettinen, accepted a full professor position at University of Eastern Finland from August 2018 onwards. The remaining members of the group are PhD students Saskia Metzler and Sanjar Karaev, both of whom are expected to graduate during 2019. Two Master’s students and one Bachelor’s student graduated from the group during the reporting period.

Vision and Research Strategy

The group’s goal is to develop data mining methods that are based on well-founded algorithmic principles. The group’s research is typically based on a three-step approach: The first step is to formulate the data mining problem and to study its computational complexity. Can the problem solved in polynomial time? If not, can it be approximated in polynomial time within a reasonable factor? Can the problem be re-formulated as an another problem? An important part of this step is to identify what makes the problem hard and is that hardness an inevitable part of the problem, or could it be avoided with a different formulation that is still essentially equivalent to the original data mining problem.

The second step utilizes the background knowledge obtained in the first step to develop algorithms that work in practice. The emphasis here is on the practical side: A theoretically justified algorithm that is never better (in empirical evaluations) than a faster heuristic is usually not desirable.

The third steps involves refining the algorithm and finding applications outside the original data mining problem. Typically this step results on more work on the previous steps, as the special properties of the new applications require refined theory and algorithms.

Research Areas and Achievements

The research of this group is broadly divided into three subfields: matrix and tensor factorizations over non-standard algebras, social network analysis, and redescription mining.

Matrix and Tensor Decompositions. Matrix – and recently also tensor – factorizations are very popular methods in data mining, but most common methods work over the standard field of real numbers or a restriction thereof, such as the semiring of nonnegative real numbers. While certainly powerful, these methods can only find certain kinds of patterns effectively, and an overarching theme in our work on non-standard algebras is to study what kinds of patterns could be found more effectively using other kinds of algebras.
The algebras we mostly work with are the Boolean algebra, that provides a clear combinatorial interpretation of the factorization; the subtropical (or max-times) algebra, that provides a “winner-takes-it-all” interpretation, where only the largest value in the components matter; and the rounding algebra, where the matrix product is thresholded to binary. The rounding algebra allows, possibly combined with the subtropical algebra, us to express complex combinatorial structures such as nested matrices, as simple rank-1 matrices.

The three steps of our research strategy are well-present in our work on matrix and tensor factorizations. We study fundamental questions on different algebras as well questions such as “when can we effectively recover a planted pattern?” We also develop efficient algorithms for finding the factorizations, and we have applied our algorithms, for example, to graph analysis.

Social Network Analysis. A clique is a typical model for a community in social networks. It (inherently) assumes that within a community everybody knows – or at least should know – everybody else. This assumption is not always correct, as social scientist have known since at least 1990’s. To accommodate the more diverse models for communities, we have developed a hyperbolic model, which allows us to characterize the members of the community, on one hand, as the core members, i.e. those who know everybody else, and tail members, i.e. those who only know members from the core. This community model provides a good match for known communities, and has allowed us to study the behaviour of community shapes over times. In order to find such communities, we can use our results and methods for tropical algebras.

Redescription Mining. The goal of redescription mining is to find two different ways to characterize the same data. For instance, we might characterize politicians on one hand by their socio-economical background, and on the other hand, by their political opinions. Such redescriptions help us to explore which combinations of political views are correlated with which socio-economical backgrounds. We have studied methods for finding the redescriptions in general, and are studying their applicability in political sciences as well as in biocology and other fields. To facilitate the use of redescription mining in other disciplines we have developed a tool for interactive and visual mining of the redescriptions.

Projects and Cooperations

Within the department, we collaborate with Daria Stepanova. Within the campus, the group collaborates with Jilles Vreeken’s group Exploratory Data Analysis and Olga Kalinina’s group Structural Bioinformatics of Protein Interactions.

Externally, our main collaborations are with Prof. Stephan Günnemann (TU Munich), Dr. Esther Galbrun (University of Eastern Finland), and Dr. James Hook (University of Bath).

Prizes and Awards

Pauli Miettinen won a Best Reviewer Award at the 2017 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).
Group Development

The research group was founded in June 2000 as a part of the Computer Graphics Department and initially it was focused on the human perception aspects in global illumination and rendering. The scope of its research gradually expanded towards High Dynamic Range Imaging (HDRI), whose goal is the precise representation of the real-world light intensity and the color gamut at all stages of image and video processing, from acquisition to display. Recently, we are interested in computational displays, which activate stereoscopic 3D vision, eye accommodation and enable motion parallax. We refer to light-fields as a suitable input data for such advanced displays. We return also to global illumination simulation but this time in the context of computational fabrication and appearance perception.

Within the reporting period (03/2017–02/2019), two new PhD students and one post-doc were hired. At present, the group consists of three PhD students, Denis Sumin, Krzysztof Wolski, Mojtaba Bemana, and three post-docs, Gurprit Singh, Okan Tarhan Tursun, and Hyeonseung Yu. Three other PhD students made internships with our group: Ana Serrano from Zaragoza University (Spain), Michał Chwesiuk and Marek Wernikowski from Westpomeranian University of Technology (Poland).

Vision and Research Strategy

The common goal of all research efforts of the group is the advancement of knowledge on image perception and the development of imaging algorithms with embedded computational models of the human visual system (HVS). This approach offers significant improvements in both the computational performance and the perceived image quality. Often, we refer to perceptual effects rather than physical effects, which puts more emphasis on the experience of the observers than physical measurements. In particular, we aim for the exploitation of perceptual effects as a means of overcoming physical limitations of display devices and enhancing the apparent image quality.

Seamless blending of visual content from both of the virtual and real world by the means of augmented reality (AR) is a long-standing goal. To this end, a faithful reproduction of visual cues such as binocular vision, motion parallax, and eye lens accommodation is of particular importance. We refer to displays with such a functionality as hyper-reality displays (H-RD). Developing an H-RD requires a joint design of hardware with non-conventional display optics and software with specialized rendering algorithms that altogether attempt to sample and reconstruct important aspects of the complex 5D plenoptic function. Such plenoptic function, also referred to as a light-field, is a complete representation of the light
distribution at every position and direction in a scene. Modeling visual perception is a principled approach to optimize the sampling and reconstruction, so that the discrepancies from the perception of a reference real-world scene are minimized. For example, the human visual system (HVS) sensitivity to brightness, contrast, and object depth may dynamically change as a function of gaze position, ambient luminance levels and accommodative state of the eye lens, which in turn might depend on the semantic scene content, object positions in screen-space, their depth levels, and observer’s motion relative to the objects. Having a model of the HVS which accounts for all these factors enables dynamic reallocation of rendering resources to achieve the best visual quality.

By capitalizing on our experience gained from building an award winning, extreme wide-field-of-view see-through AR display that supports eye lens accommodation (a joint effort with Nvidia and UNC), we investigate basic HVS mechanisms that accompany switching gaze direction between virtual and real world. Such studies could be performed for the first time without focus conflicts, and other contradictions between visual cues, such as the infamous vergence-accommodation conflict (VAC). The key objective is to develop suitable perceptual models that integrate all important visual cues, and are responsive for the possible cue deviations from the real world observation, which might be imposed by the display optics design or poor rendering quality. Since such perceptual models are able to estimate the quality of viewing experience and quantify any deviations from the reality in a perceptually meaningful way, we apply them to develop objective image quality metrics and to drive foveated rendering on H-RDs with the eye accommodation.

We consider 3D printing as another aspect of seamless transition between virtual and physical objects, we are aiming at developing a domain-specific perception model for faithful appearance reproduction in fabrication. When objects are 3D printed their physical appearance is expected to match their virtual models as closely as possible. The perceptual aspects of such appearance matching have not been investigated thoroughly so far, given that each printing technology might impose some unique constraints, such as strong light diffusion within the material that washes out texture details of the object.

One of the key goals in our research vision is spreading the awareness of the importance of HVS modeling for computer graphics and image processing. We achieve this by co-authoring textbooks, such as the one on High Dynamic Range Imaging (almost 1,900 Google Scholar citations). Another important mission of our group is training of PhD students in applied perception, who proliferate such expertise in new research areas in their future scientific activities. Many of former PhD students, for example Rafał Mantiuk (Cambridge University), Piotr Didyk (USI Lugano), Tunç Aydın (Disney Research), Dawid Pajak (Nvidia Research), and Petr Kellnhofer (CSAIL MIT), who started their independent research careers, published their work with strong perceptual components at top venues ranging from Nature to ACM SIGGRAPH/TOG (over 30 papers).

**Research Areas and Achievements**

**Gaze-contingent Rendering** Information on the eye fixation allows rendering quality to be reduced in peripheral vision, and the additional resources can be used to improve the quality in the foveal region. The adoption of foveated rendering is hampered by system latency
which is especially apparent during fast saccadic movements when the information about
gaze location is significantly delayed, and the quality mismatch can be noticed. Instead of
rendering according to the current gaze position, we predict where the saccade is likely to end
and render an image for the new fixation location as soon as the prediction is available. While
the quality mismatch during the saccade remains unnoticed due to saccadic suppression, a
correct image for the new fixation is rendered before the fixation is established. We derive
a model for predicting saccade landing positions and demonstrate how it can be used in
the context of gaze-contingent rendering to reduce the influence of system latency on the
perceived quality.

**Perception-driven Rendering for Multi-layer Accommodative Displays**  Multi-focal plane
and multi-layered light-field displays are promising solutions for addressing all visual cues
observed in the real world. Unfortunately, these devices usually require expensive optimiza-
tions to compute a suitable decomposition of the input light-field or focal stack to drive
individual display layers. A simple alternative is a linear blending strategy which decomposes
a single 2D image using depth information. This method provides real-time performance, but
it generates inaccurate results at occlusion boundaries and on glossy surfaces. We propose
a perception-based gaze-contingent hybrid decomposition technique which combines the
advantages of the above strategies and achieves both real-time performance and high-fidelity
results. We perform a complete, perception-informed analysis and develop a computational
model that locally determines which of the two strategies should be applied. The prediction
is later utilized by our new gaze-contingent synthesis method which performs the image de-
composition. The results are analyzed and validated in user experiments on our custom-built
multi-focal plane display with an eye tracker.

**2D Image and Light-field Quality Evaluation**  A large number of imaging and computer
graphics applications require localized information on the visibility of image distortions.
Existing image quality metrics are not suitable for this task as they provide a single
quality value per image. On the other hand, existing visibility metrics produce visual
difference maps, and are specifically designed for detecting just noticeable distortions but
their predictions are often inaccurate. We argue that the key reason for this problem is the
lack of comprehensive image datasets with a good coverage of possible distortions that occur
in different applications. To address the problem, we collect an extensive dataset of reference
and distorted image pairs together with user markings indicating whether distortions are
visible or not (http://visibility-metrics.mpi-inf.mpg.de/). We propose a statistical
model that is designed for the meaningful interpretation of such data, which is affected by
visual search and imprecision of manual marking. We use our dataset for training existing
metrics and we demonstrate that their performance significantly improves. Also, we show
that our dataset with the proposed statistical model can be used to train a new CNN-based
metric, which outperforms the existing solutions. A similar purpose serves also our Light-field
Archive https://lightfields.mpi-inf.mpg.de, which is a collection of dense reference
and distorted light-fields as well as the corresponding quality scores which are scaled in
perceptual units. We use this dataset to check the suitability of existing 2D image quality
metrics for light-field quality evaluation.
Appearance Reproduction in Computational Fabrication  Additive 3D printing such as poly-jetting technology, which has gained importance recently, enables object volume construction of micro-voxels. Color texture reproduction in 3D printing commonly ignores such volumetric light transport between surface points on a 3D print. Such light diffusion leads to significant blur of details and color bleeding, and is particularly severe for highly translucent photopolymerization print materials. Given their widely varying scattering properties, this cross-talk between surface points strongly depends on the internal structure of the micro-voxels surrounding each surface point. In our work, we counteract heterogeneous scattering to create the impression of a crisp albedo texture on top of the 3D print, by optimizing for a fully volumetric material distribution that preserves the target appearance. Our method employs an efficient numerical optimizer on top of a general Monte-Carlo simulation of heterogeneous scattering, supported by a practical calibration procedure to obtain scattering parameters from a given set of printer materials.

Publications  Within the reporting period, the group members published their work in ACM TOG/SIGGRAPH/SIGGRAPH Asia (5), IEEE TVCG (3), Journal of Perceptual Imaging (1), Computer Graphics Forum (1) journals, and presented at CVPR (2) and other conferences.

Projects and Cooperations

Our close collaboration with the groups of Piotr Didyk (MMCI/USI Lugano) and Tobias Ritschel (UCL) resulted in multiple joint papers. We have been also working with external partners on multiple short- and long-term projects: with Henry Fuchs (UNC Chapel Hill) and David Luebke (Nvidia) on wide-field-of-view varifocal near-eye see-through display, with Diego Gutierrez and Belen Masia (Zaragoza University) on image quality in VR, with Bernd Bickel (IST Austria), Tim Weyrich (UCL) and Jaroslav Krivanek (Charles University) appearance fabrication, and with Rafal Mantiuk (Cambridge University) data-driven image quality evaluation.

Our research is supported by two Horizon 2020 Training Networks DISTRO: Distributed 3D Object Design and RealVision: Hyperrealistic Imaging Experience, each funds one PhD student for three years; Joint Fraunhofer and Max Planck Research Grant, Perceptually-aware Light Field Capture, Processing and Display, funded by MPG and FhG, funds three researchers for four years.

Prizes and Awards

The group members were awarded with: Best Paper Award, IEEE Virtual Reality 2017 and Digital Content Association of Japan Award at ACM SIGGRAPH Emerging Technologies 2017 (together with P. Didyk, H. Fuchs, D. Luebke and their teams from UNC Chapel Hill and Nvidia). Karol Myszkowski obtained the 2016 IEEE TVCG Best Associate Editor Award (announced in December 2017). Petr Kellnhofer, a former PhD student (now at CSAIL MIT) was awarded with Dr.-Eduard-Martin-Preis 2017 and Eurographics PhD Dissertation Award 2018.
23 Gerard Pons-Moll: Real Virtual Humans

Group Development

The *Real Virtual Humans* group was established in September 2017 in the department of Computer Vision and Multimodal Computing. Its members are Gerard Pons-Moll (head) and two doctoral students Bharat Lal Bhatnagar and Verica Lazova at MPI-Inf, in addition to two external co-advised doctoral students Thiemo Alldieck (TU-Braunschweig) and Timo von Marchard (University of Hanover). Gerard Pons-Moll has been awarded an Emmy Noether starting grant (1.6 Mi) from the DFG to fund the research group for the next 6 years.

Vision and Research Strategy

Digitizing human beings would redefine the way we think and communicate (with other humans and with machines), and it is necessary for many applications—to transport people into virtual and augmented reality, for understanding people in videos, for entertainment and for medicine and psychology. While human digitization has been demonstrated using expensive multi-view studios and 3D scanners, digitization of people from low-cost and ubiquitous devices such as consumer cameras is a major open problem. Furthermore, static 3D human reconstructions can not be automatically re-animated and controlled.

In Real Virtual Humans we investigate computational algorithms to efficiently digitize people and train machines to perceive people from visual data—a task which lies at the intersection of computer vision, machine learning and computer graphics. Our research aims at answering the following two inter-related research questions: *How do we efficiently digitize humans without losing the detail that make us real?* and *How can we train machines to perceive such rich representations from visual data?*

Research Areas and Achievements

We are focusing on building realistic models of 3D people appearance and movement, and perceiving such detailed models from ubiquitous sensors like RGB-cameras.

**Generative Models of 3D People and Clothing.** One of the goals in our group is to learn a statistical model that can factor body shape, pose and clothing. The model should be highly realistic, easy to control and easy to fit to data. Instead of using traditional physics based simulation, we learn the 3D models from real captured people. From thousands of scans of people we have learned already models of pose, shape, soft-tissue and we have shown how to factor clothing from the body. Some of the components missing in the models are:
control over the surface appearance and control on the clothing geometry, appearance and dynamics. In order to scale to the real world, we are investigating learning frameworks that leverage 3D scan data as well as image and video data. The learned 3D generative models of people in clothing constitute the mental model for machines, which simplify machine perception tasks such as perceiving people in images.

**Perceiving People from Visual Data.** Current computer vision algorithms can detect people in images or estimate 2D keypoints to a remarkable accuracy. However, people are far more complex. Our goal is to recover all the details that make us human from visual data.

We combine modern deep learning techniques with geometric optimization and 3D reasoning in order to obtain detailed reconstructions from single images, videos or depth. Almost all our perception algorithms leverage the learned 3D generative models mentioned above. Since the group started, we have developed the first method to reconstruct detailed models of people in clothing from a single video, a deep learning architecture that integrates a 3D body model within the network for single image 3D pose and shape reconstruction, an online algorithm that combines model based fitting with free-form reconstruction and a real-time multi-person 3D pose estimation method.

Combining video data and measurements from Inertial Measurement Units attached at the person extremities, we have introduced a sensor fusion method capable of estimating the pose of 2 people in outdoor scenes with multiple people and complex backgrounds. With this we have made available 3DPW, the first dataset (for 3D pose evaluation) with ground truth poses in natural videos.

Since visual perception of people and 3D modeling of people can benefit from each other, we plan to integrate them more closely together in the future. This will allow us to go beyond 3D reconstruction and learn the 3D human representations from images and video. In essence, our high-level goal is to teach machines to see and perceive people.

**Projects and Cooperations**

Within MPI-Inf the group collaborates closely with Prof. Bernt Schiele’s group and Prof. Christian Theobalt’s group. The group has external collaborations with the group of Prof. Otmar Hiliges at ETH, the group of Dr. Michael Black at MPI-IS, the group of Prof. Yebin Liu at Tsinghua University. Two students are co-advised remotely, Timo von Marcard from Prof. Rosenhahn’s group at the University of Hannover, and Thiemo Alldieck from Prof. Marcus Magnor group at TU-Braunschweig.

**Prizes and Awards**

- Emmy Noether Programme (1.6 Million to fund the group for a period of 6 years)
- Best student paper award at International Conference on 3D Vision (with Mohamed Omran and Bernt Schiele)
- Facebook Reality Labs awarded the group funding to finance 1-2 positions.
- Google Faculty Research Award (funding for 1 position for a period of 1 year)
Group Development

The group *Knowledge Base Construction and Quality* was established in Spring 2018. In addition to the group head, Simon Razniewski, it currently involves three PhD students, Sreyasi Nag Chowdhury, Cuong Xuan Chu and Hiba Arnaout (jointly supervised with Gerhard Weikum), one Master student, Shrestha Ghosh, and one Postdoc, Koninika Pal. It has also hosted two interns (Julien Romero and Archit Sakhadeo) in 2018, and is currently about to host two more.

Vision and Research Strategy

The group’s goal is to develop methodologies for advanced knowledge base construction, for instance in domains such as common-sense knowledge or fictional stories, and to enable the assessment of KB quality beyond the standard precision measures. The research builds on a combination of techniques from conceptual modelling, machine learning, and natural language processing.

Research Areas and Achievements

**Knowledge base recall**  In the area of knowledge base recall assessment, we have continued developing formal foundations for composing and propagating completeness metadata into query results, enabling, for a wide range of queries, to understand their completeness w.r.t. datasources enhanced with such metadata. We have developed textual information extraction methods for obtaining counting information on knowledge base relations, which can be used to assess the recall of actually observed data. Currently, we are working on the discovery of such hidden counting quantifier information in existing knowledge bases, aiming to uncover and link these quantifiers to semantically related relations. We are also working on building recall-awareness into information extraction processes directly, this way producing recall information directly during the extraction process, and enabling targeted and resource-efficient extraction. A demonstration system for visualizing the recall of the Wikidata KB, Recoin, has been deployed as publicly available Wikidata plugin.

**Fictional KBs**  Motivated by the success of real-world knowledge bases, we have started to expand the research towards fictional universes such as Lord of the Rings or Game of Thrones. Such fictional universes are archetypes of domain-specific universes, as could also be relevant
for companies with specific jargon, hobbies, and other domain-specific universes. The plan is to proceed along three main steps, (i) taxonomy construction, (ii) entity typing, (iii) fact extraction. By combining and extending existing techniques for taxonomy construction over Wikipedia to fictional Wikis, we have achieved promising results for the first step. Currently we are working on the second step, where a major component is transfer learning towards new and unseen fictional universes, e.g., reusing type systems from LoTR for the typing of related entities in GoT. In the third step, we aim to build a knowledge bases, similar to YAGO, for specific fictional universes, although with more focus on text-based information extraction.

**Common-sense KBs** Commonsense knowledge about object properties, human behavior and general concepts is crucial for robust AI applications. However, automatic acquisition of this knowledge is challenging because of sparseness and bias in online sources. We have started expanding on earlier work on common-sense extraction, aiming to achieve improvements along three dimensions, (i) precision and relevance, (ii) recall, (iii) semantics. In current work we have targeted novel sources, in particular search engine query logs, and obtained results that clearly outperform the state of the art in (i) and (ii). Towards (iii), we next plan to look into techniques of semantic role labelling and facet extraction, in order to add aspects about temporal and spatial validity, or model common-sense statements that require more complex logical forms than triples.

**Projects and Cooperations**

The group has close ties to the DIG team at Telecom ParisTech University (Fabian Suchanek), to the KRDB group at the Free University of Bozen-Bolzano, to the University of Indonesia (Fariz Darari), and to FU Berlin (Claudia Mueller-Birn).

**Prizes and Awards**

The paper *Completeness-aware rule mining from knowledge graphs* won the best student paper runner-up at ISWC 2017 (first author: Thomas Pelissier Tanon). The paper *Enriching knowledge bases with counting quantifiers* was selected as one of 5 featured papers out of 39 full papers at ISWC 2018.
Group Development

The research group on Question Answering currently involves Rishiraj Saha Roy (head), Gerhard Weikum (mentor), Abdalghani Abujabal (Ph.D. student), Azin Ghazimatin (Ph.D. student), and Asia Biega (Ph.D. student). Three M.Sc./B.Sc. theses have been successfully completed within the research group during the reporting period, and one is in progress. Two long-term interns were also hosted by the group during this time.

Vision and Research Strategy

Research on question answering (QA) aims to provide direct answers to natural language (NL) utterances, either over structured knowledge graphs (KG) or unstructured text documents. In our group, we have tried to push the state-of-the-art in QA along multiple dimensions. The key driving criteria have been handling diversity in question formulations, complexity in information needs, and providing unsupervised, interpretable, and efficient solutions that are not constrained to specific settings and benchmarks.

In a parallel research thread, we have also explored non-standard and emerging problems in user-centric privacy and transparency in online forums like community question answering (CQA) sites and e-commerce platforms. The vision here has been to develop an advisory tool for users that could reduce their vulnerability to profiling by major service providers, and provide actionable explanations to items recommended by such providers.

Research Areas and Achievements

Diversity in question formulation. Templates are commonly used for mapping NL questions to SPARQL queries executable over a KG, as they represent a highly generalizable approach to QA. But syntactic differences in question formulation pose a stiff challenge to template-based QA systems. So we proposed the addition of a semantic layer (via a similarity function) to overcome this limitation. We also introduced the scope for human relevance feedback on the answers to prevent noisy information creeping into the training data. Our modules for template-based answering, similarity-based answering, and user feedback, are assimilated into a continuous learning framework. We published our work in:

- WWW 2018, EMNLP 2017 (demo), WWW 2017

Complexity in information needs. A second challenge stems from complexities in the information needs themselves. This can be quantified by the presence of multiple entities and relations in the questions. This problem is most pronounced when answers can be computed
only by joining evidence from multiple documents. We proposed a method that can answer complex questions directly from textual sources on-the-fly, by computing similarity joins over partial results. It first builds a noisy quasi knowledge graph of dynamically retrieved entity names and relational phrases, overlays type and alignment information, and then computes the best answers by an algorithm for Group Steiner Trees.

**Temporal questions.** Complex questions can also be categorized into various classes, and we showed that questions with temporal intent need specialized treatment. Our system TEQUILA handles temporal questions by decomposing and rewriting them into sub-questions, answers these, and then reasons with candidate answers and temporal intervals to produce high-precision results. We published our work in:

- CIKM 2018, HQA 2018 (WWW 2018 workshop)

**Answer types.** Efficiency is another prime concern that is often overlooked in developing research prototypes for QA. KG-QA systems may execute a few hundred SPARQL queries before selecting the best that is used for the final answers. Such choices are not suitable for answering with interactive response times. We developed a method that predicts and uses expected answer types for this problem, aimed primarily at complex questions, where the problem is much more severe. The solution, coined TIPI, uses clever strategies for significantly pruning the search space of possible queries. We published our work in:

- IJCNLP 2017

**QA benchmarks.** Finally, for extending research in these interesting directions, the community needs benchmarks containing challenging questions that are compositional in nature, and those that exhibit phenomena like comparisons, aggregations, temporal reasoning, and so on. To address this, we created ComQA, a large dataset (11,214 questions grouped into 4,834 paraphrase clusters) of real user questions collected from the WikiAnswers platform, that we cleaned with substantial crowdsourcing effort. We published our work in:

- NAACL 2019

**Privacy and user-utility.** We provided the first practical and formal framework for studying privacy and user-utility tradeoffs. Our framework, that created several mediator accounts between the user and the service provider, leveraged solidarity in a large community to scramble user interaction histories. We published our work in:

- SIGIR 2017, CIKM 2017

**Explanations for feed items.** Another area of high current relevance that we investigated is the feed that a user receives on social platforms like news, music, or CQA forums. We proposed the first framework that systematically discovers, ranks, and explains relationships between users’ actions and items in their social feeds. We published our work in:

- WSDM 2019

**Projects and Cooperations**

Current external collaborators include Mohamed Yahya (Bloomberg L.P., UK), Mirek Riedewald (Northeastern University, USA), Zhen Jia (Southwest Jiaotong University, China), Xiaolu Lu (RMIT University, Australia), and Yafang Wang (Alibaba Group, China).
Group Development

Gurprit Singh joined MPI-INF as a post-doctoral researcher in September, 2017. He is currently establishing the Sampling and Rendering Group in the computer graphics department. The group has one Master thesis student (Alexander Kohn), one Bachelor thesis student (Matias Klimpel) and a research assistant (Chitra Singh as HiWi). Two prospective Master thesis students (Vassilen Chizhov, Tian-Qi Fan) are scheduled to join in March 2019.

Vision and Research Strategy

Our group is dedicated towards developing a modern deep learning pipeline for sampling and rendering algorithms. We strive to provide an end-to-end workflow for applications ranging from offline production rendering to real-time devices for virtual and augmented reality (VR/AR). Each application has different requirements, however, in all cases the rendered image quality is directly affected by the underlying sampling strategy and the sample correlations. Our group aims to develop sound theoretical error formulations based on Monte Carlo and Quasi-Monte Carlo (MCQMC) literature to administer sample correlations across multiple dimensions. Consequently, this would allow designing task-specific loss functions for training purposes. Simultaneously, we equip modern deep learning architectures with the state-of-the-art quality metrics to deliver proof-of-concepts. However, existing neural network (NN) architectures are designed and optimized for either structured data like images or point clouds limited to 3D only. Towards this end, we are developing optimization techniques to scale deep NNs for our purposes that would allow progressive generation of millions of correlated point samples per second in high dimensions (100s). We hope that our research would also benefit communities from computer vision, geometry and shape analysis—that deals with unstructured data (e.g., point clouds)—by establishing a coherent exchange of knowledge.

Research Areas and Achievements

Our research areas can be classified into two broad categories: analysis of sampling patterns and rendering algorithms and deep network synthesis. We focus on understanding how different sampling schemes and their correlations affect error during MC estimation, followed by designing NN architectures from the knowledge gained from these analyzes.

Sampling and Rendering Analysis To derive optimal benefits from a deep learning rendering pipeline, we need to design loss functions that reflect not only local but also global
interactions: both within samples and the light transport. This requires analyzing sample correlations and how well a rendering algorithm explores the underlying manifold.

We recently taught a course at SIGGRAPH ASIA 2018 and are publishing an EG STAR 2019 report (conditionally accepted) that summarizes various spatial and Fourier statistical tools developed over the years to better understand the impact of sample correlations on the MC estimation error. Importance functions from which samples are derived also dramatically affect the estimation error. We have performed an in-depth theoretical analysis that analyzes importance sampling in conjunction with correlated samples (to be published as a CGF 2019 journal paper). Our spectral domain expertise have drawn attention in other fields. We partially assisted on developing a spectral measure of distortions for change detection in dynamic graphs (Complex Networks 2018) and a perception driven hybrid-decomposition model for multi-layer VR displays (IEEE VR 2019).

Concurrently, we are analyzing rendering algorithms in the gradient domain to alleviate the non-differentiable nature of the light transport integrals (due to the visibility function present within the integral).

**Deep Network Synthesis** Several sampling design principles have been established in the past few years (first authored SIGGRAPH 2017 & 2015 by G. Singh) following a rigorous theoretical and empirical analysis. Current state-of-the-art sampling algorithms are not able to realize these futuristic designs while preserving important correlations across multiple dimensions. We combine modern deep learning architectures with these sampling design principles to model loss functions that can abstractly handle local and global interactions. This results in training kernels that can produce stochastic samples with required correlations in a matter of seconds. We published the proof-of-concept deep learning based system (arXiv tech-report, 2018) that proposes several spatial and Fourier tools as a part of loss functions. In recent work, we achieve state-of-the-art quality results and pushed it further to higher dimensions. We also develop a memoryless deep learning system alleviating the constraints due to GPU memory limits. This marks another step towards our goal to provide an end-to-end learning pipeline for sampling and rendering algorithms.

**Projects and Cooperations**

We are actively collaborating with University College London (Tobias Ritschel) and are commencing new projects in collaboration with the University of Tokyo (Toshiya Hachisuka) and AutoDesk Arnold renderer (Iliyan Georgiev). We have successfully completed projects in collaboration with University of Edinburgh (Kartic Subr), Dartmouth College (Wojciech Jarosz) and Université de Lyon, LIRIS/CNRS (Victor Ostromoukhov & David Coeurjolly). We have also collaborated on an external project from LIX, Ecole Polytechnique, Paris (Maks Ovsjanikov). Internally, we have and are collaborating with MPI-INF members (Karol Myszkowski, Hans-Peter Seidel, Piotr Didyk currently at the University of Lugano), including postdocs (Hyunseung Yu, Okan Tursun) working in perception for VR/AR devices.
27 Daria Stepanova: Semantic Data

Group Development

The group *Semantic Data* was established in November, 2017. Its current members are Daria Stepanova, Mohamed Gad-Elrab and Vinh Thinh Ho. During the reporting period, Daria Stepanova has accepted a research scientist position at the Bosch company. However, she continues to supervise the PhD students Mohamed Gad-Elrab and Vinh Thinh Ho at the Max Planck Institute for Informatics. Both students are successful in their research and have published several papers at top-tier conferences related to semantic technologies.

Other members of the group during the reporting period were a master student Hai Dang Tran, who graduated in 2017, as well as Thomas Pellissier Tanon and Mahak Sarin who completed their internships in 2017 and 2018 respectively.

Vision and Research Strategy

Means for enriching data with semantics and methods for reasoning about it are at the heart of any intelligent system. Especially now, with the maturation of knowledge graphs (KGs) and their popularity in applications, such as semantic web search, the need for deductive and inductive reasoning services is strongly increasing.

Our group deals with a variety of issues in this context. We work on new methods for knowledge representation and their applicability to web data. We are concerned with handling incomplete and inconsistent information, data repair, diagnostic reasoning, knowledge revision. Furthermore, we aim at combining learning and reasoning techniques for semantically-enhanced data processing in a variety of applications. In particular, we develop inductive learning approaches for deriving hidden insights from knowledge graphs and textual resources in the form of logical rules. Moreover, we exploit such rules to perform semantic fact checking over web sources. Finally, we are also concerned with making use of declarative rule languages in the context of pattern mining.

Research Areas and Achievements

The research of the *Semantic Data* group is broadly divided into three subfields: *rule learning*, *semantic fact checking*, and *declarative pattern mining*.

**Rule Learning.** Recent developments in information extraction have enabled the automatic construction of large knowledge graphs, e.g., DBpedia or YAGO, which capture information about the world. As KGs are automatically created, they are often incomplete.
To tackle the incompleteness issue in KGs we develop sophisticated rule learning approaches by adapting inductive logic programming and relational learning techniques for our purposes. More specifically, given a KG and possibly additional meta information about it, such as the number or missing facts of certain type acquired, e.g., from the web, our goal is to extract data correlations from the KG and cast them into logical rules (possibly with negation). Such rules can be effectively applied to deduce potentially missing facts thus completing the KG. Dynamic exploitation of feedback from external sources (e.g., precomputed embedding models) makes our methods general and allows for harnessing unstructured web sources for rule learning.

**Semantic Fact Checking.** Fact checking is a crucial task for accurately populating, updating and curating knowledge graphs. Manually validating candidate facts is time-consuming. Therefore, to better support KG curators in deciding the correctness of the candidate facts, we develop novel methods for finding semantically related evidence in web sources and the underlying KG, and for computing human-comprehensible explanations for facts. The key for detecting semantic evidence is intensional background knowledge in the form of rules, which can be either manually stated by domain experts or automatically extracted from the data.

**Declarative Pattern Mining.** Availability of vast amounts of data from different domains has led to an increasing interest in the development of scalable and flexible methods for data analysis. An important feature of flexible data analysis methods is their ability to incorporate user’s background knowledge and different criteria of interest. Such criteria are often provided in the form of constraints to the valid set of answers.

In our group we develop hybrid methods for semantically-enriched data analysis which combine highly scalable dedicated algorithms for pattern retrieval with flexible rule-based declarative languages such as answer set programming for incorporation of background knowledge. We focus on various types of patterns ranging from itemsets to sequences and graphs as well as apply our techniques to more sophisticated data mining tasks such as approximate tiling.

**Projects and Cooperations**

Within the department, we collaborate with Prof. Dr. Pauli Miettinen, Dr. Simon Razniewski and Dr. Paramita Mirza.

Externally, our main collaborators are Prof. Dr. Evgeny Kharlamov (Oxford University), Dr. Sergey Paramonov (KU Leuven) and Dr. Francesca A. Lisi (University of Bari).

**Prizes and Awards**

- Distinguished student paper award at International Semantic Web Conference (ISWC 2017)
- Best poster award at Rules and Reasoning systems conference (RuleML+RR 2017)
The group Text Analysis was established in November 2017. It’s head, Jannik Strötgen, joined the Bosch Center for Artificial Intelligence in March 2018.

Vision and Research Strategy

Despite the development of knowledge bases and their improvements in recent years, most human knowledge is still available only in unstructured format, in particular as natural language text. Thus, applications such as search engines and question answering systems benefit from knowledge extracted from texts. Our group aims at developing natural language processing tools for extracting valuable information from large document collections. In particular, we tackle the extraction and interpretation of temporal information, as time is an important dimension in any information space. For instance, we keep on extending and improving the temporal tagger HeidelTime. We also work on semantically refined tasks: we develop a time-aware search engine, which allows to formulate queries with temporal constraints on the documents’ content, and perform exploratory corpus analysis. The field of digital humanities offers further opportunities. Instead of studying interesting phenomena based on manually analyzed examples, applying natural language processing techniques helps to perform large-scale analyses on big corpora and can lead to new insights in the respective research areas, such as literary science.

Projects and Cooperations

The research group maintains loose cooperations in the following fields: On temporal information extraction topics, we collaborate externally with Anne-Lyse Minard, Manuela Speranza, Bernardo Magnini (all Fondazione Bruno Kessler, Trento, Italy) and Andreas Spity and Michael Gertz (both Heidelberg University). In the area of digital humanities, we collaborate externally with Evelyn Gius (University of Hamburg), and Nils Reiter and Marcus Willand (both University of Stuttgart).

Prizes and Awards

Jannik Strötgen won a Best Reviewer Award at the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP).
29 Thomas Sturm: Arithmetic Reasoning

Group Development

Thomas Sturm is as a Research Director at French Centre national de la recherche scientifique (CNRS) since 2016. His primary affiliation is the Veridis group at LORIA Nancy, France, a joint research facility of CNRS, the University of Lorraine, and Inria. Veridis is in turn a joint Inria group with RG1 at MPI-INF. In the Arithmetic Reasoning unit, Thomas Sturm and Christoph Weidenbach are currently co-supervising two doctoral students, M. Bromberger and M. Voigt. Former members since the foundation of the unit in 2013 include M. Jaroschek (now TU Wien, Austria) and M. Košta (now Slovak Academy of Sciences).

Vision and Research Strategy

Contemporary research on arithmetic in the context of first-order automated reasoning takes place in several fields of research. On the one hand, there are algebraic decision procedures for first-order sentences with fixed signatures and fixed semantics; examples are Collins’s and Weispfenning’s procedures for the reals or Seidenberg-based methods for differential algebra. Research in SMT solving considers similar procedures in a specialized setting with a different focus. On the other hand, there is free first-order proving assigning no semantics to the occurring functions and predicates but expecting them to be axiomatized in the input. Resolution-based methods are a key approach here, for which numerous improvements have been unified within the superposition framework. The Arithmetic Reasoning unit aims at integrating expertise from all relevant communities and to break barriers in both communities and thinking. Our fundamental theoretical research is generally accompanied by robust implementations of our methods. Thomas Sturm is the principle developer of the computer logic system Redlog focussing on interpreted first-order logic. Redlog can optionally be driven by the SMT solver veriT developed at the LORIA Nancy branch of the Veridis group. The superposition prover SPASS by Christoph Weidenbach provides an additional software platform admitting also free first-order logic.

Projects and Cooperations

SYMBIONT (Symbolic Methods for Biological Networks, 2018–2021)

SYMBIONT is an interdisciplinary project ranging from mathematics via computer science to systems biology and systems medicine. The project has a clear focus on fundamental research on mathematical methods, and prototypes in software, which are in turn benchmarked against models from computational biology databases. The principal approach is to combine
symbolic deduction methods with model reduction methods for the analysis of biological networks. In order to cope more effectively with the parameter uncertainty problem we impose an entirely new paradigm replacing thinking about single instances with thinking about orders of magnitude. Our computational methods are diverse and involve various branches of mathematics such as tropical geometry, real algebraic geometry, theories of singular perturbations, invariant manifolds and symmetries of differential systems.

**ANR-17-CE40-0036/DFG-391322026 PRCI. Consortium: T. Sturm (Coordinator, Nancy, France), A. Weber (Coordinator, Bonn, Germany), F. Boulier (Lille, France), F. Fages (Saclay, France), D. Grigoriev (Lille, France), O. Radulescu (Montpellier, France), A. Schuppert (Aachen, Germany), W. Seiler (Kassel, Germany), S. Walcher (Aachen, Germany)**

**SC-SQUARE (Satisfiability Checking and Symbolic Computation, 2015–2018)**

Symbolic Computation is concerned with the algorithmic determination of exact solutions to mathematical problems. Recent developments in the area of Satisfiability Checking tackle similar problems with different algorithmic and technological approaches. Corresponding software addresses some prevailing problems with a direct effect to our society. For instance, Satisfiability Checking is an essential backend for assuring the security and the safety of computer systems. Bridges between the communities in the form of common platforms and roadmaps are necessary to foster an exchange, and to support and to direct their interaction. SC-SQUARE has initiated a wide range of corresponding activities, identified common challenges, offered global events and bilateral visits, and proposed standards.

**EU H2020-FETOPEN-2015-CSA 712689. Consortium: J. Davenport (Coordinator, Bath, UK), E. Abraham (Aachen, Germany), A. Bigatti (Genova, Italy), B. Buchberger (Linz, Austria), A. Cimatti (Trento, Italy), M. England (Coventry, UK), P. Fontaine (Nancy, France), S. Forrest (Jena, Germany), D. Kröning (Oxford, UK), W. Seiler (Kassel, Germany), T. Sturm (Saarbrücken, Germany)**

**SMART (Satisfiability Modulo Arithmetic Theories, 2014–2017)**

SMART has adapted state-of-the-art arithmetic decision procedures for use within an SMT context in order to support formal system verification. SMT solvers, are successfully used, e.g., for program verification, bounded and parameterized model checking, or test generation. Computer algebra systems allow more advanced reasoning in fragments of arithmetic. The combination of both approaches has several advantages: The arithmetic solver need not handle propositional operators; the SMT solver gets access to a sophisticated infrastructure for handling expressive arithmetic fragments; users benefit from the modular architecture of SMT solvers where arithmetic can be combined with other relevant theories.

**ANR-13-IS02-0001/DFG-242816019 Blanc. Consortium: P. Fontaine (Coordinator, Nancy, France), T. Sturm (Coordinator 2014–2016, Saarbrücken, Germany), M. Sagraloff (Coordinator 2016–2017, Saarbrücken, Germany), S. Merz (Nancy, France), L. Voisin (SYSTEREL, France)**

**Prizes and Awards**

- First place at the SMT Competition 2017 for nonlinear real arithmetic in Redlog
- Ramón y Cajal Fellowship by the Spanish Ministry of Science and Innovation (2008)
30 Paul Swoboda: Combinatorial Computer Vision

Group Development

The Combinatorial Computer Vision group was established in March 2018 as a junior research group in the department of Computer Vision and Multimodal Computing. Its members are Paul Swoboda, the group leader, together with the PhD students Jan-Hendrik Lange and Andrea Hornakova.

Vision and Research Strategy

Our aim is to advance the state-of-the-art in optimization algorithms for computer vision and machine learning. Our emphasis lies on discrete optimization problems. Our research strategy is geared towards generally applicable algorithmic design principles, which benefit a wide variety of discrete optimization problems. Towards this end our group studies new and established algorithmic procedures from a theoretical point of view and makes them usable as part of high-performance computer routines. We work on approximative techniques that work well in practice as well as on exact methods that can obtain globally optimal solutions to various optimization problems. Our work is connected to a large body of work in the combinatorial optimization, constraint satisfaction community on the one hand as well as to machine learning and computer vision research on the other hand.

Research Areas and Achievements

Our research areas can be categorized into the two subfields of efficient algorithms and optimality techniques for discrete optimization problems. We also make our effort accessible by developing generally usable algorithmic libraries and by collecting benchmark datasets for various discrete optimization problems (discrete tomography, graph matching, multicut) to advance the state-of-the-art in optimization algorithms.

Efficient Algorithms. Since optimization problems in computer vision and machine learning can have up to billions of variables, the need for algorithms with low complexity and good scalability, yet delivering high quality solutions, arises. Additionally, it is important to have general principles for designing efficient algorithms instead of proposing ad-hoc approaches that only work for specific problems.

A first line of work concerns message passing algorithms for Lagrange decompositions. We study generally applicable message passing techniques that can be used for a wide range
of discrete optimization problems. Specifically, we have validated our algorithmic design on graph matching, multigraph matching, multicut and cell tracking for biological image analysis.

Another line of work concerns proximal bundle solvers for discrete optimization problems. While in general slower than message passing, they have stronger theoretical guarantees in terms of convergence rates and benefit from recent work on efficient Frank-Wolfe methods in machine learning.

The work has been published in CVPR 2017 (3 times).

**Optimality techniques.** Almost all discrete optimization problems that appear in computer vision and machine learning are NP-hard. However, due to the structural simplicity of most problems we encounter in computer vision, we can often make verify global optimality of at least part of the solution in polynomial time. Such restricted optimality certificates establish partial optimality or persistency. We study partial optimality certificates and algorithms providing them efficiently for Markov Random Fields, the multicut, max-cut, and graph matching.

A second approach towards optimality concerns approaches to obtain globally optimal solution by combining efficient message passing algorithms and exact, but slower, branch-and-bound solvers. The underlying principle consists of identifying the easier parts of the problem that can be solved exactly with message passing and leaving the hard rest to any exact solver. The results on the subproblems solved by either algorithm can then be combined into a globally optimal solution.

The work has been published at NIPS 2013, CVPR 2014, CVPR 2015, TPAMI 2015 and AAAI 2018.

**Projects and Cooperations**

Internally, we cooperate at MPI-INF with Professor Bernt Schiele and the group of Professor Christian Theobalt at the Graphics, Vision and Video group. We also cooperate with the Professor Carsten Rother’s and Professor Stefania Petra’s groups at Heidelberg University, Björn Andres from the Combinatorial Image Analysis group at the Max-Planck Institute for Informatics and the Bosch Center for Artificial Intelligence, Florian Jug from the Max-Planck Institute of Molecular Biology and Genetics (MPI-CBG) and Vladimir Kolmogorov from the Austrian Institute of Science and Technology.
31 Christian Theobalt: Graphics, Vision & Video

Group Development

Christian Theobalt (h-index: 59, 10815 Cit. (since 2014 – h-index: 48, 6896 Cit.); Google Scholar: Feb 15, 2019), is a tenured professor at the MPI for Informatics heading the research group Graphics, Vision & Video (GVV). He is also a professor of computer science at Saarland University. Among his research awards are two ERC grants. He is a co-founder of the Captury (see Sect. 38.10.1), a spin-off from his lab, that commercializes state-of-the-art marker-less motion and performance capture techniques developed in his group.

Currently, the GVV group consists of four postdocs (Weipeng Xu, Vladislav Golyanik, Florian Bernard, Mohamed Elgharib), ten PhD students (Ikhsanul Habibie, Marc Habermann, Abhimitra Meka, Edgar Tretschk, Hyeongwoo Kim (thesis under review), Ayush Tewari, Dushyant Mehta, Franziska Mueller, Jiayi Wang, Mallikarjun Br) and a research engineer (Oleksander Sotnychenko). The team of Gerard Pons Moll is affiliated. In the reporting period, one PhD student graduated, two handed in their thesis (2017, Pablo Garrido (now postdoc at Technicolor); Nadia Robertini (2018, thesis submitted), Hyeongwoo Kim (2019, thesis submitted)). Several former GVV students/postdocs accepted new positions in the reporting period, for instance: Dr. Dan Casas (now assistant professor at Univ. Madrid), Dr. Kwang In Kim (now full professor at UNIST Korea), Dr. Michael Zollhoefer (now Visiting Assistant Profressor at Stanford), Dr. Avishek Chatterjee (now Assistant Professor at IIT Kharagpur). Four Master students, and two research immersion lab students graduated from the group. We hosted several short and long term research visitors, e.g.: Lingjie Liu (U. Hong Kong), Linjie Lv (Tsinghua University), Lucas Thies (Erlangen University), Patrick Perez (Technicolor), Tianteng Bi (Beijing Insitute of Technology), Mario Botsch (U. Bielefeld), Angela Dai (Stanford), Thabo Beeler (ETHZ), Johan Thunberg (U. Luxemburg), Wenping Wang (U. Hong Kong), Markus Oberweger (TU Graz), Lan Xu (HKUST), Michael Neff (UC Davis), Theo Thonat (INRIA).

Vision and Research Strategy

The research group Graphics, Vision & Video investigates fundamental algorithmic questions at the intersection of computer graphics, computer vision and machine learning. We investigate new ways to systematically combine the insights from computer vision on reconstruction and scene interpretation, with the expertise in graphics on the forward process of efficiently simulating and representing complex visual scene models. We also explore architectural synergies between these visual computing paradigms and machine learning concepts, even new ways to deeply integrate both in new types of end-to-end trainable architectures. We strive
for new algorithms that can learn and jointly refine their algorithmic structure and employed model representations on large corpora of sparsely or unlabeled real world data. Driving these concepts forward will lead to 3D and 4D scene reconstruction, scene interpretation and scene synthesis algorithms of previously unseen robustness, efficiency, accuracy, generality, scalability, semantically meaningful controllability, and advanced explainability. The new methods will advance computer graphics and computer vision in general, and provide new insights relevant to machine learning and human-computer interaction. But their relevance extends into a rapidly growing number of more general research and application areas that will profoundly transform our future lives. For example, they will revolutionize creative technology to create and edit visual content, as well as, future virtual and augmented realities. They will build important foundations of essential, currently unavailable, visual perception abilities needed by future intelligent computational and autonomous systems that need to understand and interact with the general human world.

**Research Areas and Achievements**

The GVV group investigates basic algorithmic problems in four primary research areas.

**Reconstructing the Real World in Motion:** We advanced the state of the art on marker-less human motion and performance capture, as well as general dense 4D reconstruction of the real world in motion, to which we made important contributions in the past, along several dimensions: generality of scenes that can be handled, the accuracy and quality, efficiency and robustness, and the simplicity of sensors needed. Notably, we researched entirely new ways of fusing and deeply integrating model-based and deep learning-based scene reconstruction.

For instance, our team developed entirely new approaches fusing machine learning and generative model-based reconstruction for state-of-the-art multi-view marker-less human motion capture. They succeed in outdoor scenes with challenging lighting using a very low number of sensors. We also presented the first approach for real-time 3D motion capture of full body motion from a single color camera. It combines a new CNN using a new tailored location map scene representation with a model-based skeleton fitting approach. Egocentric real-time pose estimation with a head-mounted fisheye camera is also feasible, for the first time, by our work. We further extended these concepts to enable new methods for multi-person 3D pose estimation from monocular video that, by means of a new posemap formulation in a CNN, achieves state-of-the-art accuracy and unseen real-time performance. The group also contributes widely used data sets (MPI-INF-3DHP, MuCo-3DHP, MarCONI) to train and test monocular 3D pose estimation methods.

In the past, we made important contributions to marker-less 3D hand tracking. In the reporting period, we developed the first approach to capture, in real-time, the 3D hand in interaction with objects in cluttered scenes with a single depth camera. We also developed a pioneering approach for real-time 3D hand motion capture from a single color camera. Further, we presented new ways for HCI and on-body interaction with hand tracking.

We presented new robust formulations for general deformable shape capture from monocular video and presented the first methods for reconstruction of high quality static shape and texture of a human template model from monocular color video. The group developed
new formulations for high-quality multi-view performance capture of humans in general apparel that succeed in more general uncontrolled scenes. Our research also resulted in the first approach for full performance capture of 3D pose and deforming surface geometry of a human in general everyday apparel from a single color video. It combines model-based and deep learning-based reconstruction. In follow-up work, we presented the first method to do full 3D human performance capture from monocular video even in real-time.

We also presented widely cited methods to capture dynamic face geometry, expression, appearance and illumination from monocular video. An entirely new way of integrating a CNN and a model-based reconstruction approach in an end-to-end-trainable architecture enabled face reconstruction at unseen speed and accuracy from single images, while even enabling training on unlabeled face images. Follow-up work showed how we can even learn a full parametric face model from community image and video data.

We also maintain one of the largest repositories of reference data sets for general static and dynamic scene reconstruction, and marker-less motion and performance capture, GVVPerf-CapEva (see Sect. 38.10.2).

Large-scale, High-quality 3D Reconstruction with Lightweight Sensors: In the reporting period we continued to contribute state-of-the-art approaches for high quality static scene reconstruction with single (depth)cameras. We presented a new, widely used, algorithm for real-time dense 3D scanning and bundle adjustment with an RGB-D camera, and presented one of the first methods to scan very thin geometric structures with a consumer RGB-D camera by using geometric curve structures as fusion primitives. We further developed a new dictionary-based approach to 3D geometry denoising, as well as a state-of-the-art method for object retrieval and pose estimation from single color images.

Computational Videography and Inverse Rendering: The group continued to contribute new methods for advanced video processing and computational videography. For instance, we showed that by means of combining advanced monocular performance capture with new methods for neural network based video synthesis, entirely new ways of generating and editing human face portrait videos, as well as videos of entire humans at previously unseen photo realism as well as computational performance can be achieved. Beyond use in creative media generation, our approaches also pave the way for new ways of photo-realistic VR and AR experiences, such as photoreal headset removal for VR telepresence, or human avatar reenactment for video-realistic telepresence. We also developed new advanced approaches for inverse rendering, i.e., estimation of lighting and reflectance from monocular color and depth video. Examples are the first real-time method for BRDF estimation from monocular RGB, as well as the first real-time approach for live user-guided intrinsic video.

Foundational Algorithms for Visual Computing and Machine Learning: We investigate foundational algorithms of cross-cutting relevance for real world reconstruction, real world synthesis and inverse rendering algorithms. For instance, we developed new scalable and efficient approaches for solving general matching problems, as well as developed new programming tools to implement efficient GPU solvers for high dimensional non-convex energy minimization problems in visual computing. We also investigate new ways to design memory-
efficient yet performant learning architectures, in particular CNNs, new ways to better understand and explain their emergent structural properties, and new ways to combine them with expert-designed algorithms and model representations in end-to-end trainable architectures.

**Publications and Grants**  In the reporting period, we published more than 45 peer-reviewed publications, including many at the top publications venues (SIGGRAPH/SIGGRAPH Asia/ACM TOG (10), EUROGRAPHICS (2), TVCG (2), CVPR (12), ICCV (2), ECCV (1), 3DV (4), PAMI (2), CHI(1), IJCV (2)). Since 2011, Christian Theobalt acquired nearly 6 Mio. Euros in external funding. In the reporting period, this included: The ERC Starting Grant *CapReal*, the ERC Consolidator Grant *4DRepLy*, and multiple projects funded through the Intel VCI, Oculus / Facebook, and Technicolor.

**Projects and Cooperations**

We closely collaborate with colleagues on Saarland Informatics Campus, e.g., the groups of: Bernt Schiele (MPI-INF), Gerard Pons-Moll (MPI-INF), Mario Fritz (MPI-INF), Jürgen Steinle (UdS), Philipp Slusallek (UdS), and Joachim Weickert (UdS).

We have collaborated with several external partners, e.g., with: Dr. Patrick Perez (Technicolor, *Advanced Videography*), Dr. Shahram Izadi (Google, *3D and 4D Reconstruction*), Prof. Matthias Niessner (TUM, *3D/4D Reconstruction, Face Capture, VR/AR*), Prof. Pat Hanrahan (Stanford, *Programming Tools*), Dr. Kalyan Sunkavalli (Adobe Research, *Face Appearance Capture*), Prof. Antti Oulasvirtta (Aalto University, *Gesture-based Human-Computer Interaction*), Prof. Marc Stamminger (Univ. Erlangen, *3D and 4D Reconstruction, Inverse Rendering, Image-based Rendering*), Prof. Michael Zollhöfer (Stanford, *4D Reconstruction, Video-based Rendering and Editing*), Prof. Maneesh Agrawala (Stanford) and Adam Finkelstein (Princeton) on *Model-guided Video Editing*, Prof. Wenping Wang (Univ. Hong Kong, *Performance Capture and Neural Rendering*), Prof. Niloy Mitra (UCL, *RGB-D Reconstruction*), Prof. Christian Richardt (Univ. Bath, *User-centric Video, Computational Videography*), Prof. Kwang In Kim (UNIST Korea, *Neural Network Design*), Prof. Pascal Fua (EPFL, *Marker-less Human Pose Estimation*), Dr. Carsten Stoll and Ron Mallet (Facebook Reality Labs) on *Deformable Capture*, Prof. Vincent Lepetit (U. Bordeaux) and Andrea Tagliasacchi (Google) and Dr. Kwang Moo Yi (U. Victoria) on *Hand Segmentation*, Prof. Marcus Magnor (TU Braunschweig, *Video-based Reconstruction of People Models*).

**Prizes and Awards**

In the reporting period, Christian Theobalt was awarded an *ERC Consolidator Grant* (2017) and received the renowned *Karl Heinz Beckurts Award* (2017). Franziska Müller received a *Google PhD Fellowship* (2017), and the national *Women STEM Award* (2017). Franziska Müller and Christian Theobalt received the *Best Academic Paper Award* at ACM ISS 2018 (together with co-authors from UdS). Marc Habermann received the *Günter Hotz Medal* of Saarland University for the best Master students in computer science (2017). Pablo Garrido received the *Eurographics PhD Dissertation Award* (2018).
Group Development

The research group Exploratory Data Analysis was established in October 2013 at the Cluster of Excellence MMCI. The group is being led by Dr. Jilles Vreeken, who during the reporting period accepted an offer for a tenured faculty position at the CISPA Helmholtz Center for Information Security. The group currently consists of seven PhD. students, Kailash Budhathoki, Sebastian Dalleiger, Jonas Fischer, Janis Kalofolias, David Kaltenpoth, Panagiotis Mandros and Alexander Marx. Between October 2015 and September 2018 Dr. Mario Boley was a post-doc in the group, who then took an offer as Lecturer (Asst. Prof) at Monash University in Melbourne, Australia.

In addition, six Master students are currently working on their thesis, and two further students are working as student research assistants in the group. During the reporting period eleven M.Sc. and B.Sc. theses have been completed.

Vision and Research Strategy

Thanks to modern technology, collecting and storing massive amounts of complex information has become fairly straightforward. Examples include astronomical observations, records of hospital patients, as well as experimental data in the life-sciences. The complexity and the sheer volume of this data has led to the situation where analysis of the data – the actual goal of collecting and storing all of it – has become increasingly difficult, if not sheer impossible.

Our group’s research goal is to solve this problem. That is, our research is in the fields of data mining and machine learning, with a focus on exploratory data analysis. We aim to develop theory and methods by which we are able to efficiently identify the most interesting and useful sub-structures in data of any size or shape. We study a wide range of problems, from discovering non-linear dependencies, mining small sets of informative patterns, to inferring causal relations and models from empirical data.

Research Areas and Achievements

Correlation and Causality  A key research topic of the group is concerned with measuring correlation and inferring causality. That is, how can we automatically and without having to set any parameters or expectations identify those attributes that interact strongly, and how can we do so on large and high-dimensional data. Correlation does not equate causation, yet one of the main goals in science is to identify causal relations. We investigate how we can reliably identify causal relations and models from empirical data. We study techniques for
telling cause from effect between bivariate pairs of random variables, discovering causal networks over joint distributions, as well as determining whether a seemingly causal dependency is actually due to a hidden confounder.

**Association and Summarisation**  The second key goal of the group is to develop theory and methods for summarizing the main structure of given datasets in easily understandable terms. Toward this goal, we are both investigating how to best define objective functions that capture this goal, as well as how to find optimise these often highly complex and non-convex scores in order to find good summaries for the data at hand. We are investigating various types of data, including simple tabular data, including binary, categorical and continuous real-valued; univariate or multivariate, continuous or discrete time-series, such as sensor or text data; as well as large graphs from various domains.

**Detection and Exploration**  The third topic we investigate is that of high-lighting what stands out in your data. This includes anomaly detection, discovering those parts of your data that most likely were generated by a different process, which has applications in error detection, preventive maintenance, as well as discovering there could be multiple processes behind your data. In addition, we study methods for interactive data exploration including techniques to steer attention to identify either those regions that are highly similar to what the user has identified as interesting, or instead, what is highly deviating and surprising given this history.

**Projects and Cooperations**

Within MPI-INF, Jilles Vreeken is collaborating with Dr. M. Schulz (D3 and MMCI, now at U. Frankfurt) on the topic of finding patterns in genomics data.

On campus, Jilles Vreeken is collaborating with Prof. D. Klakow (UdS) on finding patterns in morphologically rich languages, Prof. H.-P. Lenhof (UdS) on discovering causal graphs from genomics data, Prof. C. Rossow (CISPA) on tracing back command-and-control nodes in botnets, Prof. A. Zeller (CISPA) on discovering patterns in crash-causing inputs.

In addition, our main collaborators within Europe include Dr. F. Bonchi (ISI Foundation), Prof. T. De Bie (Ghent U), Dr. L. Ghiringhelli (FHI Berlin), Prof. A. Gionis (Aalto U), Dr. D. Janzing (Amazon), Prof. M. Scheffler (FHI Berlin), Prof. A. Siebes (Utrecht U), Prof. N. Tatti (Helsinki U), Prof. M. van Leeuwen (Leiden U), Prof. I. Zliobaite (Helsinki U), and outside Europe include Prof. L. Akoglu (CMU), Prof. M. Boley (Monash), Prof. C. Faloutsos (CMU), Dr. E. Kiciman (MSR), Prof. D. Koutra (U Michigan), Prof. B.A. Prakash (Virginia Tech), Prof. K. Yamanishi (Tokyo U).

**Prizes and Awards**

Jilles Vreeken won the IEEE ICDM 2018 *Tao Li Award* for Excellence in Research.

Panagiotis Mandros, Mario Boley, and Jilles Vreeken won the IEEE ICDM 2018 *Best Paper Award* for *Discovering Reliable Dependencies from Data: Hardness and Improved Algorithms.*
Group Development

The group **Searching, Mining, and Learning with Informal Text** was established in September 2018. The group is headed by Andrew Yates, involves the PhD students Anna Tigunova (jointly supervised with Gerhard Weikum), Ghazaleh Haratinezhad Torbati and Kashyap Popat, and collaborates with the postdocs Paramita Mirza and Erisa Terolli. The group is currently hosting one intern, Siddhant Arora (IIT Delhi), and the Master’s student Ahmad Ahmadian is currently writing his thesis in the group.

Vision and Research Strategy

In contrast to authoritative information sources, like encyclopedias, news articles, and academic papers, much of the information available on the Web is contained in informal text that requires different strategies to interpret. The group aims to develop methods for searching, mining, and learning with such text so that it may be integrated with other knowledge. This goal spans both information retrieval and natural language processing tasks, such as mining health-related claims from social media, extracting information from dialogues, and learning to identify relevant spans of text.

Research Areas and Achievements

**Extracting information from dialogues and social media** While conversational agents are becoming increasingly popular (e.g., Amazon Alexa, Apple Siri), creating agents capable of holding personalized conversations remains a challenge. Ideally, such agents would customize their interactions based upon knowledge about the user while simultaneously extracting new personal information from the user’s utterances. We are investigating methods for extracting such information with the goal of developing personal knowledge bases that can be used in downstream applications, such as conversational agents and personalized search. We have proposed a family of neural models, called Hidden Attribute Models, for inferring a user’s attribute values based upon their utterances (e.g., “I write code in Redmond” implies the author is a software engineer working for Microsoft). Our current work involves making such Hidden Attribute Models more general, in order to relax the requirement that all possible attribute values are observed during training.
Neural matching and ranking  In the area of information retrieval, we have proposed several neural retrieval models for improving ranking performance and performing new retrieval tasks. Our PACRR model was the first neural retrieval model to effectively use positional information (e.g., term dependency). We improved upon PACRR with the Co-PACRR model, which incorporates components for capturing several types of context, and adapted it to the task of complex answer retrieval in the context of the TREC 2017 and 2018 Complex Answer Retrieval track. Currently, we are exploring methods for further improving neural retrieval models through the use of information retrieval axioms and heuristics. On the application side, we are investigating methods for personalizing health search and plan to explore ways in which a personal knowledge base can be used to improve ranking.

Analyzing health forums  The goal of this project is to develop tools for studying health conditions using social media. This encompasses both methods for identifying users suffering from a health condition (“diagnosed users”) and models for differentiating between diagnosed users and control users. This project is motivated by the observation that mental health has a unique connection to language. That is, mental health problems differ from many other medical conditions in that they cannot be diagnosed with methods like imaging techniques (e.g., x-rays) or laboratory analyses (e.g., blood tests). Thus studying patients’ language use has potential for achieving a better understanding of the underlying disease. We have developed pattern-based approaches for identifying social media users who claim to be suffering from a mental health condition (e.g., PTSD, depression) and matching them with appropriate control users. Using this data, we developed a neural network-based approach for differentiating between depressed users and control users based upon differences in their language in general forums (i.e., ignoring health-related forums and statements). While this approach was successful, it is not grounded in clinical knowledge and thus is difficult for doctors to interpret. We are currently exploring methods for identifying diagnosed users that closely related to the diagnostic criteria used by clinical doctors (e.g., abnormal sleep behavior, low energy, trouble concentrating, reduced enjoyment of pleasurable activities, etc).

Projects and Cooperations

In addition to collaborations with other researchers in the department, the group maintains external collaborations with Georgetown University (Nazli Goharian’s group), University of Chinese Academy of Sciences (Ben He’s group), Blekinge Institute of Technology (Michael Unterkalmsteiner), Rutgers University (Gerard de Melo), and the U.S. National Institutes of Health (Ayah Zirkily).

Prizes and Awards

The paper Depression and Self-Harm Risk Assessment in Online Forums received a “best paper award” at EMNLP 2017. The paper SMHD: A Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions was recognized as an “area chair favorite” among the resource papers at COLING 2018. Andrew Yates received a “best reviewer award” at EMNLP 2018.
Vision and Research Strategy

As the computing landscape is being reshaped by the dramatic shift towards ubiquitous parallelism, and by the sheer scale of data, extracting performance from existing applications gives rise to formidable challenges. In digital geometry processing, this problem gets amplified by the unstructured nature of data (e.g. meshes) and by the serial nature of prevalent algorithmic solutions. As a result, the high performance promise of modern hardware seems elusive.

The thrust behind our research effort is to rethink solution development with parallelism built-in from the ground up as opposed to the current trend of looking at parallelization as a mere post processing step left for “Ninja coders” or automatic optimization software. We examine the whole solution development pipeline (problem abstraction, data structures, mathematical modeling, and implementation), with the goal of producing simple yet streamlined solutions capable of unleashing the computational power afforded by modern hardware. We seek to make high performance in digital geometry processing accessible at a high level of abstraction without requiring practitioners to delve into the low level intricacies of parallelism. To facilitate deployment, our solutions need to be independent of the underlying granularity, i.e. the same code can be operated on clusters as well as on single graphics cards (GPUs). Our efforts are guided by parsimonious memory usage, improved memory access patterns, and cache performance. To avoid data redundancy, we target seamless transition between data traversal and numerics as it is a valuable asset for coding convenience and enhanced performance.

Research Areas and Achievements

The research within the group targets three complementary areas: data structures, numerical kernels, and applications.

Data structures. For unstructured meshes, we have introduced the mesh matrix, a lean sparse matrix representation which readily brings forward memory layout. We have endowed this representation with action maps, which symbolically augment standard sparse linear algebra kernels to perform both traversal and numerical computations on meshes. For dynamic scenarios, we have developed a novel hierarchical sparse matrix data structure and we have introduced a fully-dynamic graph data structure capable of intensive edge and vertex updates within the limited memory available on the Graphics Processing Unit (GPU).
Numerical kernels. Although efficient sparse matrix linear algebra kernels exist for different hardware architectures, e.g Intel MKL and Nvidia cuSPARSE, these closed source libraries focus on standard algebra primitives. Highly efficient mesh operations based on matrices will also make use of these primitives, but they are not sufficient for carrying advanced mesh processing tasks efficiently. In our formalization, mesh operations require a sequence of matrix operations, often operate locally on data, and change mesh connectivity. These operations pose new challenges for sparse matrix linear algebra formulations as high performance requires the fusion of successive matrix operations, considering cache locality and fast on-chip memory when performing mesh-local operations, and considering alternative data representations that allow for efficient changes to the matrix and thus the mesh structure. Our starting point is the development of sparse efficient matrix-vector and sparse matrix-matrix multiplication for the GPU. Our recent results on the entire University of Florida Sparse Matrix Collection, comparing against vendor provided implementations and state-of-the-art approaches are highly favorable.

Applications. So far, we have targeted some of the most commonly used and far reaching applications across various fields. All of which are believed to be hard to parallelize. In computer graphics, we addressed the problem of subdivision surfaces, a widely used modeling paradigm where an extremely simple polygonal shape representation is enriched throughout a combination of vertex insertions and smoothing iterations, which yield visually pleasant meshes suitable for use in production. In meshing, we have targeted the generation of Voronoi diagrams on surfaces. Voronoi Diagrams are a common theme across several disciplines and are arguably one the most studied problems in computational geometry. Their extension to the surface setting further complicates vectorization as metric estimation poses serious challenges. In numerical simulation, we focus on the whole pipeline from finite elements assembly to numerical solution strategies. Numerical evidence suggests that the assembly problem weighs heavily on performance and impedes scalability. Across all these application, we broke the high performance deadlock throughout novel geometric and numerical abstractions and we have shown that it is possible to eliminate the unpleasant idle-time modelers and engineers spend waiting before seeing the outcome of their modeling and simulation efforts.

Projects and Cooperations

Within the department we collaborate with the Image and Geometry group of Renjie Chen. Externally, we collaborate with Prof. Markus Steinberger (TU Graz), and Dr. Hamish Carr (University of Leeds).

Prizes and Awards

Martin Winter, Rhaleb Zayer, Markus Steinberger: Best Paper Award IEEE-HPEC High Performance Extreme Computing, 2017
Part III

Research Units in Detail
35 D1: Algorithms and Complexity

35.1 Personnel

Director
Prof. Dr. Kurt Mehlhorn (since December 1990)

Senior Researchers and Subgroup Coordinators
Dr. Alkmini Sgouritsa (since September 2018)
Dr. Karl Bringmann (since January 2016)
Dr. Christoph Lenzen (since July 2013)
Dr. Andreas Karrenbauer (since January 2013)
Dr. Christian Ikenmeyer (until January 2019)
Dr. Michael Sagraloff (until February 2018)

Researchers (as of March 1st, 2019)
Dr. Pranabendu Misra (since February 2019)
Dr. Stefano Leucci (since January 2019)
Dr. William (Will) Rosenbaum (since September 2018)
Dr. Reuven Hodges (since September 2018)
Dr. Tim Oosterwijk (since September 2018)
Dr. Eunjin Oh (since March 2018)
Dr. Nitin Saurabh (since January 2018)
Dr. Saeed Akhoondian Amiri (since December 2017)
Dr. Marvin Künnemann (since April 2017)
Dr. Pavel Petrov Kolev (since January 2014)
Dr. Antonios Antoniadis (since January 2014)
Dr. Cosmina Croitoru (since October 2012)

Ph.D. Students
Philip Wellnitz (since October 2018)
Johannes Bund (since August 2018)
Ben Wiederhake (since August 2018)
Bhaskar Ray Chaudhury (since June 2017)
André Nusser (since June 2017)
Anurag Pandey (since September 2015)
Maximilian John (since April 2015)
Attila Kinali-Dogan (since January 2015)
Daniel Vaz (since October 2014)
Davis Issac (since August 2014)
Andreas Schmid (since April 2014)
Alexander Kobel (since October 2011)

Secretaries
Ingrid Finkler-Paul
Christina Fries

Former (Recent) Staff
- Dr. Christian Ikenmeyer (until January 2019), Research Group Leader, Max Planck Institute for Software Systems
- Dr. Emanuele Natale (until December 2018), Permanent Researcher, CNRS, I3S, Université Côte d’Azur, INRIA
- Dr. Bojana Kodric (until October 2018), Postdoc, Gran Sasso Science Institute
- Dr. Ruben Becker (until September 2018), Postdoc, Gran Sasso Science Institute
- Gorav Jindal (until September 2018), Visiting Doctoral Candidate, Department of Computer Science, Aalto University
- Dr. Yun Kuen Cheung (until August 2018), Postdoc, Singapore University of Technology and Design
- Dr. Bundit Laekhanukit (until July 2018), Associate Professor, Shanghai University of Finance and Economics
- Dr. Aruni Choudhary (until February 2018), Postdoc, FU Berlin
- Dr. Krzysztof Fleszar (until February 2018), Assistant Professor, University of Warsaw
- Dr. Sandy Heydrich (until February 2018), Research Associate, Fraunhofer Institute for Industrial Mathematics
- Prof. Dr. Michael Sagraloff (until February 2018), Professor, University of Applied Sciences Landshut
- Dr. Kevin Schewior (until October 2017), Postdoc, Universidad de Chile
- Dr. Stephan Friedrichs (until September 2017), TomTom
- Dr. Michael Dirnberger (until August 2017), Senior Data Scientist, KIANA Systems
- Dr. Hang Zhou (until August 2017), Assistant Professor, Ecole Polytechnique
- Dr. Moti Medina (until July 2017), Lecturer, Ben-Gurion University of the Negev
- Dr. Reut Levy (until July 2017), Postdoc, Weizmann Institute
- Dr. Thomas Kesselheim (until March 2017), Junior Professor, TU Dortmund
- Dr. Karteek Sreenivasaih (until February 2017), Postdoc, Saarland University

35.2 Visitors

In the time period from January 2017 to February 2019, the following researchers visited our group:
Jörg-Rüdiger Sack | 01.01.2017–25.01.2017 | CSC Carleton
Fulvio Gesmundo | 04.01.2017–10.01.2017 | University of Copenhagen
Erik Jan van Leeuwen | 20.02.2017–14.03.2017 | Utrecht University
Guy Even | 01.03.2017–04.09.2017 | Tel-Aviv University
Dror Rawitz | 13.03.2017–17.03.2017 | Bar-Ilan University
Hendrik Fichtenberger | 27.03.2017–31.03.2017 | TU Dortmund
Yadu Vasudev | 27.03.2017–31.03.2017 | IIT Madras
Chien-Chung Huang | 21.05.2017–26.05.2017 | ENS Paris
Ioana Bercea | 29.05.2017–31.08.2017 | Tel-Aviv University
Joseph Landsberg | 30.05.2017–04.06.2017 | Texas A&M University
Sunil Chandran Leela | 06.06.2017–30.11.2017 | IISc Bangalore
Ruben Hoeksma | 26.06.2017–30.06.2017 | University of Bremen
Seri Khoury | 15.07.2017–10.10.2017 | Technion
Nikhil Kumar | 16.08.2017–14.09.2017 | IIT Delhi
Francois LeGall | 20.08.2017–25.08.2017 | Kyoto University
Amir Yehudayoff | 20.08.2017–25.08.2017 | Technion
Mathew Francis | 02.09.2017–16.09.2017 | ISI Chennai
Emilio Cruciani | 04.09.2017–02.12.2017 | GSSI L’Aquila
Mohsen Ghaffari | 08.09.2017–12.09.2017 | ETH Zurich
Mayank Goswami | 22.01.2018–26.01.2018 | Queens College NY
Andy Rosen | 27.02.2018–01.03.2018 | University of Michigan
Grigorius Koumoutsos | 05.03.2018–06.03.2018 | TU Eindhoven
Meena Mahajan | 09.03.2018–22.03.2018 | CIT Chennai
Khaled Elbassioni | 29.03.2018–30.03.2018 | Masdar Institute
Alexandru Popa | 01.05.2018–31.07.2018 | University of Bucharest
Utkarsh Bajpai | 01.05.2018–29.07.2018 | VIT University
Pranjal Dutta | 09.05.2018–31.07.2018 | IIT Delhi
Supriya Tiwari | 01.06.2018–30.11.2018 | IISER Pune
Rob van Stee | 12.06.2018–13.06.2018 | University of Siegen
Fabrizio Grandoni | 12.06.2018–13.06.2018 | IDSIA Switzerland
Parinya Chalermsook | 02.07.2018–20.07.2018 | University of Helsinki
Hossein Vahidi | 02.07.2018–29.07.2018 | University of Tehran
Nikhil Balaji | 02.07.2018–07.07.2018 | RWTH Aachen
Wanchote Jiamjitrak | 03.07.2018–14.07.2018 | Aalto University
Iliad Ramezani | 05.07.2018–21.09.2018 | Sharif University
Sina Akbari | 07.07.2018–16.09.2018 | Sharif University
35.3 Group Organization

The whole group meets twice a week for a scientific talk and discussions. We also have reading and discussion groups on special topics. The whole group meets every two weeks to discuss organizational matters. The selection of new postdocs is made by all group members holding a PhD.

35.4 Algorithmic Game Theory and Online Optimization

Coordinators: Alkmini Sgouritsa and Kurt Mehlhorn

Algorithmic game theory and online algorithms study problems in the context of uncertainty. In algorithmic game theory, this uncertainty arises from the interaction of a multitude of rational users in large decentralized systems, such as road-traffic networks, auctions, markets, or economic applications on the internet. The area combines algorithmic insights with concepts and ideas from microeconomics and game theory. One of the main goals is to understand user interaction using game-theoretic models. A fundamental tool is equilibrium computation, where we compute and analyze equilibria, and characterize if and how fast agents can reach an equilibrium. Besides analyzing this centralized notion of stability concepts, we also study if and how distributed dynamics converge to equilibria. In algorithmic mechanism design we ask how to design efficient algorithms that implement desirable goals in the system in the presence of strategic agents.

In online optimization, the uncertainty stems from the fact that an algorithm has to make decisions before having observed the entire input. In this area, we study online algorithms for optimization problems and analyze them using the competitive ratio, which compares the solution quality to an optimal solution with knowledge of the complete instance. For dynamic algorithms we measure the time or the number of structural changes when maintaining a solution in a problem instance that dynamically changes over time.
35.4.1 Equilibrium Computation and Convergence

A fundamental algorithmic approach to microeconomics and game theory is equilibrium computation. Strategic games can be interpreted as descriptions of dynamic systems, and equilibria represent fixed points of these systems. The computation of fixed points is often difficult and can be quite different from the computation of optimal solutions in optimization. In many cases, notions of equilibrium are guaranteed to exist, and then equilibrium computation becomes a search problem (find an equilibrium) and not a decision or optimization problem. Efficient algorithms for equilibrium computation take a compact description of the game as input and return in polynomial time an equilibrium as output. A special case is to see if agents have some means to converge to equilibrium (in polynomial time) using suitable distributed protocols, learning behavior, or dynamics.

The above methods are helpful when a designer has some means to establish an equilibrium, to help researchers in experimental analysis of game-theoretic models, or as a plausibility check if equilibrium concepts could, in principle, always be obtained in reasonable time.

Combinatorial Algorithm for General Arrow-Debreu Markets

Investigators: Bhaskar Ray Chaudhury and Kurt Mehlhorn

The Linear Arrow-Debreu Market model is one of the fundamental market models which involves a set $B$ of $n$ rational agents and a set $G$ of $m$ divisible goods. We will refer to an individual agent by $b_i$ for $i \in [n]$ and to an individual good by $g_j$ for $j \in [m]$. Each agent comes with a basket of goods to the market or in other words agent $b_i$ owns $w_{ij}$ units of good $g_j$ (the total supply of $g_j$ is then $\sum_i w_{ij}$ units). Moreover agents have utilities over the goods and $u_{ij} \geq 0$ is the utility derived by agent $b_i$ from one unit of good $g_j$. The goal is to determine a set of prices for the goods where the total aggregate demand equals the total aggregate supply for each good. Formally we wish to determine a positive price vector $p \in \mathbb{R}^m$ such that every agent spends all of his budget (money spent by agent $i$ equals $\sum_j w_{ij} p_j$) only on the goods that give maximum utility to price ratio (agent $i$ only invests on $\{ j \mid u_{ij}/p_j = \max_{\ell \in [m]} u_{i\ell}/p_{\ell} \}$) and all the goods are completely sold (total money spent on any good $j$ by all agents equals $\sum_i w_{ij} p_j$).

Such a price vector is called a vector of equilibrium prices. The existing algorithms for determining equilibrium prices are either combinatorial or based on ellipsoid and interior point methods. Combinatorial algorithms are more interesting as they yield a deeper understanding of underlying structures in Arrow-Debreu markets. In [1] we design the fastest known algorithm for determining equilibrium prices in Arrow-Debreu markets. Our algorithm is combinatorial and improves upon the previously existing fastest algorithm [2], which is an interior point algorithm, by a factor of $\Omega(n)$. We also study bottleneck instances for all existing combinatorial algorithms. In particular, we construct instances in which the number of arithmetic operations performed by the existing combinatorial algorithms is very high.

References

Amortized Analysis of Asynchronous Price Dynamics

Investigators: Yun Kuen Cheung in cooperation with Richard Cole (Courant Institute, NYU)

In Algorithmic Game Theory and Computational Economics, one important aspect is to provide algorithmic explanation on why Nash and market equilibrium can be achieved in realistic settings. In practice, price updates on different goods in a market are often made in a distributed and non-coordinated manner. In other words, price updates are asynchronous. However, almost all prior analyses made strong synchrony assumptions. We present an amortized analysis which, for the class of CES Fisher markets, can demonstrate convergence toward market equilibrium amid general asynchronous tatonnement price updates [1].

Tatonnement price updates can be briefly described as follows: if a good is over-demanded (i.e., its demand exceeds its supply), raise the price of the good; and vice versa. It was known that in a CES Fisher market, synchronous tatonnement is equivalent to gradient descent on certain convex function, where the minimum point of the function corresponds to the market equilibrium [2]. The high-level idea in [2] is to show that each round of synchronous tatonnement updates reduce the convex function value by at least a certain fraction; we call the reduction of each round the progress.

When tatonnement price updates are asynchronous, each update might be using outdated information, so the progresses will be countered by errors. The key question to ask is whether such errors can accumulate to nullify the progresses, and to bar the price updates from converging to equilibrium. We show that this cannot happen. The key insight in our analysis is that we can use a fraction of the progresses of all updates to compensate all errors. In other words, we use an amortization of good progresses against bad errors. The remaining fraction of the progresses can then be used to demonstrate overall convergence towards the market equilibrium. Furthermore, the convergence rate is only slightly worse than its synchronous counterpart.

References


Proportional Response Dynamics in Fisher Markets

Investigators: Yun Kuen Cheung in cooperation with Richard Cole and Yixin Tao (Courant Institute, NYU)

A major goal in Algorithmic Game Theory is to justify equilibrium concepts from an algorithmic perspective. One appealing approach is to identify natural distributed algorithms that converge quickly to an equilibrium. In many prior works, a natural price update rule called tatonnement was studied. In recent years, a new type of distributed updates called proportional response dynamics, which are well motivated from applications in peer-to-peer networks, are studied. Briefly speaking, in proportional response dynamics, each buyer will distribute her spending on different goods in proportion to the utility yielded by each good in her current allocation, while each seller of a good will distribute the good to buyers in proportion to the spending that the seller receives from the buyers; this process is repeated to form a dynamic process.

Prior work [1, 3, 4] showed that proportional response dynamics converge toward market equilibrium in a range of substitute domain of the CES Fisher markets. It was an interesting open problem on whether proportional response dynamics or its variants can converge in other CES Fisher markets, i.e., those markets in the complementary domain or the domain of mixtures of substitutes and complements of CES Fisher markets. In [2], we answer positively by proposing a variant of proportional response dynamics.

Interestingly, the variant of proportional response dynamics can actually be derived from a mirror descent rule on a convex function, where its minimum point corresponds to the market equilibrium. By drawing such a connection, our technical analysis has been focusing on the convergence of mirror descent in a general class of Bregman-strongly-convex functions. This provides a strong generalization on the standard gradient descent on Lipschitz-gradient-continuous strongly-convex functions. Furthermore, we provide a simple mirror-descent-esque update method and a short analysis to show that the method converges to the saddle point in a general class of Bregman-convex-concave functions. In summary, while our original target was solving an open problem about market dynamics, we end up simultaneously making fundamental contributions to the theory of optimization.

References


Tracing Equilibrium in Dynamic Markets via Distributed Adaptation

Investigators: Yun Kuen Cheung in cooperation with Martin Hoefer and Paresh Nakhe (Goethe University Frankfurt)

In real-world decentralized multi-agent systems, agents’ actions are often coupled with changes in the environment which are out of the agents’ control. Yet, in many important domains, the existing analyses presume static environments. In [1], we bridge such a gap between existing work and reality, with a focus on markets.

Competitive (market) equilibrium is a central concept in economics with numerous applications beyond markets, such as scheduling, fair allocation of goods, or bandwidth distribution in networks. Natural and decentralized processes like tatonnement and proportional response dynamics are known to converge quickly towards equilibrium in large classes of static Fisher markets. In contrast, many large real-world markets are subject to frequent and dynamic changes. We provide the first provable performance guarantees of discrete-time tatonnement and proportional response dynamics in dynamic markets. We analyze the class of CES Fisher markets, and quantify the impact of changes in supplies of goods, budgets of agents, and utility functions of agents on the convergence of the processes to equilibrium.

Since the equilibrium becomes a dynamic object and will rarely be exactly or even approximately reached, we provide bounds expressing the distance to equilibrium that will be maintained via tatonnement and proportional response dynamics updates. Our results indicate that in many cases, the processes trace the equilibrium rather closely and quickly recover conditions of approximate market clearing. We achieve this by quantifying the impact of variation in market parameters on several potential functions which guarantee convergences in static settings. Algebraically, it amounts to showing the following type of inequality

\[ \Phi(t) \leq (1 - \epsilon) \cdot \Phi(t - 1) + E(t - 1), \]

where \( \Phi(t) \) is the potential function value at time \( t \), \( E(t - 1) \) is the impact of environmental variations on the potential function value at time \( t - 1 \), and \( \epsilon \) is some constant. Our result follows by iterating the above inequality recursively.

While we focus on market settings, our approach, as discussed above, is actually quite general and handy. We formulate our approach into two general frameworks for Lyapunov dynamical systems. They are of independent interest, which we demonstrate with the analysis of load balancing and gradient descent in dynamic environment settings.

References

Multiplicative Weights Updates with Constant Step-Size in Graphical Constant-Sum Games

Investigator: Yun Kuen Cheung

The concept of Nash Equilibrium (NE) has been central in game theory. The existential proof of NE is non-constructive, while the definition of NE itself is also static, both of which shed no insight how NE can be computed or reached. In turn, lots of researchers have devoted efforts to justify the concept of NE by providing algorithms/dynamics which might compute/reach a NE. Among them, Multiplicative Weights (MW) updates have drawn a lot of attention, due to its simplicity and naturalness, and perhaps more importantly, its distributive implementability, which is essential in games we observe in reality, where communication between players is often severely limited. However, various hardness and communication complexity results strongly indicate that MW updates cannot efficiently compute/reach NE for general games. But can MW updates do so for interesting sub-families of games?

For two-person zero-sum games, a celebrated result by Freund and Schapire [2] showed that the time-average of MW updates with diminishing step sizes converges to NE. However, beyond this positive result, little was known about the pointwise behaviour of MW updates with constant step size in zero-sum games. In [1], we prove a result about the pointwise behaviour, which is in stark contrast with the time-average convergence result: in any zero-sum game with a fully-mixed NE (i.e., at NE every strategy is played with a strictly positive probability), the pointwise MW updates actually moves away from the NE. More precisely, we show that the Kullback-Leibler divergence (a famous asymmetric distance measure) between the current point of MW updates and the fully-mixed NE is strictly increasing.

There are a number of consequences of this theoretical finding. First, not only that MW updates are diverging from the NE, but even worse, they converge to the boundary of the state space. Intuitively, this means that although the NE is fully-mixed, beyond a certain time, MW updates will always essentially ignore at least one of the strategies. Second, the regret of MW updates, an important benchmark on how well a learning dynamic performs, can be shown to be at least $\Omega(1/T)$, where $T$ is the number of steps MW updates are run.

Third, and perhaps the most fascinating (if not frustrating) one, is that when the step size is sufficiently small, there always exists two strategies, for each of them MW updates exhibit the following fluctuating phenomenon: the strategy is essentially ignored (say it will be played with probability at most $10^{-1000}$) for a long time, and then it will resurge and be played with a decent probability (say at least $10^{-1}$); and such fluctuations repeat indefinitely.

References

- [2] Y. Freund and R. E. Schapire. Game theory, on-line prediction and boosting. In A. Blum and
35.4.2 Mechanism Design

Algorithmic mechanism design studies scenarios where there is an optimization problem at hand, but rational agents control some input parameters. These parameters are unknown to the optimizer and are private values of the agents. Agents might only be interested in satisfying their own interests and therefore the outcome is the result of strategic interaction of bidders, which is modeled using game-theoretic solution concepts, such as Nash equilibria.

Notably, agents’ strategies may lead to outputs different from the global optimum. We would like to give guarantees on any equilibrium solution compared to the optimum solution. The price of anarchy is the concept that measures how the efficiency of a system degrades due to the selfish behaviour of its agents; more specifically it is the worst case ratio between the objective values of any equilibrium and an optimum solution. The optimizer then designs mechanisms, which consist of some underlying rules, under the goal of optimizing the price of anarchy. Examples of mechanisms may be a charging scheme of agents for using some facilities or, in a case of an auction, it would be the allocation rule based on which the goods are given to the agents and the payment scheme that decides the agents’ charges.

A Near Optimal Mechanism for Energy Aware Scheduling

Investigators: Antonios Antoniadis in cooperation with Andrés Cristi (Universidad de Chile)

With the increased popularity of cloud computing it is of paramount importance to understand energy-efficiency also from a game-theoretic perspective. An important question is how the operator of a server should deal with combining energy-efficiency and the particular interests of the clients. Consider a cloud server, where clients can submit programs (jobs) for processing. The satisfaction that each customer perceives is directly linked to how long she has to wait while her job is being processed. Therefore, her satisfaction is modelled by a non-decreasing function of the time that her job took to finish processing – and this function is, naturally, private information. The server has to process the jobs and charge each client while trying to optimize the social cost, defined as the energy expenditure plus the total customer satisfaction, i.e., the sum of the values of the cost functions of the clients. The operator would like to design a mechanism in order to optimize this objective, which ideally is computationally tractable and charges the users “fairly”.

In [1] we describe and analyze one such mechanism (in this case a mechanism is comprised by an algorithm for scheduling the jobs and a charging scheme) called modAVR. Mechanism modAVR relies on an adaption of the well-known Average Rate (AVR) [2] algorithm for scheduling the jobs. We prove that modAVR combines the aforementioned properties with a constant Price of Anarchy, i.e., despite the fact that it is based on an algorithm designed for optimizing the energy alone, every equilibrium it results in is near-optimal for the total social cost as well. An interesting feature of modAVR is that it is indirect: each user needs only to declare an upper bound on the completion time of her job, and not the cost function.

In addition to modAVR, we also investigate and analyze several further interesting variations of this mechanism.
Designing Networks with Good Equilibria under Uncertainty

Investigators: Alkmini Sgouritsa in cooperation with George Christodoulou (University of Liverpool)

In a network cost-sharing game a rooted undirected graph with nonnegative edge costs is given and a set of players with terminal vertices needs to establish connectivity with the root. Each player selects a path and the global objective is to minimise the cost of the used edges. These games have applications in multicast routing and in network infrastructure. The cost of an edge may represent infrastructure cost for establishing connectivity or renting expense, and needs to be covered by the users.

There are several ways to split the edge cost among its users and this is dictated by a cost-sharing protocol. Naturally, it is in the players best interest to choose paths that charge them with small cost, and therefore the solution will be a Nash equilibrium. One concept that has been extensively studied in this setting is the Price of Anarchy (PoA) which is defined as the worst case comparison between the minimum possible total cost and the total cost of any Nash equilibrium.

The seminal work of Chen, Roughgarden and Valiant [1] was the first to address design questions for this game. They thoroughly studied the PoA for the following informational assumptions.

1. The designer has full knowledge of the instance, that is, she knows both the underlying graph and the players’ terminals. They showed that there exists a cost-sharing protocol that achieves constant PoA.

2. The designer has no knowledge of the underlying graph; they showed that the PoA of any cost-sharing protocol is logarithmic in the number of vertices.

However, there are situations where the former assumption is too optimistic while the latter is too pessimistic. In [2] we propose a model that lies in the middle-ground; the designer has prior knowledge of the underlying graph metric, but knows nothing about the positions of the terminals. His goal is to process the graph and choose a universal cost-sharing protocol that has low PoA against all possible requested subsets. The main question we address is to what extent prior knowledge of the underlying metric can help in the design.

We first demonstrate that there exist graph metrics, the outerplanar graphs, where knowledge of the underlying metric can dramatically improve the performance of good network cost-sharing design. However, in our main technical result, we show that there exist graph metrics for which knowing the underlying metric does not help in the design and any universal protocol has logarithmic PoA. The result is tight and matches the bound of [1]
which ignores the graph metric. We attack this problem by developing new techniques that employ powerful tools from extremal combinatorics, and more specifically Ramsey Theory in high dimensional hypercubes.

References


**Price of Anarchy for Mechanisms with Risk-Averse Agents**

*Investigators: Thomas Kesselheim and Bojana Kodric*

Many practical, “simple” auction mechanisms are not incentive compatible, making it beneficial for agents to behave strategically. A standard example is the first-price auction, in which one item is sold to one of $n$ agents. Each of these agents is asked to report a valuation; the item is given to the agent reporting the highest value, who then has to pay what he/she reported. A common way to understand the effects of strategic behavior is to study resulting equilibria and to bound the price of anarchy. That is, one compares the social welfare that is achieved at the (worst) equilibrium of the induced game to the maximum possible welfare. Typical equilibrium concepts are Bayes-Nash equilibria and (coarse) correlated equilibria, which extend mixed Nash equilibria toward incomplete information or learning settings, respectively. A key assumption in these analyses is that agents are risk neutral: Agents are assumed to maximize their expected quasilinear utility, which is defined to be the difference of the value associated to the outcome and payment imposed to the agent. So, an agent having a value of 1 for an item would be indifferent between getting this item with probability 10% for free and getting it all the time, paying 0.9.

However, there are many reasons to believe that agents are not risk neutral. For instance, in the above example the agent might favor the certain outcome to the uncertain one. Therefore, in [1], we ask the question: What “simple” auction mechanisms preserve good performance guarantees in the presence of risk-averse agents? As already mentioned, previous work has focused on agents with quasilinear utilities, possibly with a budget. Our model subsumes this as a special case but also captures that agents might be less sensitive to payments than in the risk-neutral model. We show that many positive price-of-anarchy results proved in the smoothness framework [2] continue to hold in the more general risk-averse setting. A sufficient condition is that agents can never end up with negative quasilinear utility after playing an undominated strategy. This is true, e.g., for first-price and second-price auctions. For all-pay auctions, similar results do not hold: We show that there are Bayes-Nash equilibria with arbitrarily bad social welfare compared to the optimum.

References

35.4.3 Online Algorithms

In online algorithms, we quantify the cost of lacking knowledge about future events. In this area, algorithms typically receive input piecewise over time. Whenever a piece of information arrives (such as, e.g., a new agent request, a new packet, etc.), the algorithm has to decide immediately and irrevocably how to deal with it without having seen the rest of the optimization instance. A well-established performance benchmark is the competitive ratio: For a given instance, it is given by the ratio between the objective-function values of the solution computed by the online algorithm and an optimal offline solution. For obtaining the offline solution we not only assume knowledge of the entire instance but also unlimited computational power.

There are different ways to model the uncertainty of the input and what the algorithm knows about it. The traditional approach is a worst-case analysis, where one assumes that a hypothetical adversary defines the optimization instance as well as the order in which it is presented. Due to this strong adversary, the performance guarantees in this model are very strong and apply to many different scenarios. However, in some problems, this perspective is too pessimistic to yield meaningful results. Slightly weaker models that reduce the power of the adversary have also been studied. For example, the adversary may still define the optimization instance but not the order in which it is revealed; the order instead may be a randomly drawn permutation or satisfy some properties known to the algorithm. Another weaker model is to assume that the input variables are independently and identically distributed (i.i.d.), where the adversary can only choose the worst-case distribution.

Prophet Inequalities and Posted Prices

Investigators: Thomas Kesselheim in cooperation with Paul Dütting (London School of Economics), Michal Feldman (Tel Aviv University), and Brendan Lucier (Microsoft Research)

The prophet inequality [2] is a famous result from stopping theory. One is presented a random sequence of rewards. Whenever a reward is presented, one can choose to pick this reward and to stop the sequence or to discard this reward and to continue. This setting has the intriguing interpretation of selling one item among multiple buyers. As optimal strategies are based on thresholds, they naturally correspond to posting a price for the item. Because of this connection, algorithmic mechanism design has extended this result to combinatorial allocation problems. For example, we might have multiple items to allocate.

In [1], we present a general framework for such stochastic online maximization problems with combinatorial feasibility constraints. Our framework establishes prophet inequalities by constructing price-based algorithms, a natural extension of threshold algorithms for settings beyond binary selection. The analysis takes the form of an extension theorem: we derive sufficient conditions on prices when all weights are known in advance, then prove that the
resulting approximation guarantees extend directly to stochastic settings. The framework unifies and simplifies much of the existing literature on prophet inequalities and posted price mechanisms, and is used to derive new and improved results for combinatorial markets (with and without complements), multi-dimensional matroids, and sparse packing problems.

References


Optimal Prophet Secretaries

Investigators: Tim Oosterwijk in cooperation with José Correa (Universidad de Chile), Patricio Foncea (Massachusetts Institute of Technology), Ruben Hoeksma (University of Bremen), and Tjark Vredeveld (Maastricht University)

The classic prophet inequality states that, when faced with a finite sequence of non-negative independent random variables, a gambler who knows their distribution and is allowed to stop the sequence at any time, can obtain, in expectation, at least half as much reward as a prophet who knows the values of each random variable and can choose the largest one. Following this classic theorem from the 70s, many results have been obtained for several related optimal stopping problems. Moreover, the recently uncovered connection between prophet inequalities and posted price mechanisms, has given the area a new surge. In [2], we survey some exiting new developments and highlight some compelling open problems.

In particular, we emphasize techniques that we used in our article [1], where we consider the situation in which the sequence of random variables comes in random order. We look at both a non-adaptive and an adaptive version of the problem. In the former case the gambler sets a threshold for every random variable a priori, while in the latter case the thresholds are set when a random variable arrives. For the non-adaptive case, we obtain an algorithm computing thresholds achieving an expected reward within at least a $1 - 1/e$ fraction of the expected maximum and prove this constant is optimal. For the adaptive case with i.i.d. random variables, we obtain a tight 0.745-approximation, solving a problem posed by Hill and Kertz in 1982 and proving a follow-up conjecture of Kertz from 1986.

We also apply these prophet inequalities to posted price mechanisms. In this setting, there is a seller holding a single item to sell to a set of customers, who have independent random valuations for the item. Customers arrive sequentially and in this work the arrival sequence is assumed to be random. Upon arrival of a customer, the seller offers a price as a take-it-or-leave-it offer, and the customer either takes the item at that price or simply leaves it. The goal of the seller is to find the prices that maximize his expected revenue. These mechanisms are very appealing because of their simplicity and the fact that strategic behaviour vanishes. A common example of this practice is that of direct mail campaigns, in which the seller contacts its potential buyers directly and offers each one a certain price for the item. The item is then sold to the first consumer who accepts the offer.
We prove the same tight bounds for both a non-adaptive and an adaptive posted price mechanism when buyers arrive in random order.

References


Combinatorial Secretary Problems with Ordinal Information

Investigators: Bojana Kodric in cooperation with Martin Hoefer (Goethe University Frankfurt)

The secretary problem is a classic approach to model online decision making under uncertain input. Recently, a variety of combinatorial extensions have been studied in the computer science literature [1]. Combinatorial extensions such as matroid or matching secretary problems have become an important tool to study algorithmic problems in dynamic markets. Here the decision maker must know the numerical value of each arriving element, which can be a demanding informational assumption. Indeed, the best known algorithms for matroid or matching secretary problems rely heavily on knowing the exact weight structure of elements. They either compute max-weight solutions to guide the admission process or rely on advanced bucketing techniques to group elements based on their weight. For a decision maker, in many applications it can be quite difficult to determine an exact cardinal preference for each of the incoming candidates. In contrast, in the original problem, the optimal algorithm only needs ordinal information about the candidates [2]. This property provides a much more robust guarantee, since the numerical values can be arbitrary, as long as they are consistent with the preference order.

In [4], we initiate the study of combinatorial secretary problems with ordinal information, in which the decision maker only needs to be aware of a preference order consistent with the values of arrived elements. The goal is to design online algorithms with small competitive ratios. For a variety of combinatorial problems, such as bipartite matching, general packing LPs, and independent set with bounded local independence number, we design new algorithms that obtain constant competitive ratios. For the matroid secretary problem, we observe that many existing algorithms for special matroid structures maintain their competitive ratios even in the ordinal model. In these cases, the restriction to ordinal information does not represent any additional obstacle. Moreover, we show that ordinal variants of the submodular matroid secretary problems can be solved using algorithms for the linear versions by extending [3]. In contrast, we provide a lower bound of $\Omega(\sqrt{n}/(\log n))$ for algorithms that are oblivious to the matroid structure, where $n$ is the total number of elements. This contrasts the upper bound of $O(\log n)$ in the cardinal model and shows that the technique of thresholding is not sufficient for good algorithms in the ordinal model.
35 D1: Algorithms and Complexity

References


35.4.4 Stable and Fair Outcomes

There are situations that we are interested in outcomes that preserve some desirable properties. Determining that kind of outcomes or good approximations of them in polynomial time is in our scope of interest. Two notable examples are stable matching and fair good allocation where agents have preferences and we try to satisfy them in a stable or fair manner.

A stable matching is a matching where no coalition appears, meaning that there is no pair of agents that both strictly prefer to be matched rather than remain in the current matching. One way to obtain a stable matching is to perform some round-based iterative process that converges (in polynomial time) to a stable matching. Such a process may be a best-response dynamic, where agents choose optimal actions based on the costs they observed in previous rounds and obtain performance guarantees with respect to the best action in hindsight.

A fair good allocation can be defined in several different ways. One way would be to allocate goods to agents by trying to equally satisfy them (as much as possible). Another notion would be the envy-freeness, where no agent envies the allocation of any other player, meaning that they do not prefer the goods allocated to other agents more than their goods. We study both notions of fairness.

Dynamics in Matching and Coalition Formation with Constraints

Investigators: Daniel Vaz in cooperation with Martin Hoefer (Goethe University Frankfurt) and Lisa Wagner (RWTH Aachen)

Matching and coalition formation are fundamental aspects of the organization of many multi-agent systems. The simplest case of this class of games is stable matching: each agent is a node in a fixed network which strives to be matched to another agent, and has a complete preference list over all other agents it can be matched with. For a matching $M$, a pair of agents $\{u, v\} \notin M$ is a blocking pair if both $u$ and $v$ strictly improve their preference by matching up (and dropping any other matches in $M$). A matching without any blocking pair is a stable matching.

More generally, hedonic games allow the formation of coalitions, which serve the same purpose as pairs in matchings, but allowing more than two agents to be matched. Similarly
to matchings, each agent can be part of at most one coalition, and its preference must depend only on the agents in the coalition. A blocking coalition is a group of agents that could (all individually) improve by abandoning their current coalitions and forming a new one together. A state (set of agent-disjoint coalitions) is core-stable if there is no blocking coalition.

In our work [2], we analyze the convergence properties of improvement dynamics: in each step, a blocking coalition is selected and then resolved, by adding it to our state $M$, while at the same time removing overlapping coalitions. We are interested in studying the behaviour of such dynamics, namely how fast they converge to a (core-)stable state, or in which situations they might cycle between states and not converge. Moreover, we consider additional local constraints for coalitions, such as restricted information, visibility, or externalities, which determine whether each coalition can be formed.

We propose and study a general class of hedonic coalition formation games that we term coalition formation games with constraints. This class includes and extends many recently studied variants of stable matching, such as locally stable matching, socially stable matching, and friendship matching. Perhaps surprisingly, we show that all these variants are encompassed in a class of consistent instances. In consistent instances, for every initial state there is a sequence of polynomially many improvement steps to a stable state. Moreover, we show that for such instances there always exists a polynomial sequence to every state that is reachable via improvement dynamics. Our characterization is tight in the sense that we provide exponential lower bounds whenever at least one of the axioms of consistency is violated.

In addition, we propose extensions of correlated locally stable matching to capture visibility and dynamics for coalitions with more than two agents. We use a set of graph structures to capture which coalitions can be discovered and are available for deviation if they provide higher benefit. In this class of games, we prove a tight characterization of graph structures that guarantee convergence in polynomial number of steps.

This builds on earlier work of our group [1, 3].

References


Market Equilibria and Fair Division

Investigators: Bhaskar Ray Chaudhury, Yun Kuen Cheung, and Kurt Mehlhorn in cooperation with Jugal Garg (UIUC), Naveen Garg (IIT Delhi), and Martin Höfer (Goethe
The fair division problem asks to allocate indivisible items to a set of \( n \) agents in a fair manner. In the simplest case, each agent \( i \) has a utility \( u_{ij} \) for each good and utilities are additive, i.e., the value of a set \( S \) of goods for \( i \) is \( u_i(S) = \sum_{j \in S} u_{ij} \). Let \( x = (x_1, \ldots, x_n) \) be an allocation of the goods to the agents.

There are different notions of fairness. John Nash postulated three conditions for a fair allocation, e.g., invariance under scaling, and showed that the allocation maximizing the Nash social welfare \( \text{NSW}(x) = \left( \prod_{1 \leq i \leq n} u_i(x_i) \right)^{1/n} \) satisfies all three conditions.

Another notion of fairness is the envy-freeness which is agent-oriented. An allocation \( x \) is envy-free if for any two agents \( i \) and \( h \), \( u_i(x_h) \leq u_i(x_i) \), i.e., agent \( i \) likes his bundle at least as much as the bundle allocated to any other agent. Envy-freeness can be guaranteed for divisible goods (give every agent a fraction \( 1/n \) of every good), but is in general impossible to be achieved for indivisible goods. Consider the case of two agents and one good. The appropriate notion for indivisible goods is envy-freeness up to one good, i.e., for any two agents \( i \) and \( h \), for all good \( g \) there is a good \( j \) such that \( u_i(x_h \setminus g) \leq u_i(x_i) \), i.e., after removal of the most valuable good from another bundle, i likes his bundle at least as much as the other bundle.

For divisible goods, an allocation maximizing Nash social welfare is easy to be computed. One can simply organize a Fisher market in which each agent has the same budget. The solution of the Fisher market is then an optimal allocation.

In a sequence of papers, we [4, 6, 7] and others [5, 1, 2, 3] obtain better and better approximation algorithms for the Nash social welfare. The strongest result [7] is as follows. We have \( m \) distinct goods that are available in several copies or items; there are \( k_j \) items of good \( j \). The agents have non-increasing utilities for the different items of a good, i.e., for all \( i \) and \( j \)

\[
u_{i,j,1} \geq u_{i,j,2} \geq \ldots \geq u_{i,j,k_j}.
\]

An allocation assigns the items to the agents. For an allocation \( x \), \( x_i \) denotes the multi-set of items assigned to agent \( i \), and \( m(j,x_i) \) denotes the multiplicity of \( j \) in \( x_i \). Of course, \( \sum_i m(j,x_i) = k_j \) for all \( j \). The total utility of bundle \( x_i \) under valuation \( u_i \) is given by

\[
u_i(x_i) = \sum_j \sum_{1 \leq \ell \leq m(j,x_i)} u_{i,j,\ell}.
\]

Furthermore, each agent has a utility cap \( c_i \) and then the utility of bundle \( x_i \) for agent \( i \) is defined as

\[
\bar{u}_i(x_i) = \min(c_i, \nu_i(x_i)).
\]

We construct instances of the capped utilities where no NSW-optimal allocation is envy-free up to one good, showing that both notions of fairness cannot be achieved simultaneously by any allocation. However we design a polynomial-time algorithm to determine an allocation that approximates both notions of fairness. Our approximation algorithm returns an allocation \( x \) such that \( \text{NSW}(x^*) \leq (e^{1/e} + \varepsilon) \cdot \text{NSW}(x) \approx (1.445 + \varepsilon) \cdot \text{NSW}(x) \) for any positive \( \varepsilon \),
where $x^*$ is the optimal NSW allocation. Also our allocation is approximately envy-free up to one good, i.e., for two agents $i$ and $h$ there is a good $g$ such that $\bar{u}_i(x_h \setminus g) \leq (2 + \varepsilon) \bar{u}_i(x_i)$ for any positive $\varepsilon$.

References


35.5 Combinatorics, Computing, and Randomness

Coordinator: Karl Bringmann

In the research area “Combinatorics, Computing, and Randomness”, we investigate the structures and principles that are fundamental for the understanding and design of efficient algorithms. On combinatorial objects such as graphs and strings, we develop techniques for designing efficient algorithms, both in the classic sense of exact polynomial-time solutions and in modern directions such as streaming algorithms, approximation algorithms for polynomial-time problems, and fixed-parameter tractability. We complement algorithm design with fine-grained complexity theory, a modern approach from complexity theory that yields quantitative running time lower bounds, and thus allows us to provide evidence that our designed algorithms are best-possible. Since these ingredients are crucial for many, also more applied problems, there are many links to other research areas in D1 and beyond.\(^1\)

\(^1\)Throughout this section, for convenience the notation $\tilde{O}(T)$ denotes running times of the form $O(T \log^c T)$ for any constant $c \geq 0$. 

131
35.5.1 Fine-Grained Complexity Theory

The classic theory of NP-hardness provides a coarse-grained classification of problems into efficient (=polynomial-time solvable) and intractable (=NP-hard). In contrast, modern fine-grained complexity theory yields quantitative running time lower bounds, and thus provides a more fine-grained classification of problems according to their time complexity. Naturally, such stronger lower bounds rely on stronger hypotheses than $P \neq NP$, e.g., the Strong Exponential Time Hypothesis for Satisfiability. In our group, we develop fine-grained complexity theory by designing proof techniques for tight fine-grained lower bounds and applying them to a variety of problems. We also contribute to the foundations of fine-grained complexity theory, e.g., by studying the hypotheses that this theory relies on.

Pseudopolynomial Algorithms for Subset Sum

Investigators: Karl Bringmann in cooperation with Amir Abboud (IBM Almaden), Danny Hermelin, and Dvir Shabtay (Ben-Gurion University)

Subset Sum is a fundamental problem at the intersection of theoretical computer science, optimization, and operations research: Given a target $t$ and a set $Z$ consisting of $n$ positive integers, decide whether some subset $Y$ of $Z$ sums to $t$. A pseudopolynomial time algorithm by Bellman from 1957 solves Subset Sum in time $O(nt)$ [2]. This algorithm is being taught for decades in undergraduate algorithms courses and thus had a great influence on computer science. Since this pseudopolynomial time algorithm is a fundamental part of our curriculum, and since Subset Sum is one of the core NP-hard problems, it is a fundamental question whether the running time of $O(nt)$ can be improved.

The original motivation to revisit this problem was to prove a conditional lower bound showing optimality of Bellman’s algorithm. Instead, in [3] we presented a randomized algorithm running in time $\tilde{O}(n + t)$, ignoring logarithmic factors. This improves Bellman’s algorithm by about a factor $n$.

Our improved algorithm is likely to be essentially optimal: A conditional lower bound ruling out any significant further improvements was first shown by Cygan et al. under the non-standard Set Cover hypothesis [4], and later we showed the same lower bound under the standard Strong Exponential Time Hypothesis [1]. Together, we now know a conditionally best-possible algorithm for Subset Sum.

As a corollary of our lower bound [1], we prove a “Direct-OR” theorem for Subset Sum under the Strong Exponential Time Hypothesis, offering a new tool for proving conditional lower bounds: It is now possible to assume that deciding whether one out of $q$ given instances of Subset Sum is a YES instance requires time $(qt)^{1 - o(1)}$. As an application of this corollary, we prove a tight conditional lower bound for the classic Bicriteria Shortest Path problem, separating its complexity from that of Subset Sum: On graphs with $m$ edges and edge lengths bounded by $L$, we show that the $O(Lm)$ pseudo-polynomial time algorithm for Bicriteria Shortest Path [5] cannot be improved to $\tilde{O}(L + m)$, in contrast to our improvement for Subset Sum.
A Dichotomy for Regular Expression Membership Testing

Investigators: Karl Bringmann in cooperation with Allan Grønlund and Kasper Green Larsen (Aarhus University)

A regular expression is a term involving the operations concatenation \( \circ \), union \( | \), Kleene’s star \( * \), and Kleene’s plus \( + \). In regular expression membership testing, we are given a regular expression \( R \) and a string \( s \) and want to decide whether \( s \) is in the language described by \( R \).

A big success story of the 70s was to show that the problem can be solved in time \( O(nm) \) [3], where \( n \) is the length of the string \( s \) and \( m \) is the size of \( R \). This quite efficient running time, coupled with the great expressiveness of regular expressions, made them the standard tool that they are today.

Many special cases of membership testing have been studied that can be solved faster than in quadratic time. However, a systematic study of tractable cases was made possible only recently, with the first conditional lower bounds reported by Backurs and Indyk [1]. Viewing a regular expression as a tree where the inner nodes are labeled by \( \circ, |, * \), and \( + \), they call a regular expression homogeneous of type \( t \in \{ \circ, |, * , + \}^d \) if in each level \( i \) of the tree all inner nodes have type \( t_i \), and the depth of the tree is at most \( d \). Restricting to type \( t \), Backurs and Indyk studied \( t \)-membership testing. For any type of depth 2 or 3, they either presented an almost-linear time algorithm or a quadratic conditional lower bound, the only exception being \( +|\circ\)-membership. The latter is also known as the Word Break problem: Given a set of strings \( D \) and a text \( s \), is it possible to partition \( s \) into substrings that all belong to \( D \)?

In [2], we complete their work as follows:

- We present two new almost-linear time algorithms, for the types \( | + \circ + \) and \( | + \circ \). Our algorithms generalize all known special cases of regular expression membership testing previously known to be in almost-linear time.

- We classify all types, except for the Word Break problem, into almost-linear time or quadratic time assuming the Strong Exponential Time Hypothesis. This extends the classification from depth 2 and 3 to any constant depth.
For the Word Break problem we design an algorithm running in time $\tilde{O}(nm^{1/3} + m)$. Surprisingly, we also prove a tight conditional lower bound, restricted to "combinatorial" algorithms. This follows by a reduction from the $k$-Clique problem. We therefore establish Word Break as the only intermediate problem.

In total, we prove tight upper and lower bounds for any type of bounded-depth homogeneous regular expressions, which yields a full dichotomy for regular expression membership testing.

References


Fine-Grained Complexity of Analyzing Compressed Data

Investigators: Karl Bringmann and Marvin Künnemann in cooperation with Amir Abboud (Stanford University) and Arturs Backurs (MIT)

Can we analyze data without decompressing it? As our data keeps growing, understanding the time complexity of problems on compressed inputs, rather than in convenient uncompressed forms, becomes more and more relevant. Suppose we are given a compression of size $n$ of data that originally has size $N$, and we want to solve a problem with time complexity $T(\cdot)$. The naïve strategy of “decompress-and-solve” gives time $T(N)$, whereas “the gold standard” is time $T(n)$: analyzing the compression as efficiently as if the original data was small.

Motivated by an open problem raised by Oren Weimann [2], in [1] we give the first quantitative barriers for improvements over the “decompress-and-solve” strategy. Specifically, we restrict our attention to data in the form of a string (text, files, genomes, etc.) and study the most ubiquitous tasks. While the challenge might seem to depend heavily on the specific compression scheme, most methods of practical relevance (Lempel-Ziv-family, dictionary methods, and others) can be unified under the elegant notion of Grammar-Compressions. A vast algorithmic literature, across many disciplines, establishes this as the “right” notion for a fundamental study.

Our main result is a framework for proving (conditional) lower bounds in this field, allowing us to assess whether decompress-and-solve can be improved, and by how much. Among others, we prove:

- The $O(nN \sqrt{\log N/n})$ bound for LCS and the $O(\min\{N \log N, nM\})$ bound for Pattern Matching with Wildcards are optimal up to $N^{o(1)}$ factors, under the Strong Exponential Time Hypothesis. (Here, $M$ denotes the uncompressed length of the compressed pattern.)
Decompress-and-solve is essentially optimal for Context-Free Grammar Parsing and RNA Folding, under the $k$-Clique conjecture.

We give an algorithm showing that decompress-and-solve is not optimal for Disjointness.

References


Multivariate Fine-Grained Complexity of Longest Common Subsequence

Investigators: Karl Bringmann and Marvin K"unnemann

We revisit the classic combinatorial pattern matching problem of finding a longest common subsequence (LCS) of two strings $x$ and $y$ of length at most $n$. Assuming the Strong Exponential Time Hypothesis (SETH), we could recently show that the textbook $O(n^2)$-time dynamic programming algorithm is essentially optimal in the worst case [2] (an independent proof was obtained by Abboud et al. [1]). Yet still, extensive theoretical and practical work breaks this quadratic time bound for well-behaved inputs, which invites a much more detailed investigation beyond the worst-case complexity.

Specifically, we consider the variety of algorithms produced for LCS and its variations during more than 40 years of intense scientific interest. Particular attention was put into identifying and exploiting input parameters that yield strongly subquadratic time algorithms for special cases of interest, e.g., differential file comparison (see, e.g., the UNIX diff utility). This line of research was successfully pursued until 1990, at which time significant improvements came to a halt. This raises the question: Can we (1) justify the lack of further improvements and (2) determine whether some special cases of LCS admit faster algorithms than currently known?

To this end, in [3] we provide a systematic study of the multivariate complexity of LCS, taking into account all parameters previously discussed in the literature: the input size $n := \max\{|x|, |y|\}$, the length of the shorter string $m := \min\{|x|, |y|\}$, the length $L$ of an LCS of $x$ and $y$, the numbers of deletions $\delta := m - L$ and $\Delta := n - L$, the alphabet size, as well as the numbers of matching pairs $M$ and dominant pairs $d$. For any class of instances defined by fixing each parameter individually to a polynomial in terms of the input size, we prove a SETH-based lower bound matching one of three known algorithms (up to lower order factors of the form $n^{o(1)}$). Specifically, we determine the optimal running time for LCS under SETH as $(n + \min\{d, \delta \Delta, \delta m\})^{1+o(1)}$. Polynomial improvements over this running time must necessarily refute SETH or exploit novel input parameters. We establish the same lower bound for any constant alphabet of size at least 3. For binary alphabet, we show a SETH-based lower bound of $(n + \min\{d, \delta \Delta, \delta M/n\})^{1-o(1)}$ and, motivated by difficulties to improve this lower bound, we design an $O(n + \delta M/n)$-time algorithm, yielding again a matching bound.
Figure 35.1: Illustration of the edit operations of tree edit distance.

Our systematic approach yields a comprehensive perspective on the well-studied multivariate complexity of LCS, which might inspire similar studies of multivariate complexity landscapes for further polynomial-time problems.

References


Tree Edit Distance Cannot be Computed in Strongly Subcubic Time

Investigators: Karl Bringmann in cooperation with Paweł Gawrychowski, Oren Weimann (University of Haifa), and Shay Mozes (IDC Herzliya)

The edit distance between two rooted ordered trees with $n$ nodes labeled from an alphabet $\Sigma$ is the minimum cost of transforming one tree into the other by a sequence of elementary operations consisting of deleting and relabeling existing nodes as well as inserting new nodes, see Figure 35.1 for illustrations. Here, each operation has a cost depending on the label of the involved node, and these costs are given as part of the input. Tree edit distance is a well known generalization of string edit distance. The fastest known algorithm for tree edit distance runs in $O(n^3)$ time and is based on a similar dynamic programming solution as string edit distance [3]. In our work [2], we show that a truly subcubic $O(n^{3-\epsilon})$-time algorithm for tree edit distance is unlikely: For $|\Sigma| = \Omega(n)$, a truly subcubic algorithm for tree edit distance implies a truly subcubic algorithm for the All Pairs Shortest Paths problem. For $|\Sigma| = O(1)$, a truly subcubic algorithm for tree edit distance implies an $O(n^{k-\epsilon})$ algorithm for finding a maximum weight $k$-clique.

Thus, while in terms of upper bounds string edit distance and tree edit distance are highly related, in terms of lower bounds string edit distance exhibits the hardness of the Strong Exponential Time Hypothesis [1] whereas tree edit distance exhibits the hardness of All
Pairs Shortest Paths. Our result provides a matching conditional lower bound for one of the last remaining classic dynamic programming problems.

References


More Consequences of Falsifying SETH and the Orthogonal Vectors Conjecture

Investigators: Karl Bringmann in cooperation with Amir Abboud (IBM Almaden), Holger Dell (Saarland University), and Jesper Nederlof (Eindhoven University of Technology)

The Strong Exponential Time Hypothesis and the Orthogonal Vectors hypothesis are two popular hardness assumptions used to prove a plethora of lower bounds, especially in the realm of polynomial-time algorithms. In our work [1], we strengthen the evidence for these hardness assumptions. In particular, we show that if the Orthogonal Vectors hypothesis fails, then two problems for which we are far from obtaining even tiny improvements over exhaustive search would have surprisingly fast algorithms. If the Orthogonal Vectors hypothesis is false, then there is a fixed $\varepsilon > 0$ such that:

- For all $d$ and all large enough $k$, there is a randomized $O(n^{(1-\varepsilon)k})$-time algorithm for the Zero-Weight-$k$-Clique and Min-Weight-$k$-Clique problems on $d$-hypergraphs with $n$ vertices. As a consequence, the Orthogonal Vectors hypothesis is implied by the Weighted Clique hypothesis. This adds to the very few connections among the hypotheses of fine-grained complexity theory.

- For all $c$, the Satisfiability of sparse $TC^1$ circuits on $n$ inputs (that is, circuits with $cn$ wires, depth $c\log n$, and negation, AND, OR, and threshold gates) can be solved in time $O((2 - \varepsilon)^n)$. This yields strong evidence in favor of the Orthogonal Vectors hypothesis, since for Satisfiability of sparse $TC^1$ circuits not even algorithms running in time $O(2^n/n^{10})$ are known.

References

Tighter Connections Between Formula-SAT and Shaving Logs

Investigators: Karl Bringmann in cooperation with Amir Abboud (IBM Almaden)

For a problem that can be solved in time $O(n^2)$, the typical goal in fine-grained complexity theory is to show that there is no $O(n^{2-\varepsilon})$-time algorithm under a reasonable hypothesis. Such a lower bound rules out any further significant improvements, however, it does not cover improvements by logarithmic factors. For instance, the standard $O(n^2)$ dynamic programming algorithm for Longest Common Subsequence was improved to time $O(n^2 / \log^2 n)$ by Masek and Paterson [3]. The long line of research developing such improvements is known as the art of shaving log-factors, and until very recently lacked tools for proving negative results.

The only approach for lower bounds concerning log-factor improvements was suggested in a recent paper of Abboud et al. [2], who blame the hardness of shaving logs on the hardness of solving Satisfiability on Boolean formulas (Formula-SAT) faster than exhaustive search. In particular, they show that an $O(n^2 / \log^{1000} n)$-time algorithm for Longest Common Subsequence would imply a major advance in circuit lower bounds. Whether this approach can lead to tighter barriers was unclear.

In our work [1], we push this approach to its limit and, in particular, prove that a well-known barrier from complexity theory stands in the way for shaving five additional log-factors for fundamental combinatorial problems. For Longest Common Subsequence, Regular Expression Pattern Matching, as well as the Fréchet distance from computational geometry, we show that any $O(n^2 / \log^{7+\varepsilon} n)$-time algorithm would imply new Formula-SAT algorithms.

Our main result is a reduction from Satisfiability on formulas of size $s$ over $n$ variables to Longest Common Subsequence on sequences of length $N = 2^n/2 \cdot s^{1+o(1)}$. Our reduction is essentially as efficient as possible, and it greatly improves the previously known reduction for LCS with $N = 2^n/2 \cdot s^c$, for some $c \geq 100$ [2].

References


Fine-Grained Complexity of Longest Common Increasing Subsequence

Investigators: Marvin Künnemann in cooperation with Lech Duraj and Adam Polak (Jagiellonian University)

We consider the canonical generalization of the well-studied Longest Increasing Subsequence problem to multiple sequences, called $k$-LCIS: Given $k$ integer sequences $x_1, \ldots, x_k$ of length
at most \( n \), the task is to determine the length of the longest common subsequence of \( x_1, \ldots, x_k \) that is also strictly increasing. Especially for the case of \( k = 2 \) (called LCIS for short), several algorithms have been proposed that require quadratic time in the worst case.

Assuming the Strong Exponential Time Hypothesis (SETH), in [3, 4] we prove a tight lower bound, specifically, that no algorithm solves LCIS in (strongly) subquadratic time. This puts LCIS into a growing list of problems having a natural quadratic-time dynamic programming solution that is essentially optimal under SETH. Interestingly, the proof makes no use of normalization tricks common to hardness proofs for similar problems, such as the longest common subsequence problem (LCS) [1, 2].

Investigating the problem’s complexity more closely, we further strengthen this lower bound (1) to rule out \( O((nL)^{1-\varepsilon}) \)-time algorithms, where \( L \) denotes the solution size, (2) to rule out \( O(n^{k-\varepsilon}) \)-time algorithms for \( k \)-LCIS, and (3) to follow already from weaker variants of SETH. For the related Longest Common Weakly Increasing Subsequence problem, in which we search for the longest common subsequence that is weakly increasing, we obtain the same conditional lower bounds.

References


Clique-Based Lower Bounds for Parsing Tree-Adjoining Grammars

Investigators: Karl Bringmann and Philip Wellnitz

Tree-adjoining grammars (TAGs) are a system to manipulate certain trees to arrive at strings. Being more powerful than context-free grammars, TAGs capture various phenomena of human languages which require more formal power and are thus popular in computational linguistics.

The prime algorithmic problem on TAGs is parsing (sometimes called recognition): Given a TAG \( \Gamma \) and a string \( s \) of length \( n \), decide whether \( \Gamma \) can generate \( s \). Rajasekaran and Yooseph’s parser [3] solves this problem in time \( O(n^{2\omega}) \), where \( \omega < 2.373 \) is the matrix multiplication exponent. The best algorithms avoiding fast matrix multiplication require time \( O(n^6) \).

A limited explanation for the complexity of TAG parsing was given by Satta [4], who designed a reduction from Boolean matrix multiplication to TAG parsing, showing that any
TAG parser running faster than $O(|\Gamma| \cdot n^6)$ on grammars of size $|\Gamma| = \Theta(n^{12})$ yields a Boolean matrix multiplication algorithm running faster than $O(n^3)$.

Following an approach by Abboud et al. [1] for context-free grammar recognition, in [2] we resolve many of the disadvantages of the previous lower bound. We show that, even on constant-size grammars, any improvement on Rajasekaran and Yooseph’s parser would imply a breakthrough for the $k$-Clique problem. This establishes tree-adjoining grammar parsing as a practically relevant problem with the unusual running time of $n^{2\omega}$, up to lower order factors.

References


Fine-Grained Inapproximability

Investigators: Bundit Laekhanukit in cooperation with Karthik C.S. (Weizmann Institute of Science) and Pasin Manurangsi (UC Berkeley)

Studying the approximability of computational problems has been an active area of research for decades, as approximation algorithms are one of the most important tools in coping with NP-hard optimization problems. Complementing the design of approximation algorithms, there is research to understand the limitations of computing an approximate solution. While the approximabilities of most fundamental NP-hard optimization problems are now well-understood, little is known about the approximability of polynomial-time solvable problems. We studied the inapproximability of polynomial-time solvable problems and proved hardness results based on various computational assumptions.

In [1], we study the parameterized complexity of approximating the $k$-Dominating Set problem, where the input is an integer $k$ and a graph $G$ on $n$ vertices, and the goal is to select a set of $k$ vertices such that every remaining vertex is a neighbor of some selected vertex. An $F(k)$-approximation algorithm for the $k$-Dominating Set problem computes a dominating set of size at most $F(k) \cdot k$ whenever the input graph $G$ has a dominating set of size $k$. Among other results, we show that for any computable functions $F,T$ and any $\varepsilon > 0$, no $F(k)$-approximation algorithm can run in time $T(k) \cdot n^{\varepsilon(k)}$ under the Exponential Time Hypothesis, or in time $T(k) \cdot n^{k - \varepsilon}$ under the Strong Exponential Time Hypothesis. This considerably strengthens previous lower bounds for the problem.
35.5.2 Modern Aspects of Combinatorial Problems

Classic combinatorial problems such as computing the longest common subsequence of two strings, multiplying two matrices, or pattern matching of a pattern string in a text, have had a great influence on theoretical computer science, because their standard solutions are taught in undergraduate algorithms courses and thus they belong to the community’s common knowledge base. We study these classic problems from modern viewpoints, at the same time yielding new insights into these problems and developing modern branches of computer science. For instance, for longest common subsequence we study streaming and sketching algorithms, for matrix multiplication we study error-correcting algorithms, and for pattern matching we study the problem on compressed texts.

Algorithms and Data Structures: The Basic Toolbox

Investigators: Kurt Mehlhorn in cooperation with Peter Sanders (KIT, Karlsruhe), Martin Dietzfelbinger (TU Ilmenau), and Roman Dementiev (Intel)

We have completed the second edition of our textbook [1]

Sequential and Parallel Algorithms and Data Structures
The Basic Toolbox

The book now covers sequential and parallel algorithms and data structures. It will appear in the summer of 2019. We quote from the preface.

The first edition of this book treated only sequential algorithms and data structures. Today, almost every computer, be it a desktop, a notebook, or a smartphone, has multiple cores, and sequential machines have become an exotic species. The reason for this change is that sequential processors have ceased to get proportional performance improvements from increased circuit complexity. Although the number of transistors in an integrated circuit (still) doubles every two years (Moore’s law), the only reasonable way to use this transistor budget is to put multiple processor cores on a chip. The consequence is that nowadays every performance-critical application has to be parallelized. Moreover, big data – the explosion of data set sizes in many applications – has produced an enormous demand for algorithms that scale to a large number of processors. This paradigm shift has profound effects on teaching algorithms. Parallel algorithms are no longer a specialized topic reserved for a small percentage of students. Rather, every student needs some exposure to parallel algorithms, and parallel solution paradigms need to be taught early on. As a consequence, parallel algorithms should be integrated tightly and early into algorithms courses. We therefore
decided to include parallel algorithms in the second edition of the book. Each chapter now has some sections on parallel algorithms. The goals remain the same as for the first edition: a careful balance between simplicity and efficiency, between theory and practice, and between classical results and the forefront of research. We use a slightly different style for the sections on parallel computing. We include concrete programming examples because parallel programming is still more difficult than sequential programming (the programs are available at github.com/basic-toolbox-sample-code/basic-toolbox-sample-code/).

A first edition of this book was published in 2008. Since then the book has been translated into Chinese, Greek, Japanese, and German. Martin Dietzfelbinger translated the book into German. Actually, he did much more than a translation. He thoroughly revised the book and improved the presentation at many places. He also corrected a number of mistakes. Thus, the book gained through the translation, and we decided to make the German edition the reference for any future editions. It is only natural that we asked Martin to become an author of the German edition and any future editions of the book.

Soon after the publication of the German edition, we started working on the revised English edition. We decided to expand the book into the parallel world for the reasons indicated in the preface. Twenty years ago, parallel machines were exotic, nowadays, sequential machines are exotic. However, the parallel world is much more diverse and complex than the sequential world, and therefore algorithm-engineering issues become more important. We concluded that we had to go all the way to implementations and experimental evaluations for some of the parallel algorithms. We invited Roman Dementiev to work with us on the algorithm engineering aspects of parallel computing. Roman received his PhD in 2006 for a thesis on “Algorithm Engineering for Large Data Sets”. He now works for Intel, where he is responsible for performance engineering of a major database system.

References

(Weighted) Longest Common Subsequence in Sketching and Streaming Settings

Investigators: Karl Bringmann and Bhaskar Ray Chaudhury

The longest common subsequence of two strings is the longest string that is a subsequence of both. The problem of computing the longest common subsequence has been studied extensively and has numerous applications in bioinformatics, natural language processing, file comparisons etc. In [1] we study the Longest Common Subsequence problem in sketching and streaming settings that are motivated by big data applications. We focus on strings defined on a small alphabet set, e.g., binary files (Σ = \{0,1\}) or DNA sequences (Σ = \{A,G,C,T\}).
Sketching: Alice is given string $x$ and Bob is given string $y$. Both also are given a number $L$. Alice and Bob compute sketches $sk_L(x)$ and $sk_L(y)$ and send them to a third person, the referee, who decides whether $x$ and $y$ have a common subsequence of length at least $L$. The task is to minimize the size of the sketch (i.e., the number of bits transferred) as well as the running time of Alice and Bob (encoding) and of the referee (decoding).

Streaming: We are given $L$, and we scan the string $x$ from left to right once, and then the string $y$ from left to right once. After that, we need to decide whether $x$ and $y$ have a common subsequence of length at least $L$. We want to minimize the space usage as well as running time.

For both the settings we obtain a sketch size and streaming space usage of $O(L|\Sigma|^{-1} \log L)$, where $|\Sigma|$ is the alphabet size, and we also prove matching unconditional lower bounds.

As an application, we study a variant of the Longest Common Subsequence problem called Weighted Longest Common Subsequence, where each alphabet symbol is equipped with a weight that is given as input and the task is to compute a common subsequence of maximum total weight. We present a $O(\min(nm, n + m|\Sigma|))$-time algorithm when $x, y$ have length $n, m$, respectively, with $n \geq m$. We also prove that this time bound is optimal assuming the Strong Exponential Time Hypothesis.

In total, we achieve optimality of encoding time, the number of bits transferred, and decoding time for (weighted) longest common subsequence of strings defined on a small alphabet in sketching and streaming settings.

References


Dynamic Search Trees

Investigators: Kurt Mehlhorn in cooperation with Parinya Chalermsook (Aalto University), Mayank Goswami (CUNY), Lászlo Kozma (FU Berlin), and Thatchaphol Saranurak (KTH Sweden)

Multi-finger binary search trees (BSTs) are a far-reaching extension of the classical BST model, with connections to the well-studied $k$-server problem. Finger search is a popular technique for speeding up BST operations when a query sequence has locality of reference. BSTs with multiple fingers can exploit more general regularities in the input. In our work [1], we consider the cost of serving a sequence of queries in an optimal (offline) BST with $k$ fingers, a powerful benchmark against which other algorithms can be measured. We show that the $k$-finger optimum can be matched by a standard dynamic BST (having a single root-finger) with an $O(\log k)$ factor overhead. This result is tight for all $k$, improving the $O(k)$ factor implicit in earlier work. Furthermore, we describe new online BSTs that match this bound up to a $(\log k)^{O(1)}$ factor. Previously only the “one-finger” special case was known to hold for an online BST [4, 2]. Splay trees, assuming their conjectured optimality [5], would
have to match our bounds for all $k$. Our online algorithms are randomized and combine techniques developed for the $k$-server problem with a multiplicative-weights scheme for learning tree metrics. To our knowledge, this is the first time when tools developed for the $k$-server problem are used in BSTs. As an application of our $k$-finger results, we show that BSTs can efficiently serve queries that are close to some recently accessed item. This is a (restricted) form of the unified property [3] that was previously not known to hold for any BST algorithm, online or offline.

References


Few Matches or Almost Periodicity: Faster Pattern Matching with Mismatches in Compressed Texts

Investigators: Karl Bringmann, Marvin Kühnemann, and Philip Wellnitz

A fundamental problem on strings in the realm of approximate string matching is pattern matching with mismatches: Given a text $t$, a pattern $p$, and a number $k$, determine whether some substring of $t$ has Hamming distance at most $k$ to $p$; such a substring is called a $k$-match.

As real-world texts often come in compressed form, we study the case of searching for a small pattern $p$ in a text $t$ that is compressed by a straight-line program. This grammar compression is popular in the string community, since it is mathematically elegant and unifies many practically relevant compression schemes such as the Lempel-Ziv family, dictionary methods, and others. We denote by $m$ the length of $p$ and by $n$ the compressed size of $t$. While exact pattern matching, that is, the case $k = 0$, is known to be solvable in near-linear time $\tilde{O}(n + m)$ [3], despite considerable interest in the string community, the fastest known algorithm for pattern matching with mismatches runs in time $\tilde{O}(n\sqrt{m}\text{poly}(k))$ [2], which is far from linear even for very small $k$.

In [1], we obtain an algorithm for pattern matching with mismatches running in time $\tilde{O}((n + m)\text{poly}(k))$. This is near-linear in the input size for any constant (or slightly
We obtain analogous running times for counting and enumerating all $k$-matches.

Our algorithm is based on a new structural insight for approximate pattern matching, essentially showing that either the number of $k$-matches is very small or both text and pattern must be almost periodic. While intuitive and simple for exact matches, such a characterization is surprising when allowing $k$ mismatches.

References


Towards Nondeterministic Verification of Matrix Products

Investigator: Marvin Künnemann

Freivalds’ algorithm [1] yields a surprisingly simple randomized $O(n^2)$-time solution for verifying whether an $n \times n$ matrix $C$ is the product of two $n \times n$ matrices $A$ and $B$. It is a long-standing open question whether this algorithm can be derandomized, i.e., whether matrix products can be verified deterministically in near-quadratic time. Starting from an interest in exploring the power of randomness, certifying algorithms and barriers for fine-grained reductions, in [3] we investigate a relaxation of this question: Can we derandomize Freivalds’ algorithm using additional nondeterministic guesses?

We discuss consequences of a positive or negative answer to this relaxed question and provide potential avenues towards resolving it. Particularly, we show that sufficiently fast nondeterministic verifiers for 3SUM or univariate polynomial identity testing yield faster nondeterministic verifiers for matrix multiplication. Furthermore, we present the partial algorithmic progress that distinguishing whether an integer matrix product is correct or contains between 1 and $n$ erroneous entries can be performed in time $\tilde{O}(n^2)$ – interestingly, the difficult case of deterministic matrix product verification is not a problem of “finding a needle in the haystack”, but rather cancellation effects in the presence of many errors.

Our main technical contribution is a deterministic algorithm that corrects an integer matrix product containing at most $t$ errors in time $\tilde{O}(\sqrt{tn^2} + t^2)$. To obtain this result, we show how to compute an integer matrix product with at most $t$ nonzeroes in the same running time. This improves upon known deterministic output-sensitive integer matrix multiplication algorithms [2, 4] for $t = \Omega(n^{2/3})$ nonzeroes, which is of independent interest.
A Mastermind Variant

Investigators: Kurt Mehlhorn in cooperation with Peyman Afshani (Aarhus University), Manindra Agrawal (IIT Kanpur), Benjamin Doerr (École Polytechnique, Paris), Carola Doerr (Sorbonne, Paris), and Kasper Green Larsen (Aarhus University)

We study the query complexity of a permutation-based variant of the guessing game Mastermind. In this variant, the secret is a pair \((z, \pi)\) which consists of a binary string \(z \in \{0,1\}^n\) and a permutation \(\pi\) of \([n]\). The secret must be unveiled by asking queries of the form \(x \in \{0,1\}^n\). For each such query, we are returned the score

\[
f_{z,\pi}(x) = \max\{i \in [0..n] \mid \text{for all } j \leq i: \ z_{\pi(j)} = x_{\pi(j)}\};
\]

i.e., the score of \(x\) is the length of the longest common prefix of \(x\) and \(z\) with respect to the order imposed by \(\pi\). The goal is to minimize the number of queries needed to identify \((z, \pi)\).

We [1] prove matching upper and lower bounds for the deterministic and randomized query complexity of this game, which are \(\Theta(n \log n)\) and \(\Theta(n \log \log n)\), respectively.

References


35.5.3 Algorithms on Networks and Graph Theory

Graphs and networks are ubiquitous in the sciences, as they can be used to model many types of relations in physical, biological, social, and information systems. Our work tackles a wide range of algorithmic questions from traditional tasks (such as computing a cycle of minimum cost-to-time ratio) to modern problems (such as estimating the \(k\)-step random walk distribution). Since many algorithmic developments on graphs are rooted in advances in graph theory, we also study structural questions (such as proving bounds on spanning tree congestion).
Scaling-Free Strongly Polynomial Approximation Algorithms

Investigators: Karl Bringmann and Marvin Künemann in cooperation with Karol Wegrzycki (University of Warsaw)

To bypass long-standing time barriers, such as the cubic time barrier for the All Pairs Shortest Path (APSP) problem, remarkable work developed techniques for designing fast approximation algorithms in the polynomial-time regime. In particular, using the scaling technique, Zwick [2] obtained a \((1 + \varepsilon)\)-approximation algorithm for APSP running in time \(\tilde{O}(\frac{n^\omega}{\varepsilon} \log W)\), where \(\omega \leq 2.373\) is the exponent of matrix multiplication and \(W\) is the largest weight. This algorithm, in turn, can be used to approximate several graph characteristics including the diameter, radius, median, minimum-weight triangle, and minimum-weight cycle in the same time bound.

Unfortunately, as a by-product of the scaling technique, Zwick’s algorithm has a factor \(\log W\) in the running time – thus, for large floating-point input numbers it is preferable to use the exact cubic-time algorithm. This raises the question: Can we avoid using the scaling technique and find efficient strongly polynomial approximation schemes for APSP and related problems, i.e., algorithms whose number of arithmetic operations is independent of \(W\)?

In [1], we answer this question in the affirmative. Our main results are as follows.

- We design approximation schemes in strongly polynomial time \(\tilde{O}(\frac{n^\omega}{\varepsilon})\) for APSP on undirected graphs as well as for the graph characteristics diameter, radius, median, minimum-weight triangle, and minimum-weight cycle on directed or undirected graphs.

- For APSP on directed graphs we design an approximation scheme in strongly polynomial time \(\tilde{O}(n^{\frac{\omega+3}{2}} \varepsilon^{-1})\). This is significantly faster than the best exact algorithm.

- We explain why our approximation scheme for APSP on directed graphs has a worse exponent than \(\omega\): Any improvement over our exponent \(\frac{\omega+3}{2}\) would improve the best known algorithm for \((\min, \max)\)-product. In fact, we prove that approximating directed APSP and exactly computing the \((\min, \max)\)-product are equivalent.

Our techniques yield a framework for approximation problems over the \((\min, +)\)-semiring that can be applied more generally. In particular, we obtain the first strongly polynomial approximation scheme for \((\min, +)\)-convolution in strongly subquadratic time, and we prove an equivalence of approximate \((\min, +)\)-convolution and exact \((\min, \max)\)-convolution.

References


Improved Algorithms for Computing the Cycle of Minimum Cost-to-Time Ratio

Investigators: Karl Bringmann in cooperation with Thomas Dueholm Hansen (Aarhus University) and Sebastian Krinninger (University of Salzburg)

In this project, we have revisited the classic problem of computing a cycle of minimum cost-to-time ratio in a given directed graph with \(n\) nodes and \(m\) edges, where each edge is labelled with a cost and a time per unit flow. A typical instance of this problem is ship routing: A vessel for hire can make profit \(p(i,j)\) by going from port \(i\) to port \(j\) taking time \(t(i,j)\); eventually it must complete a cycle with total profit \(P\) and total lapsed time \(T\), where \(P\) is the sum of the profits and \(T\) is the sum of the times on the edges of the cycle. For maximum rate of return, the shipowners should find the cycle that maximizes the cost-to-time ratio; by negating the costs, the maximization problem can be turned into a minimization problem. This minimum cost-to-time ratio problem has a long history in combinatorial optimization and has recently received attention in the context of quantitative verification of hardware and software.

Using Lawler’s binary search approach [3], an optimal solution can be computed by calling a negative cycle detection algorithm for at most \(O(\log(nW))\) times if the costs and times are integers of maximum range \(W\). Most notably, Goldberg’s negative cycle detection algorithm [2] implies a running time of \(\tilde{O}(m\sqrt{n}\log^2 W)\) for the minimum cost-to-time ratio problem. However, such running times are only weakly polynomial and depend logarithmically on the size of the input numbers. They are thus not suitable for computational models that allow symbolic operations on real values. The state of the art in terms of strongly polynomial algorithms is Meggido’s parametric search approach [4] which achieves a running time of \(\tilde{O}(n^3)\). Our new result [1] is an improvement upon this running time for sparse graphs: we develop an algorithm with running time \(\tilde{O}(m^{3/4}n^{3/2})\). This is the first improvement for this long-standing open problem in more than 30 years.

References


Computing Tutte Paths in 2-Connected Planar Graphs

Investigators: Andreas Schmid in cooperation with Jens M. Schmidt (TU Ilmenau)

Tutte paths are one of the most successful tools for attacking problems on long cycles in planar graphs. Unfortunately, existence results based on them are non-constructive, as their
proofs inherently use an induction on overlapping subgraphs and these overlaps prevent any attempt to bound the running time by a polynomial.

For special cases however, computational results of Tutte paths are known: For 4-connected planar graphs, Tutte paths are in fact Hamiltonian paths and Chiba and Nishizeki [1] showed how to compute such paths in linear time. For 3-connected planar graphs, Tutte paths have a significantly more complicated structure, and we have only recently been able to show that they can be computed in polynomial time [3]. However, Tutte paths are defined for general 2-connected planar graphs and this is what most applications need. In this unrestricted setting, no computational results for Tutte paths are known.

In [4] we give the first efficient algorithm that computes a Tutte path (in this unrestricted setting). One of the strongest existence results about such Tutte paths is due to Sanders [2], which allows one to prescribe the end vertices and an intermediate edge of the desired path. Encompassing and strengthening all previous computational results on Tutte paths, we show how to compute such a special Tutte path efficiently. Our method refines both existence results of Thomassen [5] and Sanders [2], and avoids that the subgraphs arising in the inductive proof intersect in more than one edge by using a novel iterative decomposition along 2-separators. Finally, we show that our algorithm runs in time $O(n^2)$.

References


On Computing the Graph Diameter

Investigators: Karl Bringmann in cooperation with Thore Husfeldt (ITU Copenhagen, Lund University), Måns Magnusson (Lund University), and Sebastian Krinninger (University of Salzburg)

The diameter of an edge-weighted graph is the largest distance between any two nodes. In a sparse graph, the diameter can be computed in time $\tilde{O}(n^2)$ by running Dijkstra from every node. This matches a conditional lower bound based on the Strong Exponential Time Hypothesis which rules out truly subquadratic time [2]. On the other hand, for very sparse graphs the diameter can be computed in linear time, e.g., the diameter of a star can be computed by finding the two largest edge weights. Looking more closely, the lower bound even holds restricted to graph with treewidth $O(\log n)$ [2], but the best known algorithms require
treewidth $O(\log n / \log \log n)$ to compute the diameter in truly subquadratic time [2, 5]. Closing this gap was considered to be a possibly fruitful open problem.

In our work [3], we show that the diameter can be computed essentially in linear time in $n$ times a factor of the form $\left(\frac{\text{tw} + \log n}{\text{tw}}\right)$, where tw is the treewidth of the input graph. This yields truly subquadratic time up to treewidth $\delta \log n$ for some constant $\delta > 0$, and thus closes the gap. Our result is an improved analysis of an algorithm of Cabello and Knauer [5] in the regime of non-constant treewidth by revisiting the analysis of orthogonal range searching, improving bounds of the form $\log d n$ to $(d + \log d n)$. We also show how to compute the eccentricities, radius, and Wiener index in the same running time, and we investigate the parameterization by vertex cover number instead of treewidth.

A streamlined exposition of the conditional lower bound for computing the diameter can be found in our work [4]. In particular, we present a unified proof that shows the known lower bounds for diameter in two settings: (1) under the Strong Exponential Time Hypothesis in the random access machine model [2] and (2) unconditionally in the Congest model from distributed computing [1]. This also yields a better hardness of approximation result in the distributed setting.

References


Density Independent Algorithms for Sparsifying k-Step Random Walks

Investigators: Gorav Jindal and Pavel Kolev in cooperation with Richard Peng and Saurabh Sawlani (Georgia Tech)

The transition matrix of a $k$-step random walk in a graph again forms a graph and arises as an intermediate object in a variety of graph algorithms [2, 3]. In [4] we consider the problem of computing a sparse approximation of the the transition matrices of $k$-step random walks on undirected and weighted graphs. Our improvements are based on a better understanding
of processes that sample such walks [1], as well as tighter bounds on key weights underlying these sampling processes.

Given a graph with \( n \) vertices and \( m \) edges, our algorithm produces a graph with about \( n \log n \) edges that approximates the \( k \)-step random walk graph in about \( m + k^2 n \log^4 n \) time. In order to obtain this running time bound, we also revisit “density independent” algorithms for sparsifying graphs whose runtime overhead is expressed only in terms of the number of vertices.

References


Rainbow Coloring

Investigators: Davis Issac in cooperation with Pinar Heggernes, Paloma T. Lima (University of Bergen), Juho Lauri (Nokia Bell Labs, Dublin), Erik Jan van Leeuwen (Utrecht University), Anita Das (Infosys Bangalore, India), and L. Sunil Chandran (Indian Institute of Science)

A well-studied coloring problem is to assign colors to the edges of a graph \( G \) so that, for every pair of vertices, all edges of at least one path between them receive different colors. The minimum number of colors necessary in such a coloring is the rainbow connection number \( rc(G) \) of the graph. A more restricted coloring called strong rainbow coloring requires that there is at least one shortest path between every pair of vertices. The minimum number of colors necessary in such a coloring is the strong rainbow connection number \( src(G) \) of the graph.

When proving upper bounds on \( src(G) \), it is natural to prove that a coloring exists where, for every shortest path between every pair of vertices in the graph, all edges of the path receive different colors. Therefore, in [1], we introduce and formally define this more restricted edge coloring number, which we call very strong rainbow connection number \( vsrc(G) \). We also give upper bounds on \( vsrc(G) \) for several graph classes, some of which are tight. These immediately imply new upper bounds on \( src(G) \) for these classes, showing that the study
of $\text{vsrc}(G)$ enables meaningful progress on bounding $\text{src}(G)$. Then we study the complexity of the problem of computing $\text{vsrc}(G)$, particularly for graphs of bounded treewidth, and show that this is an interesting problem in its own right. We prove that $\text{vsrc}(G)$ can be computed in polynomial time on cactus graphs; in contrast, this question is still open for $\text{src}(G)$. We also observe that deciding whether $\text{vsrc}(G) = k$ is fixed-parameter tractable in $k$ and the treewidth of $G$. Finally, on general graphs, we prove that there is no polynomial-time algorithm to decide whether $\text{vsrc}(G) \leq 3$ nor to approximate $\text{vsrc}(G)$ within a factor $n^{1-\epsilon}$, unless P=NP.

In [2], we study the vertex variants of rainbow coloring. Given a graph with colors on its vertices, a path is called a rainbow vertex path if all its internal vertices have distinct colors. We say that the graph is rainbow vertex-connected if there is a rainbow vertex path between every pair of its vertices. We study the problem of deciding whether the vertices of a given graph can be colored with at most $k$ colors so that the graph becomes rainbow vertex-connected. Although edge-colorings have been studied extensively under similar constraints, there are significantly fewer results on the vertex variant that we consider. In particular, its complexity on structured graph classes was explicitly posed as an open question. We show that the problem remains NP-complete even on bipartite apex graphs and on split graphs. The former can be seen as a first step in the direction of studying the complexity of rainbow coloring on sparse graphs, an open problem which has attracted attention but has seen limited progress. We also give hardness of approximation results for both bipartite and split graphs. To complement the negative results, we show that bipartite permutation graphs, interval graphs, and block graphs can be rainbow vertex-connected optimally in polynomial time.

References


Spanning Tree Congestion and Connected Partitioning

Investigators: Davis Issac and Yun Kuen Chueng in cooperation with L. Sunil Chandran (Indian Institute of Science)

We study a natural problem in graph sparsification, the Spanning Tree Congestion problem. Informally, the Spanning Tree Congestion problem seeks a spanning tree with no tree-edge routing too many of the original edges. The root of this problem dates back to at least 30 years ago, motivated by applications in network design, parallel computing and circuit design. Variants of the problem have also seen algorithmic applications as a preprocessing step of several important graph algorithms.
For any general connected graph with \( n \) vertices and \( m \) edges, we show that its spanning tree congestion is at most \( \mathcal{O}(\sqrt{mn}) \), which is asymptotically optimal since we also demonstrate graphs with spanning tree congestion at least \( \Omega(\sqrt{mn}) \). We present a polynomial-time algorithm which computes a spanning tree with congestion \( \mathcal{O}(\sqrt{mn} \cdot \log n) \). We also present another algorithm for computing a spanning tree with congestion \( \mathcal{O}(\sqrt{mn}) \); this algorithm runs in sub-exponential time when \( m = \omega(n) \).

For achieving the above results, an important intermediate theorem is the generalized Győri-Lovász theorem, for which only a non-constructive proof was known. We give the first elementary and constructive proof by providing a local search algorithm with running time \( \mathcal{O}^*(4^n) \), which is a key ingredient of the above-mentioned sub-exponential time algorithm. We also show that for any graph which satisfies certain expansion properties, its spanning tree congestion is at most \( \mathcal{O}(n) \), and a corresponding spanning tree can be computed in polynomial time. We then use this to show that a random graph has spanning tree congestion \( \Theta(n) \) with high probability.

References


Geometric Inhomogeneous Random Graphs and the Algorithmic Small-World Phenomenon

Investigators: Karl Bringmann in cooperation with Ralph Keusch, Johannes Lengler (ETH Zurich), Yannic Maus, and Anisur Molla (University of Freiburg)

Real-world networks, like social networks or the internet infrastructure, have structural properties such as large clustering coefficients that can best be described in terms of an underlying geometry. This is why the focus of the literature on theoretical models for real-world networks shifted from classic models without geometry, such as Chung-Lu random graphs [3], to modern geometry-based models, such as hyperbolic random graphs [5].

In our work [1], we contribute to the theoretical analysis of these modern, more realistic random graph models. Instead of studying directly hyperbolic random graphs, we use a generalization that we call geometric inhomogeneous random graphs. Since we ignore constant factors in the edge probabilities, geometric inhomogeneous random graphs are technically simpler (specifically, we avoid hyperbolic cosines), while preserving the qualitative behaviour of hyperbolic random graphs.

We prove the following fundamental structural and algorithmic results on geometric inhomogeneous random graphs. As our main contribution we provide a sampling algorithm that generates a random graph from our model in expected linear time, improving the best-known sampling algorithm for hyperbolic random graphs by a substantial factor \( \mathcal{O}(\sqrt{n}) \).

We establish that geometric inhomogeneous random graphs have clustering coefficients in \( \Omega(1) \), we prove that they have small separators, i.e., it suffices to delete a sublinear number
of edges to break the giant component into two large pieces, and we show how to compress these graphs using an expected linear number of bits.

Based on these insights into the basic properties of geometric inhomogeneous random graphs, we study the algorithmic small-world phenomenon in [2]. This phenomenon describes that in the human acquaintance graph very short paths exist between two typical persons, and people are able to find such paths without global knowledge of the network. It was empirically established by Milgram’s letter forwarding experiments from the 60s [6], and theoretically explained by Kleinberg in 2000 [4]. However, from today’s perspective Kleinberg’s model has several severe shortcomings that limit the applicability to real-world networks. In order to give a more convincing explanation of the algorithmic small-world phenomenon, we study greedy routing in geometric inhomogeneous random graphs, which overcomes some of the previous shortcomings.

We show that greedy routing succeeds with constant probability, and in case of success almost surely finds a path that is an almost shortest path. Our results are robust to changes in the model parameters and the routing objective. Moreover, since constant success probability is too low for technical applications, we study natural local patching methods augmenting greedy routing by backtracking and we show that such methods can ensure success probability 1 in a number of steps that is close to the shortest path length.

These results also address the question of Krioukov et al. [5] whether there are efficient local routing protocols for the internet graph. There were promising experimental studies, but the question remained unsolved theoretically. Our results give for the first time a rigorous and analytical answer, assuming our random graph model.

References


Dimensionality of Graphs

Investigators: Bundit Laekhanukit in cooperation with Roee David (Datorama) and Karthik C.S. (Weizmann Institute of Science)

Understanding the dimensionality of a graph in various measurements is an important question in both theory and practice as these measurements show hidden parameters of a graph that are important in developing algorithms for graph and network problems. In [1], we study two notions of the dimension of a graph, the sphericity and contact dimension, defined as follows. Every graph \( G \) can be represented by a collection of equi-radii spheres in a \( d \)-dimensional metric \( \Delta \) such that there is an edge \( uv \) in \( G \) if and only if the spheres corresponding to \( u \) and \( v \) intersect. The smallest integer \( d \) such that \( G \) can be represented by a collection of spheres (all of the same radius) in \( \Delta \) is called the sphericity of \( G \), and if the collection of spheres are non-overlapping, then the value \( d \) is called the contact dimension of \( G \). We study the sphericity and contact dimension of the complete bipartite graph \( K_{n,n} \) in various \( L^p \)-metrics. We also connect this to the complexity of the monochromatic closest pair and bichromatic closest pair problems.

References


35.5.4 Miscellaneous

Improved Bounds on Fourier Entropy and Min-Entropy

Investigators: Nitin Saurabh in cooperation with Srinivasan Arunachalam (MIT), Sourav Chakraborty (ISA Kolkata), Michal Koucký (Charles University Prague), and Ronald de Wolf (CWI Amsterdam)

A longstanding conjecture by Mansour [4] states that for a Boolean function computable by a DNF formula with \( m \) terms, most of its Fourier mass is concentrated on only poly(\( m \))-many coefficients. This would imply a polynomial-time agnostic learning algorithm for DNF’s [3] answering a major open question in computational learning theory.

Mansour’s conjecture is about approximation by sparse polynomials, but in our work [1] we use a different approach. In particular, we consider the Fourier entropy-influence conjecture (FEIC) posed by Friedgut and Kalai [2] which states that the Fourier entropy of a Boolean function is at most the influence of the function. It is known that bounding the Fourier entropy of a DNF by the bottom fan-in (or the logarithm of the top fan-in) suffices to imply Mansour’s conjecture. We take exactly this approach and verify Mansour’s conjecture for “unambiguous” DNFs. More generally, we give a better upper bound on the Fourier entropy in terms of the average unambiguous certificate complexity of a Boolean function. Informally, the unambiguous certificate complexity is similar to the standard certificate complexity measure, except the collection of certificates are now required to be unambiguous, i.e., every input should be consistent with a unique certificate. By the average unambiguous certificate
complexity we mean the expected number of bits set by an unambiguous certificate on a uniformly random input.

We also study the Fourier min-entropy-influence conjecture (FMEIC). This is a weaker version of the FEIC stating that the min-entropy of the Fourier distribution is bounded by the total influence of the function. Although weaker, it suffices to imply most of the consequences of the FEIC (cf. [5]). We verify the FMEIC for Read-$k$ DNFs.

We also study consequences of the FEIC on the structure of polynomials that pointwise approximate a Boolean function. In this work we ask, suppose the FEIC were true, what can be said about approximating polynomials? For instance, are these approximating polynomials $p$ sparse in their Fourier domain? Do approximating polynomials have small spectral norm? We prove that a restricted class of polynomials called flat block-multilinear polynomials of degree $d$ and sparsity $2^{o(d)}$ cannot $1/3$-approximate any Boolean function.

References


35.6 Combinatorial Optimization

Coordinator: Andreas Karrenbauer

Optimization problems are ubiquitous. Whenever there is a choice between several alternatives, we are faced with an optimization problem. In many cases, the number of feasible solutions is overwhelming. However, if the set of feasible solutions is discrete, e.g., the product of binary decisions, it can often be described by combinatorial objects such as paths or cycles in a graph. Various methods have been developed to tackle such problems: integer programming, parameterized and exact algorithms, approximation algorithms, and combinatorial algorithms, among others. Researchers from D1 contribute by inventing new techniques to design more efficient algorithms, to obtain sharper analyses, and to tighten upper and lower bounds on the computational complexity of various optimization problems. In this area, our group was visible at many top venues such as STOC, FOCS, SODA, ICALP, IPCO, and NIPS.
35.6.1 Approximation Algorithms and Hardness of Approximation

The study of approximation algorithms provides a more detailed view to differentiate combinatorial optimization problems that are equivalent in terms of NP-hardness. To this end, we investigate the complexity of finding solutions that are provably not far from the optimum. Formally, we say that an algorithm is an \(\alpha\)-approximation algorithm for a certain optimization problem, if for any instance of the problem, the algorithm outputs a solution whose objective value is within a factor \(\alpha\) of the optimum. If not stated otherwise, the algorithm is supposed to run in time polynomial in the size of the input. In some cases, we allow quasi-polynomial time, for example if this allows for making \(\alpha\) arbitrarily close to 1.

Besides our quest for designing algorithms with improved approximation factors, we also analyze the complexity of approximation problems, i.e., we derive bounds for the best \(\alpha\) any algorithm for a certain problem can achieve assuming some complexity hypothesis (typically \(P \neq NP\)). Our work spans different areas, including for instance graph problems, packing problems, and problems stemming from geometry.

Furthermore, we also consider approximation algorithms for problems that are solvable in polynomial time because already an algorithm with quadratic runtime might be prohibitive in the context of big data, whereas a \(1 + \epsilon\)-approximation running in quasi-linear time could be sufficient.

**Approximating Geometric Knapsack via L-packings**

*Investigators: Sandy Heydrich in cooperation with Waldo Gálvez, Fabrizio Grandoni, Salvatore Ingala, Arindam Khan (IDSIA, USI-SUPSI, Switzerland), and Andreas Wiese (Universidad de Chile)*

We study the two-dimensional geometric Knapsack problem in which we are given a set of \(n\) axis-aligned rectangular items, each one with an associated profit, and an axis-aligned square knapsack. The goal is to find a (non-overlapping) packing of a maximum profit subset of items inside the knapsack (without rotating items). The best-known polynomial-time approximation factor for this problem (even just in the cardinality case where all items have equal profit) is \(2 + \epsilon\) [2]. In our work we break the 2-approximation barrier, achieving a polynomial-time \(\frac{17}{16} + \epsilon < 1.89\)-approximation, which improves to \(\frac{558}{325} + \epsilon < 1.72\) in the cardinality case [1].

Essentially all prior work on the two-dimensional geometric Knapsack problem approximation packs items inside a constant number of rectangular containers, where items inside each container are packed using a simple greedy strategy. We deviate for the first time from this setting: we show that there exists a large profit solution where items are packed inside a constant number of containers plus one L-shaped region at the boundary of the knapsack, which contains items that are high and narrow and items that are wide and thin. The items of these two types possibly interact in a complex manner at the corner of the L.

The above structural result is not enough however: the best-known approximation ratio for the subproblem in the L-shaped region is \(2 + \epsilon\) (obtained via a trivial reduction to one-dimensional Knapsack by considering tall or wide items only). Indeed this is one of the simplest special settings of the problem for which this is the best known approximation factor. As a second major, and the main algorithmic contribution, we present a PTAS for
this case. We believe that this will turn out to be useful in future work in geometric packing problems.

We also consider the variant of the problem with rotations, where items can be rotated by 90 degrees. Also in this case the best-known polynomial-time approximation factor (even for the cardinality case) is $2 + \epsilon$ [2]. Exploiting part of the machinery developed for the variant without rotation plus a few additional ideas, we obtain a polynomial-time $\frac{3}{2} + \epsilon$-approximation for the variant with rotation, which improves to $\frac{4}{3} + \epsilon$ in the cardinality case [1].

References


Stabbing Rectangles by Line Segments – How Decomposition Reduces the Shallow-Cell Complexity

*Investigators: Krzysztof Fleszar in cooperation with Timothy M. Chan (UIUC), Thomas C. van Dijk, Joachim Spoerhase, and Alexander Wolff (University of Würzburg)*

We initiate the study of the following natural geometric optimization problem [1]. The input is a set of axis-aligned rectangles in the plane. The objective is to find a set of horizontal line segments of minimum total length so that every rectangle is stabbed by some line segment. A line segment stabs a rectangle if it intersects its left and its right boundary (see Fig. 35.2). The problem, which we call STABBING, can be motivated by a resource allocation problem and has applications in geometric network design [3]. To the best of our knowledge, only special cases of this problem have been considered so far.

STABBING is a weighted geometric set cover problem, which we show to be NP-hard. A constrained variant of STABBING turns out to be even APX-hard. While for general set cover the best possible approximation ratio is $\Theta(\log n)$, it is an important field in geometric approximation algorithms to obtain better ratios for geometric set cover problems. Chan et al. [2] generalized Varadarajan’s results based on quasi-uniform sampling [4] to obtain sub-logarithmic performances for a broad class of weighted geometric set cover instances that are characterized by having low shallow-cell complexity. The shallow-cell complexity of STABBING instances, however, can be high so that a direct application of the framework of
Chan et al. gives only logarithmic bounds. We still achieve a constant-factor approximation by decomposing general instances into what we call laminar instances that have low enough complexity. Our decomposition technique yields the same approximation factor also for the variant where the rectangles can be stabbed by horizontal and vertical segments.

Our results provide natural examples for the fact that the property of having low shallow-cell complexity is not closed under the union of the set families. In spite of this, constant-factor approximations are still possible. Our results also show that the representation as a union of low-complexity families may not be obvious at first glance. We therefore hope that our approach helps to extend the reach of quasi-uniform sampling beyond the concept of low shallow-cell complexity also in other settings.

References


Unsplittable Flow on a Path

*Investigators: Hang Zhou in cooperation with Fabrizio Grandoni (IDSIA Switzerland), Tobias Mönke (Saarland University), and Andreas Wiese (University of Chile).*

In the problem of *Unsplittable Flow on a Path (UFP)*, we are given a path with non-negative edge capacities and a set of tasks, each one characterized by a subpath, a demand, and a profit. Our goal is to select a subset of tasks of maximum total profit so that the total demand of the selected tasks on each edge does not exceed the respective edge capacity. UFP naturally captures several applications in bandwidth allocation, job scheduling, and caching. The long-standing open question is whether there is a polynomial-time approximation scheme for UFP. In our work [2], we studied the power of resource augmentation for UFP by developing a new technique to create slack, some free space, to help the algorithmic design. In later work [3], we introduced a new packing method where small tasks are placed into boxes, and we improved the approximation for UFP from the previous best ratio of \((2 + \epsilon) [1]\) to the ratio of \((5/3 + \epsilon)\).
Airports and Railways

**Investigators:** Antonios Antoniadis in cooperation with Anna Adamaszek (University of Copenhagen), Amit Kumar (IIT Delhi), and Tobias Mömke (Saarland University)

Airports and Railways is a framework of problems, useful for modelling a wide range of connectivity problems, which was first introduced by Adamaszek, Antoniadis and Mömke in [2], and combines capacitated facility location with network design. In this framework one is given a graph with weights on the vertices and on the edges, together with a parameter \( k \). The vertices of the graph represent cities, and weights denote the costs of opening airports in the cities and building railways that connect pairs of cities, respectively. The parameter \( k \) can be thought of as the capacity of an airport. The goal is to construct a minimum cost network of airports and railways connecting the cities, where each connected component in the network spans at most \( k \) vertices, contains an open airport, and the network satisfies some additional requirements specific to the problem in the framework.

Building upon our previous work on the problem, in [1] we were able to obtain the first bicriteria approximation algorithm for Airports and Railways in the general metric case, which yields a 4-approximate solution with a resource augmentation of the airport capacity \( k \) by a factor 2. More generally, for any parameter \( 0 < p \leq 1 \), where \( pk \) is an integer, we develop a \((4/3)(2 + 1/p)\)-approximation algorithm for metric Airports and Railways with a resource augmentation by a factor of \( 1 + p \). Furthermore, we obtain the first constant factor approximation algorithm that does not resort to resource augmentation for Airports and Railways in the Euclidean plane. Additionally, for the Euclidean setting we provide a quasi-polynomial time approximation scheme for the same problem with a resource augmentation by a factor of \( 1 + \mu \) on the airport capacity, for any fixed \( \mu > 0 \).

**References**

A QPTAS for the General Scheduling Problem with Identical Release Dates

Investigators: Antonios Antoniadis in cooperation with Ruben Hoeksma (University of Bremen), Julie Meißner (TU Berlin), José Verschae (Universidad Catolica, Santiago, Chile), and Andreas Wiese (Universidad de Chile)

The General Scheduling Problem generalizes numerous interesting scheduling problems with objective functions given by the sum of the respective job costs. Two important examples are minimizing the total weighted flow time or the total weighted tardiness.

Given a set of jobs with processing times, release dates, and job dependent cost functions, one seeks to find a minimum cost preemptive schedule on a single machine. The best known algorithm for this problem is an $O(\log \log P)$-approximation (where $P$ denotes the range of the job processing times) [2], while the best lower bound shows only strong NP-hardness.

An interesting special case of the General Scheduling Problem is when the release dates are identical. Here, the problem remains strongly NP-hard and the best known approximation algorithm has a ratio of $e + \varepsilon$ (running in quasi-polynomial time). In [1] we reduce the latter gap by giving a quasi-polynomial-time approximation scheme (QPTAS) if the numbers in the input are quasi-polynomially bounded. This rules out the existence of an APX-hardness proof under common complexity-theoretic assumptions. Our techniques are based on the QPTAS known for the covering version of the so-called unsplittable-flow on a path problem (UFP-Cover), a particular case of the General Scheduling problem where we must pick a subset of intervals (jobs) on the real line with associated heights and costs. If an interval is selected, its height will help cover a given demand on any point contained within the interval. We reduce our problem to a generalization of UFP-Cover and use a sophisticated divide-and-conquer procedure with interdependent non-symmetric subproblems.

Additionally, we also present a QPTAS for two further variants of UFP-Cover. For the case of agreeable intervals we give an algorithm based on a new dynamic programming approach which might be useful for other problems of this type. The second one is a resource augmentation setting where we are allowed to slightly enlarge each interval.

References


Optimization of Bootstrapping in Circuits

Investigators: Hang Zhou in cooperation with Fabrice Benhamouda (IBM Research), Tancrède Lepoint (Google), and Claire Mathieu (ENS Paris)

In 2009, Gentry proposed the first Fully Homomorphic Encryption (FHE) scheme [2], an extremely powerful cryptographic primitive that enables to perform computations, i.e., to evaluate circuits, on encrypted data without decrypting it first. This has many applications, particularly in cloud computing.

In all currently known FHE schemes, encryptions are associated with some (non-negative integer) noise level. At each evaluation of an AND gate this noise level increases. This increase is problematic because decryption succeeds only if the noise level stays below some maximum level \( L \) at every gate of the circuit. To ensure that property, it is possible to perform an operation called bootstrapping to reduce the noise level. Though critical, bootstrapping is a time-consuming operation. This expense motivates a new problem in discrete optimization: minimizing the number of bootstrappings in a circuit while still controlling the noise level.

In our work [1], we (1) formally define the bootstrap problem, (2) design a polynomial-time \( L \)-approximation algorithm using a novel method of rounding of a linear program, and (3) show a matching hardness result: \( (L - \epsilon) \)-inapproximability for any \( \epsilon > 0 \).

References


A PTAS for Euclidean TSP with Hyperplane Neighborhoods

Investigators: Antonios Antoniadis, Krzysztof Fleszar, and Kevin Schewior in cooperation with Ruben Hoeksma (University of Bremen)

The Traveling Salesperson Problem (TSP) is commonly regarded as one of the most important problems in combinatorial optimization. In TSP, a salesperson wishes to find a tour that visits a set of clients in the shortest way possible. A very natural generalization of TSP is motivated by clients that are not static (as in TSP) but willing to move in order to meet the salesperson: In the Traveling Salesperson Problem with Neighborhoods, one is given a collection of geometric regions in some space. The goal is to output a minimum length tour that visits at least one point in each region.

Even in the Euclidean plane, TSP with Neighborhoods is known to be APX-hard [2], which gives rise to studying more tractable special cases of the problem. In [1] we focused on the fundamental special case of regions that are hyperplanes in the \( d \)-dimensional Euclidean space. While for \( d = 2 \) an exact algorithm with running time \( O(n^3) \) is known [5], settling the exact approximability of the problem for \( d = 3 \) has been repeatedly posed as an open
question over the past 15 years (see, e.g., [3, 4]). To date, only an approximation algorithm with guarantee exponential in \(d\) was known [4], and NP-hardness remains open.

Our main result is a Polynomial Time Approximation Scheme (PTAS) for any arbitrary fixed dimension \(d\) that works for both the tour and path version of the problem. Our algorithm is based on approximating the convex hull of the optimal tour by a convex polytope of bounded complexity. Such polytopes are represented as solutions of a sophisticated LP formulation, which we combine with the enumeration of crucial properties of the tour. In the analysis of our approximation scheme, we show that our search space includes a sufficiently good approximation of the optimum. To do so, we develop a novel and general sparsification technique to transform an arbitrary convex polytope into one with a constant number of vertices and, in turn, into one of bounded complexity in the above sense. Hereby, we maintain important properties of the polytope.

References


Approximability of Network Design Problems

Investigators: Bundit Laekhanukit in cooperation with Marek Cygan (University of Warsaw), Fabrizio Grandoni (IDSIA), Guy Kortsarz (Rutgers University), and Shi Li (University at Buffalo)

Network design is a class of problems that has been studied for decades due to its importance in both theory and practice. Unfortunately, most of problems in this classes are NP-hard, which motivates to study approximation algorithms. We investigated their approximability of these problems under various different running time restrictions.

In [4], we study the approximability of the infamous Directed Steiner Tree problem in the regime of quasi-polynomial-time algorithms and prove tight approximability of this problem. In the Directed Steiner Tree problem, we are given an \(n\)-vertex directed edge-weighted graph, a root \(r\), and a collection of \(k\) terminal nodes. Our goal is to find a minimum-cost set of edges that contains a directed path from \(r\) to every terminal. We present an \(O(\log^2 k/\log \log k)\)-approximation algorithm for DST that runs in quasi-polynomial-time. By adjusting the parameters in the hardness result of Halperin and Krauthgamer, we complement our algorithm by a matching lower bound of \(\Omega(\log^2 k/\log \log k)\) for the class...
of quasi-polynomial-time algorithms. This is the first improvement on the Directed Steiner Tree problem since the classical quasi-polynomial-time $O((\log^3 k))$-approximation algorithm by Charikar et al. [1].

The area of subexponential-time algorithms has recently become a trend of research attracting many researchers. In [2], we study the inapproximability of Directed Steiner Tree and relate problems in the regime of subexponential-time algorithms. We prove almost tight (super-polynomial) conditional lower bounds, for achieving desired approximation ratios for various problems. To illustrate our results, let us consider the Set Cover problem with $n$ elements and $m$ sets. Suppose that we want to approximate Set Cover to a factor of $(1 - \alpha) \ln n$, for a given parameter $0 < \alpha < 1$. What is the best possible running time for achieving such an approximation? This question was answered implicitly in the work of Moshkovitz [5]: Assuming both the Projection Games Conjecture and the Exponential Time Hypothesis, any $((1 - \alpha) \ln n)$-approximation algorithm for Set Cover must run in time $\geq 2^{\Omega(n^c)}$, for some constant $0 < c < 1$. Our work extends this line of research. First, we show that under the Projection Games Conjecture and the Exponential Time Hypothesis, any $((1 - \alpha) \ln n)$-approximation for Set Cover requires time $2^{\Omega(n^\alpha)}$. This (almost) matches the running time of $2^{O(n^\alpha)}$ for approximating Set Cover to a factor $(1 - \alpha) \ln n$ by Cygan et al. [3]. Our result is tight up to the constant multiplying the $n^\alpha$ term in the exponent. Our lower bound applies to all generalizations of Set Cover, e.g., Group Steiner Tree, Directed Steiner Tree, Covering Steiner Tree, and Connected Polymatroid. We also show that in almost exponential time, these generalizations reduce back to Set Cover: We design $((1 - \alpha) \ln n)$-approximation algorithms for all these problems that run in time $2^{n^\alpha \log n \text{poly}(m)}$.

References


Survivable Network Design for Group Connectivity

Investigators: Bundit Laekhanukit and Daniel Vaz in cooperation with Parinya Chalermsook (Aalto University), Syamantak Das (IIIT Delhi), and Guy Even (Tel-Aviv University)

The GROUP STEINER TREE (GST) problem is a classical problem in combinatorial optimization and theoretical computer science. In the edge-cost variant of the GST problem, we are given an undirected graph $G = (V, E)$ on $n$ vertices with edge costs $c : E \to \mathbb{R}_{\geq 0}$, a
source vertex \( s \) and a collection of \( h \) subsets of vertices, called groups, \( S_1, \ldots, S_h \subseteq V \). The goal is to find a minimum-cost tree \( H \subseteq G \) that connects \( s \) to some vertex from each group \( S_i \), for all \( i = 1, 2, \ldots, h \). In our work, we consider a fault-tolerant variant of GST which we call (Restricted Rooted) Group SNDP. In this setting, each group \( S_i \) has a demand \( k_i \in [k], k \in \mathbb{N} \), and we wish to find a minimum-cost subgraph \( H \subseteq G \) such that, for each group \( S_i \), there is a vertex in the group that is connected to the root via \( k_i \) (vertex or edge) disjoint paths.

While GST admits an \( O(\log^2 n \log h) \) approximation, its higher connectivity variants are known to be Label-Cover hard [6], and for the vertex-weighted version, the hardness holds even when \( k = 2 \). This implies that the problem admits no \( 2^{\log^{1-\epsilon} n} \)-approximation (unless \( \textbf{NP} \subseteq \textbf{DTIME}(n^{\log^{o(1)} n}) \)), and it is widely believed that it has no subpolynomial approximation in polynomial time. The GST problem, which corresponds to the case of \( k = 1 \), admits no \( O(\log^{2-\epsilon} h) \)-approximation (unless \( \textbf{NP} \subseteq \textbf{ZTIME}(n^{\log^{o(1)} n}) \)) [5].

Previously, positive results were known only for the edge-weighted case when \( k = 2 \) [4, 6] and for a relaxed variant where \( k_i \) disjoint paths from \( s \) may end at different vertices in a group [3], for which the authors gave a bicriteria approximation where connectivity demands are approximately met. For \( k \geq 3 \), there was no non-trivial approximation algorithm known for Group SNDP, except for the special case of the relaxed variant on trees, which follows from results for GST.

In our paper [1], we present an \( O(\log n \log h) \) approximation algorithm for Group SNDP that runs in time \( n^{f(k,w)} \), where \( w \) is the treewidth of the input graph. Our algorithm works for both edge and vertex weighted variants, and the approximation ratio nearly matches the lower bound when \( k \) and \( w \) are constants. The key to achieving this result is a framework that non-trivially extends the results of our earlier work [2].

This framework first embeds all feasible solutions to the problem into a dynamic programing table. However, finding the optimal solution in this table remains intractable. We formulate a linear program relaxation for the dynamic program and obtain an approximate solution via randomized rounding. This framework also allows us to systematically construct dynamic programs for high-connectivity problems (without group constraints). As a result, we present new exact algorithms for several variants of survivable network design problems in low-treewidth graphs.

References

Disjoint paths problems are elementary problems with a long history and significant connections to optimization and structural graph theory. In our research, we studied the classical NP-hard versions of finding maximum-size subsets from given sets of $k$ terminal pairs that can be routed via edge-disjoint paths (MAXEDP) or node-disjoint paths (MAX NDP) in a given graph. The approximability of MAXEDP/MAX NDP is currently not well understood. There is a significant gap between the best known lower [2] and upper bound [1], and closing this gap is currently one of the big open problems in approximation algorithms.

As an outcome of our research [3], we strengthen fundamental results for these problems. We provide new bounds formulated in terms of the feedback vertex set number $r$ of a graph, which measures its vertex deletion distance to a forest. In particular, we obtain the following results:

- For MAXEDP, we give an $O(\sqrt{r} \log kr)$-approximation algorithm. Up to a logarithmic factor, our result strengthens the best known ratio $O(\sqrt{n})$ by Cherkuri et al. [1], as $r \leq n$.

- Further, we show how to route $\Omega(OPT^*)$ pairs with congestion $O(\log kr/\log \log kr)$, strengthening the bound obtained by the classic approach of Raghavan and Thompson [4].

- For MAX NDP, we give an algorithm that gives the optimal answer in $(k + r)^{O(r)} \cdot n$ time. This result is a substantial improvement on the run time of $2^k r^{O(r)} \cdot n$, which can be obtained via an algorithm by Scheffler [5].

We complement these positive results by various hardness bounds.

Our main insights are twofold. First, the parameter $r$ seems to correlate quite well with the “difficulty” of disjoint paths problems. Second, it suffices to study graphs with close to linear feedback vertex set number in order to improve the best known ratio of Chekuri et al.

References

Triangular Cacti and the Maximal Planar Subgraph Problem

Investigators: Andreas Schmid in cooperation with Parinya Chalermsook (Aalto University)

In the Maximum Planar Subgraph problem, we are given a graph $G$ and our goal is to find a planar subgraph $H$ of $G$ with maximum number of edges. Besides being a basic problem in graph theory, Maximum Planar Subgraph has many applications including, e.g., circuit design, factory layout, and graph drawing, so it has received significant attention in both theoretical and empirical literature. Since the problem is NP-hard, past research has focused on approximation algorithms. An approximation ratio of $\frac{1}{3}$ can be easily achieved by simply returning any spanning tree of $G$. The current best known approximation ratio is $\frac{4}{9}$ obtained two decades ago [1] based on computing a maximum cactus subgraph. A cactus graph is a graph in which any edge is contained in a cycle and any two cycles are edge-disjoint.

In [2] we investigate how well a cactus subgraph performs for the Maximum Planar Triangles problem. Here instead of maximizing the number of edges in the solution, we are looking for a planar subgraph with a maximum number of triangular faces. We present a constructive proof of the fact that any plane graph $G$ contains a cactus subgraph $C$ where $C$ contains at least a $\frac{1}{6}$-fraction of the triangular faces of $G$. We also show that this ratio cannot be improved by showing a tight lower bound. Together with an algorithm for linear matroid parity, our bound implies two approximation algorithms for computing “dense planar structures” inside any graph: (i) A $\frac{1}{6}$-approximation algorithm for, given any graph $G$, finding a planar subgraph with a maximum number of triangular faces; this improves upon the previous $\frac{1}{11}$-approximation for Maximum Planar Triangles; (ii) An alternate (and arguably more illustrative) proof of the $\frac{4}{9}$-approximation algorithm for Maximum Planar Subgraph.

Our bound is obtained by analyzing a natural local search strategy and heavily exploiting the exchange arguments, which provides an example for the power of local search.

References


Steiner Point Removal

Investigator: Yun Kuen Cheung

Graph sparsification generally describes a transformation of a large graph into a smaller graph that preserves, either exactly or approximately, certain features (e.g., distance, cut, flow) of the large graph. Its algorithmic value is clear, since the compressed graph can be computed in a preprocessing step of an algorithm, so as to reduce subsequent running time and memory requirement.

Here, we study a graph sparsification problem called Steiner Point Removal (SPR). Given a weighted graph $G = (V,E,w)$ with a set of $k$ terminals $T \subset V$, the SPR problem seeks for a new graph $G'$ which is a minor of $G$, such that the vertex set of $G'$ is $T$ (i.e., no non-terminals in the minor) and the distance between every pair of terminals is preserved within a multiplicative distortion $\alpha$:

$$\forall t_i, t_j \in T, \quad \text{dist}_G(t_i, t_j) \leq \text{dist}_{G'}(t_i, t_j) \leq \alpha \cdot \text{dist}_G(t_i, t_j).$$

The target is to minimize $\alpha$. The requirement that $G'$ is a minor of $G$ is crucial since minor operations preserve certain structural similarities, e.g., planarity, of the input graph $G$. Such preservation is important for applying certain algorithms, since they might only work or be efficient for planar graphs or other specific input graphs.

Kamma et al. [2] used a ball-growing algorithm to show that the distortion is at most $O(\log^5 k)$ for general graphs. Briefly speaking, the algorithm repeatedly grows balls around the $k$ terminals with radii chosen randomly according to certain exponential distributions. The ball of a terminal keeps “absorbing” non-terminals until all non-terminals are gone. By showing that not too many close-by non-terminals are absorbed by different terminals, together with some extra arguments, the bound of $O(\log^5 k)$ can be achieved with high probability.

We improve the distortion bound to $O(\log^2 k)$ [1]. We use essentially the same algorithm as Kamma et al., but we pre-process the input graph using an algorithm by Krauthgamer et al. [3]. The pre-processing step outputs a new graph with at most $O(k^4)$ vertices, while preserving all terminal distances exactly. By having only $\text{poly}(k)$ vertices left in this new graph, we are able to show that there are only $O(k^3) = \text{poly}(k)$ bad events that might lead to a large distortion, and we show that each of them occur with probability at most $O(1/k^6)$.

References

The Itinerant List-Update Problem

Investigators: Kevin Schewior in collaboration with Neil Olver, René Sitters, Leen Stougie (CWI and VU Amsterdam), and Kirk Pruhs (University of Pittsburgh)

We introduce the itinerant list-update problem (ILU) problem [2], which models the memory-management decisions that have to be made in Domain Wall Memory [3], an emerging and efficient memory technology. The online version can be interpreted as a relaxation of the classic list update problem [4] in which the pointer no longer has to return to a home location after each request. The classic list update problem was one of the first problems studied in online algorithms. We consider both approximation algorithms and online algorithms (via competitive analysis) for this problem. Depending on the application, both versions may be relevant in practice.

We show that offline ILU is essentially equivalent to a dynamic variation of the classical minimum linear arrangement problem (MLA), which we call DMLA. Both ILU and DMLA are very natural, but have not been studied before. We then give an $O(\log^2 n)$-approximation algorithm for DMLA, which carries over to ILU. While the approach is based on well-known divide-and-conquer approaches for the standard MLA problem [1], the dynamic nature of these problems introduces substantial new difficulties.

We also show an $\Omega(\log n)$ lower bound on the competitive ratio for any randomized online algorithm for ILU. This shows that online ILU is more difficult (with respect to competitive ratio) than the classical list update problem, for which $O(1)$-competitive algorithms, like Move-To-Front, are known.

It remains open whether our offline or online bounds are asymptotically tight.

References


Algorithms for Energy-Efficient Scheduling

Investigators: Antonios Antoniadis and Sebastian Ott in cooperation with Neal Barcelo, Michael Nugent, Kirk Pruhs (University of Pittsburgh), Mario Consuegra (Florida International University, Miami), Peter Kling (Hamburg University), and Sören Riechers (Paderborn University)

Energy consumption is increasingly becoming a crucial concern in computing environments. The global energy consumption of such computing environments is nowadays comparable
to that of the aviation industry and is currently growing at a rate of 10-12% per year. Limitations in battery technology further highlight the importance of being able to develop algorithms that are not only time- and space-efficient, but energy-efficient as well. To this end, chip manufacturers increasingly incorporate energy-saving functionalities to their processors. One of the most common such functionalities is dynamic speed scaling, where the operating speed of the processor can be dynamically adjusted over time. A higher speed implies a higher performance, but this performance comes at the cost of a higher energy consumption. In practice, the power function $P(s) = s^3$ provides a good approximation on the true power consumption when the processor is run at speed $s$.

One of the simplest algorithmic problems arising in this context is to schedule a given number of jobs and control the processing speed, in order to minimize the overall energy consumption while respecting the release times and deadlines of the jobs. The theoretical foundation of such problems was laid by Yao, Demers and Shenker [3], who designed a polynomial time algorithm for the deadline-based setting.

In [2] we extend their model to also consider both the maximum allowed speed of the processor and the energy costs may vary continuously over time. This is a very natural problem arising with the expansion of renewable energies and smart-meters. Theoretical algorithm design for speed scaling problems often tends to discretize problems, as our tools in the discrete realm are often better developed or understood. Using the above speed scaling variant with variable, continuous maximal processor speeds and energy prices as an example, we demonstrate that a more direct approach via tools from variational calculus can not only lead to a very concise and elegant formulation and analysis, but also avoids the “explosion of variables/constraints” that often comes as a result of discretizing. Using well-known tools from calculus of variations, we derive combinatorial optimality characteristics for our continuous problem and provide a quite concise and simple correctness proof.

We also look into the problem where jobs can be partitioned into arbitrarily small parts and the quality of a schedule is given by the tradeoff between how much energy it consumes and how long each such part “spends in the system” [1]. In other words, we study problem of computing an optimal energy and fractional weighted flow trade-off schedule for a speed-scalable processor with discrete speeds, and give a polynomial time algorithm for it. Our algorithm uses a geometric approach that is based on structural properties obtained from a primal-dual formulation of the problem.

References


A Tight Lower Bound for Online Convex Optimization with Switching Costs

Investigators: Antonios Antoniadis and Kevin Schewior

A natural online problem arising when rightsizing data centers is online convex optimization with switching costs (first introduced by Lin et al. [5]). Consider a data-center consisting of several machines, such that each incurs some cost when switched on or off. Whenever a new request arrives, the data center operator needs to decide how many machines to activate/deactivate for processing the request, and then pay the corresponding processing cost incurred by these machines for the current request.

More formally, consider a server initially located at a point $p_0$ on the real line which is presented with an online sequence of non-negative convex functions $f_1, f_2, \ldots, f_n : \mathbb{R} \to \mathbb{R}^+$. In response to each function $f_i$, the server moves to a new position $p_i$ on the real line, resulting in cost $|p_i - p_{i-1}| + f_i(p_i)$. The total cost is the sum of costs of all steps. Since this is an online problem, one is interested in designing competitive algorithms.

The best known deterministic online algorithm attains a competitive-ratio of 2 [1, 2, 4], and it has been conjectured that $(2 - \epsilon)$-competitive algorithms exist for some $\epsilon > 0$ [4].

In [3], we solve the problem in the classical sense: We give a lower bound of 2 on the competitive ratio of any possibly randomized online algorithm, thus matching the best known competitive ratio. Our lower-bound construction is surprisingly simple, in that it only considers two specific and very simple, linear input functions.

References


Lower Bounds for Online Matching on the Line

Investigators: Antonios Antoniadis in cooperation with Carsten Fischer and Andreas Tönnis (University of Bonn)

Online matching on the line captures many scenarios where items arrive one by one over time and have to be immediately assigned/matched to a different set of items. More formally, the task is to match a set of requests $R$ online to a given set of servers $S$. The similarity/distance metric between any two points in $R \cup S$ is a line metric and the objective for the online algorithm is to minimize the sum of distances between matched server-request pairs. Since this is an online problem, one is interested in designing competitive algorithms.

Online matching on the line is a well-studied problem and – despite recent improvements – there is still a large gap between the best known lower and upper bounds: The best known deterministic algorithm for the problem is $O(\log n)$-competitive, while the best known deterministic lower bound is 9.001. The lower and upper bounds on the competitive ratio for randomized algorithms are 4.5 and $O(\log n)$, respectively.

In [1] we prove that any deterministic online algorithm which in each round: (i) is local, i.e., it bases the matching decision only on information local to the current request, and (ii) is symmetric (in the sense that the decision corresponding to the mirror image of some instance $I$ is the mirror image of the decision corresponding to instance $I$), must be $\Omega(\log n)$-competitive. We then extend the result by showing that it also holds when relaxing the symmetry property so that the algorithm might prefer one side over the other, but only up to some degree. This proves a barrier of $\Omega(\log n)$ on the competitive ratio for a large class of “natural” algorithms. This is of particular interest because this class of local and relaxed-symmetric algorithms includes all deterministic online algorithms found in the literature so far which implies that sufficiently new ideas/techniques would be required in order to break the $\Omega(\log n)$-barrier.

Furthermore, we show that our result can be extended to randomized algorithms that locally induce a symmetric distribution over the chosen servers. The $\Omega(\log n)$-barrier on the competitive ratio holds for this class of algorithms as well.

References


Handling Commitment in Online Admission Control

Investigators: Kevin Schewior in collaboration with Lin Chen (University of Houston), Franziska Eberle, Nicole Megow (University of Bremen), and Clifford Stein (Columbia University)

In many applications of scheduling, such as cloud services, jobs arrive over time and require to be scheduled until a certain deadline. As the computation power is limited, not all jobs can always be finished. So it is a common objective of the owner of the machine to maximize the number (or the weight) of finished jobs, that is, to maximize throughput.
Throughput Scheduling has been widely studied (e.g., [1, 3, 4]). However, job owners may require some sufficiently early feedback on the processing of their jobs, leading to different commitment models, which have not been studied as rigorously. Also, for some existing models, best-possible results have not been obtained yet. We address these shortcomings in our work [2].

In our model, jobs with processing requirements and deadlines arrive online over time at their release dates, and there is a single machine to schedule them. We consider online algorithms and use competitive analysis to evaluate them. To circumvent known impossibility results [1, 3], we make a standard slackness assumption by which the feasible time window for scheduling a job is at least $1 + \varepsilon$ times its processing time, for some $\varepsilon > 0$.

We quantify the impact that different provider commitment requirements have on the performance of online algorithms. Our main contribution is one universal algorithmic framework for online job admission both with and without commitments. Without commitment, our algorithm with a competitive ratio of $O(1/\varepsilon)$ is the best possible (deterministic) for this problem. For commitment models, we give the first non-trivial performance bounds. If the commitment decisions must be made before a job’s slack becomes less than a $\delta$-fraction of its size, we prove a competitive ratio of $O(\varepsilon/(\varepsilon - \delta)^2)$, for $0 < \delta < \varepsilon$. When a provider must commit upon starting a job, our bound is $O(1/\varepsilon^2)$. Finally, we observe that for scheduling with commitment the restriction to the “unweighted” throughput model is essential; if jobs have individual weights, we rule out competitive deterministic algorithms.

References


Distributed Algorithms for Network Flows

Investigators: Ruben Becker, Andreas Karrenbauer, and Christoph Lenzen in cooperation with Mohsen Ghaffari (ETH Zurich), Sebastian Krinninger (University of Salzburg), Fabian Kuhn (University of Freiburg), and Boaz Patt-Shamir (Tel Aviv University)

The undirected transshipment and single-source shortest paths (SSSP) problems are fundamental problems in theoretical computer science with a diverse set of applications. These problems are particularly challenging when studied from the point of view of distributed or parallel models of computation. In this line of work we achieve interesting results for
certain distributed and streaming models of computation by applying methods stemming from continuous optimization to these classical combinatorial optimization problems. In [1], we present a tailored gradient descent algorithm that computes solutions that are optimal up to a multiplicative error of $1 + \varepsilon$. Denoting with $n$ the number of nodes in the graph, this method takes $\varepsilon^{-O(1)} \log^{O(1)} n$ iterations, while in each iteration solving an instance of the transshipment problem up to a multiplicative error of $\log^{O(1)} n$. Such a solution of poly-logarithmic error can in turn be obtained by computing a solution on a sparse spanner of logarithmic stretch. This approach leads to interesting results in various distributed and streaming models of computation. We highlight them in the following:

1. **Broadcast Congest model:** $(1 + \varepsilon)$-approximate SSSP using $\tilde{O}((\sqrt{n} + D)\varepsilon^{-O(1)})$ rounds,\(^2\) where $D$ is the (hop) diameter of the network.

2. **Broadcast Congested clique model:** $(1 + \varepsilon)$-approximate shortest transshipment and SSSP using $\tilde{O}(\varepsilon^{-O(1)})$ rounds.

3. **Multipass streaming model:** $(1 + \varepsilon)$-approximate shortest transshipment and SSSP using $\tilde{O}(n)$ space and $\tilde{O}(\varepsilon^{-O(1)})$ passes.

The previously fastest SSSP algorithms for these models leverage sparse hop sets. In the Broadcast Congest model for example, for constant $\varepsilon$, the previously best known bound is $O((\sqrt{n} + D) \cdot n^{o(1)})$ [3], whereas our bound scales like $\tilde{O}(\sqrt{n} + D)$. We achieve our result by bypassing the hop set construction; computing a sparse spanner is sufficient in our case. The above bounds assume non-negative integer edge weights that are polynomially bounded in $n$; for general non-negative weights, running times scale with the logarithm of the maximum ratio between non-zero weights. Our algorithms also apply for the case of asymmetric costs for traversing an edge in opposite directions. In this case, running times scale with the maximum ratio between the costs of both directions over all edges.

Furthermore, we designed a near-optimal distributed algorithm computing a $(1 + o(1))$-approximation of single-commodity maximum flow in undirected weighted networks that runs in $(D + \sqrt{n}) \cdot n^{o(1)}$ communication rounds in the Congest model [2]. This was the first improvement over the trivial bound of $O(n^2)$, and it nearly matches the $\tilde{\Omega}(D + \sqrt{n})$ round complexity lower bound. The development of our algorithm entails two sub-results of independent interest:

(i) A $(D + \sqrt{n}) \cdot n^{o(1)}$-round distributed construction of a spanning tree of average stretch $n^{o(1)}$.

(ii) A $(D + \sqrt{n}) \cdot n^{o(1)}$-round distributed construction of an $n^{o(1)}$-congestion approximator consisting of the cuts induced by $O(\log n)$ virtual trees. The distributed representation of the cut approximator allows for evaluation in $(D + \sqrt{n}) \cdot n^{o(1)}$ rounds.

All our algorithms make use of randomization and succeed with high probability.

**References**

A number of recent works have studied algorithms for entrywise $\ell_p$-low rank approximation. Given an $n \times n$ matrix $A$ and an integer $k$, these algorithms output a rank-$k$ matrix decomposition consisting of an $n \times k$ matrix $U$ and a $k \times n$ matrix $V$ minimizing the error measure $\|A - UV\|_p = \sum_{i,j} |A_{i,j} - (UV)_{i,j}|^p$ if $p > 0$, or $\|A - B\|_0 = \sum_{i,j} [A_{i,j} \neq (UV)_{i,j}]$ if $p = 0$.

When $p = 2$, the error measure coincides with the Frobenius norm, which can be solved in polynomial time using the singular value decomposition. For $p \in (0, 2]$, the error measure is often considered more robust, since it pays less for noisy entries as one does not square the distances but instead raises them to a smaller power. The case $p = 0$ corresponds to minimizing the number of disagreements, and is known as robust principal component analysis [3].

There are various problem variants depending on the field or semi-ring that the entries of the matrices $A, U, V$ are chosen from, e.g., the reals, a finite field, or the Boolean semi-ring, and depending on the corresponding notion of matrix multiplication $(UV)_{i,j} = \langle U_{i,:}, V_{:,j} \rangle$.

The problem is NP-hard for $p \in \{0, 1\}$, even for $k = 1$. On the algorithmic side, a constant factor approximation algorithm was known, albeit running in time $n^{\text{poly}(k)}$ [4]. It was left open whether there is a PTAS for $\ell_p$-low rank approximation for any $p \in [0, 2)$. In our work [1], we give an affirmative answer to this question for constant $k$ and $\varepsilon$. For $p \in (0, 2)$ and over the reals, we design the first approximation scheme, computing a $(1 + \varepsilon)$-approximation in time $n^{\text{poly}(k/\varepsilon)}$. For $p = 0$ and over the Boolean semi-ring, we design the first approximation scheme, running in time $(1/\varepsilon)^{\text{poly}(k)} \cdot n^{2+o(1)}$; this is almost-linear in $n$ for any constant $k, \varepsilon$. Moreover, for $p = 0$ and over any finite field, we obtain an approximation scheme running in time $n^{\text{poly}(k/\varepsilon)}$.

In the paper [2], we focus on approximation algorithms running in polynomial time $\text{poly}(n, k)$. We present novel $\ell_0$-low rank approximation algorithms which significantly improve the runtime and the approximation factor of prior works. Over the reals, we design a bicriteria algorithm that computes a $O(k^2 \log(n/k))$-approximation, but outputs a decomposition of rank $O(k \log(n/k))$. To the best of our knowledge, this is the first algorithm with provable guarantees for $k > 1$. For the case when $k = 1$, we give a $(2 + \varepsilon)$-approximation algorithm which for dense matrices runs in sublinear time.
References


35.6.2 Combinatorial Algorithms and Algorithm Engineering

In addition to approximation algorithms considered in the previous section, researchers from the Algorithms and Complexity Group also develop algorithms to solve combinatorial problems exactly. This includes efficient combinatorial algorithms running in polynomial time and the solution of integer linear programs, e.g., by branch-and-bound, which is guaranteed to find an optimum solution in finite time and provides rigorous per-instance optimality gaps when stopped prematurely. Furthermore, we combine theoretical and experimental research in algorithm engineering processes to tackle optimization problems that are directly motivated by applications.

Testing Substitutability

Investigators: Cosmina Croitoru and Kurt Mehlhorn

A choice function on a finite set $U$ of alternatives is any function $f$ from subsets of $U$ to subsets of $U$ that maps any set $A$ to a subset of itself, i.e., $f(A) \subseteq A$ for all $A \subseteq U$. A choice function $f$ is substitutable if

$$A \subseteq B \implies f(B) \cap A \subseteq f(A) \text{ for all } A, B \subseteq U,$$

i.e., the additional alternatives provided by $B$ do not promote any $x \in A - f(A)$ to the set of selected elements.

We are interested in choice functions induced by preference lists $Y$ on subsets of $U$. A preference list $Y$ is simply an ordered list of subsets of $U$ and the associated choice function $f_Y$ maps any subset $A$ of $U$ to the first element on the list that is contained in $A$.

As an example, consider $U = \{a, b, c, d\}$ and $Y = (\{a, b\}, \{a, c, d\}, \{a, c\}, \{a\}, \{c\}, \emptyset)$. Then we obtain $f_Y(\{a, b, c\}) = \{a, b\}$. Note that the function $f_Y$ is not substitutable since $d \in (f_Y(\{a, c, d\}) \cap \{d\}) - f_Y(\{d\})$. We refer to [3] for a discussion of the role of substitutable choice functions in economics.
The papers [3] and [1] propose algorithms for testing whether the choice function induced by a (strict) preference list of length \( N \) over a universe \( U \) is substitutable. The running time of these algorithms is \( O(|U|^3 \cdot N^3) \) and \( O(|U|^2 \cdot N^3) \), respectively. In our work [2], we present an algorithm with running time \( O(|U|^2 \cdot N^2) \).

The \( O(N) \) factor speed-up over the existing algorithms is significant since \( N \) is exponential in the size of the largest member of \( Y \), hence exponential in the size \( |U| \) of the universe. As future work we will study whether our algorithm applies also for weak preferences.

References


**From DQBF to QBF by Dependency Elimination**

_Investigators: Ruben Becker and Andreas Karrenbauer in cooperation with Ralf Wimmer, Christoph Scholl, and Bernd Becker (University of Freiburg)_

Dependency quantified Boolean formulas (DQBFs) are an important concept in logics, formal methods, and verification. They are a generalization of ordinary quantified Boolean formulas (QBFs). While the latter have the restriction that every existential variable depends on all universal variables in whose scope it is, DQBFs allow arbitrary dependencies, which are explicitly specified in the formula. This makes DQBFs more expensive to solve than QBFs – for DQBF the decision problem is \( \text{NEXPTIME}\)-complete, for QBF ‘only’ \( \text{PSPACE}\)-complete. However, there are practically relevant applications that require the higher expressiveness of DQBFs for a natural and tremendously more compact modeling.

In our work [2], we propose the elimination of dependencies to convert a given dependency quantified Boolean formula (DQBF) to an equisatisfiable QBF. The conversion to QBF in general leads to an exponential blow-up of the formula: If one needs to eliminate \( n \) dependencies of an existential variable \( y \), then the resulting formula contains up to \( 2^n \) different copies of this variable. Therefore, it is important to carefully select the dependencies that are eliminated in order to keep this blow-up as small as possible. We show how to select a set of dependencies to eliminate such that we arrive at a smallest equisatisfiable QBF in terms of existential variables that is achievable using dependency elimination. This problem of choosing an optimal set of dependencies to eliminate can be stated in a natural way as the problem of determining a cost-minimal set of edges to flip in order to break all cycles in a bipartite tournament graph, also called the dependency graph. A (non-linear) cost function takes into account the number of existential variables after eliminating a set of dependencies. We first show that this optimization problem is \( \text{NP}\)-complete and then give an exact and efficient solution for this combinatorial optimization problem that is based on integer linear programming with dynamically added constraints in a cutting plane like fashion. This approach has been implemented and included in the state-of-the-art DQBF.
solver HQS [1]. Experiments show that dependency elimination is clearly superior to previous methods using variable elimination.

References


Computational Support for Functionality Selection in Interaction Design

Investigators: Andreas Karrenbauer in cooperation with Antti Oulasvirta, Anna Maria Feit, and Perttu Lähteenlahti (Aalto University)

Designing interactive technology entails several objectives, one of which is identifying and selecting appropriate functionality. Given candidate functionalities such as “print,” “bookmark,” and “share,” a designer has to choose which functionalities to include and which to leave out. Such choices critically affect the acceptability, productivity, usability, and experience of the design. However, designers may overlook reasonable designs because there is an exponential number of functionality sets and multiple factors to consider. In [1], we are the first to formally define this problem and propose an algorithmic method to support designers to explore alternative functionality sets in early stage design. Based on interviews of professional designers, we mathematically define the task of identifying functionality sets that strike the best balance among four objectives: usefulness, satisfaction, ease of use, and profitability. We develop an integer linear programming solution that can efficiently solve very large instances (set size over 1,300) on a regular computer. We present realistic special cases that can be solved by combinatorial algorithms in polynomial time. Further, we build on techniques of robust optimization to search for diverse and surprising functionality designs. Empirical results from a controlled study and field deployment are encouraging. Most designers rated computationally created sets to be of comparable or superior quality than their own. Designers reported gaining better understanding of available functionalities and the design space. In [2], we review further applications of combinatorial optimization in the context of Human-Computer-Interaction.

References


35.6.3 Algorithms in Nature and Algorithms Inspired by Nature

We investigated algorithms in nature and algorithms inspired by nature. We continued our work on the slime mold Physarum polycephalum. We showed that the dynamics can not only solve shortest path problems and transportation problems but the general basis pursuit problem, i.e., given a positive $m$-vector $c$ and an $n \times m$ matrix $A$ and an $n$-vector $b$, the goal is to find the vector $x$ minimizing $c^T|x|$ subject to the constraint $Ax = b$. We also showed that the dynamics is robust with respect to perturbations thus making our results more relevant to biology as biological systems cannot be expected to compute exactly. We also did wet-lab work with Physarum.

We consider the ANTS problem in which a group of $n$ agents collaboratively search for a target in a two-dimensional plane. Because this problem is inspired by the behavior of biological species, we argue that in addition to studying the time complexity of solutions it is also important to study the selection complexity $\chi$, a measure of how likely a given algorithmic strategy is to arise in nature due to selective pressures. Intuitively, the larger the $\chi$ value, the more complicated the algorithm, and therefore the less likely it is to arise in nature.

The computation of electrical flows is a crucial primitive for many recently proposed optimization algorithms on weighted networks. The ability to perform this task in a fully decentralized way is implicit in a number of biological systems. Thus, a natural question is whether this task can provably be accomplished in an efficient way by a network of agents executing a simple protocol. We provide a positive answer.

Improved Convergence Results of the Physarum Dynamics

*Investigators: Ruben Becker, Andreas Karrenbauer, Pavel Kolev and Kurt Mehlhorn in cooperation with Vincenzo Bonifaci (IASI-CNR)*

[6] showed, in a series of wet-lab experiments, that the slime mold Physarum polycephalum possesses the ability to solve shortest path problems, [8] proposed a mathematical model, a system of differential equations, for the slime’s adaption process, and [2] showed that the process convergences to the shortest path for all graphs. In the reporting period, we strengthened our result in several directions.

- We showed the the Physarum dynamics converges for a much larger class of problems [1]. In particular, it can find optimum solutions to basis pursuit problems.

Let $A$ be an $n \times m$ matrix, $c$ a positive $m$-vector and $b$ an $n$-vector. The *basis pursuit problem* asks for the $n$-vector $x$ of minimum weighted one-norm $c^T x$ subject to the constraints $Af = b$ and $|f| \leq x$. The *Physarum dynamics* evolves an $n$-vector $x$ according to the dynamics 

$$\dot{x}_e = \left( |q_e| - x_e \right),$$

where $q$ minimizes the energy $\sum_e (c_e/x_e)q_e^2$ subject to the constraints $Aq = b$ and $\text{supp}(q) \subseteq \text{supp}(x)$ and the components of $x$ and $e$ are indexed by $e$.

- For the Euler discretization of the dynamics, convergence was previously shown by Straszak and Vishnoi [7]. We give an improved convergence result for the discretization [1]. We can start the process from an strongly dominating starting point, i.e.,

179
any point $x(0)$ such that $Af = b$ for some $f$ with $|f| \leq x(0)$. They required a feasible starting point. In order to obtain an $\epsilon$-approximation of the optimal objective one can work with a step size independent of $\epsilon$ (they had linear dependence) and a number of steps depending logarithmically on $1/\epsilon$ and quadratically on $\text{opt}/\Phi$; they had quartic dependence on $\text{opt}/(\epsilon \Phi)$, where $\Phi$ is the difference between the minimal non-optimal cost and the optimal cost ($\text{opt}$).

– Nature cannot be expected to follow the differential equation exactly. In [5], we show convergence to the optimum solution for the basis pursuit problem for a much more general dynamics

$$\dot{x}_e = a_e(x, t) (|q_e| - x_e)$$

where $a_e(x, t)$ is any Lipschitz-continuous function that is bounded away from zero and from infinity, i.e., there is a positive constant $C$ such that $1/C \leq a_e(x, t) \leq C$ for all $x$ and $t$.

**Characterizing networks formed by Physarum polycephalum**

_Investigators: Michael Dirnberger and Kurt Mehlhorn in cooperation with Tim Mehlhorn (Saarbrücken)_

The work [3, 4] is somewhat unusual for the group as it concerns experimental work in the wet-lab carried out by Kurt’s son Tim Mehlhorn followed by evaluation carried out by Michael Dirnberger and myself. We have created a data collection of Physarum networks and extracted abstract graphs from these networks together with characteristic properties, e.g., edge length and edge width. The collection allows us to test conjectures, e.g., that the capacity of cuts orthogonal to the direction of growth is approximately constant. Figure 35.3 shows a graph extracted from a Physarum network.

**References**

Searching the Plane with Limited Memory

Investigators: Christoph Lenzen in collaboration with Nancy Lynch (MIT), Calvin Newport (Georgetown University), and Tsvetomira Radeva (MIT)

We consider the ANTS problem [1] in which a group of $n$ agents collaboratively search for a target in a two-dimensional plane. Because this problem is inspired by the behavior of biological species, in [2] we argue that in addition to studying the time complexity of solutions it is also important to study the selection complexity $\chi$, a measure of how likely a given algorithmic strategy is to arise in nature due to selective pressures. Intuitively, the larger the $\chi$ value, the more complicated the algorithm, and therefore the less likely it is to arise in nature.

In more detail, for algorithm $A$ we define $\chi(A) = b + \log \ell$, where $b$ is the number of memory bits used by each agent and $\ell$ bounds the fineness of available probabilities (agents use probabilities of at least $1/2^\ell$). Note that using a fair coin one needs exactly $\ell$ coin flips to simulate a biased coin that shows head with probability $1/2^\ell$, which requires $\ell$ different states (and thus at least $\log \ell$ memory bits) in a state machine simulating such a biased coin based on fair coin flips. This motivates the combined measure $\chi$ relating memory and fineness of probability.

In [2], we study the trade-off between the standard performance metric of speed-up, which measures how the expected time to find the target improves with $n$, and our new selection metric. Our goal is to determine the thresholds of algorithmic complexity needed to enable efficient search. Concretely, consider $n$ agents searching for a treasure located within some distance $D$ from the origin (where $n$ is sub-exponential in $D$). For this problem, we identify the threshold $\log \log D$ to be crucial for our selection complexity metric. We first prove a new upper bound that achieves a near-optimal speed-up for $\chi(A) \approx \log \log D + O(1)$. In particular, for $\ell \in O(1)$, the speed-up is asymptotically optimal. By comparison, the existing results for this problem [1] that achieve similar speed-up require $\chi(A) \in \Omega(\log D)$. We then show that this threshold is tight by describing a lower bound showing that if $\chi(A) < \log \log D - \omega(1)$,
then with high probability the target is not found in $D^2 - o(1)$ moves per agent. Hence, there is a sizable gap with respect to the straightforward $\Omega(D^2/n + D)$ lower bound in this setting.

**References**


**Distributed Natural Algorithms for Electrical Flow Estimation**

Investigators: Emanuele Natale in cooperation with Luca Becchetti (Sapienza University of Rome) and Vincenzo Bonifaci (National Research Council)

The computation of electrical flows is a crucial primitive for many recently proposed optimization algorithms on weighted networks. While typically implemented as a centralized subroutine, the ability to perform this task in a fully decentralized way is implicit in a number of biological systems [2]. Thus, a natural question is whether this task can provably be accomplished in an efficient way by a network of agents executing a simple protocol. In [1], we provide a positive answer, proposing two distributed approaches to electrical flow computation on a weighted network: a deterministic process mimicking Jacobi’s iterative method for solving linear systems, and a randomized token diffusion process, based on revisiting a classical random walk process on a graph with an absorbing node. We show that both processes converge to a solution of Kirchhoff’s node potential equations, derive bounds on their convergence rates in terms of the weights of the network, and analyze their time and message complexity.

**References**


**35.7 Geometry and Algebra**

*Coordinators: Christian Ikenmeyer and Michael Sagraloff*

Geometry and algebra are two deeply interconnected fundamental pillars of modern mathematics that have important connections to theoretical computer science, e.g., in computer algebra, coding theory, cryptography, and algebraic complexity theory. The interplay between geometry and algebra has a wide range of real-world applications, for example in robotics,
computer-aided design, geometric modeling, signal processing, control theory, computer graphics, computer vision, and geographical information, to name a few. Many geometric questions have a natural algebraic analog that permits a rigorous and also a computational treatment of geometric properties, while many algebraic questions can benefit greatly from geometric intuition and methods. The whole field of algebraic geometry is focused on this connection between geometry and algebra and indeed the main geometric objects of study – algebraic varieties – are defined as zero sets of systems of polynomials, which is a purely algebraic notion.

Our research covers problems from both algebra and geometry, where our results cover the whole spectrum from real-world applications to the analysis of algorithms and computational complexity lower bounds.

35.7.1 Algorithms and Data Structures in the Plane

Many real-world problems can naturally be phrased as problems in planar Euclidean geometry. The main objects in this classical area of mathematics include positions, angles, lengths, distances, areas, perimeters, curves, and polygons. Many problems in planar Euclidean geometry are so fundamental that their algorithms are used as building blocks in numerous applications. Hence solving them efficiently is of paramount importance.

Fast Fencing

Investigators: Karl Bringmann in cooperation with Mikkel Abrahamsen, Anna Adamaszek, Alan Roytman, Mikkel Thorup (University of Copenhagen), Vincent Cohen-Addad (Sorbonne Université), Mehran Mehr (TU Eindhoven), and Eva Rotenberg (Technical University of Denmark)

We study clustering with the objective of minimizing the clusters’ perimeters. This is motivated by “fence enclosure” scenarios, where the cost is proportional to the total fence length. Specifically, given $n$ points in the plane, we aim at finding a set of closed curves such that (1) each point is enclosed by a curve and (2) the total length of the curves is minimized. We consider two main variants. In the first variant, we pay a unit cost per curve in addition to the total length of the curves. An equivalent formulation of this version is that we have to enclose $n$ unit disks, paying only the total length of the enclosing curves, so we refer to this variant as unit disk fencing. In the other variant, we are allowed to use at most $k$ closed curves and pay no cost per curve, and we refer to this as $k$-cluster fencing.

Capoyleas et al. [3] showed how to solve $k$-cluster fencing in $n^{O(k)}$ time. Arkin et al. [2] used this to solve unit disk fencing in exponential time. At the time, they conjectured that the problem with $k$ curves is NP-hard when $k$ is part of the input.

We refute this conjecture (unless P equals NP) by presenting an algorithm that is polynomial in both $n$ and $k$ for $k$-cluster fencing. Moreover, for unit disk fencing we present a near-linear time algorithm. Our algorithm for unit disk fencing can be generalized to the case where the input consists of objects that are allowed to be disks with different diameters or polygonal objects that we want to be fenced. For this variant, our running time increases by a factor that is logarithmic in the ratio between the maximum and the minimum object diameter.
Minimizing Distance-to-Sight in Polygonal Domains

Investigator: Eunjin Oh

In the quickest pair-visibility problem we are given a starting point $s$ and a target point $t$ amidst polygonal obstacles in the plane. Our goal is to compute the minimum distance that the two points need to travel in order to see each other. Here, there are two variants of the problem, one for minimizing the maximum of the two travel distances and one for minimizing the sum of the two travel distances. This problem is motivated from the line-of-sight communication model where two robots are required to be visible to each other in order to establish communication.

Wynters and Mitchell studied this problem for both variants [3]. For the min-max variant, they gave an $O(n^3 \log n)$-time algorithm using $O(n^2)$ space. For the min-sum variant, they gave an $O(nm)$-time algorithm using $O(m)$ space, where $m$ is the number of edges in the visibility graph of the polygonal obstacles. Note that $m = \Theta(n^2)$ in the worst case. Very recently, Ahn et al. [1] considered a simpler version of the quickest pair-visibility problem in which two points are given in a simple polygon with no holes, and presented linear-time algorithms for both the min-max and the min-sum variants of the problem.

In [2], we presented an algorithm for the min-max variant of the problem which takes $O(n \log^2 n + h^3 \log^4 h)$ time using $O(n \log n)$ space, where $h$ is the number of obstacles and $n$ is their total number of vertices. This substantially improves the algorithm by Wynters and Mitchell. Also, we showed that this problem is likely to be optimal up to polylogarithmic factors by a reduction from the 3SUM problem.

References


Dynamic Point Location

Investigators: Eunjin Oh in cooperation with Hee-Kap Ahn (POSTECH)

Given a planar subdivision, a point location query asks for a given query point specified by its coordinates to find the face of the subdivision containing the query point. In many situations such point location queries are made frequently and subdivisions change dynamically. Thus, it is desirable to keep track of the changes by maintaining a data structure so that point location queries can be answered efficiently. This problem is called the dynamic point location problem, and it is one of the fundamental problems in computational geometry.

Since the early 1990’s, this problem has been extensively studied [1, 2]. The query and update times have been progressively improved. However, all of these data structures only work for connected subdivisions (i.e., the underlying graphs of subdivisions are connected.) Thus it is natural to ask if one can achieve a point location data structure for general (not necessarily connected) subdivisions supporting sublinear query and update times. Indeed, this problem was explicitly posed by Snoeyink in the Handbook of Computational Geometry [5].

In [3, 4], we present point location data structures for general subdivisions. More specifically, in [4] we resolve an open problem posed by Snoeyink by presenting a data structure for supporting polylogarithmic query time and sublinear update time. However, the update time is larger than the update times of the previously known data structures for connected subdivisions. Thus, it is natural to ask if there is a special case where polylogarithmic update and query times can be achieved simultaneously. In [3], we focus on the special case that only the insertion of edges is allowed, and show that our data structure supports polylogarithmic query and update times in this special case.

References


Geodesic Voronoi Diagrams in Simple Polygons

*Investigator: Eunjin Oh*

Given any two points contained in a simple polygon, the *geodesic distance* between the two points is defined as the length of the shortest path contained in the simple polygon connecting the two points. Since the early 1980s, many classical geometric problems have been studied in the geodesic setting in a simple polygon including the problem of computing the geodesic Voronoi diagram. The geodesic (nearest-point) Voronoi diagram of a set of $m$ point sites in a simple $n$-gon is defined as the subdivision of the simple polygon into cells, exactly one cell per site, such that every point in a cell has the same nearest site under the geodesic distance.

The first algorithm for this problem was presented in 1989 [1] and takes $O(n \log^2 n + m \log m \log n)$ time. Also, this paper shows that $\Omega(n + m \log m)$ is a lower bound of the computation time. Since then, it has been asked if the geodesic Voronoi diagram can be computed in an optimal time complexity. This question was explicitly posed by Mitchell [2] and by Aronov [1]. In spite of the progress made by several researchers, the optimal running time for computing the geodesic Voronoi diagram was not known prior to our work.

In [3], we presented an $O(n + m \log m)$-time algorithm for computing the geodesic Voronoi diagram of $m$ points in a simple $n$-gon, which matches the known lower bound. This resolves a 30-years-old open problem.

**References**


A fast implementation of near neighbors queries for Fréchet distance

*Investigators: Karl Bringmann in cooperation with Julian Baldus (Saarland University)*

The Fréchet distance is a popular similarity measure on curves. Roughly speaking, given curves $\pi$ and $\sigma$ it measures the minimal length of a leash connecting a dog to its owner as they walk without backtracking along $\pi$ and $\sigma$, respectively. As a natural measure of curve similarity, it has broad applications in geographic information systems. In theory, the problem of computing the Fréchet distance of two given curves with $n$ vertices is well understood, since there is a classic algorithm running in time $O(n^2 \log n)$ [1] and a conditional lower bound ruling out any significant improvements [3].

However, from a practical perspective the problem is much less understood. In particular, until recently there was no publicly available implementation of a Fréchet distance algorithm, except for the naive solution. This was the motivation for the ACM SIGSPATIAL GIS Cup
to ask for implementations of near neighbor data structures for the Fréchet distance. In this setting, after a preprocessing phase on a database of curves, a query consists of a curve $\sigma$ and a distance threshold $\delta$ and the task is to report every curve $\pi$ in the database that has Fréchet distance at most $\delta$ to the query curve $\sigma$. Our approach [2] is to use a quadtree data structure to enumerate all curves in the database that have similar start and endpoints as the query curve. On these curves we run positive and negative filters to narrow the set of potential results. Only for the curves where these heuristics fail, we compute the Fréchet distance exactly, by running a novel recursive variant of the classic quadratic-time algorithm.

Among 28 submissions to the ACM SIGSPATIAL GIS Cup 2017, our submission [2] made the first place.

References


Fréchet Distance Under Translation: Conditional Hardness and an Algorithm via Offline Dynamic Grid Reachability

Investigators: Karl Bringmann, Marvin Künnemann, and André Nusser

The discrete Fréchet distance is a popular measure for comparing sequences of points or polygonal curves. An important variant is the *discrete Fréchet distance under translation*, which is invariant under translations and thus enables detection of similar movement patterns in different spatial domains. For sequences of $n$ points in the plane, the previously fastest known algorithm for the discrete Fréchet distance under translation runs in time $\tilde{O}(n^5)$ [1]. This was achieved by constructing a certain arrangement of disks of size $O(n^4)$, and then traversing the faces of this arrangement while updating reachability in a directed grid graph of size $N = O(n^2)$, which can be done in time $\tilde{O}(\sqrt{N}) = \tilde{O}(n)$ per update [3].

Our work [2] improves on previous results as follows:

- Although it is a well-known open problem to solve dynamic reachability in directed grid graphs faster than in time $\tilde{O}(\sqrt{N})$, we improve this part of the algorithm: We observe that an *offline* variant of dynamic s-t-reachability in directed grid graphs suffices, and we solve this offline variant in amortized time $\tilde{O}(N^{1/3})$ per update. This results in an improved running time of $\tilde{O}(n^{14/3}) = \tilde{O}(n^{4.66...})$ for the discrete Fréchet distance under translation.

We provide evidence that constructing the arrangement of size $O(n^4)$ is necessary in the worst case, by proving a conditional lower bound of $n^{4-o(1)}$ on the running time for the discrete Fréchet distance under translation, assuming the Strong Exponential Time Hypothesis. This is surprising, since – to the best of our knowledge – exhaustively enumerating such a large arrangement is not known to be necessary for any other geometric problem.

References


35.7.2 High-dimensional Euclidean Geometry

Data is often naturally modeled as a set of points in a high-dimensional Euclidean vector space. Many algorithmic challenges arise from this perspective, which involve for example the development of new data structures for specific database queries, clustering algorithms, and the extraction of topological information about the dataset.

Range-Clustering Queries

Investigators: Eunjin Oh in cooperation with Hee-Kap Ahn (POSTECH)

Range searching asks to preprocess a set of objects and to build a data structure so that all objects intersecting a given query range can be reported quickly. There are classical variants of this problem such as computing the number of objects intersecting a query range, checking whether an object intersects a query range, and finding the closest pair of objects intersecting a query range. However, there might be a large number of objects intersecting a query range in many real-world applications, and thus it takes a long time to report all of them. In this case, one might want to obtain a property of such objects (such as a clustering cost) instead of obtaining all such objects.

In [3], we present a data structure for answering range-clustering queries efficiently. More specifically, we study the $(1 + \varepsilon)$-approximate range-clustering problem for three variants of clustering with a set of points in $d$-dimensional Euclidean space with $d \geq 2$ and axis-parallel rectangular range queries: $k$-median, $k$-means, and $k$-center range-clustering query problems. There are several previous results on this problem. Nekrich and Smid [2] presented a data structure for answering range $k$-center, $k$-median, and $k$-means clustering queries. However, their algorithm uses $k$ and $\varepsilon$ in constructing the data structure for the clusterings, and therefore, $k$ and $\varepsilon$ are fixed over range-clustering queries. Very recently, Abrahamsen et
al. [1] considered the $k$-center range-clustering query problem for points in $d$-dimensional space. Their data structure allows $k$ and $\varepsilon$ to be given as a query, but their approach works only for the $k$-center clustering.

In [3], we improved their results in the sense that our data structure allows $k$ and $\varepsilon$ to be given as a query and our query algorithm is faster than the query algorithm by Abrahamsen et al. Moreover, our approach also works for the $k$-median and $k$-means clusterings while the approach of Abrahamsen et al. works only for the $k$-center clustering.

References


Maximum Volume Subset Selection for Anchored Boxes

Investigators: Karl Bringmann in cooperation with Sergio Cabello (University of Ljubljana) and Michael Emmerich (LIACS, Leiden University)

An anchored box is an axis-parallel box in $d$-dimensional Euclidean space that has one corner at the origin. We study the volume selection problem for anchored boxes: Given $n$ anchored boxes and an integer $k$, select $k$ boxes that maximize the volume of the union of the selected boxes. This geometric problem is motivated by applications in skyline queries for databases and in multicriteria optimization, where the problem is known as the hypervolume subset selection problem. In the plane, i.e., in dimension $d = 2$, the problem can be solved in polynomial time [1]. However, in any dimension $d \geq 3$ all known algorithms essentially enumerate all $\binom{n}{k}$ subsets.

In our work [2], we show that the problem is NP-hard already in 3 dimensions, thus explaining the big difference in the state of the art. Previously, NP-hardness was only known when $d$ is part of the input and thus can be as large as $n$.

Moreover, in 3 dimensions we avoid the enumeration of all subsets, and obtain an algorithm running in time $n^{O(\sqrt{k})}$. To this end, we project the 3-dimensional problem to a 2-dimensional variant and therefore use planar separator techniques.

Finally, for any constant dimension $d$, we give an efficient polynomial-time approximation scheme. In particular, for any constant $d, \varepsilon$ our algorithm computes a $(1 + \varepsilon)$-approximation in time $O(n(k + \log n))$. 

189
Complications of Delaunay Simplices

Investigators: Aruni Choudhary in cooperation with Arijit Ghosh (Indian Statistical Institute Kolkata), Siargey Kachanovich, and Mathijs Wintraecken (INRIA Sophia-Antipolis)

Simplicial meshes are standard methods to approximate geometric objects. These meshes are used in algorithms for several tasks, including numerically solving partial differential equations, finite element approximation of functions and computational dynamical systems. The quality of approximation of these algorithms depends on the goodness of the mesh. We discuss three measures to capture goodness: the first is the thickness of a simplex, which is the ratio of the minimal height to the maximal edge length of the simplex. The second measure is the aspect ratio of a simplex, which is the ratio of the minimal height to the diameter of its circumsphere. In the context of Delaunay triangulations, recently Boissonnat, Dyer and Ghosh [1] introduced a new quality measure called protection: intuitively, this measures how far a Delaunay triangulation is from being degenerate. The protection is the supremum of the amount by which each circumball of a $d$-simplex in the Delaunay triangulation can be inflated, so that it does not contain any other point in its interior.

We concern ourselves with a family of lattices, which we obtain by a distortion of the integer grid in $\mathbb{R}^d$ along the principal diagonal direction $(1, \ldots, 1)$. This family was first studied in [4] by Edelsbrunner and Kerber, with an ulterior motive to do topological analysis of high-dimensional image data. We study the quality measures of this lattice [2]. In particular, we observe that this family contains both $A^*_d$ and its dual, the $A_d$ lattice; the former of which has a high value of protection.

Coxeter triangulations are triangulations of Euclidean space based on a single simplex. This means that given an individual simplex we can recover the entire triangulation of Euclidean space by inductively reflecting in the faces of the simplex. We establish that the quality of the simplices in all Coxeter triangulations is a fraction of $O(1/\sqrt{d})$ of the quality of the regular simplex [3]. We further investigate the protection for these triangulations. In particular, one family of Coxeter triangulations achieves a protection value of $O(1/d^2)$. We conjecture that both bounds are optimal for triangulations in Euclidean space.

References

Approximate Simplicial Filtrations

Investigators: Aruni Choudhary in cooperation with Michael Kerber (TU Graz) and Sharath Raghvendra (Virginia Tech)

Topological data analysis attempts to extract relevant topological information of a data set by interpreting data as a shape and understanding the connectivity of this shape, that is, studying the presence of connected components, tunnels, voids, and other high-dimensional topological features. A common way to represent data is with a set $P$ of $n$ points in $d$-dimensional Euclidean space $\mathbb{R}^d$. Čech complexes and Rips complexes are popular building blocks used in this process. At any given scale $\alpha \geq 0$, these are simplicial complexes that capture the connectivity of the union of balls of radius $\alpha$ centered at the input points. A simplex on $P$ is a subset of $P$. The Čech complex is a collection of simplices such that for each simplex, the balls centered at the set of points intersect. The Rips complex includes all simplices of the Čech complex, and adds in simplices corresponding to cliques in the intersection graph. By varying the scale parameter $\alpha$, one gets a nested sequence of these complexes called filtrations, which can then be analyzed to reveal the underlying topological features. A very desirable feature of this process is that the topological information obtained is stable to noise in the input data.

Unfortunately, these filtrations pose a great computational challenge for high dimensional point sets. The $k$-skeleton of these complexes containing only simplices up to dimension $k$ has size $n^{O(k)}$, which is already quite large even for relatively small values of $k$. We study the possibility of approximating the topological information while avoiding this expensive computation. There have been several works in this area, all arriving at $(1+\epsilon)$-approximations that have worst-case size $n \cdot (1/\epsilon)^{O(d^2)}$ for $n$ points in $d$-dimensional space, see, e.g., [7, 6]. These approaches differ in the filtrations, the approximation guarantees, and the applied methods, and some of the techniques are more suited for lower dimensional skeletons and metric spaces with low intrinsic dimension.

We present several techniques to reduce the complexity further. We present an $O(d)$-approximation scheme that has worst-case size $n \cdot 2^{O(d \log d)}$ [4]. We improve the approximation factor to $2^{d^{1/4}}$ with the same size bound [2]. In conjunction with dimension reduction techniques [8], both of these schemes give the first approximation results with truly polynomial size in $n$, independent of the dimension $d$. We additionally give a lower bound result on the size of approximations: for $n$ points in $\Theta(\log n)$ dimensions, we get a lower bound of $n^{\Omega(\log \log n)}$ for sufficiently small $\epsilon$. More recently, we presented a $(1 + \epsilon)$-approximation [3] that has size $n \cdot (1/\epsilon)^{d2^{O(d \log d + \log d)}}$, bringing down the exponent of $1/\epsilon$ for the first time. An extension of this scheme further reduces the size to $n \cdot (1/\epsilon)^{d2^{O(d \log d)}}$, which matches our
lower bound result. Thus, this result brings the problem towards a closure for point sets in Euclidean space.

We also give results for non-Euclidean space; see [1] for more details. Some of our results use the integer grid in $d$-dimensions, while others use a link to high-dimensional lattice geometry: we make use of the $A^*$ lattice, whose Voronoi polytopes are more commonly known as permutahedra [5].

References


35.7.3 Computer Algebra, Algebraic Complexity Theory, and Algebraic Geometry

Efficient algebraic algorithms are the backbone of modern computer algebra, which is a fundamental tool in many real-world applications. The capabilities and the limits of algebraic computations in different algebraic computational models are studied in the field of algebraic complexity theory. Algebraic geometry with its algebro-geometric dictionary forms the strong bridge between geometry and algebra, connecting these two fundamental pillars of modern mathematics.
ANewDsc: An Approximate Arithmetic Newton-Descartes Method for Univariate Polynomial Real Root Finding

Investigators: Alexander Kobel and Michael Sagraloff in cooperation with Fabrice Rouillier (INRIA Paris) and Maplesoft Waterloo

Finding the zeros of a univariate polynomial function is a prolific task in mathematics and engineering, often called the “fundamental problem of computer algebra”, in reference to the fundamental theorem of algebra, which states that each polynomial has as many roots over the complex numbers as its degree, counted with multiplicity.

According to the celebrated Abel-Ruffini impossibility theorem, dating back to 1799, there are no closed-form solutions for the roots of univariate polynomials of degree higher than four. This implies that polynomial root finding is intrinsically a numerical task. However, traditional approaches from numerics, such as Newton iteration, lack certificates on the output or even reliable termination. For guarantees on correctness and completeness, more elaborate subdivision schemes are employed. In most realizations, such solvers rely on predicates to count or bound the number of roots inside a region of the real line or complex plane, and recursively refine a suitable initial region until the predicates are tight.

However, only recently formulations have appeared that do not intrinsically require the (unrealistic) Real-RAM model of infinite precision arithmetic; yet, the cost of evaluation of well-established predicates based on Sturm sequences or Descartes’ Rule of Signs becomes prohibitive already for inputs of moderate size. Moreover, in most formulations, convergence to roots is only linear, at least in the worst case. Those issues have been addressed in [5], where Sagraloff and Mehlhorn combined the classical bisection scheme for real root finding based on Descartes’ Rule with a variant of Newton iteration to achieve quadratic convergence, as well as modifications that allow the algorithm to work with finite, but arbitrary, precision arithmetic and for inputs with arbitrary computable coefficients.

Those improvements mark the culmination of many years of active research in this area; in particular, for the first time a dedicated real root solver matches the worst-case bit complexity of Pan’s seminal algorithm for factorization and complex root finding [3].

The latter is of merely theoretical interest: in the author’s own words, its “implementation would require a non-trivial work”, which is substantiated by the fact that no corresponding effort has been made in the past fifteen years. In contrast, Sagraloff and Mehlhorn’s algorithm ANewDsc is amenable for real-world use. During the past report period [1], we provided a proof-of-concept implementation of the method, based on Rouillier and Zimmermann’s RS library [4]. It demonstrated that the gain of up to three orders of magnitude in worst-case complexity over previous methods, predicted in theory, translates into practice without substantial impairments of the performance for small or well-conditioned instances.

Over the course of the past two years, we significantly overhauled ANewDsc to industrial strength. Extending the collaboration with INRIA and Maplesoft Waterloo, the new solver has been integrated in the MAPLE product line. At the time of writing, it is designated to become the default routine for univariate polynomial real root solving in the upcoming MAPLE 2019 distribution [2]. The revised version affirms the improvements in performance, in particular for ill-conditioned instances; but more importantly, up to our knowledge it is the first time that an algorithm for finding the roots of polynomials with arbitrary computable coefficients is exposed to end users.
The Geometry of Fast Matrix Multiplication

Investigators: Christian Ikenmeyer in cooperation with Jon Hauenstein (University of Notre Dame), Luca Chiantini (University of Siena), Joseph Landsberg (Texas A&M University), Giorgio Ottaviani (University of Florence), and Nicholas Ryder (UC Berkeley)

Strassen’s seminal paper [4] started a long line of research to understand fast matrix multiplication. His result can be interpreted as the realization that the $2 \times 2$ matrix multiplication tensor

$$M_2 := \sum_{i,j,k=1}^2 e_{i,j} \otimes e_{j,k} \otimes e_{k,i} \in (\mathbb{C}^{2\times 2})^\otimes 3$$

has rank 7 instead of 8. In [3] we investigate geometric and representation-theoretic properties of this decomposition. In particular, we apply a base change to Strassen’s algorithm to establish a beautiful rank 7 decomposition, which shows a large number of the tensor’s symmetries: If we let

$$a := \left( \begin{array}{cc} \omega & 0 \\ 0 & \omega^2 \end{array} \right), \quad b := \left( \begin{array}{cc} \sigma & \bar{i} \\ i & \bar{\sigma} \end{array} \right), \quad c := \left( \begin{array}{cc} i & i \\ \bar{i} & \bar{i} \end{array} \right)$$

for $\omega := \exp\left(-\frac{2\pi i}{3}\right)$, $\nu := \frac{i}{\sqrt{3}}$, $\sigma := \exp\left(\frac{2\pi i}{12}\right)/\sqrt{3}$, and define

$$\varrho : \mathbb{C}^{2\times 2} \rightarrow \mathbb{C}^{2\times 2}, \quad \varrho(X) = aXa^{-1},$$

then

$$M_2 = a^{\otimes 3} + b^{\otimes 3} + (\varrho(b))^{\otimes 3} + (\varrho^2(b))^{\otimes 3} \quad = \quad c \otimes \varrho(c) \otimes \varrho^2(c) + \varrho(c) \otimes \varrho^2(c) \otimes c + \varrho^2(c) \otimes c \otimes \varrho(c).$$

Note the invariance under the cyclic shift of the three tensor positions, but also the invariance under the application of $\varrho$. Also note here that the decomposition consists of 4 cubes and only 3 other rank-1 tensors. These 3 rank-1 tensors form an orbit under the action of cyclic shifts.
of tensor positions. In [1] we go one step further and focus on $M_3 := \sum_{i,j,k=1}^3 e_{i,j} \otimes e_{j,k} \otimes e_{k,i} \in (\mathbb{C}^{3\times 3})^{\otimes 3}$. Its rank is known to lie between 19 and 23. We find a rank 23 decomposition that consists of 11 cubes and one orbit of the action of a group $\mathbb{Z}_3 \times \mathbb{Z}_3$, where the $\mathbb{Z}_3$ part is again the cyclic shift of tensor positions and the $\mathbb{Z}_4$ is the canonical generalization of the map $\varphi$ from our analysis of $M_2$. Observing this high number of cubes in the decomposition initiated the paper [2]. In this paper we prove that the matrix multiplication exponent does not change if instead of the tensor rank of $M_n$ we study the so-called Waring rank of the homogeneous degree 3 matrix multiplication polynomial $p_n := \sum_{i,j,k=1}^n x_{i,j}x_{j,k}x_{k,i} \in \mathbb{C}[x_{1,1}, \ldots, x_{n,n}]^3$. Here the Waring rank of $p_n$ is the smallest number of summands that are needed to obtain $p_n$ as a sum of cubes of homogeneous linear polynomials. Moreover, in [2] we define numerous other models of computation that give the matrix multiplication exponent and analyze the first small instances with the help of numerical algebraic geometry.

References


A PTAS for the Commutative Rank of Symbolic Matrices of Constant Degree Polynomials

Investigators: Gorav Jindal and Anurag Pandey in cooperation with Vishwas Bhargava (Rutgers University) and Markus Bläser (Saarland University)

We consider matrices whose entries are given by multivariate polynomials over some field $\mathbb{F}$ (i.e., for a matrix $M$, $M_{ij} \in \mathbb{F}[x_1, \ldots, x_m]$). The commutative-rank of such a symbolic matrix is defined to be the maximum number $r$ such that at least one of the $r \times r$ minors is a non-zero polynomial. Over large enough fields, this is the same as the maximum possible rank of the evaluated $M$ (i.e., evaluating the entries by fixing the variables $x_1, \ldots, x_m$ to some constants from $\mathbb{F}$) over $\mathbb{F}$. This is a fundamental problem, as it generalizes several computational problems from algebra, geometry and combinatorics. For instance, the very special case of the problem when the entries are given by homogeneous linear forms, already subsumes the problems such as testing for perfect matching in graphs and identity testing of algebraic branching programs. Its higher degree cases capture more general problems of algebro-geometric nature, like testing algebraic independence of polynomials (over fields of zero characteristic) using the Jacobian matrix, and finding the dimension of the dual varieties of hypersurfaces using the Hessian matrix.

Although the problem has a simple and efficient randomized algorithm using the Schwartz-Zippel lemma, an efficient deterministic computation of the commutative-rank is a major
open problem, since a deterministic algorithm for exactly computing the rank in the linear case is already equivalent to the celebrated Polynomial Identity Testing (PIT) problem which itself would imply circuit complexity lower bounds [6].

Recently, there has been a series of results on computing the non-commutative-rank in the homogeneous linear case in deterministic polynomial time [4, 5]. Since the non-commutative-rank of the linear case is at most twice the commutative-rank, one immediately gets a deterministic $\frac{1}{2}$-approximation algorithm for the computation of the commutative-rank. This leads to a natural question of whether this approximation ratio can be improved.

In our work [2, 3], we answer this question affirmatively. More specifically, given an $n \times n$ matrix $M$ with entries given by homogeneous linear forms, and a rational number $0 < \epsilon < 1$, we present a simple greedy algorithm that runs in time $O(n^{4+\frac{1}{2}})$ and outputs an assignment $(\lambda_1, \ldots, \lambda_m) \in \mathbb{F}^m$ for the variables, such that the rank of the evaluated $M$, i.e.,

$$\text{rank}(\left((M(\lambda_1, \ldots, \lambda_m))\right) \geq (1 - \epsilon)\cdot \text{commutative-rank}(M(x_1, \ldots, x_m)).$$

The proof of correctness uses the so-called Wong sequences.

The next question is whether one could generalize this algorithm to the non-linear case, i.e., when the entries are given by polynomials of higher degrees. Such a higher degree generalization was already known in the non-commutative world, where the more general case in which the entries of the matrix are given by poly-sized formulas, reduces to the case where the entries are given by linear polynomials, using Higman’s trick. Higman’s trick however only preserves the co-rank, hence it cannot be used to reduce the problem of commutative-rank approximation to the case when the matrix entries are linear polynomials.

In our second work [1], we give a deterministic PTAS also for the higher degree case when the entries are given by polynomials with degrees bounded by a constant $d$, and hence we take a step towards bridging the knowledge gap between the non-commutative world and the commutative world. More specifically, given an $n \times n$ matrix $M(x_1, \ldots, x_m) = (M_{ij})_{n \times n}$, whose entries $M_{ij}$ are homogeneous forms of constant degree $d$ over a field $\mathbb{F}$ with $|\mathbb{F}| > nd$ and a constant $0 < \epsilon < 1$, our algorithm computes an assignment $(\lambda_1, \ldots, \lambda_m) \in \mathbb{F}^m$ such that

$$\text{rank}(M(\lambda_1, \ldots, \lambda_m)) \geq (1 - \epsilon)\cdot \text{commutative-rank}(M(x_1, \ldots, x_m)).$$

This algorithm runs in time $O\left((nmd)^{O(d^2/\epsilon)} \cdot M(n)\right)$, where $M(n)$ is the time required to compute the rank of an $n \times n$ matrix over $\mathbb{F}$. Clearly, the above running time is polynomial when $d$ is a constant. For the proof of correctness in the higher degree case, the Wong sequences are not useful because it is not even clear how to define a suitable Wong sequence in this case. To tackle the problem, we find a new way to analyze the low-degree components of the minors of the matrix obtained in the greedy step, which allows us to use the same algorithm strategy as in the linear case above for higher degree forms as well.

References


Complexity of Symmetric Polynomials

Investigators: Gorav Jindal in cooperation with Markus Bläser (Saarland University)

Arithmetic circuits model computations over a field. Boolean circuits have $\lor, \land, \neg$ gates. Analogously, arithmetic circuits have $+, -, \times$ gates. Each gate of such an arithmetic circuit computes a multivariate polynomial in the natural way. If the output gate of an arithmetic circuit $C$ computes the polynomial $g$, then we say that $C$ computes the polynomial $g$. The size of an arithmetic circuit $C$ is defined as the number of gates in $C$. The arithmetic complexity $L(f)$ of a polynomial $f \in \mathbb{C}[x_1, x_2, \ldots, x_n]$ is defined as the minimum size of any circuit which computes $f$.

Let $\mathfrak{S}_n$ be the symmetric group defined as the set of all permutations of the set $\{1, 2, \ldots, n\}$. A Boolean function $f : \{0, 1\}^n \to \{0, 1\}$ is said to be symmetric if $f(x_1, x_2, \ldots, x_n) = f(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$ for all $(x_1, x_2, \ldots, x_n) \in \{0, 1\}^n, \sigma \in \mathfrak{S}_n$. It is easy to see that all symmetric Boolean functions can be computed by constant depth threshold circuits, that is, are contained in the class $\text{TC}^0$. The notion of symmetric polynomials can also be defined similarly. It is natural to ask whether symmetric polynomials can also be computed efficiently, i.e., whether the arithmetic complexity of symmetric polynomials is also small? This is the question we studied in this work [3]. It was left as open question in [4].

The fundamental theorem of symmetric polynomials states that for a symmetric polynomial $f_{\text{Sym}} \in \mathbb{C}[x_1, x_2, \ldots, x_n]$, there exists a unique “witness” $f \in \mathbb{C}[y_1, y_2, \ldots, y_n]$ such that $f_{\text{Sym}} = f(e_1, e_2, \ldots, e_n)$, where the $e_i$’s are the elementary symmetric polynomials. In this work, we study the arithmetic complexity $L(f)$ of the witness $f$ as a function of the arithmetic complexity $L(f_{\text{Sym}})$ of $f_{\text{Sym}}$. We show that the arithmetic complexity $L(f)$ of $f$ is bounded by $\text{poly}(L(f_{\text{Sym}}), \deg(f), n)$. To the best of our knowledge, prior to this work only exponential upper bounds were known for $L(f)$ ([2, 1]). As a corollary of this result, we show that if $\text{VP} \neq \text{VNP}$ then there exist symmetric polynomial families which have super-polynomial arithmetic complexity.

References


**Geometric Complexity Theory**

Investigators: Karl Bringmann, Christian Ikenmeyer, and Gorav Jindal in cooperation with Markus Bläser, Vladimir Lysikov (Saarland University), Peter Bürgisser (TU Berlin), Fulvio Gesmundo (University of Copenhagen), Stefan Mengel (CRIL), Ketan Mulmuley (University of Chicago), Greta Panova (University of Southern California), Michael Walter (University of Amsterdam), and Jeroen Zuiddam (IAS Princeton)

Geometric Complexity Theory was initiated by Mulmuley and Sohoni [9, 10] in order to resolve Valiant’s famous determinant vs permanent conjecture [11]. This ambitious approach consists of several layers of precise conjectures on certain representation theoretic multiplicities. Given three partitions (i.e., three finite lists of nonincreasing natural numbers) \(\lambda, \mu, \nu\), the Kronecker coefficient \(k(\lambda, \mu, \nu)\) is the multiplicity of the Specht module \([\lambda]\) in the tensor product of Specht modules \([\mu] \otimes [\nu]\). In [7] we proved that it is NP-hard to decide whether \(k(\lambda, \mu, \nu)\) is positive, which disproved a conjecture of Mulmuley. On the other hand, we used our proof to construct many triples \((\lambda, \mu, \nu)\) with vanishing Kronecker coefficient. These were conjectured by Mulmuley and Sohoni to be useful to separate the \(GL_n\)-orbit closures of the \(n \times n\) determinant and the padded \(m \times m\) permanent \((m < n)\). We disproved this hope in [8] by showing that all Kronecker coefficients relevant to the geometric complexity theory program are positive. This was strengthened in [4] from Kronecker coefficients to the multiplicities in the coordinate ring of the orbit closure of the determinant. This is a major no-go result in geometric complexity theory.

Due to these no-go results, in [5] we initiated the study of other models of computation that replace the determinant but that have nicer geometric behavior. In [5] we study homogeneous polynomials that arise as the trace of a power of a matrix whose entries are homogeneous linear polynomials. The main result of [5] is comparable to [8]: the vanishing of representation-theoretic coefficients here can also not separate the desired complexity classes. Ultimately one wants to study products of different matrices whose entries are homogeneous linear polynomials, because the corresponding complete polynomial is characterized by its symmetries. There is still the hope that the vanishing of representation-theoretic coefficients can separate complexity classes in that model.

We take a step back and ask some more high-level questions:

In geometric complexity theory the ground field is the complex numbers and the reduction notion is the group orbit closure. As an aside, in [6] we prove that different classical reduction notions in algebraic complexity theory differ in power. We also show that VNP-completeness depends on the ground field in a subtle way.

In [3] we take the classical algebraic computational model of iterated \(3 \times 3\) matrix multiplication and analyze its \(2 \times 2\) counterpart. Allender and Wang [1] showed that not
every polynomial can be computed as an iterated $2 \times 2$ matrix multiplication of affine linear polynomials, but we prove that the geometric counterpart is a valid computational model: Every polynomial can be approximated by a $2 \times 2$ matrix multiplication of affine linear polynomials. Moreover the length of this product expression is polynomially equivalent to the border formula size, a well-known measure of complexity in geometric complexity theory. Our result also initiates the study of a new complete polynomial, the continuant, which is ubiquitous in the theory of continued fractions.

In [2] we investigate the fundamental capabilities and limitations of algebraic proofs for complexity lower bounds. We provide a setting in which superpolynomial algebraic lower bounds proofs require superpolynomial circuit size under a reasonable Boolean complexity assumption. Moreover, we show how geometric complexity theory can in principle break this barrier by encoding hard functions by concise representation-theoretic labels. As a side result from our techniques we prove that there is a constant multiplicative error to which tensor rank is NP-hard to approximate.

References


A Classification of Spherical Schubert Varieties in the Grassmannian

Investigators: Reuven Hodges in cooperation with Venkatramani Lakshmibai (Northeastern University)

In a previous work [3], we studied the actions of certain reductive subgroups of the general linear group $GL_N(\mathbb{C})$ on Schubert subvarieties of the Grassmannian. In particular, any parabolic subgroup $Q$ of the general linear group that acts on a Schubert variety $X(w)$ by left multiplication will decompose into the semidirect product of a reductive group $L$ and a unipotent group $U$. The action of $Q$ on $X(w)$ induces an action of $L$ on $X(w)$, which in turn induces an action of $L$ on the homogeneous coordinate ring $\mathbb{C}[X(w)]$ of $X(w)$. The main result of [3] is a combinatorial description of the decomposition of $\mathbb{C}[X(w)]$ into irreducible $L$-modules. This decomposition is given in terms of tensor products of irreducible general linear group representations since any such $L$ may be written as a product of smaller general linear groups embedded diagonally in $GL_N(\mathbb{C})$.

In our work [2], we apply our decomposition result towards the classification of those Schubert varieties that are also spherical varieties. When a reductive group $R$ acts on a normal, irreducible variety, we call it a spherical $R$-variety if it has a dense open orbit for a Borel subgroup of the reductive group. These varieties were first studied by Luna and Vust [5], and saw considerable development in the 1980s and 1990s by many authors including Brion, Knop, Luna, and Vust [1, 4]. The first result of our paper shows that a Schubert variety in the Grassmannian is a spherical $L$-variety if and only if the decomposition of $\mathbb{C}[X(w)]$ into irreducible $L$-modules is multiplicity free; that is, each irreducible $L$-representation shows up at most once in the decomposition. The rest of the paper is then devoted to giving the combinatorial criteria for a pair $X(w)$ and $L$ such that the decomposition of $\mathbb{C}[X(w)]$ into irreducible $L$-modules is multiplicity free.

References


35.8 Theory of Distributed and Embedded Systems

Coordinator: Christoph Lenzen

Broadly speaking, distributed computing concerns systems that consist of multiple agents that act based on local information. The main challenge is typically how agents gain access to
sufficient information to collaboratively solve a task quickly, despite limits on communication, faults, or inaccurate data. One of the key differences to many other models of computing is the possibility that the system may tolerate faults of individual agents. Our group in particular applies this paradigm to hardware, taking the view of, e.g., a multi-core processor as a distributed system.

35.8.1 Parallel and Distributed Distance Problems

Naturally, solving well-known computational problems involving distances such as All Pairs Shortest Paths, Single Source Shortest Paths, Steiner Trees and Forests, or flow-type problems is of great utility in distributed systems, too. We tackle these problems with all available means: randomization, approximation, decomposition, and so on. The key differences to centralized algorithms are that each node can perform its own computations, is initially only aware of its neighborhood in the graph, and needs to learn its local share of the output only (e.g., its distances from other nodes). To this end, nodes communicate over the edges of the graph; this communication is the main cost factor. For parallel algorithms, there is more flexibility: all nodes have access to a shared memory, eliminating the locality and coordination issues arising from the more stringent distributed model.

Further results on this topic are discussed in Section 35.6.1.

Distributed Distance Computation and Routing with Small Messages

Investigators: Christoph Lenzen in collaboration with Boaz Patt-Shamir (Tel Aviv University) and David Peleg (Weizmann Institute of Science)

In the Congest model of distributed computation, the system is represented by a graph whose nodes are computational devices and whose edges are communication links. Computations are performed in synchronous compute-send-receive rounds, where over each edge $O(\log n)$ bits can be sent per round. Initially, nodes know “their” share of the input (their unique $O(\log n)$-bit identifiers and, e.g., edge weights) and need to output “their” share of the output (e.g., their distances to other nodes).

In a sequence of works [1, 2, 4], we studied All Pairs Shortest Path computations and related tasks in this model. This lead to the identification of key primitives and techniques that form a toolset for designing efficient algorithms in this model.

1. Source detection: Given a subset of the nodes that are sources, let each node find the up to $\sigma$ closest sources (ties are broken by identifiers) within distance at most $h$.

2. Approximate source detection: As above, but for distance values that are accurate up to factor $1 + \varepsilon$. More precisely, the output is consistent with some distance values that satisfy this approximation guarantee w.r.t. the actual distances.

3. Skeleton graphs: A virtual graph obtained by sampling roughly $\sqrt{n}$ nodes in an $n$-node graph uniformly at random and adding an edge between sampled nodes $v$ and $w$ of weight (roughly) equal to their distance if they are connected by a shortest path of $\tilde{O}(\sqrt{n})$ hops; such a graph can be computed using the above primitives.

In a recent journal article [3], we present these concepts in a common context and provide algorithms we prove to be asymptotically optimal or near-optimal. We demonstrate the
utility of our techniques by several applications such as distance computations, distance oracles, and construction of routing tables. We believe that this article offers a careful exposition of a number of core techniques for distance computations and related tasks in the Congest model.

References


Parallel Metric Tree Embeddings

Investigators: Stephan Friedrichs and Christoph Lenzen

Tree embeddings are a powerful tool in the area of graph approximation algorithms. Roughly speaking, they transform problems on general graphs into much easier ones on trees. Fakcharoenphol, Rao, and Talwar [3] presented a probabilistic tree embedding that transforms \( n \)-node metrics into (probability distributions over) trees, while stretching each pairwise distance by at most an \( O(\log n) \) factor in expectation, which is is optimal.

We show how to achieve this for arbitrary weighted graphs \( G = (V, E) \) by a parallel algorithm, using polylog \( n \) depth and \( \tilde{O}(m^{1+\epsilon}) \) work, where \( n = |V|, \ m = |E|, \) and \( \epsilon > 0 \) is an arbitrarily small constant [4]. Our result improves upon Blelloch et al. [1] who require \( \tilde{O}(n^2) \) work as well as a metric (instead of a graph) as input.

Besides hop sets [2], our main tool in deriving these parallel algorithms is an algebraic characterization of a generalization of the classic Moore-Bellman-Ford algorithm. We consider this framework, which subsumes a variety of previous “Moore-Bellman-Ford-like” algorithms, to be of independent interest and discuss it in depth. In our algorithm, we leverage it for providing efficient query access to an approximate metric that allows sampling the tree using polylog \( n \) depth and \( \tilde{O}(m) \) work. We illustrate the generality and versatility of our techniques by various examples and a number of additional results.

Direct consequences of our techniques are that, on arbitrary weighted graphs, we

- solve \((1 + o(1))-\)approximate All Pairs Shortest Paths in polylog \( n \) depth and \( \tilde{O}(nm^{1+\epsilon}) \) work,
- improve the best previous tree-embedding algorithm [5] in the Congest model, and
- obtain better approximation algorithms for \( k \)-Median and Buy-at-Bulk Network Design.
References


35.8.2 Design of Fault-tolerant Hardware

Today’s Very Large Scale Integrated (VLSI) circuits operate on such small length and time scales that they become distributed systems in their own right. Small subcircuits take the role of nodes that communicate via wires – the “edges” of the communication graph – and billions of individual components render fault-tolerance mission-critical. Hence, insights from the area of distributed computing are useful for designing reliable hardware. However, VLSI circuits add unique challenges. In low-level hardware, even simple computations like an addition require substantial effort, and violation of timing constraints may induce so-called metastability, states of the logic that lie outside the abstraction of binary logic. We seek to adapt algorithmic ideas from distributed computing and other areas to this environment, with the goal of devising hardware with strong, provable fault-tolerance properties. One key challenge in this process is the design of clock generation and propagation schemes that are simultaneously highly resilient, accurate and precise, yet require few resources.

Self-stabilizing Byzantine Clock Synchronization

*Investigator: Christoph Lenzen in collaboration with Joel Rybicki (IST Austria)*

We study fault-tolerant algorithms for establishing synchrony in distributed systems in which each of the n nodes has its own clock. Our algorithms operate in a very strong fault model: we require self-stabilization, i.e., the initial state of the system may be arbitrary, and there can be up to \( f < n/3 \) ongoing Byzantine faults, i.e., nodes that deviate from the protocol in an arbitrary manner. Furthermore, we assume that the local clocks of the nodes may progress at different speeds (clock drift) and communication has bounded delay \( d \). In this model, we consider the pulse synchronization problem, where the task is to guarantee that eventually all correct nodes generate well-separated local pulse events (i.e., unlabeled logical clock ticks) in a synchronized manner.

We present a generic recursive approach that converts any synchronous (non-stabilizing) consensus algorithm \( A \) into a corresponding pulse synchronization algorithm \( P \) [2], where the guarantees of the consensus algorithm readily translate to guarantees for the new algorithm:
If $A$ tolerates (optimal) $f < n/3$ Byzantine faults, so does $P$.

If $A$ runs for $R$ rounds (in expectation), $P$ stabilizes in $O(dR \log n)$ time (w.h.p.).

If $A$ uses messages of $M$ bits, $P$ sends $O(M \log n)$ bits over each link every $d$ time.

Key ideas of our approach are derived from our prior work on *synchronous counting* [1]. This task can be seen as a simplified variant of the problem in the synchronous model, which we studied with the goal of obtaining a better understanding of the structure of the problem before turning to the more general setting.

Plugging state-of-the-art consensus routines into our framework, we achieve *exponential* improvements in stabilization time and the number of communicated bits compared to prior work, and give the first sublinear-time algorithm for the problem:

- In the deterministic setting, the state-of-the-art solutions stabilize in time $\Theta(f)$ and have each node broadcast $\Theta(f \log f)$ bits per time unit. We exponentially reduce the number of bits broadcasted per time unit to $\Theta(\log f)$ while retaining the same stabilization time.

- In the randomized setting, the state-of-the-art solutions stabilize in time $\Theta(f)$ and have each node broadcast $O(1)$ bits per time unit. We exponentially reduce the stabilization time to polylog $f$ while each node broadcasts polylog $f$ bits per time unit.

A revised full version of the article has been submitted to JACM and the referees have recommended acceptance subject to minor revisions.

**References**


**Robust Routing Made Easy**

*Investigators: Christoph Lenzen and Moti Medina*

Designing routing schemes is a multidimensional and complex task that depends on the objective function, the computational model (centralized vs. distributed), and the amount of uncertainty (online vs. offline). Nevertheless, there are quite a few well-studied general techniques, for a large variety of network problems. In contrast, in our view, practical techniques for designing robust routing schemes are scarce; while fault-tolerance has been studied from a number of angles, existing approaches are concerned with dealing with faults after the fact by rerouting, self-healing, or similar techniques. We argue that this comes at a high burden for the designer, as in such a system any algorithm must account for the effects of faults on communication.

With the goal of initiating efforts towards addressing this issue, we showcase in [1] simple and generic transformations that can be used as a blackbox to increase resilience against (independently distributed) faults. Given a network and a routing scheme, we
determine a reinforced network and corresponding routing scheme that faithfully preserves
the specification and behavior of the original scheme. We show that reasonably small
constant overheads in terms of size of the new network compared to the old are sufficient
for substantially relaxing the reliability requirements on individual components. The main
message in this paper is that the task of designing a robust routing scheme can be decoupled
into (i) designing a routing scheme that meets the specification in a fault-free environment,
(ii) ensuring that nodes correspond to fault-containment regions, i.e., fail (approximately)
independently, and (iii) applying our transformation to obtain a reinforced network and a
robust routing scheme that is fault-tolerant.

References

  Stabilization, Safety, and Security of Distributed Systems (SSS 2017), Boston, MA, USA, 2017,

Fault Tolerant Gradient Clock Synchronization

Investigators: Johannes Bund, Christoph Lenzen, and Will Rosenbaum

Clock synchronization is a fundamental problem in distributed computing. In many applica-
tions, it is vital that different computational entities (approximately) agree on a current
time. However, achieving perfect clock synchronization is impossible for a variety of reasons:
physical clocks are inherently noisy; there may be uncertainty in the transit times of messages
between entities; computers or links may crash, or even behave adversarially. The past 30
or so years have witnessed a tremendous growth in our theoretical understanding of clock
synchronization in two largely separate regimes. Fault Tolerant clock synchronization has
focused on the situation where all processes communicate with one another in a highly con-
nected network, but some processes are faulty, exhibiting arbitrary behavior [2, 5]. Gradient
clock synchronization (GCS) considers the case where each process communicates with only
a few neighbors, and only neighboring clocks need to be closely synchronized. Several works
have established matching upper and lower bounds on best possible synchronization for the
gradient clock synchronization problem [3, 4]. The recent work [3] for dynamic networks
implies similar bounds for clock synchronization when some process may crash. Yet there
were no known solutions for the GCS problem that tolerate even a single adversarial fault.

In [1], we describe an algorithm that achieves gradient clock synchronization in the
presence of adversarial faults. The performance of our algorithm matches the bounds of
the state-of-the-art GCS algorithm without faults. The trade-off for achieving this fault-
tolerance is that nodes in the original network are replicated. However, with modest overhead
(duplicating each node $O(\log n)$ times), the network can tolerate a constant fraction of faulty
nodes, so long as the faults are randomly distributed. Thus, our algorithm offers comparable
fault-tolerance with the best clock synchronization algorithms for a complete network, and
achieves the best possible gradient synchronization for general networks.
References


Time and Frequency Metrology

*Investigators: Attila Kinali in cooperation with Nicolas Boucquey (CERN)*

Time and frequency measurement is a fundamental part of many experiments. Being also the quantity we are able to measure with highest accuracy, it became the cornerstone of the SI unit system with its lastest revision [3]. Currently, the state of the art in single event time measurement is $0.5\text{ ps} = 0.5 \cdot 10^{-12}\text{ s}$, whereas most systems only reach about $2\text{ ps}$, with more precise systems usually having a slower measurement rate.

In order to have a measurement system to accurately characterize developed clock synchronization schemes, we were in need of a precise and high rate measurement system. We collaborated with CERN to develop a low-cost system that uses a variation of the time-interval counter design originally developed by Panek et al. [4]. Our first prototype reached a precision of $2\text{ ps}$ with an achievable rate of up to 100,000 or more measurements per second [1].

In the course of the project, we also developed a general software defined radio based system as a platform to develop time and frequency metrology measurement systems [2].

References


Noise in Electronic Systems

*Investigator: Attila Kinali*

All physical systems, weather patterns, the flow of rivers, migration patterns of animals, etc., exhibit noise. In electronics this noise places fundamental limits on the performance that can be achieved. In particular, it limits the precision and accuracy of measurement systems or the power consumption and data rate of communication systems. A better understanding of how noise is generated and propagates through the system directly leads to an improved performance of, e.g., battery life of cell phones.

While advances in production technology have reduced the noise in electronics by several orders of magnitude in the last 5 decades, it is unlikely that we will see much further improvement from this source, as noise levels are hitting fundamental limits (e.g., Johnson noise of resistors). Thus further improvement of noise performance must come from structural changes in circuit design by limiting noise propagation and amplification within the circuit. Applying standard tools from signal theory readily leads to a good description of linear time-invariant (LTI) systems. While LTI systems account for a large part of analog electronics, an ever increasing part of analog electronics falls outside of the LTI category. Unfortunately, only a few select non-LTI systems have descriptions that are amenable to noise analysis.

Starting from a methodology used to describe noise oscillators [1], we have developed a model [2] that can be adapted to all periodically excited electronic circuits and gives accurate values for the output noise of the system using only the signal and its noise level, noise levels of extrinsic and intrinsic sources, and the sensitivity to extrinsic sources.

**References**


### 35.8.3 Hazard-freedom and Metastability-containment

One specific challenge when dealing with low-level hardware (and clocking in particular) is metastability, an unstable third equilibrium state of bistable elements such as latches and registers. The standard solution is to use synchronizers whenever signals cross clock domain boundaries or performing analog-to-digital conversions. However, this takes a couple of clock cycles, which has negative impact when time is critical, e.g., in control loops operating at very high frequencies. Instead, one can contain metastability, i.e., accept it as normal part of the operation and study circuit behavior in three-valued *Kleene logic*. The goal is to develop small and low-latency circuits in this setting, posing plenty of circuit design and complexity questions.
Metastability-Containing Circuits

Investigators: Stephan Friedrichs and Christoph Lenzen in cooperation with Matthias Függer (CNRS & ENS Paris-Saclay)

In digital circuits, metastable signals have voltages strictly between logical 0 and logical 1, breaking the abstraction of Boolean logic. Unfortunately, any way of reading a signal from an unsynchronized clock domain or performing an analog-to-digital conversion incurs the risk of a metastable result; no physical implementation of a digital circuit can deterministically avoid, resolve, or detect metastability [2].

Traditionally, the only countermeasure is to write a potentially metastable signal into a synchronizer – a bistable storage element like a flip-flop – and wait. Synchronizers exponentially decrease the odds of maintained metastability over time, i.e., the waiting time determines the probability to resolve to logical 0 or 1. Accordingly, this approach delays subsequent computations and does not guarantee success.

In [1], we propose a fundamentally different approach: It is possible to contain metastability by fine-grained logical masking so that it cannot infect the entire circuit. This technique guarantees a limited degree of metastability in – and uncertainty about – the output. We justify our model by showing that it preserves the impossibilities shown in [2]. On the positive side, we characterize the functions that can be implemented in this setting. Moreover, we show that masking registers, which “hide” internal metastability from the outside world except when stabilizing internally (resulting in an observed late transition at the output), induce a strict computability hierarchy depending on the number of clock cycles used by the circuit to compute its output.

References


Metastability-Containing Sorting Networks

Investigators: Johannes Bund, Christoph Lenzen, and Moti Medina

When setup/hold times of bistable elements are violated they may enter a transient metastable state, that is neither digital 0 nor 1. In general metastability can not be avoided, detected, nor (deterministically) resolved. Nonetheless, prior work showed that it is possible to sort inputs possibly containing metastable bits [3].

Metastability-containing sorting networks are of interest in a hardware implementation of the clock synchronization algorithm of Lynch & Welch. By using sorting networks the problem of sorting multiple $B$-bit Gray code words, each possibly containing a metastable bit, boils down to sorting two such inputs.

- In [1], we show that there is a sorting circuit with asymptotically optimal $O(\log B)$ depth. However, the recursive design has (only) near-optimal $O(B \log B)$ size.
The question whether there is a size and depth-optimal sorting circuit is answered in [2] by showing that there is a circuit with depth $O(\log B)$ and size $O(B)$. The improvement is achieved by leveraging a parallel prefix computation circuit construction.

Synthesis to a standard-cell library and subsequent layout shows the improvements in size and delay for the respective chips. These simulations show that [2] improves up to 48.46% in delay and by 71.58% in size over [1].

References


Complexity of Hazard-free Circuits

Investigators: Johannes Bund, Christian Ikenmeyer, Christoph Lenzen, and Moti Medina in collaboration with Balagopal Komarath, Vladimir Lysikov (Saarland University), Andrey Mokhov (Newcastle University), and Karteek Sreenivasaih (IIT Hyderabad)

As mentioned earlier, metastability-containing circuits need to have small delay. Phrased in the language of circuit complexity, this means that we require circuits of low depth. Moreover, such circuits are not simply analyzed in terms of Boolean logic, but rather according to the 3-valued logic introduced by Kleene [4, §64], in which metastable (or any other kind of unstable or transitioning) signals are represented by a third logical value $u$. Combinational circuits in this model have already been studied in various contexts, with the goal of avoiding so-called hazards. A circuit (with one output) has a hazard at input $x \in \{0, 1, u\}^n$ iff for any possible string $y \in \{0, 1\}^n$ that coincides with $x$ at all positions $i$ such that $x_i \neq u$ it holds that the circuit outputs the same value $b \in \{0, 1\}$, yet it outputs $u$ when given input $x$.

Hazard-free combinational circuits are then exactly metastability-containing combinational circuits.

Prior general constructions of metastability-containing circuits suffer from size $\Omega(2^n)$, even if the implemented function admits a circuit of polynomial size. This even holds when considering 1-bit hazards only, i.e., hazards for which $x$ contains but a single $u$. In [3], we present a generic construction that yields a $k$-bit-hazard-free circuit implementing $f : \{0, 1\}^n \to \{0, 1\}$ of size $n^{2k}C(f)$, where $C(f)$ is the size of an arbitrary circuit implementing $f$ (which may exhibit $k$-bit-hazards). The main result of this work, however, is a striking connection between monotone circuit complexity and hazard-free circuits: For any monotone Boolean function $f$, the smallest hazard-free circuit has exactly the same size as the smallest monotone circuit (i.e., a circuit using only AND and OR gates) implementing $f$. Due to known exponential separations between standard and monotone circuit complexity for explicit functions, this in
particular establishes the same for metastability-containing circuits. By a padding argument, it also follows that the exponential dependence on \( k \) is unavoidable when constructing \( k \)-bit-hazard-free circuits.

Guided by this negative result, in recent work we circumvent the lower bound for circuits which are induced by transducers [2]. A transducer is a finite state machine transcribing an input sequence into an output sequence. This class of functions includes such central arithmetic operations as addition or taking the maximum. For the latter, previous work [1] (see Section 35.8.3) proved that in restricted settings very efficient circuits exist and suggested a more general principle. Astonishingly, we were able to show that any function induced by a constant-size transducer with constant-sized alphabets has hazard-free circuits of size \( O(n) \). However, this is only of use if the encoding of inputs and outputs is chosen in a way that is both efficient (e.g., using unary encoding makes hazard-free addition trivial) and useful, i.e., preventing hazards is sufficient to ensure that the output maintains the same degree of certainty on the encoded numerical value as the input.\(^4\) Accordingly, in addition to the generic construction of hazard-free circuits for transducer functions, we provide a family of encodings that can be used to safely add numbers with a total uncertainty of at most \( k \) in the result, yet has only \( 2^k - 1 \) redundant bits (at least \( k - 1 \) are provably needed).

We hope that this work provides a basis for devising circuits offering meaningful arithmetics despite some degree of uncertainty on the input values.

References


Frequency Adaptation Circuit

Investigators: Attila Kinali, Christoph Lenzen, and Ben Wiederhake in cooperation with Matthias Függer (CNRS & ENS Paris-Saclay)

In synchronous circuits, supply voltage is often assumed to be perfectly constant. This pushes the responsibility of handling power issues to the hardware designer. Especially in the context of supply voltage droops, which can happen in the order of tens of nanoseconds, this leads to overly conservative bounds on clock speeds. These result in poor performance, even if droops are rare. Adaptive strategies detect such potentially hazardous events and either initiate a rollback to a previous state or proactively reduce clock speed in order to

\(^4\)For instance, standard binary encoding fails dramatically: adding 0111 (7) and 000\(u\) (0 or 1) must result in \(uuuu\) (0 to 15), regardless of the circuit performing the computation.
prevent timing violations. However, the performance of state-of-the-art solutions is limited by synchronization delay to avoid that the clock signal is affected by metastability.

We present an all-digital circuit [1] that can respond to droops within a fraction of a clock cycle. This response consists of delaying individual clock pulses, specifically increasing the period between them. This approach is only possible due to careful metastability containment, as the decision (delay or no delay) is based on measured values that are still in the process of stabilization while travelling through a delay chain that simultaneously acts as a synchronizer chain. A simple digital counter at the end of the chain accumulates the delay in the classical, digital way, exploiting that the measurement values have stabilized by the time when they arrive at the counter.

We verify our solution by formally proving correctness, complemented by VHDL and Spice simulations of a 65 nm ASIC design confirming the theoretically obtained results.

Overall, our contribution enables hardware to be clocked faster while using the same voltage supply, or use less power while running at the same amortized clock frequency, or an arbitrary trade-off in-between.

References


35.8.4 Sublinear Algorithms and Property Testing

When dealing with very large graphs, no single machine can process or even hold the entire graph in its memory. This necessitates a different model of computing, where the graph is probed locally. Nonetheless, queries like “is node \(v\) part of the maximal independent set?” assume an underlying consistent graph structure. Concretely, when performing multiple independent and arbitrarily ordered queries of the above type, the yes/no answers should be correct with respect to a single maximal independent set. The goal here is to minimize the query complexity, i.e., the number of queries of the form “what are the neighbors of node \(v\)?” or “what is neighbor number 3 of \(v\)?” etc. is to be minimized.

Conceptually closely related is property testing in the Congest model. Here, the goal is to, e.g., separate the cases of an input graph being triangle-free or “far from” triangle-free (i.e., many edges would have to be removed to make it triangle-free), where any answer is feasible for graphs in between. In contrast to the above setting, each part of the graph can be inspected by some node. However, communication between nodes is limited by an \(O(\log n)\) message size restriction and the optimization target of small running time in the Congest model, implying that each node can only learn about a small part of the graph. The algorithm then produces an answer at each node, where “yes” instances should be accepted by all nodes, whereas “no” instances need to be detected as such by at least one node.
Partitions for Testing Graph Properties and Local Algorithms

Investigators: Christoph Lenzen and Reut Levi in cooperation with Hendrik Fichtenberger (TU Dortmund), Yadu Vasudev (IIT Madras), and Maximilian Wötzel (UPC Barcelona)

Constructing a sparse spanning subgraph is a fundamental primitive in graph theory. In [3], we study this problem in the Centralized Local model, where the goal is to decide whether an edge is part of the spanning subgraph by examining only a small part of the input; yet, answers must be globally consistent and independent of prior queries. Unfortunately, maximally sparse spanning subgraphs, i.e., spanning trees, cannot be constructed efficiently in this model. Therefore, we settle for a spanning subgraph containing at most $(1 + \epsilon)n$ edges (where $n$ is the number of vertices and $\epsilon$ is a given approximation/sparsity parameter). We achieve a query complexity of $\tilde{O}(\text{poly}(\Delta/\epsilon)n^{2/3})$, where $\Delta$ is the maximum degree of the input graph and $\tilde{O}$ hides polylogarithmic factors in $n$. Our algorithm is the first to do so on arbitrary graphs of non-constant degree. Moreover, we achieve the additional property that our algorithm outputs a spanning subgraph of bounded stretch, i.e., distances are approximately preserved. With high probability, for each deleted edge there is a path of $O(\log n \cdot (\Delta + \log n)/\epsilon)$ hops in the output that connects its endpoints.

Building on the partition technique developed in [3], we obtain the following results for testing minor-closed properties of graphs in [2]. We consider one-sided error property testing of $F$-minor freeness in bounded-degree graphs for any finite family of graphs $F$ that contains a minor of $K_{2,k}$, the $k$-circus graph, or the $(k \times 2)$-grid for any $k$ in $N$. This includes, for instance, testing whether a graph is outerplanar or a cactus graph. The query complexity of our algorithm in terms of the number of vertices in the graph, $n$, is $\tilde{O}(n^{2/3}/\epsilon^5)$. Czumaj et al. [1] showed that cycle-freeness and $C_k$-minor freeness can be tested with query complexity $\tilde{O}(\sqrt{n})$ by using random walks, and that testing $H$-minor freeness for any $H$ that contains a cycles requires $\Omega(\sqrt{n})$ queries. In contrast to these results, we analyze the structure of the graph and show that either we can find a subgraph of sublinear size that includes the forbidden minor $H$, or we can find a pair of disjoint subsets of vertices whose edge-cut is large, which induces an $H$-minor.

References


Sublinear Algorithms for Edge Sampling

Investigators: Will Rosenbaum in cooperation with Talya Eden and Dana Ron (Tel Aviv University)

Graph algorithms are one of the fundamental objects of study throughout computer science. Indeed, graphs are powerful abstractions of networks, and appear throughout the social, biological, physical, and mathematical sciences. In the age of Big Data, our ability to produce vast amounts of data—often represented as a graph—is rapidly exceeding our ability to disseminate said data. Thus, efficiency in processing data is becoming more and more important. In order to understand the theoretical complexity of analyzing large graphs (and other structures), several computational models have been introduced that allow access to graphs via elementary “queries.” The complexity of a problem is then measured to be the number of queries that are necessary and sufficient to solve the problem. One of the best-studied query models for graphs, known as the general graph model allows an algorithm to access a graph via queries of the following types: (1) a degree query reports the number of neighbors of a node, (2) a neighbor query returns the identity of a particular (or random) neighbor of a node, and (3) a pair query determines whether or not a given pair of nodes are connected. In this model, several works have proved matching upper and lower bounds for a variety of problems such as estimating the number edges, triangles, and cliques in a graph, and estimating the moments of the degree distribution.

The recent work of Eden and Rosenbaum [2] addressed the following basic problem theoretically: How many queries (in the general graph model) are required in order to sample an (almost) uniformly random edge in a graph? They proved that roughly $\Theta(n/\sqrt{m})$ queries are necessary and sufficient in the worst case. However, the construction that gives the lower bound in [2] is, for many applications, unrealistic, as “hard” examples contain large dense subgraphs that do not occur in many “naturally occurring” graph families. In [1], we give a more fine-grained analysis of the problem of sampling edges in graphs. In particular, we parametrize the complexity of this task by the arboricity, $\alpha$, of the graph. We show that for graphs of arboricity $\alpha$, the complexity is roughly $\Theta(n\alpha/m)$. In the extremal case, $\alpha = \sqrt{m}$, these bounds match the bounds of [2]. For many natural families of graphs, however, $\alpha$ can be much smaller. For example, for planar graphs and random preferential attachment graphs, $\alpha = O(1)$. In such cases, the $\Theta(n\alpha/m)$ bound gives an exponential improvement over the previous state-of-the-art.

References


Faster and Simpler Distributed Algorithms for Testing and Correcting Graph Properties in the Congest Model

Investigators: Reut Levi and Moti Medina in cooperation with Guy Even (Tel-Aviv University)

In [3], we present distributed testing algorithms of graph properties in the Congest model [1], derived from one-sided error testing algorithms in the general graph model. We first describe a general procedure for converting $f(D)$-round $\epsilon$-testers, where $D$ denotes the diameter of the graph, to $\epsilon$-testers with $O((\log n)/\epsilon) + f((\log n)/\epsilon)$ rounds, where $n$ is the number of processors of the network. We then apply this procedure to obtain an optimal tester, in terms of $n$, for testing bipartiteness, whose round complexity is $O(\epsilon^{-1} \log n)$. This improves over the poly$(\epsilon^{-1} \log n)$-round algorithm by Censor-Hillel et al. [1]. Moreover, for cycle-freeness, we obtain a corrector of the graph that locally corrects the graph so that the corrected graph is acyclic. Note that, unlike a tester, a corrector needs to mend the graph in many places in case it is far from having the property. In the second part of the paper we design algorithms for testing whether the network is $H$-free for any connected $H$ of size up to four with round complexity of $O(\epsilon^{-1})$. This improves over the $O(\epsilon^{-2})$-round algorithms for testing triangle freeness by Censor-Hillel et al. [1] and for testing excluded graphs of size 4 by Fraigniaud et al. [4]. In the last part we generalize the global tester by Iwama and Yoshida [5] of testing $k$-path freeness to testing the exclusion of any tree of order $k$. We then show how to simulate this algorithm in the Congest model in $O(k^{k^2+1} \cdot \epsilon^{-k})$ rounds.

This paper got merged and published in [2].

References


Sublinear Random Access Generators for Preferential Attachment Graphs

Investigators: Reut Levi and Moti Medina in cooperation with Guy Even (Tel-Aviv University) and Adi Rosén (CNRS & Université Paris Diderot)

In [1], we consider the problem of generating random graphs in evolving random graph models. In the standard approach, the whole graph is chosen randomly according to the distribution of the model before answering queries to the adjacency lists of the graph. Instead, we propose to answer queries by generating the graphs on-the-fly while respecting the probability space of the random graph model. We focus on two random graph models: the Barabási-Albert Preferential Attachment model and the random recursive tree model. We present sublinear randomized generating algorithms for both models. Per query, the running time, the increase in space, and the number of random bits consumed are polylog(n) with probability 1 − 1/poly(n), where n denotes the number of vertices. This result shows that, although the Preferential Attachment random graph model is defined sequentially, random access is possible without chronological evolution. In addition to a conceptual contribution, on-the-fly generation of random graphs can serve as a tool for simulating sublinear algorithms over large Preferential Attachment graphs.

References


35.8.5 Computational Dynamics

Dynamics are simple stochastic processes on networks, in which agents update their own state according to a symmetric function of the state of their neighbors and of their current state, with no dependency on time or on the topology of the network. In previous decades, this kind of systems has been investigated from a computability point of view, attracting the interest of mathematicians and physicists. Recently it has been subject to a renewed interest from computer scientists, as new techniques for analyzing this class of processes have made it possible to answer questions regarding their efficiency and capability as distributed algorithms.

Dynamics for Community Detection

Investigators: Emanuele Natale and André Nusser in cooperation with Giacomo Scornavacca (University of L’Aquila) and Emilio Cruciani (Gran Sasso Science Institute)

We investigate the behavior of networks of agents, which update their state according to the 2-Choices dynamics, a simple majority-based opinion dynamics [4] in which, at each discrete time step, each agent observes the state of 2 random neighbors: if the two neighbors have the same state, the agent copies it; otherwise, the agent keeps its current state. Despite its apparent simplicity, 2-Choices has been analytically characterized only on networks with a strong expansion property, under assumptions on the initial configuration that establish it as
a fast majority consensus protocol [1]. In [2], we aim at contributing to the understanding of the 2-Choices dynamics by considering its behavior on a class of networks with core-periphery structure, a well-known topological assumption in social networks. In a nutshell, assume that a densely-connected subset of agents, the core, holds a different opinion from the rest of the network, the periphery. Then, depending on the strength of the cut between the core and the periphery, a phase-transition phenomenon occurs: Either the core’s opinion rapidly spreads among the rest of the network, or a metastability phase takes place, in which both opinions coexist in the network for superpolynomial time. By looking at the 2-Choices dynamics as a simplistic model of competition among opinions in social networks, our results sheds light on the influence of the core on the rest of the network, as a function of the core’s connectivity towards the latter.

In [3], we assume that the possible interactions among agents exhibit a community structure, that is, the agents can be grouped in two communities such that interactions of agents belonging to the same community are more likely to happen than interactions among agents belonging to different communities. By leveraging the fact that the behavior of the 2-Choices dynamics can be expressed in terms of the eigenvalues of the underlying interaction network, we prove that, when the states of the nodes are randomly initialized, the system rapidly and stably converges to a configuration in which the communities maintain internal consensus on different states. Our result is the first rigorous result on the behavior of dynamics for non-consensus problems on non-complete topologies, and it has several implications in different contexts in which dynamics are adopted for computational and biological modeling purposes. In the context of Label Propagation Algorithms, a class of widely used heuristics for community detection, it represents the first theoretical result on the behavior of a distributed label propagation algorithm with quasi-linear message complexity. In the context of evolutionary dynamics on graphs, our result shows that, when the probability of adoption of a given mutation by a node of the evolutionary graph depends super-linearly on the frequency of the mutation in the neighborhood of the node and the underlying evolutionary graph exhibits a community structure, there is a non-negligible probability for species differentiation to occur.

References


35.8.6 Dynamics for Consensus

Investigators: Emanuele Natale in cooperation with Andrea Clementi, Luciano Gualà, Francesco Pasquale (University of Rome Tor Vergata), Petra Berenbrink (University of Hamburg), Robert Elsässer, Peter Kling (University of Salzburg), Mohsen Ghaffari (ETH Zurich), Frederik Mallmann-Trenn (MIT), and Giacomo Scornavacca (University of L’Aquila)

In this project, we study consensus dynamics on the complete graph of $n$ nodes. Initially, each node supports one from up to $n$ opinions. Nodes randomly and in parallel sample the opinions of constantly many nodes. Based on these samples, they use an update rule to change their own opinion. The goal is to reach consensus, a configuration where all nodes support the same opinion. In [1], we compare two well-known dynamics: 2-Choices and 3-Majority dynamics. In the former, each node samples two nodes and adopts their opinion if they agree. In the latter, each node samples three nodes: If an opinion is supported by at least two samples the node adopts it, otherwise it randomly adopts one of the sampled opinions. Known results for these dynamics focus on initial configurations with a limited number of colors (say $n^{1/3}$), or typically assume a bias, where one opinion has a much larger support than any other. For such biased configurations, the time to reach consensus is roughly the same for 2-Choices and 3-Majority dynamics. Interestingly, we prove that this is no longer true for configurations with a large number of initial colors. In particular, we show that 3-Majority dynamics reaches consensus with high probability in $O(n^{3/4} \log^{7/8} n)$ rounds, while 2-Choices dynamics can need $\Omega(n / \log n)$ rounds. We thus get the first unconditional sublinear bound for 3-Majority dynamics and the first result separating the consensus time of these processes. Along the way, we develop a framework that allows a fine-grained comparison between consensus processes from a specific class.

In [2], we consider the Undecided-State dynamics, another well-known update rule for distributed consensus. Again, previous work on the Undecided-State dynamics only considers initial color configurations with no undecided nodes and a large bias (i.e., $\Theta(n)$) towards the majority color. An interesting open question is whether this dynamics always (i.e., starting from an arbitrary initial configuration) reaches consensus quickly (i.e., within a polylogarithmic number of rounds). We analyze it in the binary case (every node can either support one of two possible colors, or be in the undecided state), and provide an unconditional analysis of the Undecided-State dynamics that answers this question in the affirmative. More precisely we prove that, starting from any initial configuration, the process reaches a monochromatic configuration within $O(\log n)$ rounds, with high probability. This bound turns out to be tight. Our analysis also shows that, if the initial configuration has bias $\Omega(\sqrt{n \log n})$, then the dynamics converges toward the initial majority color, with high probability.

References


**Community Detection in Population Protocols**

*Investigators: Emanuele Natale in cooperation with Luca Becchetti (Sapienza University of Rome), Andrea Clementi, Francesco Pasquale (University of Rome Tor Vergata), Pasin Manurangsi, Prasad Raghavendra, and Luca Trevisan (UC Berkeley)*

In this project, we consider the following asynchronous, opportunistic communication model over a graph $G$: in each round, one edge is activated uniformly and independently at random and (only) its two endpoints can exchange messages and perform local computations. Under this model, also known as *population protocols*, we study the following random process: The first time a vertex is an endpoint of an active edge, it chooses a random number, say $\pm 1$ with probability $1/2$; then, in each round, the two endpoints of the currently active edge update their values to their average. In [1], we show that, if $G$ exhibits a two-community structure (for example, two expander graphs connected by a sparse cut), the values held by the nodes will collectively reflect the underlying community structure over a suitable phase of the above process, allowing efficient and effective recovery in important cases.

In more detail, we first provide a first moment analysis showing that, for a large class of almost-regular clustered graphs, which includes the stochastic block model (the most famous model of random graph with a clustered structure), the expected values held by all but a negligible fraction of the nodes eventually reflect the underlying cut. We prove that this property emerges after a mixing period of length $O(n \log n)$. We further provide a second-moment analysis for a more restricted class of regular clustered graphs that includes the regular stochastic block model. For this case, we are able to show that most nodes can efficiently and locally identify their community of reference over a suitable time window. This results in the first opportunistic protocols that approximately recover community structure using only polylogarithmic work per node. Our second moment analysis requires new concentration bounds on the product of certain random matrices that are of independent interest.

**References**


**Noisy Consensus**

*Investigators: Emanuele Natale in cooperation with Amos Korman (IRIF, CNRS) and Lucas Boczkowski (IRIF)*

Biological systems can share and collectively process information to yield emergent effects, despite inherent noise in communication. While man-made systems often employ intricate
structural solutions to overcome noise, the structure of many biological systems is more amorphous. It is not well-understood how communication noise may affect the computational repertoire of such groups. To approach this question, in [1] we consider the basic collective task of rumor spreading, in which information from few knowledgeable sources must reliably flow into the rest of the population. We study the effect of communication noise on the ability of groups that lack stable structures to efficiently solve this task. More precisely, we consider synchronous and asynchronous communication models on the complete graph (Uniform Push and Population Protocols), and we present an impossibility result which strongly restricts reliable rumor spreading in such groups. Namely, we prove that, in the presence of even moderate levels of noise that affect all facets of the communication (any message could be mistaken by the receiver for another one with at least some small probability), no scheme can significantly outperform the trivial one in which agents have to wait until directly interacting with the sources – a process which requires linear time in the population size. Our results suggest that in order to achieve efficient rumor spread a system must exhibit either some degree of structural stability or, alternatively, some facet of the communication which is immune to noise. We corroborate this claim by providing new analyses of experimental data regarding recruitment in Cataglyphis niger desert ants. In light of our theoretical results, we also discuss insights on the strategies to overcome noise in other biological systems.

References


35.8.7 Miscellaneous

**Dominating Set in Synchronous Distributed Models**

*Investigators: Saeed Akhoondian Amiri in cooperation with Patrice Ossona de Mendez (CAMS, EHESS Paris), Roman Rabinovich (TU Berlin), and Sebastian Siebertz (HU Berlin)*

The minimum dominating set problem asks for finding the smallest size set $S$ of vertices of the input graph such that every other vertex in the graph is a neighbor of a vertex in $S$. Domination problems are among the most important problems in computer science; they are relevant both in theory and practice. There are several problem variants, for instance:

- **Connected Domination:** where the subgraph induced by the vertices in $S$ is required to form a connected graph,
- **Distance $r$ Domination:** where we replace the standard neighborhood of a vertex by all vertices within distance $r$.

These problems have been studied in different models of computation such as distributed, centralized, and parallel.

We consider the dominating set, distance $r$-dominating set, and connected dominating set problems in the Congest model [1]. In these tasks, every node of the graph needs to determine its dominator (which is defined in a problem-specific way). In each (synchronous) round of the Congest model, neighboring nodes can exchange $O(\log n)$ bits; the task is to provide an approximate solution in few number of communication rounds. We provide
a constant-factor approximation for all of the mentioned problems in $O(\log n)$ rounds on the class of bounded expansion graphs. This is a class of sparse graphs more general than planar, bounded genus, and excluded minor graphs, which is a proper subclass of bounded degeneracy graphs.

There are several works on this topic in similar models of computation and we can categorize them into two groups: those approaching the problem in general graphs and those approaching the problem in a restricted class of graphs (in turn providing better bounds). Our paper [1] falls in the second category. A randomized $O(\log n)$-round algorithm for dominating set on the more general class of graphs of bounded degeneracy was given in [2]. However, our algorithm improves over this algorithm on bounded expansion graphs by being deterministic and solving both the distance-$r$ dominating set and connected dominating set problems. To the best of our knowledge, this is the first non-trivial algorithm addressing the $r$-domination and connected $r$-domination problems.

References


Routing in Software Defined Networks

Investigators: Saeed Akhoondian Amiri in cooperation with Szymon Dudycz (University of Wroclaw), Klaus-Tycho Foerster, Mahmoud Parham, Stefan Schmid (University of Vienna), Riko Jacob (ITU Copenhagen), and Sebastian Wiederrecht (TU Berlin)

Software Defined Networks (SDNs) are designed to separate the data plane from the control plane. This separation takes away, from the underlying hardware, the complex tasks of configuring, optimizing, and securing the network. In an SDN the whole network is managed by software and, as a consequence, it is more flexible and more responsive to dynamic changes.

A typical SDN consists of a controller (or a set of controllers), switches, other network elements, and connection links between them. The task of managing the network rests on the shoulder of the controller and hence we have to program it according to the requirements of the network. Among all of the tasks a controller handles, the most important one is to send routing commands to switches and other network elements.

The need for routing commands can arise for different reasons. First, we may find new routes between a couple of pairs of nodes and announce them to the corresponding switches. Second, due to drastic changes in traffic of the network (e.g., due to a DoS attack), due to the security policy changes, etc., we may have to reroute communication. In modern networks, for both cases there are several challenges to be addressed. In the following, we will explain the issues that we addressed in our recent papers.
For finding a route between a pair of nodes there are well-known classical algorithms, such as Ford-Fulkerson, Dijkstra, etc., or even to find disjoint paths between nodes there are several modern algorithms, see e.g. [7, 8]. However, requirements of real networks are a bit more complex than such static routing solutions. For instance, we might have to find a path/walk between a pair of nodes such that the packets go through specific middle nodes, namely, waypoints. Such a requirement has several practical applications, one of them appearing in search engines where requests are routed along several chosen hops, not necessarily the shortest path, such that they can later customize the advertisements received from these hops. This problem has been studied in theory, however, the focus of researchers was to provide a path or cycle that goes through specific vertices [5]. In [2, 3] we investigated a broader and more practical case by considering walks that pass through a specific set of nodes. We provide both algorithmic and hardness results for various network structures, such as for graphs of bounded treewidth and planar graphs.

As discussed earlier, another routing task a controller is supposed to perform is to provide a rerouting scheme. At first glance, it seems that this is an easy task: just apply standard routing algorithms and then, based on the outcome of the algorithm, update forwarding rules of switches. However, the rerouting operation must not cause a misordering of packets. For instance, a temporary loop could result in huge packet losses, we do not want to add extra overhead to packet headers, not all elements of the network are synchronous in regard to receiving and performing commands of the controller, etc. There is a considerable amount of work on rerouting flows in SDNs [4, 6], but all previous works were for single flows. We expand this direction of the research by considering a larger scale, studying the a more realistic case of multiple flows [1]. We discovered that unlike single flows, already finding a feasible solution, let alone an optimal one, for the problem is hard even for two flows. The problem remains hard even if the underlying graph is a DAG. On the positive side, on DAGs we provide a polynomial time algorithm for the case that the number of flows is small.

References


Distributed Set Cover Approximation: Primal-Dual with Optimal Locality

Investigators: Moti Medina in cooperation with Guy Even (Tel-Aviv University) and Mohsen Ghaffari (ETH Zurich)

In [2], we present a deterministic distributed algorithm for computing an $f(1 + \epsilon)$ approximation of the well-studied minimum set cover problem, for any constant $\epsilon > 0$, in $O(\log(f\Delta) / \log \log(f\Delta))$ rounds. Here, $f$ denotes the maximum element frequency and $\Delta$ denotes the cardinality of the largest set. This $f(1 + \epsilon)$ approximation almost matches the $f$-approximation guarantee of standard centralized primal-dual algorithms, which is known to be essentially the best possible approximation for polynomial-time computations. The round complexity almost matches the $\Omega(\log \Delta / \log \log \Delta)$ lower bound of Kuhn et al. [3], which holds even for $f = 2$ and any polylog($\Delta$) approximation. Our algorithm also gives an alternative way to reproduce the time-optimal $2(1 + \epsilon)$-approximation of vertex cover, with round complexity $O(\log \Delta / \log \log \Delta)$, as presented by Bar-Yehuda et al. [1] for weighted vertex cover. Our method is quite different and it can be viewed as a locality-optimal way of applying a primal-dual scheme in the more general case of set cover.

References


35.9 Academic Activities

35.9.1 Journal Positions

Kurt Mehlhorn:

35.9.2 Conference and Workshop Positions

Membership in Program Committees

Antonios Antoniadis:
- 13th Workshop on Models and Algorithms for Planning and Scheduling Problems (MAPSP), Kloster Seeon, Germany, June 2017,
- 46th International Conference on Parallel Processing (ICPP), Bristol, UK, August 2017.

Karl Bringmann:
- 19th Genetic and Evolutionary Computation Conference (GECCO), Berlin, Germany, July 2017,
- 12th International Symposium on Parameterized and Exact Computation (IPEC), Vienna, Austria, September 2017,
- 44th International Colloquium on Automata, Languages, and Programming (ICALP), Warsaw, Poland, July 2017,
- 9th Annual Innovations in Theoretical Computer Science (ITCS), Cambridge, Massachusetts, USA, January 2018,
- 34th International Symposium on Computational Geometry (SoCG), Budapest, Hungary, June 2018,
- 4th Highlights of Algorithms (HALG), Amsterdam, The Netherlands, June 2018,
- 36th International Symposium on Theoretical Aspects of Computer Science (STACS), Berlin, Germany, March 2019.

Cosmina Croitoru:

Christian Ikenmeyer:

Thomas Kesselheim:

Bundit Laekhanukit:
- 12th International Frontiers of Algorithmics Workshop (FAW), Guangzhou, China, May 2018.

Christoph Lenzen:
- 44th International Colloquium on Automata, Languages, and Programming (ICALP), Warsaw, Poland, July 2017,
Eunjin Oh:
- 34th European Workshop on Computational Geometry (EuroCG), Berlin, Germany, March 2018.

Alkmini Sgouritsa:
- 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Montreal, Canada, May 2019.

Membership in Organizing Committees

Christian Ikenmeyer:
- MPI-INF and MPI-MiS joint workshop on Theoretical Computer Science and Algebraic Geometry, Saarbrücken, Germany, January 2019.

Christoph Lenzen:
- 1st Workshop on Hardware Design and Theory (HDT), Vienna, Austria, October 2017.

35.9.3 Invited Talks and Tutorials

Kurt Mehlhorn:
- Reflections on Theory and Practice, Leopoldina Lecture 2018, Saarbrücken, April 2018,
- Reflections on Theory and Practice, Invited Talk, 29th International Symposium on Algorithms and Computation (ISAAC), Jiaoxi, Yilan County, Taiwan, December 2018,
- On Fair Division of Indivisible Goods, Invited Talk, 25th International Colloquium on Structural Information and Communication Complexity (SIROCCO), Ma’ale Ha-Hamisha, Israel, June 2018,
- Physarum Solves Positive Undirected LPs, Invited Talk, 13th International Computer Science Symposium in Russia (CSR), Moscow, Russia, June 2018,
- Certifying Algorithms, Gean Course at IIT Delhi, India, March 2018
- Two Topics in Discrete Algorithms: Certifying Algorithms and Slime Mold Computations, Summer School speaker, XIII Summer School in Discrete Mathematics, Valparaiso, Chile, January 2018

Karl Bringmann:


Reuven Hodges:

- A classification of spherical Schubert varieties in the Grassmannian, Invited Talk, A Conference of Commutative Algebra and Representation Theory, New Orleans, USA, November 2018,

Christian Ikenmeyer:

- Plethysms and Kronecker coefficients in geometric complexity theory, Invited Talk, Lie Groups and Representation Theory Seminar at the University of Tokyo, Japan, March 2018,
- Geometric Complexity Theory: Complexity Lower Bounds Using Algebraic Geometry and Representation Theory, Invited Tutorial, Lower Bounds in Computational Complexity Boot Camp, Simons Institute, Berkeley, California, USA, August 2018,
- Recent Progress on Representation Theoretic Multiplicities in GCT, Invited Talk, Workshop on Algebraic Methods at Simons Institute, Berkeley, California, USA, December 2018.

Emanuele Natale:

- What can be Computed in a Simple Chaotic Way?, Invited talk, 18th Italian Conference in Theoretical Computer Science (ICTCS), Naples, Italy, September 2017.

We would like to note that regular talks at on-invitation-only workshops (e.g. Dagstuhl, Oberwolfach) are not listed.

35.9.4 Other Academic Activities

- ADOCS 2017 organized by Christian Ikenmeyer and Michael Sagraloff.
  Topic: Algebraic Complexity Theory and Computer Algebra
  Invited Speakers:
  Markus Bläser, Saarland University
  Francois Le Gall, Kyoto University
  Amir Yehudayoff, Technion
ADFOCS 2018 organized by Karl Bringmann, Marvin Künnemann, and Cosmina Croitoru.

Topic: Fine-Grained Complexity and Algorithms

Invited Speakers:

Ramamohan Paturi, University of California San Diego
Danupon Nanongkai, KTH Royal Institute of Technology
Karl Bringmann, Max Planck Institute for Informatics

Kurt Mehlhorn was the chairman of the Scientific Board of the Institute of Science and Technology Austria (2009–2017).

Kurt Mehlhorn is the co-director of the Indo Max Planck Center for Computer Science (IMPECS) (since 2010).

Kurt Mehlhorn was a member of the Award Committee of the Gödel Prize (2014–2017).

Kurt Mehlhorn was a member of the Scientific Advisory Board of the Simons Institute for the Theory of Computing, Berkeley, USA (2015–2017).

Kurt Mehlhorn was the Chairman of the Scientific Advisory Board of INRIA (2015–2017).

Kurt Mehlhorn was a member of the the Award Committee of the ESA Test-of-Time Award (2015–2017).

Kurt Mehlhorn is a member of the European Research Council (since 2016).

Bojana Kodric visited the National Institute of Informatics, Tokyo, Japan, from December 2017 to May 2018 as a research internship under the supervision of Prof. Dr. Katsumi Inoue.

### 35.10 Teaching Activities

**Summer Semester 2017**

**Courses**
- Optimization (Andreas Karrenbauer with tutors Davis Issac and Pavel Kolev)
- Approximation Algorithms (Tobias Mömke and Hang Zhou with tutor Daniel Vaz)
- Introduction to Geometric Complexity Theory (Christian Ikenmeyer in cooperation with Markus Bläser)

**Seminars:**
- Reading Group Algorithms (Emanuele Natale, Kurt Mehlhorn, and Ruben Becker)

**Winter Semester 2017/2018**

**Courses**
- Ideen und Konzepte der Informatik (Antonios Antoniadis and Kurt Mehlhorn)
- Computer Algebra (Michael Sagraloff with tutor Anurag Pandey)
Algorithmic Game Theory, Mechanism Design and Computational Economics (Yun Kuen Cheung with tutor Bhaskar Ray Chaudhury)
Fine-Grained Complexity Theory (Karl Bringmann and Marvin Künnemann with tutor Philip Wellnitz)
Theory of Distributed Systems (Christoph Lenzen with tutor Ben Wiederhake)
Geometric Complexity Theory 2 (Christian Ikenmeyer in cooperation with Markus Bläser)

Summer Semester 2018

Courses
Keeping Time in Distributed Systems (Christoph Lenzen with tutor Saeed Amiri)
A First Introduction to Geometric Complexity Theory (Christian Ikenmeyer)
Optimization (Andreas Karrenbauer with tutor Maximilian John)

Seminars:
Reading Group Algorithms (Kurt Mehlhorn and Daniel Vaz)
Selected Topics in Fine-Grained Complexity Theory (Karl Bringmann and Marvin Künnemann)

Winter Semester 2018/2019

Courses
Ideen und Konzepte der Informatik (Kurt Mehlhorn)
Multivariate Algorithmics (Karl Bringmann in cooperation with Holger Dell)
Randomized and Approximation Algorithms (Antonios Antoniadis and Marvin Künnemann)
Algorithms on Digraphs (Saeed Amiri and Will Rosenbaum)
Theory of Distributed Systems (Christoph Lenzen with tutor Ben Wiederhake)

Seminars:
Reading Group Algorithms (Kurt Mehlhorn and Daniel Vaz)

Master Theses
Alexander Anisimov, Implementing a Gradient Descent Algorithm for Phase Unwrapping as an Undirected Transshipment Problem, 2018
Peter Matthias Manderscheid, An Optimization Framework for Shift Scheduling, 2017
Paul Nicholas Manderscheid, SAR Phase Unwrapping by Undirected Shortest Transshipment, 2018
Ben Wiederhake, Towards Lightweight Fault-tolerant Pulse Synchronization, 2017
Bachelor Theses

Johannes Bund, *Metastability-Containing Sorting*, 2017
Julian Dörfler, *An efficient data structure for finding Pareto-Optimal points*, 2018
Thomas Haslbauer, *Polynomial Identity Testing of Read-k Formulas*, 2017
Matthias Leinen, *Efficient Generation of $1/f^\alpha$ Noise in Matlab Simulink*, 2018
Philip Wellnitz, *Clique-Based Lower Bounds for Parsing Tree-Adjoining Grammars*, 2017

35.11 Dissertations, Habilitations, Offers, Awards

35.11.1 Dissertations

– Cosmina Croitoru, *Graph Models for Rational Social Interaction*, 2017
– Stephan Friedrichs, *Metastability-Containing Circuits, Parallel Distance Problems, and Terrain Guarding*, 2017
– Sandy Heydrich, *A tale of two packing problems: Improved algorithms and tighter bounds for online bin packing and the geometric knapsack problem*, 2018
– Eugenia Holm, *Optimierungsprobleme mit Cliquen und Bicliquen in Graphen*, 2018
– Bojana Kodric, *Incentives in Dynamic Markets*, 2018
– Pavel Kolev, *Algorithmic Results for Clustering and Refined Physarum Analysis*, 2018

35.11.2 Offers for Faculty Positions

– Christian Ikenmeyer, University of Liverpool, UK, Senior Lecturer
– Thomas Kesselheim, TU Braunschweig, Germany, Junior Professor
– Thomas Kesselheim, TU Dortmund, Germany, Junior Professor
– Moti Medina, Ben-Gurion University of the Negev, Israel, Lecturer
– Emanuele Natale, I3S Laboratory of Université Côte d’Azur, France, CNRS Chargé de Recherche
– Tim Oosterwijk, Maastricht University, The Netherlands, Assistant Professor
– Michael Sagraloff, University of Applied Sciences Landshut, Germany, Professor
– Alkmini Sgouritsa, University of Liverpool, UK, Lecturer
– Karteek Sreenivasasaih, Indian Institute of Technology Hyderabad, India, Assistant Professor
– Hang Zhou, Ecole Polytechnique, Paris, France, Assistant Professor
35.11.3 Awards

- Karl Bringmann received an EATCS Presburger Award in 2019.
- Karl Bringmann received a Heinz Maier-Leibnitz-Prize in 2019.
- Christian Ikenmeyer received a Research Fellowship of the Simons Institute for the Theory of Computing, Berkeley, California, USA in 2018.
- Marvin Künnemann received an Otto Hahn Medal of the Max Planck Society in 2017.
- André Nusser received the “Publikationspreis der Universität Stuttgart” in 2018 (together with Stefan Funke and Sabine Storandt).
- Kurt Mehlhorn received an Honorary Doctorate Degree from the University of Patras in 2017.
- Kurt Mehlhorn was appointed as Honorary Member of the “Gesellschaft zur Förderung des Forschungstransfers” in 2018.
- Emanuele Natale received a Research Fellowship of the Simons Institute for the Theory of Computing, Berkeley, California, USA in 2017.
- Eunjiin Oh received the Lise Meitner Award from the MPI-INF in 2018.
- Kevin Schewior received a Dissertation Award of the German Operations Research Society (GOR) in 2017.
- Kevin Schewior received a Postdoctoral Researchers International Mobility Experience (PRIME) Fellowship from the German Academic Exchange Service (DAAD) in 2017.

35.12 Grants

35.13 Publications

Journal articles and book chapters


35 D1: Algorithms and Complexity


**Conference articles**


Technical reports and arXiv/ECCC articles


PhD theses


36 D2: Computer Vision and Multimodal Computing

36.1 Personnel

Head of Group
Prof. Dr. Bernt Schiele

Senior Researchers and Research Group Leaders
Dr. Zeynep Akata (affiliated since April 2017)
Dr. Björn Andres (until September 2017)
Dr. Andreas Bulling (until October 2018)
Dr. Mario Fritz (until June 2018)
Dr. Gerard Pons-Moll (since September 2017)
Dr. Paul Swoboda (since March 2018)

Researchers
Dr. Michael Xuelin Huang (until October 2018)
Dr. Qiuhong Ke (since May 2018)
Dr. Qianru Sun (full time until April 2018, affiliated since October 2018)
Dr. Shanshan Zhang (affiliated since July 2017)

Ph.D. Students
Bharat Lal Bhatnagar (since July 2018)
Apratim Bhattacharyya
Moritz Böhle (since March 2019)
Alina Dima (until December 2018)
Yang He
Xudong Hong (since January 2019)
Andrea Hornakova
Jan Hendrik Hosang (until June 2017)
Eldar Insafutdinov
Anna Khoreva (until May 2018)
Jan-Hendrik Lange
Verica Lazova (since February 2019)
Evgeny Levinkov (until October 2018)
Wenbin Li (until May 2018)
Max Losch (since September 2017)
Philipp Müller
Seong Joon Oh (until August 2018)
Mohamed Omran
Tribhuvanesh Orekondy
Farzaneh Rezaeianaran (since March 2019)
Anna Rohrbach (until June 2017)
Hosnieh Sattar
Rakshith Shetty
Julian Steil
David Stutz (since October 2017)
Yongqin Xian
Ning Yu (since September 2018)
Xucong Zhang (until November 2018)

Interns
Saurabh Sharma (since August 2018)

Secretaries
Connie Balzert

36.2 Visitors
From March 2017 to February 2019, the following researchers visited our group:

<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Felipe Yanez</td>
<td>07.04.2017</td>
<td>INSEAD European Institute of Business Administration</td>
</tr>
<tr>
<td>Daniel Cremers</td>
<td>18.04.2017</td>
<td>Technical University of Munich</td>
</tr>
<tr>
<td>David Stutz</td>
<td>28.04.2017</td>
<td>RWTH Aachen University</td>
</tr>
<tr>
<td>Vittorio Ferrari</td>
<td>03.05.2017</td>
<td>University of Edinburgh and Google Research</td>
</tr>
<tr>
<td>Mike Roberts</td>
<td>05.05.2017</td>
<td>Stanford University</td>
</tr>
<tr>
<td>Christoph Lampert</td>
<td>22.05.2017</td>
<td>IST Institute of Science and Technology Austria, Vienna</td>
</tr>
<tr>
<td>Guido Montafur</td>
<td>20.06.2017</td>
<td>Max Planck Institute for Mathematics in the Sciences, Leipzig</td>
</tr>
<tr>
<td>Max Losch</td>
<td>10.07.2017</td>
<td>University of Amsterdam</td>
</tr>
<tr>
<td>Manuela Schuler</td>
<td>24.08.2017</td>
<td>University of Stuttgart</td>
</tr>
<tr>
<td>Stephen Batifol</td>
<td>04.09.2017</td>
<td>Paris-Sud University</td>
</tr>
<tr>
<td>Andreu Girbau</td>
<td>25.09.2017</td>
<td>UPC Polytechnic University of Catalonia, Barcelona</td>
</tr>
<tr>
<td>Christopher Sprague</td>
<td>06.11.2017</td>
<td>European Space Agency</td>
</tr>
<tr>
<td>Bharat Lal Bhatnagar</td>
<td>11.01.2018</td>
<td>Microsoft Research India</td>
</tr>
<tr>
<td>Vikram Voleti</td>
<td>26.02.2018</td>
<td>IIIT International Institute of Information Technology, Hyderabad</td>
</tr>
</tbody>
</table>
36.3 Group Organization

The group currently consists of the following seven research groups, led by a research group leader (most of which are also senior researcher):

- **Computer Vision** (Bernt Schiele)
- **Scalable Learning and Perception** (Mario Fritz)
- **Perceptual User Interfaces** (Andreas Bulling)
- **Combinatorial Image Analysis** (Bjoern Andres)
- **Multimodal Deep Learning** (Zeynep Akata)
- **Real Virtual Humans** (Gerard Pons-Moll)
- **Combinatorial Computer Vision** (Paul Swoboda)
The entire group meets once per week for a scientific talk and discussions. Any organizational matters are also discussed at these meetings. Senior researchers and research group leaders regularly meet with their respective research group to discuss more specific topics related to the research of their group.

PhD-students and PostDocs have a weekly or bi-weekly scientific meeting with their respective supervisors (Zeynep Akata, Bjoern Andres, Andreas Bulling, Mario Fritz, Gerard Pons-Moll, Paul Swoboda, and Bernt Schiele). For many of the PhD-students there are also other collaborating researchers present during the discussions.

Following the tradition from previous research groups of Bernt Schiele (TU Darmstadt and ETH Zurich) the group organizes two retreats per year. During these retreats every PhD-student and researcher gives a presentation about his or her current scientific work with an emphasis on the current and future work for the next six to twelve months. An important component during the retreats is that the entire group gives feedback and discusses openly. While there is a general rule that the presentations should be between 20–25 min, the discussions are open-ended (and typically last an additional 30 min to 1 h). About a third up to half the time of every retreat is dedicated to social activities.

The selection of PhD-students and PostDocs includes discussions and demos with a significant number of the current students, researchers, and research group leaders. Decisions are taken by the research group leaders and Bernt Schiele.

36.4 Human Pose and Tracking

*Coordinators: Eldar Insafutdinov, Qianru Sun, and Bernt Schiele*

Humans are arguably the most important category in the visual object recognition. In this section we describe our work on articulated human pose estimation, the task of localizing key skeletal joints of humans in natural images and videos. Human pose estimation and tracking are crucial for a number of important applications, such as human activity recognition, virtual and augmented reality, human-robot interaction. In this section we discuss our results on several fronts that involve a fine grained analysis of humans in the visual data. In Sect. 36.4.1 we present results on generation of images of humans, where we take advantage of compactness and expressiveness of human pose representation to guide image generation process. Sect. 36.4.2 details our research on human action recognition that spans a contribution of a new dataset with densely annotated action labels to an effective action recognition model as well as an action anticipation method, addressing an important problem of predicting actions over short- and long-term time horizons. Sect. 36.4.3 describes our work on human pose estimation in images and videos. First of all, we propose an approach that relies on a graph-based formulation that is general enough to address detection, tracking and estimation of body articulations of multiple people in unconstrained real-world scenes. It also has an appealing property to automatically infer the number of people in the scene without relying on an explicit prior on person count. Secondly, we discuss our results on 3D pose estimation. We present an approach to markerless motion capture that has been developed in collaboration with the computer graphics group (D4) at MPI-INF. Our approach allows to accurately estimate 3D body pose in video footage captured with a few handheld cameras and is applicable in outdoor conditions with dynamically changing illumination.
and background. Additionally, we develop a method for single-view 3D pose estimation that utilizes the robustness of the 2D keypoint localization with 2.5D depth-ordering constraints and hypothesis generation and testing. In Sect. 36.4.4 we present powerful models that go beyond mere skeletons and provide a richer representation of human body shapes that which can be learnt either from body scan databases or from still images. Finally, in Sect. 36.4.5 we discuss our work on multiple people tracking in monocular video sequences that achieves state of the art results on a challenging object tracking benchmark.

36.4.1 Conditioned Image Generation

Investigators: Qianru Sun, Mario Fritz, Bernt Schiele in cooperation with Liqian Ma (KU Leuven), Xu Jia (KU Leuven), Stamatios Georgoulis (KU Leuven), Tinne Tuytelaars (KU Leuven), Luc Van Gool (KU Leuven, ETH)

Generating realistic-looking images is of great value for many applications such as face editing, movie making and image retrieval based on synthesized images. From the application perspective, users typically have a particular intention in mind such as changing the background, an object’s category, its color or viewpoint. In this section, we introduce our works that aim to guide the generation process explicitly by an appropriate representation of that intention to enable direct control over the generation process [1, 2]. There are many interesting applications derived from this task. For example, in movie making, we can directly manipulate a character’s human body to a desired pose or, for human pose estimation, we can generate training data for rare but important poses.

In [1], we propose the novel Pose Guided Person Generation Network (PG$^2$) that allows to synthesize person images in arbitrary poses, based on an image of that person and a novel pose. Our generation framework PG$^2$ utilizes the pose information explicitly and consists of two key stages: pose integration and image refinement. In the first stage the condition image and the target pose are fed into a U-Net-like network to generate an initial but coarse image of the person with the target pose. The second stage then refines the initial and blurry result by training a U-Net-like generator in an adversarial way. Extensive experimental results on both 128×64 re-identification images and 256×256 fashion photos show that our model generates high-quality person images with convincing details. In this work, we need paired data which contain a reference image and a target image from the identical person. While, in real application scenarios, we may not have such paired data for training. In order to solve this problem, we propose the self-supervised learning paradigm in [2].

Specifically in [2], we aim at generating such images based on a novel, two-stage reconstruction pipeline that learns a disentangled representation of the aforementioned image factors and generates novel person images at the same time. First, a multi-branched reconstruction network is proposed to disentangle and encode the three factors into embedding features, which are then combined to re-compose the input image itself. Second, three corresponding mapping functions are learned in an adversarial manner in order to map Gaussian noise to the learned embedding feature space, for each factor, respectively. Using the proposed framework, we can manipulate the foreground, background and pose of the input image, and also sample new embedding features to generate such targeted manipulations, that provide more control over the generation process. Experiments on the Market-1501 and Deepfashion datasets show that our model does not only generate realistic person images
Figure 36.1: Left: image sampling results on Market-1501. Three factors, i.e. foreground, background and pose, can be sampled independently (1st-3rd rows) and jointly (4th row). Right: similar joint sampling results on DeepFashion. This dataset contains almost no background, so we only disentangle the image into appearance and pose factors.

with new foregrounds, backgrounds and poses, but also manipulates the generated factors and interpolates the in-between states. Another set of experiments on Market-1501 shows that our model can also be beneficial for the person re-identification task. Generated images on two datasets are shown in Figure 36.1.

References


36.4.2 Human Action Recognition

Investigators: Mykhaylo Andriluka, Qiuhong Ke, Qianru Sun, Mario Fritz, Bernt Schiele in cooperation with Serena Yeung (Stanford University), Olga Russakovsky (Stanford University), Ning Jin (Stanford University), Greg Mori (Simon Fraser University), Li Fei-Fei (Stanford University), Mengyuan Liu (Nanyang Technological University), Runwei Ding (Peking University), Hong Liu (Peking University)

Human action recognition remains challenging in realistic videos, where scale and viewpoint changes make the problem complicated. In [1], we propose a simple and effective method which is robust to those changing factors. In [4], we pay attention to a more comprehensive understanding of human activity that each moment of the activity gets an accurate action label. In [2], we introduce a new method for efficient and effective long-term action anticipation.

Many complex models have been developed for human action recognition in diverse video conditions. In [1], we explore using low-level features and typical linear classifiers to achieve top performance. Our method is based on the popular baseline model bag-of-words model [3] that has been shown high efficiency but ignores the arrangement of local features. We propose to encode the relative position of visual words into bag-of-words model using a simple and compact method called Sliding Coordinates Coding (SCC) [1]. The key observation is that the relative position is robust to the variations of video scales and view angles. We also design a temporal cutting scheme to define the margin of coding within video clips with the intuitive observation that visual words farther away from each other have less relationship. When considering more comprehensive human activity in long videos, we know that human activity is continual, and every minute is filled with potential labeled actions. Further, we humans are great at multi-tasking: we can be walking while talking on the phone while holding a cup of coffee. In [4], we claim that a comprehensive understanding of human activity in video requires labeling every frame according to the actions occurring, placing multiple labels densely over a video sequence. To study this problem we extend the existing THUMOS dataset and introduce MultiTHUMOS, a new dataset of dense labels over unconstrained internet videos. Modeling multiple, dense labels benefits from temporal relations within and across classes. We define a novel variant of long short-term memory (LSTM) deep networks for modeling these temporal relations via multiple input and output connections. We show that this model improves action labeling accuracy and further enables deeper understanding tasks ranging from structured retrieval to action prediction.

Human action anticipation is very important in many real-world applications such video surveillance and human-robot interaction. Ideally, in the real-world applications, future actions should not only be predicted with high accuracy but also at arbitrary and variable time-horizons ranging from short to long-term predictions. Current work mostly focuses on predicting the next action and thus long-term prediction is achieved by recursive prediction of each next action, which is both inefficient and accumulates errors. In [2], we propose a novel time-conditioned method for efficient and effective long-term action anticipation. The main idea is to perform action anticipation by incorporating the time parameter to the observation. Therefore, our method directly anticipates the action at the future time in a one-shot fashion, and thus avoids anticipating all actions in the time period before t. Moreover, we introduce an attended temporal feature and time-conditioned skip connections
in order to effectively use the time-conditioned observation for action anticipation. We evaluate our method on the large-scale Epic-Kitchen Dataset and the 50Salads Dataset and make comparison with respect to the state-of-the-art methods and various baselines. We improve on the state-of-the-art and in particular in accuracy of the anticipation of the long-term future actions.

References


36.4.3 Human Pose Estimation


In this subsection we discuss our results in articulated human pose estimation. First, we adress the challenging task of localising 2D body poses in the scenes in the wild with multiple people. We extend the state of the art approach for multi-person pose estimation to video sequences, thus defining a new task of articulated multi-person tracking. It encompasses three well established computer vision problems: person detection, pose estimation and tracking. Our method jointly solves them and is able to exploit the synergy between the tasks. To enable further research on this problem we have collected a large-scale pose estimation dataset in videos. Additionally, we develop an effective algorithm to refine 2D body pose predictions which can be used with any existing method. In the second part we provide an outlook of our work on recovering 3D body poses as well as tracking them over time. We propose methods that can perform pose estimation from either from a single view or from two or more views.

2D Pose Estimation and Tracking in Scenes with Multiple People

Human body motion reveals important information for analysis of human activities, interactions between people, and interactions of people with the environment. As a step towards accurately measuring human motion even in complex unconstrained videos we propose a novel approach [7]. We build on our expertise in human pose estimation but address a
more difficult problem of correctly inferring entire trajectories of body joints instead of single-frame estimates. Our starting point is a model that resembles existing architectures for single-frame pose estimation but is several orders of magnitude faster. We achieve this in two ways: (1) by simplifying and sparsifying the body-part relationship graph and leveraging recent methods for faster inference, and (2) by offloading a substantial share of computation onto a feed-forward convolutional architecture that is able to detect and associate body joints of the same person even in clutter. We use this model to generate proposals for body joint locations and formulate articulated tracking as spatio-temporal grouping of such proposals, see Figure 36.2. This allows to jointly solve the association problem for all people in the scene by propagating evidence from strong detections through time and enforcing constraints that each proposal can be assigned to one person only. We report results on a public MPII Human Pose benchmark [2] and on a new dataset of videos with multiple people. We demonstrate that our model achieves state-of-the-art results while using only a fraction of time and is able to leverage temporal information to improve state-of-the-art for crowded scenes.

In the work described above we introduced the new task of articulated multi-person tracking and an algorithm that solves it. However, existing datasets annotated with human poses were not suited for training or evaluating such approaches; they either contain multiple people in standalone images [8, 2], video sequences of single isolated individuals [9] or are small-scale [7]. To address this shortcoming we created a PoseTrack: a new large-scale benchmark for video-based human pose estimation and articulated tracking [1]. Overall, the dataset contains 550 video sequences with 66,374 frames. The length of the majority of the
sequences in our dataset ranges between 41 and 151 frames with 30 annotated frames in the middle of the sequence. In addition, we densely annotate validation and test sequences with a step of four frames. In total, we provide around 23,000 labeled frames with 153,615 pose annotations. To the best of our knowledge this makes PoseTrack the largest multi-person pose estimation and tracking dataset released to date. In order to enable timely and scalable evaluation on the held-out test set, we provide a centralized evaluation server. To sample the initial interest of the computer vision community and to obtain feedback we have organized a workshop and a competition at ICCV 2017 inviting research groups to submit results of their methods. We strongly believe that the proposed benchmark will prove highly useful to drive the research forward by focusing on remaining limitations of the state of the art.

To this end we analyzed the difficulty of the benchmark. We split the sequences by the performance according to the Multiple Object Tracking Accuracy [3] and found out that the methods mostly agree on the difficulty of sequences. Currently, failures lie in the sequences that include crowds, extreme proximity of people to each other, rare poses and strong camera motions. We hosted a second challenge at ECCV 2018, for which we doubled the size of the dataset. The relative improvement to the state of the art in articulated tracking and multi-person pose estimation was 18.5% and 13% respectively, when compared to best results in the previous challenge. This certifies that our dataset proved to be useful for the community in advancing the state of the art. The benchmark is freely accessible at https://posetrack.net/.

Despite great progress in the task of human pose estimation there remain a lot of challenges, where pose estimators fail due to person-person occlusions, rare body configurations, partial visibility or cluttered backgrounds. To address these challenges cases we propose [6] a refinement step that corrects errors in the localization of keypoints and can be easily applied to predictions coming from any existing pose estimation method. More specifically, given an RGB image and a potentially noisy 2D body pose estimate, we aim to output a refined human pose by exploiting the dependencies between the image and the inherent structure of the provided body pose. To achieve this, we propose modeling this refinement function using a fully convolutional network, modified to accept not only 3 RGB channels as input, but also a rendered estimated pose configuration. This input body pose is essentially a noisy version of the target pose, so we synthesize it from the ground truth by adding several transformations that simulate the noise usually introduced by pose estimators. Evaluation on four human pose estimation benchmarks demonstrates consistent improvement after applying the proposed refinement network to pose predictions given by various state-of-the-art approaches across different tasks, showing the effectiveness and generality of the proposed framework.

3D Pose Estimation and Tracking

Here we discuss our results on estimating 3D body pose. In [5] we discuss our multi-view method for recovering body poses and motions that is able to operate with as few as two cameras and is robust to difficult imaging conditions and dynamic background clutter.

Markerless motion capture has seen great progress, but most state-of-the-art approaches fail to reliably track articulated human body motion with a very low number of cameras, let alone when applied in outdoor scenes with general background. In our work [4] we propose a method for accurate markerless capture of articulated skeleton motion of
several subjects in general scenes, indoors and outdoors, even from input filmed with as few as two cameras. The new algorithm combines the strengths of a discriminative image-based joint detection method with a model-based generative motion tracking algorithm through a unified pose optimization energy. The discriminative part-based pose detection method is implemented using convolutional networks (ConvNets) and estimates unary potentials for each joint of a kinematic skeleton model. These unary potentials serve as the basis of a probabilistic extraction of pose constraints for tracking by using weighted sampling from a pose posterior that is guided by the model. In the final energy, we combine these constraints with an appearance-based model-to-image similarity term. Poses can be computed very efficiently using iterative local optimization, since joint detection with a trained ConvNet is fast, and since our formulation yields a combined pose estimation energy with analytic derivatives. In combination, this enables to track full articulated joint angles at state-of-the-art accuracy and temporal stability with a very low number of cameras. Our method is efficient and lends itself to implementation on parallel computing hardware, such as GPUs. We test our method extensively and show its advantages over related work on many indoor and outdoor data sets captured by ourselves, as well as data sets made available to the community by other research labs. The availability of good evaluation data sets is paramount for scientific progress, and many existing test data sets focus on controlled indoor settings, do not feature much variety in the scenes, and often lack a large corpus of data with ground truth annotation. We therefore further contribute with a new extensive test data set called MPI-MARCOonI for indoor and outdoor marker-less motion capture that features 12 scenes of varying complexity and varying camera count, and that features ground truth reference data from different modalities, ranging from manual joint annotations to marker-based motion capture results. Our new method is tested on these data, and the data set is made available to the community.

References


a low number of cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, Boston, MA, USA, 2015, pp. 3810–3818. IEEE.


### 36.4.4 Human Body Models

**Investigators: Leonid Pishchulin, Gaurav Sharma, Christian Theobalt and Bernt Schiele, in cooperation with Stefanie Wohrer, Cordelia Schmid (INRIA Grenoble Rhône-Alpes), Frédéric Jurie (Université de Caen Basse-Normandie), and Thomas Helten (GoalControl GmbH)**

Statistical models of 3D human shape and pose learned from scan databases have developed into valuable tools to solve a variety of vision and graphics problems. Unfortunately, most publicly available models are of limited expressiveness as they were learned on very small databases that hardly reflect the true variety in human body shapes. In this project we contribute by rebuilding a widely used statistical body representation from the largest commercially available scan database, and making the resulting model available to the scientific community.\(^1\) As preprocessing several thousand scans for learning the model is a challenge in itself, we contribute by developing robust best practice solutions for scan alignment that quantitatively lead to the best learned models. We make implementations of these preprocessing steps publicly available. We extensively evaluate the improved accuracy and generality of our new model, and show its improved performance for human body reconstruction from sparse input data. The results of our work are described in [1].

On the task of human attributes and action recognition, we introduce an Expanded Parts Model (EPM) for recognizing human attributes (e.g. young, short hair, wearing suits) and actions (e.g. running, jumping) in still images [2]. An EPM is a collection of part templates which are learnt discriminatively to explain specific scale-space regions in the images (in human centric coordinates). This is in contrast to current models which consist of a relatively few (i.e. a mixture of) ‘average’ templates. EPM uses only a subset of the parts to score an image and scores the image sparsely in space, i.e. it ignores redundant and random background in an image. To learn our model, we propose an algorithm which automatically mines parts and learns corresponding discriminative templates together with their respective locations from a large number of candidate parts. We validate our method on three recent

\(^1\)Please visit [https://humanshape.mpi-inf.mpg.de/](https://humanshape.mpi-inf.mpg.de/) for more details.
challenging datasets of human attributes and actions. We obtain convincing qualitative and state-of-the-art quantitative results on the three datasets.

References


36.4.5 Multi-Person Tracking

*Investigators: Siyu Tang, Mykhaylo Andriluka, Bjoern Andres, and Bernt Schiele*

Multi person tracking is a problem studied intensively in computer vision. While continuous progress has been made, false positive detections, long-term occlusions and camera motion remain challenging, especially for people tracking in crowded scenes. Tracking-by-detection is commonly used for multi person tracking where a state-of-the-art person detector is employed to generate detection hypotheses for a video sequence. In this case tracking essentially reduces to an association task between detection hypotheses across video frames. In this section, we propose a novel formulation for the multi people tracking task.

A common strategy for tracking is to first select hypotheses spatially and then to link these over time while maintaining disjoint path constraints. In crowded scenes multiple hypotheses will often be similar to each other making selection of optimal links an unnecessary hard optimization problem due to the sequential treatment of space and time.

Embracing this observation, we propose to link and cluster plausible detections jointly across space and time. Specifically, we state multi-target tracking as a Minimum Cost Lifted Multicut Problem [1].

This mathematical abstraction has several advantages. Firstly, the number of persons is not fixed or biased by definition of the problem, but is estimated in an unbiased fashion from the video sequence and is determined by the solution of the problem. Secondly, multiple detections of the same person in the same frame are effectively clustered, which eliminates the need for heuristic spatial non-maxima suppression.

Furthermore, in order to avoid that distinct but similar looking people are assigned to the same track, a distinction must be made between edges that define possible connections (i.e., a feasible set) and edges that define the costs or rewards for assigning the incident nodes to distinct tracks (i.e., an objective function). We introduce two types of edges (regular and lifted edges) into the graph. The regular edges define the set of feasible solutions in the graph, namely, which pair of nodes can be joint/cut. The lifted edges add additional long range information to the objective on which nodes should be joint/cut without modifying the set of feasible solutions. Our formulation encodes long-range information, yet penalizes long-term false joints (e.g., similar looking people) by forcing valid paths in the feasible solution in a unified and rigorous manner. We show that tracks defined by local optima of this optimization problem define a new state-of-the-art for the MOT16 benchmark.
36.5 Object Detection and Recognition

Coordinators: Shanshan Zhang and Bernt Schiele

This section focuses on recognizing and localizing object classes in images. We present research on general objects as well as its canonical case of people.

In Section 36.5.1, we introduce some work on learning 3D object representations with weak supervision. In Section 36.5.2, we investigate one component of typical detection pipelines to enable true end-to-end learning of object detectors.

One of the most important object classes is humans and the research group has a long tradition of research on pedestrian detection. Section 36.5.3 reports work on failure analysis and also corresponding solutions for major challenges, including occlusion and low illumination. In Section 36.5.4, we present a new method employing spatial attention for person re-identification.

36.5.1 3D Object Recognition

Investigators: Eldar Insafutdinov, Qianru Sun, David Stutz, Weipeng Xu, and Christian Theobalt in cooperation with Andreas Geiger (Max Planck Institute for Intelligent Systems and University of Tübingen), Alexey Dosovitskiy (Intel Labs), and Zaw Lin Kyaw, Tat-Seng Chua (National University of Singapore)

For many applications, e.g., path planning in robotic applications such as autonomous driving, object detection and recognition limited to the 2D image plain is not sufficient. In these cases, detailed knowledge about the object’s 3D shape and pose is essential. 3D reconstruction and pose estimation are, however, inherently ill-posed inverse problems where many configurations of shape, color, texture and lighting may result in the very same observations, e.g., images or point clouds. Still, learning-based approaches – especially deep neural networks applied to 3D voxel grids or point clouds – have demonstrated remarkable results in spite of the necessity of large datasets with ground truth shape and pose.

We model observations, i.e., images and point clouds, jointly with their corresponding shapes and poses and utilize weakly- as well as unsupervised approaches to overcome these challenges.

Learning 3D Shape Completion under Weak Supervision

In [9] and [10], we address the problem of 3D shape completion from sparse and noisy point clouds as illustrated in Fig. 36.3. Recent approaches are mostly data-driven or learning-based: Data-driven approaches rely on a shape model whose parameters are optimized to fit the observations; learning-based approaches, in contrast, avoid the expensive optimization step by learning to directly predict complete shapes from incomplete observations in a fully-supervised manner.
Figure 36.3: Left: 3D shape completion results on ShapeNet [14], KITTI [4], ModelNet [12] and Kinect [14] as presented in [9, 10]. Our approach utilizes synthetic shapes to learn a category-specific shape prior which can subsequently be used to learn shape completion without 3D supervision. Right: 3D reconstruction results from [7] on ShapeNet [14]. Our approach predicts both camera pose and 3D shape while only requiring 2D supervision; additionally, the used point cloud representation allows to predict high fidelity shapes.

setting. However, full supervision is often not available in practice. Instead, we propose a weakly-supervised but learning-based approach to 3D reconstruction from point clouds, avoiding both the slow optimization of data-driven approaches and the required supervision of learning-based approaches. Specifically, we first learn a shape prior on synthetic shapes using a denoising variational auto-encoder [8, 6]. 3D shape completion is then formulated as maximum likelihood problem; instead of maximizing the likelihood independently for distinct observations, however, we learn to directly predict maximum likelihood solutions. This is achieved by training a new encoder which embeds the point cloud observations in the same latent space using an unsupervised maximum likelihood loss. This allows to learn 3D reconstruction on challenging real-world datasets such as KITTI [4]. Based on two novel, synthetic benchmarks derived from ShapeNet [1] and ModelNet [12] as well as on KITTI and Kinect [14], we demonstrate that our approach outperforms data-driven [3] and rivals learning-based [2] approaches while significantly reducing inference time and required supervision.

Unsupervised Learning of Shape and Pose with Differentiable Point Clouds

In [7], we move from weakly-supervised to unsupervised 3D reconstruction. In particular, we jointly learn 3D shape and camera pose estimation by minimizing the reprojection error: given multiple views of an object, the projections of the reconstructed shape to the predicted camera pose should match the provided views. To this end, we train an ensemble of camera pose predictors, each represented as convolutional neural network, using the hindsight error [5], i.e., only penalizing the best network; this allows the networks to specialize on a subset of poses. In parallel, the ensemble is distilled into a student model, always trained using the best model from the ensemble. Shape is also predicted using a convolutional neural network; in contrast to our weakly-supervised approach [9, 10], however, we predict point clouds instead of voxel grids. The predicted point clouds can be projected to the predicted camera.
poses using our differentiable point cloud renderer. The predicted points are first transformed into the standard coordinate frame, using the predicted camera poses, represented as smooth functions and discretized into a 3D occupancy grid. Then, summation along the depth dimension results in a differentiable projection operation. The resulting projection can be compared to the provided views to obtain a signal to back-propagate. This formulation allows to train the model both on silhouette views as well as color images. On ShapeNet [1], we demonstrate that our approach outperforms related work [13, 11] both quantitatively and qualitatively; qualitative results are shown in Fig. 36.3.

References


36.5.2 Non-Maximum Suppression

Investigators: Jan Hosang, Rodrigo Benenson, and Bernt Schiele

Virtually all state-of-the-art object detection pipelines consist of three steps: 1) propose a set of windows (via sliding window or object proposals), 2) score each window via a trained classifier, 3) remove overlapping detections (non-maximum suppression). Both object proposals and detection classifiers have received enormous attention, while nonmaximum suppression (NMS) has been seldom addressed. NMS is also the only component that is not present at training time and is not learned at all.

In [1], we cast NMS as a rescoring problem and design a convnet that can replace the typical GreedyNMS. It operates on hand-designed features built from detection scores and overlaps between detection.

Furthermore, in [2], we remove this restriction by designing a network that operates on arbitrary detection features. It considers pairs of detections, their latent feature representation and geometrical properties to update their representation. Repeating this building block to update detection features several times allows us to improve over GreedyNMS on the challenging real-world datasets PETS and COCO, without using additional information. This architecture opens the door for joint training of detectors and NMS without a handcrafted feature representation between the two.

References


36.5.3 People Detection

Investigators: Shanshan Zhang, Rodrigo Benenson, Mohamed Omran, Jan Hosang and Bernt Schiele in cooperation with Jian Yang (Nanjing University of Science and Technology), and Lukas Neumann, Andrea Vedaldi, Andrew Zisserman (Univ. of Oxford), and Michelle Karg, Christian Scharfenberger, Eric Piegert, Sarah Mistr, Olga Prokofyeva, Robert Thiel (Continental)

People detection is considered as a canonical case of object detection, and has become a very popular research topic in the computer vision community, due to its wide applications in the industry, e.g. autonomous driving, video surveillance, robotics, etc.
We started with analyzing failure cases in [2] and have found out that a major challenge for people detection is occlusion. Therefore, we have made great efforts to improve the performance under heavy occlusion, including exploring channel-wise attention [4]. We have also proposed a new dataset namely CityPersons [3], which serves as a more diverse training and testing database by providing a large number of unique persons and occlusion cases. Moreover, we investigate particular challenges at night based on the NightOwls dataset [1], which consists of a large amount of video data captured by a RGB camera at night.

Analyzing and Improving Pedestrian Detectors

Encouraged by the rapid progress in pedestrian detection, in [2] we investigate the gap between current state-of-the-art detectors and the human baseline, which is considered as a “perfect single frame detector”. By manually grouping the failure cases, we characterize both localization and background-versus-foreground errors.

To address the localization errors, we study the impact of mis-aligned training annotations on the detector performance, and show that we can improve localization even with a small portion of aligned training data. As of the background-versus-foreground errors, we find convnets extract more representative features, and thus enhance the discrimination ability.

Other than our in-depth analysis, we report top results on the Caltech dataset, and also provide a new sanitized set of training and testing annotations.

CityPersons: A Diverse Dataset for Pedestrian Detection

Convnets have enabled significant progress in pedestrian detection recently, but there are still open questions regarding suitable architectures and training data. In [3], we revisit CNN design and point out key adaptations, enabling plain FasterRCNN to obtain state-of-the-art results on the Caltech dataset.

To achieve further improvement from more and better data, we introduce CityPersons, a new set of person annotations on top of the Cityscapes dataset. The diversity of CityPersons allows us for the first time to train one single CNN model that generalizes well over multiple benchmarks. Moreover, with additional training with CityPersons, we obtain top results using FasterRCNN on Caltech, improving especially for more difficult cases (heavy occlusion and small scale) and providing higher localization quality.

Occluded Pedestrian Detection through Guided Attention in CNNs

Pedestrian detection has progressed significantly in the last years. However, occluded people are notoriously hard to detect, as their appearance varies substantially depending on a wide range of occlusion patterns. In [4], we aim to propose a simple and compact method based on the FasterRCNN architecture for occluded pedestrian detection.

We start with interpreting CNN channel features of a pedestrian detector, and we find that different channels activate responses for different body parts respectively. These findings motivate us to employ an attention mechanism across channels to represent various occlusion patterns in one single model, as each occlusion pattern can be formulated as some specific combination of body parts. Therefore, an attention network with self or external guidances is proposed as an add-on to the baseline FasterRCNN detector. When evaluating on the heavy...
occlusion subset, we achieve a significant improvement of 8pp to the baseline FasterRCNN detector on CityPersons and on Caltech we outperform the state-of-the-art method by 4pp.

**Pedestrian Detection at Night**

Detecting pedestrians at night exhibits some particular challenges. In [1] we introduce a comprehensive public dataset, NightOwls, for pedestrian detection at night. In comparison to daytime conditions, pedestrian detection at night is more challenging due to variable and low illumination, reflections, blur, and changing contrast. NightOwls consists of 279k frames in 40 sequences recorded at night across 3 countries by an industry-standard camera, including different seasons and weather conditions. All the frames are fully annotated and contain additional object attributes such as occlusion, pose and difficulty, as well as tracking information to identify the same object across multiple frames.

As a baseline for pedestrian detection at night time, we compare the performance of ACF, Checkerboards, Faster R-CNN, RPN+BF, and SDS-RCNN. In particular, we demonstrate that state-of-the-art pedestrian detectors do not perform well at night, even when specifically trained on night data, and we show there is a clear gap in accuracy between day and night detections.

**References**


**36.5.4 Person Recognition**

*Investigators: Yue Fan and Bernt Schiele, in cooperation with Haoran Wang, Zexin Wang, and Licheng Jiao (Xidian University)*

The problem of person detection (see Sect. 36.5.3) involves localizing people in challenging scenes. In this section, we address the additional challenge of recognizing peoples’ identities.

Global average pooling (GAP) allows to localize discriminative information for recognition. While GAP helps the convolution neural network to attend to the most discriminative
features of an object, it may suffer if that information is missing e.g. due to camera viewpoint changes. To circumvent this issue, we argue that it is advantageous to attend to the global configuration of the object by modeling spatial relations among high-level features. In [1], we propose a novel architecture for Person Re-Identification, based on a novel parameter-free spatial attention layer introducing spatial relations among the feature map activations back to the model. Our spatial attention layer consistently improves the performance over the model without it. Results on four benchmarks demonstrate a superiority of our model over the state-of-the-art achieving rank-1 accuracy of 94.7% on Market-1501, 89.0% on DukeMTMC-ReID, 74.9% on CUHK03-labeled and 69.7% on CUHK03-detected.

References


36.6 Learning

Coordinator: Zeynep Akata

Our primary research focus here is learning with the lack of labeled training data, i.e. zero-shot learning and learning in scarcely labeled datasets, i.e. few-shot learning. Learning with limited labels has been an important field of research in computer vision and machine learning in the past couple of years as it is unrealistic to collect sufficient amounts of labeled data for every object. This challenge is not specific to image classification hence we aim for proposing solutions that are generalizable to other domains.

As a foundation, our group has proposed a benchmark that define a new benchmark by unifying both the evaluation protocols and data splits, compared a significant number of the state-of-the-art methods in depth, both in the classic zero-shot setting but also in the more realistic generalized zero-shot setting. The limitations discussed in this benchmark paper built the basis of the series of papers that follow this benchmark. For instance, to circumvent the need for labeled examples of unseen classes, we propose to synthesizes CNN features conditioned on class-level semantic information, offering a shortcut directly from a semantic descriptor of a class to a class-conditional feature distribution. We proposed different models that use Generative Adversarial Networks (GAN), Variational Autoencoders (VAE) and their combination.

Although, image classification is one of the most prominent applications in low shot learning, the potential applications are not limited to classification. We extended the field of application to zero-shot sketch-based image retrieval, i.e. an emerging task in computer vision, allowing to retrieve natural images relevant to sketch queries that might not been seen in the training phase, and zero-shot semantic image segmentation, both accomplished without seeing any labeled examples of the target classes, i.e. semantic image segmentation by assigning a label to every pixel even though either no or only a few labeled sample of that class was present during training.
36.6.1 Few-shot Learning

Investigators: Yongqin Xian, Qianru Sun, Saurabh Sharma, Subhabrata Choudhury, Zeynep Akata, Bernt Schiele in cooperation with Edgar Schoenfeld (UvA), Sayna Ebrahimi (UC Berkeley), Samarth Sinha (U Toronto), Trevor Darrell (UC Berkeley)

Deep learning has achieved great success in visual tasks. Despite its breakthrough, for best performance, training a deep neural network requires a large amount of labeled data which is costly and tedious. In almost all scenarios, there is an exponentially decay in terms of number of samples per class, i.e. only a few classes contain a large number of samples whereas most classes are sparsely populated. In this section, we show that compelling results can be achieved on categories with only few training examples. In [5], we learn a generative model which synthesizes CNN features from class-level semantic information, e.g. attributes; the proposed generative model combines the strength of VAE and GAN, and leverages unlabeled data to synthesize highly discriminative features that are well suited for augmenting classifier training. In [1], we propose a model where a shared latent space of image features and class embeddings is learned by modality-specific aligned variational autoencoders. In [2], we propose a few-shot learning method called meta-transfer which adapts a deep NN for few-shot learning tasks by learning scaling and shifting functions of DNN weights. In [3], we introduce zero- and few-label semantic segmentation tasks which aim to make pixel-wise prediction on the classes that have zero or only few labeled training examples; our proposed semantic projection network (SPNet) solves those tasks by mapping each pixel to the shared semantic word embedding space and transferring knowledge from base classes to novel classes.

Feature Generating Frameworks for Any-Shot Learning

When labeled training data is scarce, a promising data augmentation approach is to generate visual features of unknown classes using their attributes [4]. To learn the class conditional distribution of CNN features, these models [4] rely on pairs of image features and class attributes. Hence, they can not make use of the abundance of unlabeled data samples. In this paper, we tackle any-shot learning problems i.e. zero-shot and few-shot, in a unified feature generating framework that operates in both inductive and transductive learning settings. We develop a conditional generative model that combines the strength of VAE and GANs and in addition, via an unconditional discriminator, learns the marginal feature distribution of unlabeled images. We empirically show that our model learns highly discriminative CNN features for five datasets, i.e. CUB, SUN, AWA and ImageNet, and establish a new state-of-the-art in any-shot learning, i.e. inductive and transductive (generalized) zero- and few-shot learning settings. We also demonstrate that our learned features are interpretable: we visualize them by inverting them back to the pixel space and we explain them by generating textual arguments of why they are associated with a certain label.

Our main contributions in this work are as follows. (1) We propose the $f$-VAEGAN-D2 model that consists of a conditional encoder, a shared conditional decoder/generator, a conditional discriminator and a non-conditional discriminator. The first three networks aim to learn the conditional distribution of CNN image features given class embeddings optimizing VAE and WGAN losses on labeled data of seen classes. The last network
learns the marginal distribution of CNN image features on the unlabeled features of novel classes. Once trained, our model synthesizes discriminative image features that can be used to augment softmax classifier training. (2) Our empirical analysis on CUB, AWA2, SUN, FLO, and large-scale ImageNet shows that our generated features improve the state-of-the-art in low-shot settings, i.e (generalized) zero- and few shot learning in both the inductive and transductive settings. (3) We demonstrate that our generated features are interpretable by inverting them back to the raw pixel space and by generating visual explanations.

On the other hand, many approaches in generalized zero-shot learning rely on cross-modal mapping between the image feature space and the class embedding space. As labeled images are rare, one direction is to augment the dataset by generating either images or image features. However, the former misses fine-grained details and the latter requires learning a mapping associated with class embeddings. In [1], we take feature generation one step further and propose a model where a shared latent space of image features and class embeddings is learned by modality-specific aligned variational autoencoders. This leaves us with the required discriminative information about the image and classes in the latent features, on which we train a softmax classifier. The key to our approach is that we align the distributions learned from images and from side-information to construct latent features that contain the essential multi-modal information associated with unseen classes. Our results on ImageNet with various zero-shot splits show that our latent features generalize well in large-scale settings.

Our contributions in this work are as follows. (1) We propose a model that learns shared cross-modal latent representations of multiple data modalities using simple VAEs via distribution alignment and cross alignment objectives. (2) We extensively evaluate our model using conventional benchmark datasets, i.e. CUB, SUN, AWA1 and AWA2, developed for zero-shot learning and extended to few-shot learning. Our model establishes the new state-of-the-art performance on generalized zero-shot and few-shot learning settings on all these datasets. Furthermore, we show that our model can be extended easily to more than two modalities trained simultaneously. (3) Finally, we show that the latent features learned by our model improve the state of the art in the truly large-scale ImageNet dataset in all splits for the generalized zero-shot learning task.

Meta-Transfer Learning for Few-Shot Learning

Apart from generating synthesized features for novel classes [5], an effective alternative method for the few-shot learning problem is meta-learning. The key idea is to leverage a large number of similar few-shot tasks in order to learn how to adapt a base-learner to a new task for which only a few labeled samples are available. As deep neural networks (DNNs) tend to overfit using a few samples only, meta-learning typically uses shallow neural networks (SNNs), thus limiting its effectiveness. In this paper, we propose a novel few-shot learning method called meta-transfer learning (MTL) [2] which learns to adapt a deep NN for few-shot learning tasks. Specifically, “meta” refers to training multiple tasks, and “transfer” is achieved by learning Scaling and Shifting functions of DNN weights for each task. In addition, we introduce the hard task (HT) meta-batch scheme as an effective learning curriculum for MTL. We conduct experiments using (5-class, 1-shot) and (5-class,
Figure 36.4: We propose (generalized) zero- and few-label semantic segmentation tasks, i.e. segmenting classes whose labels are not seen by the model during training or the model has a few labeled samples of those classes. To tackle these tasks, we propose a model that transfers knowledge from seen classes to unseen classes using side information, e.g. semantic word embedding trained on free text corpus.

5-shot) recognition tasks on two challenging few-shot learning benchmarks: miniImageNet and Fewshot-CIFAR100. Extensive comparisons to related works validate that our meta-transfer learning approach trained with the proposed HT meta-batch scheme achieves top performance. An ablation study also shows that both components contribute to fast convergence and high accuracy.


Another important visual task where few-shot learning can be beneficial is semantic segmentation because pixel-level labelling in this context is particularly expensive. There have been several attempts to reduce the annotation effort such as learning from image level labels and bounding box annotations. In this paper we take this one step further and focus on the challenging task of zero- and few-label learning of semantic segmentation. As shown in Figure 36.4, we define this task as image segmentation by assigning a label to every pixel even though either no labeled sample of that class was present during training, i.e. zero-label semantic segmentation, or only a few labeled samples were present, i.e. few-label semantic segmentation. Our goal is to transfer the knowledge from previously seen classes to novel classes. Our proposed semantic projection network (SPNet) achieves this goal by incorporating class-level semantic information into any network designed for semantic segmentation, in an end-to-end manner. We also propose a benchmark for this task on the challenging COCO-Stuff and PASCAL VOC12 datasets. Our model is effective in segmenting
novel classes, i.e. alleviating expensive dense annotations, but also in adapting to novel classes without forgetting its prior knowledge, i.e. generalized zero- and few-label semantic segmentation.

References


36.6.2 Zero-shot Learning

Investigators: Yongqin Xian, Zeynep Akata, Bernt Schiele, in collaboration with Christoph Lampert (IST Austria), Anjan Dutta (UBC), Sarkhan Badirli (Purdue University), Murat Dundar (Purdue University)

Human are highly capable of recognizing novel objects categories using some form of external information, without seeing any actual visual example of that category. Enabling computers with this capability has been introduced as zero-shot learning task in the intersection of computer vision and machine learning. Zero-shot learning can been formally posed as follows: labeled images are provided for certain visual classes during training and the task is to learn a model that can make predictions for novel classes at test time where auxiliary side information is given to associate seen and unseen classes. In this section, we show our contributions to this area by establishing a new benchmark and developing novel methods. In [4] and [2], we introduce a new zero-shot learning benchmark and show the state-of-the-art by evaluating influential zero-shot learning approaches under the same evaluation protocol. In [3], we tackle the challenging generalized zero-shot learning by generating CNN features of unseen classes. In [1], we propose a semantically aligned paired cycle-consistent generative (SEM-PCYC) model for zero-shot sketch based image retrieval.

Zero-shot learning – The Good, the Bad and the Ugly

Due to the importance of zero-shot learning, the number of proposed approaches has increased steadily recently. We argue that it is time to take a step back and to analyze the
Figure 36.5: Our $f$-CLSWGAN: we propose to minimize the classification loss over the generated features and the Wasserstein distance with gradient penalty.

status quo of the area. The purpose of this work [4] is three-fold. First, given the fact that there is no agreed upon zero-shot learning benchmark, we first define a new benchmark by unifying both the evaluation protocols and data splits. This is an important contribution as published results are often not comparable and sometimes even flawed due to, e.g. pre-training on zero-shot test classes. Second, we compare and analyze a significant number of the state-of-the-art methods in depth, both in the classic zero-shot setting but also in the more realistic generalized zero-shot setting. Finally, we discuss limitations of the current status of the area which can be taken as a basis for advancing it.

In the journal extension of [2] we made the following changes. First, we introduce Animal with Attributes 2 (AWA2) dataset: it inherits the same 50 classes and attributes annotation from the original Animal with Attributes (AWA), but consists of 37,322 different images accompanied with publicly available licenses. Second, we additionally evaluate a group of generative zero-shot learning approaches and study the transductive zero-shot learning setting where unlabeled images from unseen classes are available. Third, we show and analysis the results of zero-shot and generalized zero-shot learning on the large-scale ImageNet. Finally, we provide a more comprehensive review of the area by significantly extending the related works.

Feature Generating Networks for Zero-Shot Learning

One of the key observations from our benchmark papers [4, 2] is that most of existing state-of-the-art approaches fail to achieve satisfactory results for the challenging generalized zero-shot learning task, which can be explained by the extreme training data imbalance between seen and unseen classes. To circumvent the need for labeled examples of unseen classes, we propose a novel generative adversarial network (GAN) that synthesizes CNN features conditioned on class-level semantic information, offering a shortcut directly from a semantic descriptor of a
class to a class-conditional feature distribution. Our proposed approach, pairing a Wasserstein GAN with a classification loss, is able to generate sufficiently discriminative CNN features to train softmax classifiers or any multimodal embedding method. The overall architecture of our generative model i.e. f-CLSWGAN, is shown in Figure 36.5. Our experimental results demonstrate a significant boost in accuracy over the state of the art on five challenging datasets – CUB, FLO, SUN, AWA and ImageNet – in both the zero-shot learning and generalized zero-shot learning settings.

In this work, we argue that the zero-shot learning scenario is a great testbed for evaluating the robustness and generalization of generative models. In particular, if the generator learns discriminative visual data with enough variation, the generated data should be useful for supervised learning. Hence, one contribution of our paper is a comparison of various existing GAN-models and another competing generative model, i.e. GMMN, for visual feature generation. In particular, we look into both zero-shot learning (ZSL) where the test time search space is restricted to unseen class labels and generalized zero-shot learning (GZSL) for being a more realistic scenario as at test time the classifier has to decide between both seen and un- seen class labels. In this context, we propose a novel GAN-method, namely f-CLSWGAN that generates features instead of images and is trained with a novel loss improving over alternative GAN-models.

We summarize our contributions as follows. (1) We propose a novel conditional generative model f-CLSWGAN that synthesizes CNN features of unseen classes by optimizing the Wasserstein distance regularized by a classification loss. (2) Across five datasets with varying granularity and sizes, we consistently improve upon the state of the art in both the ZSL and GZSL settings. We demonstrate a practical application for adversarial training and propose GZSL as a proxy task to evaluate the performance of generative models. (3) Our model is generalizable to different deep CNN features, e.g. extracted from GoogleNet or ResNet, and may use different class-level auxiliary information, e.g. sentence, attribute, and word2vec embeddings.

Semantically Tied Paired Cycle Consistency for Zero-Shot SBIR

Zero-shot sketch-based image retrieval (SBIR) is an emerging task in computer vision, allowing to retrieve natural images relevant to sketch queries that might not been seen in the training phase. Existing works either require aligned sketch-image pairs or inefficient memory fusion layer for mapping the visual information to a semantic space. In [1], we propose a semantically aligned paired cycle-consistent generative (SEM-PCYC) model for zero-shot SBIR, where each branch maps the visual information to a common semantic space via an adversarial training. Each of these branches maintains a cycle consistency that only requires supervision at category levels, and avoids the need of highly-priced aligned sketch-image pairs. A classification criteria on the generators’ outputs ensures the visual to semantic space mapping to be discriminating. Furthermore, we propose to combine side information from text-based and hierarchical models via a feature selection auto-encoder that selects discriminating side information within a same end-to-end model. Our results demonstrate a significant boost in zero-shot SBIR performance over the state-of-the-art on the challenging Sketchy and TU-Berlin datasets.

The main contributions of this work are: (1) We propose the SEM-PCYC model for
zero-shot SBIR task, that maps sketch and image features to a common semantic space with the help of adversarial training. The cycle consistency constraint on each branch of the SEM-PCYC model facilitates bypassing the requirement of aligned sketch image pairs. (2) Within a same end-to-end framework, we combine different side information via a feature selection guided auto-encoder which effectively choose side information that minimizes intra-class variance and maximizes inter-class variance. (3) We evaluate our model on two datasets (Sketchy and TU-Berlin) with varying difficulties and sizes, and provide an experimental comparison with latest models available for the same task, which further shows that our proposed model consistently improves the state-of-the-art results of zero-shot SBIR on both datasets.

References


36.6.3 Multilabel and Weakly Supervised Learning

*Investigators: Maksim Lapin, Gaurav Sharma, Bernt Schiele, in cooperation with Matthias Hein (Univ. of Saarland) and Karan Sikka (SRI International)*

While the previous sections addressed learning in scarcity of labeled datasets, we now look at challenges posed by weakly and ambiguously labeled data.

**Weakly Supervised Learning for Video Classification**

In [2], we address the problem of video classification for facial analysis and human action recognition. We propose a novel weakly supervised learning method that models the video as a sequence of automatically mined, discriminative sub-events (e.g. onset and offset phase for “smile”, running and jumping for “highjump”). The proposed model is inspired by the recent works on Multiple Instance Learning and latent SVM/HCRF – it extends such frameworks to model the ordinal aspect in the videos, approximately. We obtain consistent improvements over relevant competitive baselines on four challenging and publicly available video based facial analysis datasets for prediction of expression, clinical pain and intent in dyadic conversations, and on three challenging human action datasets. We also validate the method with qualitative results and show that they largely support the intuitions behind the method.
Multiclass, Top-k, and Multilabel Learning

In [1], we suggest extensions of the established multiclass loss functions to address top-k error minimization. In particular, we propose smooth top-k SVM and two top-k versions of the softmax loss, top-k entropy and truncated top-k entropy. We also introduce the notion of classification calibration for the top-k error, and analyze which of the multiclass methods are top-k calibrated.

We recognize a close relationship between multilabel methods, label ranking, and top-k classification in the multiclass setting. Furthermore, we provide a review of the employed performance metrics and the surrogate loss functions, where we also propose a smoothed version of the multilabel SVM loss.

Finally, we discuss efficient optimization schemes based on SDCA and contribute a set of novel algorithms for computing the proximal maps that can be used to train classifiers with the considered multiclass, top-k, and multilabel loss functions.

An extensive empirical evaluation on multiclass and multilabel benchmarks leads to interesting insights, including the overall competitive performance of the softmax loss as well as the improvements in top-k error with the suggested surrogate losses.

References


36.7 Privacy and Security

Coordinators: Mario Fritz and Bernt Schiele

As techniques from machine learning and computer vision continue to mature, many of these approaches have made their way into products and systems. With this transition, they also have an increasing societal impact and contribute to the overall attack surface of our IT infrastructure. Not surprisingly, we see an increasing risk of attacks on such systems that threaten privacy and security. On the one hand, advanced image and text analysis algorithm acquire a detailed understanding of multi-modal content and thereby also infer private information of us. We seek a better understanding of private information in these modalities as well as provide methods to inform users about privacy risk and approaches how to mitigate them. On the other hand, future intelligent system are subject to security risk as they have shown to be susceptible to inference attacks and manipulations. Evasion attacks by adversarial perturbations have shown how the outcome of classifications can be tempered with by inducing small, imperceptible perturbations onto the image data and inference attacks can lead to leakage of sensitive information on models and the associated intellectual property.
We find that these question-answer pairs tend to appear in- 
none of them can succeed in fooling the victim VQA model. 
answer pairs, we have also tried other attack methods and 
nor the CW attack can succeed. In fact, for these question-
there exist a few question-answer pairs where neither ours 
provide more evidence to confirm this hypothesis.

We also observe that the attack success rate with respect to 
tack generated on the N2NMN model is significantly lower 
than the MCB model. This further shows that N2NMN 
also take adversarial perturbations and the generalizability.

**36.7.1 Adversarial Perturbations**

Investigators: Mario Fritz, Seong Joon Oh, Anna Rohrbach, Bernt Schiele, David Stutz, and Edgar Tretschk, in cooperation with Xinyun Chen (University of California, Berkeley), Trevor Darrell (University of California, Berkeley), Matthias Hein (Saarland University and University of Tübingen), Chang Liu (Stanford University), Dawn Song (University of California, Berkeley), and Xiaojun Xu (University of Illinois at Urbana-Champaign)

Adversarial perturbations are small (often imperceptible) modifications on the input image that leads to a greatly altered output on a learned model. It takes only an unnoticeable modification on the image to completely fool the system. They pose potentially grave threats to security- and safety-critical systems, such as self-driving cars and automatic medical operations. As worst-case assessment of threats is crucial for those applications, the research field has investigated diverse types of perturbations applicable to different tasks and under various constraints. This subsection presents our contributions on extending the threat models from traditional classifier setups to either a multi-modal (vision and language) model or an agent acting in an environment to accumulate rewards. We then present adversarial perturbations designed for privacy protection against automatic person recognition systems. We wrap up the subsection with our studies on the relationship between a model’s robustness to adversarial perturbations and the generalizability.

**Fooling Vision and Language Models Despite Localization and Attention Mechanism**

While adversarial attacks are known to succeed on classifiers, it has been an open question whether more complex vision systems are vulnerable. In this work, we study adversarial examples for vision and language models, which incorporate natural language understanding and complex structures such as attention, localization, and modular architectures [9]. Specifically, we investigate targeted attacks on a dense captioning model and two visual question answering (VQA) models. Our evaluation shows that we can generate adversarial examples with a high success rate (i.e., > 90%) for these models. Moreover, we show that attention, bounding box localization, and compositional internal structures are vulnerable to adversarial attacks. In Figure 36.6 we demonstrate the attention heatmaps for two source
Sequential Attacks on Agents for Long-Term Adversarial Goals

Deep reinforcement learning is increasingly important for real-world tasks like autonomous driving. For such tasks, security is crucial. However, neural networks are known to be vulnerable to adversarial attacks that end up manipulating the victim network’s output. To properly defend against such attacks in a deep reinforcement learning setting, we first need to be aware of potential attacks.

In [6], we show that an adversary can be trained to control a deep reinforcement learning agent. Our technique works on fully trained victim agents and makes them pursue an alternative, adversarial goal when under attack. See Figure 36.7 for a schematic. In contrast to traditional attacks on e.g. image classifiers, our setting involves adversarial goals that may not be immediately reachable but instead may require multiple steps to be achieved. Thus, the adversary needs to plan for the long-term future by optimizing for the accumulative future adversarial reward. Since this is the problem tackled by deep reinforcement learning, our solution can rely on established techniques like Deep Q Networks. At training time, a feed-forward adversary learns to manipulate the victim agent’s input using barely perceptible perturbations. This avoids the need for a time-consuming iterative generation of adversarial

---

Figure 36.7: Rewards pursued by unaffected (left) versus attacked (right) agents. When the perturbation module (ATN) is inserted at the input stream, the agent is misguided to follow actions that lead to increasing adversarial reward \( r^A \) over time, rather than the original reward \( r^O \).
Figure 36.8: Adversarial perturbations can protect privacy. We propose a game theoretic framework for dealing with the uncertain recognition system.

Adversarial Image Perturbation for Privacy Protection – A Game Theory Perspective

While adversarial perturbations are often studied in the context of attacks on security-critical applications, it is also possible to exploit them for user privacy protection. Users like sharing personal photos with others through social media. At the same time, they might want to make automatic identification in such photos difficult or even impossible. Classic obfuscation methods such as blurring are not only unpleasant but also not as effective as one would expect. Recent studies on adversarial perturbations suggest that it is possible to confuse recognition systems effectively without unpleasant artifacts. However, in the presence of counter measures against adversarial perturbations, it is unclear how effective they would be in particular when the choice of counter measure is unknown. Game theory provides tools for studying the interaction between agents with uncertainties in the strategies. [3] introduces a general game theoretical framework for the user-recogniser dynamics (Figure 36.8), and present a case study that involves current state of the art adversarial perturbations and person recognition techniques. We derive the optimal strategy for the user that assures an upper bound on the recognition rate independent of the recogniser’s counter measure.

Disentangling Adversarial Robustness and Generalization

Obtaining deep networks that are robust against adversarial examples and generalize well is an open problem. A recent hypothesis [5, 7] even states that both robust and accurate models are impossible, i.e., adversarial robustness and generalization are conflicting goals. In an effort to clarify the relationship between robustness and generalization, [4] studies adversarial robustness in the context of the underlying, low-dimensional data manifold and explicitly distinguishes between regular, unconstrained adversarial examples and adversarial examples constrained to the manifold, so-called on-manifold adversarial examples, as illustrated in Figure 36.9. Based on this distinction, it is shown that 1. regular adversarial examples
leave the manifold; 2. adversarial examples constrained to the manifold, i.e., on-manifold adversarial examples, exist; 3. on-manifold adversarial examples are generalization errors, and on-manifold adversarial training boosts generalization; 4. and regular robustness is independent of generalization. These claims are confirmed through extensive experiments on synthetic data with access to the true manifold as well as on EMNIST [1], Fashion-MNIST [8] and CelebA [2] with approximated manifolds. Our findings imply that both robust and accurate models are possible, i.e. adversarial robustness and generalization are not contradictory objectives. However, different models (architectures, training strategies etc.) can exhibit different robustness and generalization characteristics.

References


36.7.2 Attacks and Defenses for Machine Learning

Investigators: Max Augustin, Mario Fritz, Seong Joon Oh, Tribhuvanesh Orekondy, Bernt Schiele, and Ning Yu, in cooperation with Michael Backes (CISPA), Larry Davis (U. Maryland), Kathrin Grosse (CISPA), Lucjan Hanzlik (CISPA), Ahmed Salem (CISPA) and Yang Zhang (CISPA)

Machine Learning (ML) models and especially deep neural networks are deployed to improve productivity or experience e.g., photo assistants in smartphones, image recognition APIs in cloud-based internet services, and for navigation and control in autonomous vehicles. Here, we investigate the privacy and security implications of such deployments.

Large-scale deployments of deep learning models in the wild has motivated us to ask: can someone abuse the model solely based on blackbox access? In particular, we first address model stealing, wherein an adversary aims to infer details of a ‘victim’ blackbox ML model. Such attacks circumvent the victim’s time, money, and human effort – ranging from collecting a massive annotated dataset to tuning the right model for the task. To this end, via a sequence of blackbox queries to the victim model, we studied stealing various aspects: attributes [2] (e.g., architecture, optimisation procedure) and functionality [3]. Remarkably, we found our approaches to be highly effective. Our approach kennen-io [2] can consistently infer various aspects of the victim model architecture, with up to 99.5% accuracy (or $3.6 \times$ chance-level). Furthermore, we also find the functionality of the victim can be effectively transferred into a ‘knockoff’ [3] assuming no knowledge of the model attributes nor the training data; we find the knockoffs recover $0.81-0.96 \times$ victim’s performance across a diverse range of datasets.

To mitigate various forms of attacks, we additionally explore defenses. To this end, we first propose ‘MLCapsule’ [1], a guarded offline deployment of machine learning as a service. MLCapsule executes the ML model locally on the user’s client and therefore the data never leaves the client. Meanwhile, MLCapsule offers the service provider the same level of control and security of its model as the commonly used server-side execution. In addition, MLCapsule is applicable to offline applications that require local execution. Beyond protecting against direct model access, we demonstrate that MLCapsule allows for implementing defenses against advanced attacks on machine learning models such as model stealing/reverse engineering and membership inference. While MLCapsule proposes countermeasures against traditional ‘inference attacks’, we in addition make a step towards discerning fake images. In [4], we ask: if, and to what extent, a generated fake image can be attributed to a particular Generative Adversarial Networks (GANs) of a certain architecture and trained with particular data and random seed. Our analysis shows single samples from GANs carry highly characteristic fingerprints which make attribution of images to GANs...
possible. Surprisingly, this is even possible for GANs with same architecture and same training that only differ by the training seed.

To summarize, in this direction, we highlighted vulnerabilities [2, 3] of ML models deployed ‘in the wild’, such as by model stealing attacks. To mitigate attacks and thereby taking a step towards safe deployments of ML models, we also investigated various countermeasures [1, 4, 3]. We hope the techniques studied here enables practitioners to understand risks and implement methods to defend against various attacks in an increasingly ML-driven world.

References


36.7.3 Enforcing Privacy

Investigators: Mario Fritz, Seong Joon Oh, Tribhuvanesh Orekondy, Bernt Schiele, Qianru Sun, Ayush Tewari, Christian Theobalt, and Weipeng Xu in cooperation with Liqian Ma (KU-Leuven/PSI), and Luc Van Gool (KU-Leuven/PSI and ETH Zürich)

A massive amount of data is uploaded to the Internet everyday, often containing a broad range of private information. It is becoming increasingly crucial to aid users enforce their privacy preferences on such data. In this line of work, we propose techniques which allows identifying and controlling these aspects across data spanning multiple modalities.

Obfuscating Private Pixels

We first look at techniques which allow users enforce privacy in visual data via obfuscation. In [2], we propose a method which identifies a broad range of private information (e.g., physical disability, location) and in addition, enforce privacy by automatically redacting corresponding pixels by blacking them out. We find our approach is effective at achieving various privacy-utility trade-offs within 83% of the performance of redactions based on ground-truth annotation.

While [2] obfuscates by blacking-out relevant pixels, we also propose replacement as an alternate strategy. In particular, we study novel head inpainting approaches [5, 6] which replaces the region by synthesizing corresponding pixels and thereby, obfuscating the identity of the depicted person. In [5], we approach the task by generating facial landmarks (from image context), followed by head inpainting conditioned on the landmarks. We extend this by introducing a parametric face model [6] to overcome unfitting artifacts from a lack of
controllability in [5]. To this end, we first replace the identity related information of the depicted person, followed by synthesizing the complete head given the rendered face and an obfuscated region around the head as conditional inputs. We find both these approaches significantly outperform baseline methods on publicly available datasets with low recognition rate and high image quality.

User Attribution

While we previously discussed protecting user identity revealed in visual data, we now focus on other modalities. First, we observe an adversary can identify privacy relevant attributes such as gender or age based on a natural language text’s author or user. In [4], we propose an automatic method called Adversarial Author Attribute Anonymity Neural Translation (A\textsuperscript{4}NT) to combat such text-based adversaries. A\textsuperscript{4}NT involves a sequence-to-sequence language model and an adversarial training framework to transform the input text to obfuscate author attributes without paired data. We find our method learns to make minimal changes to the input to successfully fool author attribute classifiers, while preserving semantics of the input text.

We now move the focus to another mode which compromises users’ identity: model updates resulting from distributed training of ML models. Federated Learning methods [1] enables users to train machine learning (ML) models by sharing model parameter updates computed locally on clients (e.g., mobile phones), while the raw private data on clients is never shared. To truly preserve users’ privacy, it is crucial that the model updates do not inadvertently encode user-specific information. However, our study [3] indicate otherwise. We perform linkability attacks on model updates and find them to be highly effective – achieving 20\times-175\times chance-level performance. To mitigate them, we propose strategies based on augmenting with domain-specific data; we find that these strategies offer substantial protection against linkability threats with little effect to training utility.

References


36 D2: Computer Vision and Multimodal Computing


36.7.4 Understanding Privacy

*Investigators: Seong Joon Oh, Qianru Sun, Tribhuvanesh Orekondy, Rodrigo Benenson, Mario Fritz and Bernt Schiele*

Today, major part of our social life is captured via social media. As we communicate through multi-modal channels such as Facebook or Twitter, we leave traces that explicitly and implicitly capture people’s biological profile [2, 3], social relations [5] and other privacy-related attributes [4] in texts, blogs, images and video. As we are approaching a future, where intelligent and potential autonomous systems become our assistants and coworkers, we not only want them to be proficient at their task, but also enable them to blend in and act appropriately in different situations of our – human – life. Additionally, through better understanding about such privacy-related hidden information we would like to inform users about potential privacy risks.

**Person Recognition**

Using machines to automatically recognize people in personal photos may greatly enhance user convenience by easing photo album organization. For human identification task, however, traditional focus of computer vision has been face recognition and pedestrian re-identification. Person recognition in social media photos sets new challenges for computer vision, including non-cooperative subjects (e.g. backward viewpoints, unusual poses) and great changes in appearance. To tackle this problem, we build a simple person recognition framework that leverages convnet features from multiple image regions (head, body, etc.) [2, 3]. We propose new recognition scenarios that focus on the time and appearance gap between training and testing samples. We present an in-depth analysis of the importance of different features according to time and viewpoint generalizability. In the process, we verify that our simple approach achieves the state of the art result on the PIPA benchmark, arguably the largest social media based benchmark for person recognition to date with diverse poses, viewpoints, social groups, and events.

**Social Relation Recognition**

Social relations are the foundation of human daily life. Developing techniques to analyze such relations from visual data bears great potential to build machines that better understand us and are capable of interacting with us at a social level. Previous investigations have remained partial due to the overwhelming diversity and complexity of the topic and consequently have only focused on a handful of social relations. In [5], we argue that the domain-based theory from social psychology is a great starting point to systematically approach this problem. The theory provides coverage of all aspects of social relations and equally is concrete and predictive about the visual attributes and behaviors defining the relations included in each domain. We provide the first dataset built on this holistic conceptualization of social life.
Figure 36.10: We investigate the recognition of social relations in a domain-based approach. Our study is based on Bugental’s social psychology theory [1] that partitions social life into 5 domains from which we derive 16 social relations.

that is composed of a hierarchical label space of social domains and social relations (see Figure 36.10). We also contribute the first models to recognize such domains and relations and find superior performance for attribute based features. Beyond the encouraging performance of the attribute based approach, we also find interpretable features that are in accordance with the predictions from social psychology literature. Beyond our findings, we believe that our contributions more tightly interleave visual recognition and social psychology theory that has the potential to complement the theoretical work in the area with empirical and data-driven models of social life.

**Understanding and Predicting Privacy Risks**

Except user identity and social relation [2, 3, 5], other explicit content, such as GPS data, devices (e.g. mobile phones) as well as web services (e.g. Facebook) offer to set privacy settings in order to enforce the users’ privacy preferences. In [4], we propose the first approach that extends this concept to image content in the spirit of a Visual Privacy Advisor. First, we categorize personal information in images into 68 image attributes and collect a dataset, which allows us to train models that predict such information directly from images. Second, we run a user study to understand the privacy preferences of different users w.r.t. such attributes. Third, we propose models that predict user specific privacy score from images in order to enforce the users’ privacy preferences. Our model is trained to predict the user specific privacy risk and even outperforms the judgment of the users, who often fail to follow their own privacy preferences on image data.

**References**


36.8 Scalable Learning and Perception

Coordinator: Mario Fritz

With the advances of deep learning techniques, recent success has spurred the hope to address more holistic challenges that encompass not only scene understanding but also include light, surfaces, physics and multiple agents. While such complex tasks – in particular graphics and physics simulation – have traditionally been the domain of model-based approaches, we have seen increased success in a data-driven paradigm. In particular, as such tasks often involve complex underdetermined problems, modeling and resolving uncertainty is a key issue. In order to make progress towards these broader challenges, we research advanced machine learning approaches that model the interaction between surfaces and light and thereby give rise to new approaches to estimate and recreate surface appearances as well as environmental light. Equally, we seek learning techniques that can successfully approximate physical phenomena, e.g. for use in manipulation scenarios. In order to equip deep learning techniques with the ability to reason under uncertainty, we investigate novel approaches the phrase deep learning techniques as bayesian model and account for different types of uncertainty.

36.8.1 Deep Learning for Graphics and Physics

Investigators: Mario Fritz, Anna Rohrbach, Maxim Maximov, Rakshith Shetty, Ning Yu in cooperation with Sayna Ebrahimi, Trevor Darrell (UC Berkeley), Stamatios Georgoulis, Tinne Tuytelaars (KU Leuven), Luc Van Gool (ETHZ), Tobias Ritschel (UCL), Konstantinos Rematas (University of Washington), Efstratios Gavves (University of Amsterdam), Jeanette Bohg (Stanford), Ales Leonardis (University of Birmingham)

The success of recent machine learning techniques has had strong impact beyond “simple” recognition problems. We investigate data-driven solutions to complex problems such as control, modeling physical events or interactions with light and surfaces.

Machine Learning for Content Manipulation

Taking an image of an object is at its core a lossy process. The rich information about the three-dimensional structure of the world is flattened to an image plane and decisions such as viewpoint and camera parameters are final and not easily reversible. As a consequence, possibilities of changing viewpoint are limited. Given a single image depicting an object,
novel-view synthesis is the task of generating new images that render the object from a different viewpoint than the one given. The main difficulty is to synthesize the parts that are disoccluded; disocclusion occurs when parts of an object are hidden by the object itself under a specific viewpoint. We show how to improve novel-view synthesis by making use of the correlations observed in 3D models and applying them to new image instances. We propose a technique to use the structural information extracted from a 3D model that matches the image object in terms of viewpoint and shape [7]. For the latter part, we propose an efficient 2D-to-3D alignment method that associates precisely the image appearance with the 3D model geometry with minimal user interaction. Our technique is able to simulate plausible viewpoint changes for a variety of object classes within seconds (Fig. 36.11). Additionally, we show that our synthesized images can be used as additional training data that improves the performance of standard object detectors.

While great progress has been made recently in automatic image manipulation, it has been limited to object-centric images like faces or structured scene datasets. We take a step towards general scene-level image editing by developing an automatic interaction-free object removal model [8]. Our model learns to find and remove objects from general scene images using image-level labels and unpaired data in a generative adversarial network (GAN) framework. We achieve this with two key contributions: a two-stage editor architecture consisting of a mask generator and image in-painter that co-operate to remove objects, and a novel GAN based prior for the mask generator that allows us to flexibly incorporate knowledge about object shapes. We experimentally show on two datasets that our method effectively removes a wide variety of objects using weak supervision only (Fig. 36.12).

Finally, we address the problem of interpolating visual textures [9]. We formulate the problem of texture interpolation by requiring (1) by-example controllability and (2) realistic and smooth interpolation among an arbitrary number of texture samples. To solve it we propose a neural network trained simultaneously on a reconstruction task and a generation
Machine Learning for Modeling the Interaction of Surfaces and Light

We also present a method that estimates reflectance and illumination information from a single image depicting a single-material specular object from a given class under natural illumination [3]. We follow a data-driven, learning-based approach trained on a very large dataset, but in contrast to earlier work we do not assume one or more components (shape, reflectance, or illumination) to be known. We propose a two-step approach, where we first estimate the object’s reflectance map, and then further decompose it into reflectance and illumination. For the first step, we introduce a Convolutional Neural Network (CNN) that directly predicts a reflectance map from the input image itself, as well as an indirect scheme that uses additional supervision, first estimating surface orientation and afterwards inferring the reflectance map using a learning-based sparse data interpolation technique. For the second step, we suggest a CNN architecture to reconstruct both Phong reflectance parameters and high-resolution spherical illumination maps from the reflectance map. We also propose new datasets to train these CNNs. We demonstrate the effectiveness of our approach for both steps by extensive quantitative and qualitative evaluation in both synthetic and real data as well as through numerous applications, that show improvements over the state-of-the-art (Fig. 36.13).

We have extended this line of research to an approach with can deal with multiple materials and asked more generally “How much does a single image reveal about the environment it was taken in?” [2]. We investigate how much of that information can be retrieved from a foreground object, combined with the background (i.e. the visible part of the environment). Assuming it is not perfectly diffuse, the foreground object acts as a complexly shaped and far-from-perfect mirror. An additional challenge is that its appearance confounds the light coming from the environment with the unknown materials it is made of. We propose a
learning-based approach to predict the environment from multiple reflectance maps that are computed from approximate surface normals. The proposed method allows us to jointly model the statistics of environments and material properties. We train our system from synthesized training data, but demonstrate its applicability to real-world data (Fig. 36.14). Interestingly, our analysis shows that the information obtained from objects made out of multiple materials often is complementary and leads to better performance.

Most recently, we propose a deep representation of appearance, i.e. the relation of color, surface orientation, viewer position, material and illumination [6]. Previous approaches have used deep learning to extract classic appearance representations relating to reflectance model parameters (e.g. Phong) or illumination (e.g. HDR environment maps). We suggest to directly represent appearance itself as a network we call a deep appearance map (DAM). This is a 4D generalization over 2D reflectance maps, which held the view direction fixed. First, we show how a DAM can be learned from images or video frames and later be used to synthesize appearance, given new surface orientations and viewer positions. Second, we demonstrate how another network can be used to map from an image or video frames to a DAM network to reproduce this appearance, without using a lengthy optimization such as stochastic gradient descent (learning-to-learn). Finally, we show the example of an appearance estimation-and-segmentation task, mapping from an image showing multiple materials to multiple deep appearance maps.

**Machine Learning for “intuitive physics”**

Understanding physical phenomena is a key competence that enables humans and animals to act and interact under uncertain perception in previously unseen environments containing novel objects and their configurations. Developmental psychology has shown that such skills are acquired by infants from observations at a very early stage. We contrast a more traditional approach of taking a model-based route with explicit 3D representations and physical simulation by an end-to-end approach that directly predicts stability from appearance [5]. We ask the question if and to what extent and quality such a skill can directly be acquired in a data-driven way—bypassing the need for an explicit simulation at run-time. We present a learning-based approach based on simulated data that predicts stability of towers comprised of wooden blocks under different conditions and quantities related to the potential fall of the towers. We first evaluate the approach on synthetic data and compared the results to human judgments on the same stimuli. Further, we extend this approach to reason about future states of such towers that in return enables successful stacking (Fig. 36.15).
Figure 36.15: We propose an end-to-end approach to propose physical events about stability that we utilize in robotic manipulation.

Furthermore, we study how an artificial agent can autonomously acquire this intuition through interaction with the environment [4]. We created a synthetic block stacking environment with physics simulation in which the agent can learn a policy end-to-end through trial and error. Thereby, we bypass to explicitly model physical knowledge within the policy. We are specifically interested in tasks that require the agent to reach a given goal state that may be different for every new trial. To this end, we propose a deep reinforcement learning framework that learns policies for stacking tasks which are parametrized by a target structure – departing from conventional approaches based on simulation and planning. We validated the model on a toy example navigating in a grid world with different target positions and in a block stacking task with different target structures of the final tower. In contrast to prior work, our policies show better generalization across different goals.

**Gradient-free policy architecture search and adaptation**

The goal of this work is to learn a policy for an autonomous driving task minimizing crashes and other safety violations while training [1]. To this end we propose a gradient-free optimization algorithm which learns to generate an optimal network architecture from demonstration using a new reward function that optimizes accuracy and model size simultaneously. By learning from both demonstration and environmental reward we develop a model that can learn with relatively few early catastrophic failures. We show that our method can adapt the model learned by demonstration to a new domain relying on target environmental rewards. Experimental evaluation shows that our model reduces the number of crashes incurred while learning to drive, compared to baselines based only on reward or demonstration but not both, or compared to previously proposed fixed architectures that were not optimized for the domain.
36 D2: Computer Vision and Multimodal Computing

References


36.8.2 Uncertainty in Deep Learning

*Investigators: Apratim Bhattacharyya, Mario Fritz, Bernt Schiele*

For autonomous agents to successfully operate in the real world, anticipation of future events and states of their environment is a key competence. We begin this section with an investigation on the limits of video prediction – direct prediction of the future of observed scenes. We see that its is difficult to accurately predict the future in the long-term as futures become increasingly uncertain. Therefore, we show that it is necessary to explicitly model the uncertainty associated with the future in a probabilistic framework. Our models lead to state-of-the-art results on a variety of tasks ranging from pedestrian trajectory prediction to the prediction of the future of street scenes.
Long Term Image Boundary Prediction

Here, in [4] we explore the limits of video prediction with respect to time-horizon – how far into the future we can predict without an explicit model of uncertainty. As video prediction directly in RGB space suffers from blurriness problems, we explore an intermediate representation – boundaries. Boundary estimation in images and videos has been a very active topic of research, and organizing visual information into boundaries and segments is believed to be a cornerstone of visual perception. Scene boundaries capture the important structure and extents of objects. Moreover, they can be accurately estimated [6]. Prediction of future scene boundaries requires understanding of object dynamics and motion patterns including an intuitive understanding of physical laws or “intuitive physics”.

While prior work has focused on estimating boundaries for observed frames, our work aims at predicting boundaries of future unobserved frames. This requires our model to learn about the fate of boundaries and corresponding motion patterns – including a notion of “intuitive physics”. We focus on two particular scenarios involving motion and local interactions (Figure 36.16). The first one, which we call physics-based motion, can fully be described by the laws of physics, e.g. dynamics of billiard balls. The second one, which we call agent-based motion, also involves understanding of intentions, e.g. dynamics of an ice-skater. We show that compared to RGB prediction, boundary prediction allows for long-term prediction in both the scenarios. Furthermore, we show that fusion of RGB and boundary prediction leads to improved RGB predictions.

Long-Term On-Board Prediction of People in Traffic Scenes under Uncertainty

In [4] (above), we show that it is difficult to predict agent-based motion in the long-term. However, it is necessary for important several (semi-) autonomous systems inducing assisted and autonomous driving. Progress towards advanced systems for assisted and autonomous driving is leveraging recent advances in recognition and segmentation methods. Yet, we are still facing challenges in bringing reliable driving to inner cities, as those are composed of highly dynamic scenes observed from a moving platform at considerable speeds. Anticipation of the long-term motion of agents e.g. pedestrians and vehicles becomes a key element in order to react timely and prevent accidents.

We propose the first approach [2] to predict people (pedestrians including cyclists) trajectories from on-board cameras over such long-time horizons with uncertainty estimates. Due to the particular importance for safety, we are focusing on the people class. While pedestrian
trajectory prediction has been approached in prior work, we propose the first approach for on-board prediction. As predictions are made with respect to the moving vehicle, we formulate a novel two stream model for long-term person bounding box prediction and vehicle ego motion (odometry). Our model consists of two specialized streams for prediction of pedestrian motion and odometry. The odometry specialist stream predicts the most likely future vehicle odometry sequence. The bounding box specialist stream consists of a novel Bayesian RNN encoder-decoder architecture to predict odometry conditioned distributions over pedestrian trajectories and to capture epistemic and aleatoric uncertainty. In contrast to prior work, we model both aleatoric (observation) uncertainty and epistemic (model) uncertainty in order to arrive at an estimate of the overall uncertainty. Bayesian probability theory [5] provides us with a theoretically grounded approach to dealing with both types of uncertainties.

Our experimental results show that it is indeed possible to predict people trajectories at the desired time horizons and that our uncertainty estimates are informative of the prediction error (Figure 36.17). We also show that both sequence modeling of trajectories as well as our novel method of long term odometry prediction are essential for best performance.

**Accurate and Diverse Sampling of Sequences based on a “Best of Many” Sample Objective**

As argued in [2], for autonomous agents to successfully operate in the real world, anticipation of future events and states of their environment is a key competence. In many scenarios, problem can be formalized as a sequence extrapolation problem, where a number of observations are used to predict the sequence into the future. Real-world scenarios demand a model of uncertainty of such predictions, as predictions become increasingly uncertain – in particular on long time horizons. While impressive results have been shown on point estimates, scenarios that induce multi-modal distributions over future sequences remain challenging.

In this work [1], we address these challenges in a Gaussian Latent Variable model for sequence prediction. The Conditional Variational Auto-Encoder (CVAE) framework [8] is popularly used for learning of the Gaussian Latent Variables. We identify two key limitations of this CVAE framework. First, the currently used objectives hinder learning of diverse
samples due to a marginalization over multi-modal futures. Second, a mismatch in latent variable distribution between training and testing leads to errors in model fitting. We overcome both challenges which results in more accurate and diverse samples – better capturing the true variations in data. Our core contribution is a “Best of Many” sample objective that leads to more accurate and more diverse predictions that better capture the true variations in real-world sequence data (Figure 36.18). Compared to the CVAE objective, the recognition network has multiple chances to draw samples with high posterior probability. This encourages diversity in the generated samples. Furthermore, the data log-likelihood estimate in our objective is tighter. Therefore, this bound loosens the constrains on the recognition network and allows it more closely match the latent variable distribution.

Our models also empirically outperform prior work. We demonstrate improved accuracy as well as diversity of the generated samples on three diverse tasks: stroke completion, Stanford Drone Dataset and HKO weather data. On all three datasets we consistently outperform the state of the art and baselines.

**Bayesian Prediction of Future Street Scenes using Synthetic Likelihoods**

Dropout based Bayesian inference [5] provides a computationally tractable, theoretically well grounded approach to learn likely hypotheses/models to deal with uncertain futures and make predictions that correspond well to observations – are well calibrated. However, it turns out that such approaches fall short to capture complex real-world scenes, even falling behind in accuracy when compared to the plain deterministic approaches. This is because, when the underlying data distribution is multimodal and the model set under consideration do not have explicit latent state/variables (as most popular deep deep neural network architectures), the approach of [5] is unable to recover the true model uncertainty [7]. This is because this approach is known to conflate risk and uncertainty. This limits the accuracy of the models over a plain deterministic (non-Bayesian) approach. The main cause is the data log-likelihood maximization step during optimization – for every data point the average likelihood assigned by all models is maximized.
This forces every model to explain every data point well, pushing every model in the distribution to the mean. We address this problem through an objective leveraging synthetic likelihoods which relaxes the constraint on every model to explain every data point, thus encouraging diversity in the learned models to deal with multi-modality.

In this work [3], we propose a novel Bayesian formulation (Figure 36.19) for anticipating future scene states which leverages synthetic likelihoods that encourage the learning of diverse models to accurately capture the multi-modal nature of future scene states. We show that our approach achieves accurate state-of-the-art predictions and calibrated probabilities through extensive experiments for scene anticipation on Cityscapes datasets. Moreover, we show that our approach generalizes across diverse tasks such as digit generation and precipitation forecasting.

References

36.9 Virtual Humans

Coordinator: Gerard Pons-Moll

The world is shifting towards a digitization of everything – music, books, movies and news in digital form are common in our everyday lives. Digitizing human beings has the potential to redefine the way we think and communicate (with other humans and with machines), and it is necessary for many applications – for example, to analyze people in videos, to transport people into virtual and augmented reality, for entertainment and special effects in movies, and for medicine and psychology.

Creating digital models of people is extremely challenging – the task requires many components of artificial intelligence such as natural language processing, emotion and gesture recognition, social signal processing, motion analysis and 3D appearance modeling. In our group, we focus on modelling and capturing the human appearance and motion. Works at the intersection of Computer Graphics (CG) and Computer Vision (CV) have shown that human performances can be digitized using expensive multi-camera equipment–but capturing and learning models of clothing and interaction remains an open problem. More importantly, such specialized recording setups are unpractical, are not available to the regular user, and do not scale.

In Real Virtual Humans, we are focusing on two inter-related aspects of human digitization. The first is to learn realistic models of 3D people appearance and movement – what can be understood as building a mental model of people. Instead of hand-crafting the models, we learn them by capturing real people in clothing using different sensors such as 4D scanners, RGB cameras, and depth and inertial sensors. The second is to perceive people in 3D from ubiquitous sensors like RGB-cameras or IMUs–for this challenging task, we leverage a generative human model, and investigate geometric and learning algorithms. Perception and model learning benefit from each other and therefore we explore ways of addressing them jointly.

36.9.1 Human Pose Estimation from Video and IMU

Investigators: Gerard Pons-Moll in cooperation with Timo Von Marcard (University of Hannover), Roberto Henschel (University of Hannover), Bodo Rosenhahn (University of Hannover), Yinghao Huang (MPI for Intelligent Systems), Michael Black (MPI for Intelligent Systems), Manuel Kaufmann (ETH Zuerich), Emre Aksan (ETH Zuerich), Otmar Hiliges (ETH Zuerich)

The recording of human motion is necessary for modelling, understanding and automatically animating full-body human movement. Traditional marker-based optical Motion Capture (MoCap) systems are intrusive and restrict motions to controlled laboratory spaces. Therefore, simple daily activities like biking, or having coffee with friends cannot be recorded with such systems. Image based motion capture methods offer an alternative, but they are still not accurate enough, and require direct line of sight with the camera.

To address these issues and to be able to record human motion in everyday natural situations, we leverage Inertial Measurement Units (IMUs), which measure local orientation
Figure 36.20: IMU capture. Top: Deep Inertial Poser [1] is the first real-time human motion capture method which requires only 6 IMU sensors attached at the lower-arms, lower-legs, back and head. We synthesized a large dataset of IMU data from MoCap (Marker-based) data using a body model, and learned to regress the full human pose with a deep model that propagates information forward and backward in time. Bottom: Video Inertial Poser (VIP [2]) combines IMU sensors and a single moving camera to recover the pose of people in very complex scenes. With VIP we could record the first dataset with “ground truth” poses in the wild, which provides a benchmark for monocular 3D human pose estimation methods.

and acceleration. IMUs provide cues about the human motion without requiring external cameras, which is desirable for outdoor recordings where occlusions occur often.

However, existing IMU systems are intrusive because they require a large number of sensors (17 or more), worn on the body. In previous work [3] (SIP), we have demonstrated an optimization-based approach which can recover full body motion from only 6 IMUs attached to wrists, lower-legs waist and head.

While less intrusive, SIP is inherently offline, which limits a lot of applications. In recent work, [1] we present a Deep Learning based real-time algorithm for full body reconstruction from 6 IMUs alone. We found that propagation of information forward and backward in time is crucial for reconstructing natural human motion, for which we use a bi-directional Recursive Neural Network, see Figure 36.20-Top. To achieve good generalization, we synthesize IMU readings with their corresponding poses—obtained by fitting the SMPL body model to marker-based datasets.

In contrast to visual measurements, IMU cannot provide absolute joint position information. This makes pure IMU based methods inaccurate for certain types of motions. Hence, in recent work, we introduce VIP [2] which combines IMUs and a single moving camera, to robustly recover human pose in challenging outdoor scenes. The moving camera, sensor heading drift, cluttered background, occlusions and many people visible in the video make the problem very hard. We associate 2D pose detections in each image to the corresponding IMU-equipped persons by solving a novel graph-based optimization problem that forces 3D to 2D coherency within a frame and across long range frames. Given these associations, we
jointly optimize the pose of the SMPL body model, the camera pose and heading drift using a continuous optimization.

Using VIP, we collected the 3DPW dataset, which includes videos of humans in challenging scenes with accurate 3D parameters that provide, for the first time, the means to quantitatively evaluate monocular methods in difficult scenes and stimulate new research in this area, see Figure 36.20-Bottom.

References


36.9.2 Perceiving 3D People from Visual Data


Humans are incredibly good at perceiving people from visual data. Without even thinking about it, we quickly perceive the body shape, posture, facial expressions and clothing of other people. We believe all these details are crucial to understand humans from visual data, and to have natural communication.

Computer Vision methods in this area have seen a lot of progress thanks to the availability of large scale datasets with crowd sourced 2D annotations such as 2D joints and segmentation masks. Deep learning methods can predict such annotations because they are very effective at recognizing patterns. However, humans are far more complex than 2D joints or segmentation masks and unfortunately it is practically impossible to manually annotate in images full 3D geometry, human motion or clothing.

Our approach to this problem is to infer and learn a powerful representation of people in 3D space. Intuitively, such representation encodes the machine mental model of people. Given an image, the inference algorithms should predict the full detail in 3D, which should be consistent with learned 3D human shape priors and its projection should overlap with the image observations. This opens the door for semi-supervised learning because unlabeled images alone can be used to infer properties about the 3D world.

Following this paradigm, we introduced methods to reconstruct 3D human shape and pose from images, human shape and clothing from videos, and non-rigid deformations from video.
Direct prediction of 3D body pose and shape remains a challenge even for highly parameterized deep learning models. Mapping from the 2D images space to the 3D space is difficult due to perspective ambiguities, and lack of training data with 3D annotations.

To address this, we introduced (Neural Body Fitting (NBF) [8]), which integrates a statistical body model (SMPL) within a CNN, leveraging reliable bottom-up semantic body part segmentation and robust top-down body model constraints, see Figure 36.21-Top. NBF is fully differentiable, and can be trained using 2D and 3D annotations. In detailed experiments, we analyze how the components of our model affect performance, especially the use of part segmentations as an explicit intermediate representation, and present a robust, efficiently trainable framework for 3D human pose estimation from 2D images with competitive results on standard benchmarks. Using a similar bottom-up-top-down architecture [6] (unpublished-submission), we propose a network architecture that comprises a new disentangled hidden space encoding explicit 2D and 3D features—it achieves state-of-the-art accuracy on challenging in the wild data.

In [7], we introduced one of the first methods for 3D human pose estimation of multiple people. Estimating multiple in 3D requires novel architectures and output representations in order to deal with the varying number of people and occlusions. In [7], we introduce occlusion-robust pose-maps (ORPM) which enable full body pose inference even under strong partial occlusions by other people and objects in the scene. Their key idea is to output a fixed number of maps which encode the 3D joint locations of all people in the scene, which are associated to person identities in a second stage, see Figure 36.21-Bottom. This allows to estimate the pose of multiple people at once without explicit bounding box detection.
Clothing: Estimating 3D Humans in Clothing from Images and Videos.

Understanding human behavior is not only about motion and body shape. The type of clothing people wear is another form of expression. People use clothing to express their political views, age, gender or social status. Instead of inferring body pose and shape while being invariant to clothing, we aim at perceiving and capturing human body shape along with clothing (category, appearance and shape) from images.

In [3] we describe a method to obtain accurate 3D body models alongside clothing and texture of arbitrary people from a single, monocular video in which a person is moving—we achieve a reconstruction accuracy of 4.5mm. This is the first method capable of reconstructing people and their clothing from a single RGB video without using a pre-scanned templates. At the core of our approach is the transformation of dynamic body pose into a canonical frame of reference. Our main contribution is a method to transform the silhouette cones corresponding to dynamic human silhouettes to obtain a visual hull in a common reference frame. This enables efficient estimation of a consensus 3D shape, texture and implanted animation skeleton based on a large number of frames. Results on 4 different datasets demonstrate the effectiveness of our approach to produce accurate 3D models, see Figure 36.22-Top. Requiring only an RGB camera, our method enables everyone to create their own fully animatable digital double, e.g., for social VR applications or virtual try-on for online fashion shopping.

We further extended the approach in [2] by integrating shading cues and a graph based optimization for body texture generation from a single RGB-video, see Figure 36.22-Bottom. While accurate, these approaches [3, 2] are based on non-linear optimization, which can fail when they are not initialized correctly, and are typically slow. Hence, in recent work [1] we
propose an approach that combines the benefits of deep learning and model based fitting together. The network takes multiple video frames as input and produces bottom-up a single coherent shape and a 3D pose for each of the frames. The estimates are optimized top-down in order to maximize the silhouette overlap and minimize the re-projection error. This results in accurate predictions in less than 10 seconds.

Understanding people clothing preferences according to body shape is another important problem with lots of practical applications—the most prominent one perhaps is for effective clothing recommendation in order to minimize returns. Our key idea is to leverage the large photo-collections in the internet to study the correlation between body shape and clothing preference [10]. To that end, we propose a novel inference technique capable of estimating body shape under clothing given a few images of the same person in different clothing. With these shape estimates and the predicted clothing categories from the images, we compute a distribution of clothing category conditioned on body shape. Using our method, we found out that clothing is indeed correlated with body shape, and we can predict clothing categories based on automatic shape estimation.

Non-rigid Tracking from Depth and Video

While human motion is mostly articulated, clothing dynamics are non-rigid and require special treatment. Although non-rigid reconstruction and tracking has been demonstrated using free-form volumetric methods like DynamicFusion, the methods fail to robustly track humans and clothing. Model-based methods which fit a statistical body model to depth data are more robust but are limited to capturing only articulation.

In DoubleFusion [11], we combine free-form volumetric dynamic reconstruction with model based fitting. DoubleFusion, is the first real-time system which can simultaneously reconstruct detailed geometry, non-rigid motion and the body shape under clothing from a single depth camera. A pre-defined node graph on the body surface parameterizes the non-rigid deformations near the body, and a free-form dynamically changing graph parameterizes the outer surface layer far from the body, which allows more general reconstruction. Moreover, the inner body shape is optimized online and is forced to fit inside the outer surface layer. Overall,
our method enables increasingly denoised, detailed and complete surface reconstructions, fast motion tracking performance and plausible inner body shape reconstruction in real-time. A real-time demo was shown at CVPR’18.

Although DoubleFusion can track human motion and clothing jointly, it has two limitations. The first one is that it uses a single surface to represent the visible skin and the clothing. The second is that it can only track non-rigid deformations that are visible in the depth image. In SimulCap [12], we draw inspiration from our previous work [9], and model the garments and body as separate surfaces. In addition, we track clothing motion by means of physics based simulation. This marries physics based simulation with pure capture methods, which allows to incorporate physical constraints, and allows to recover the motion of the invisible parts (e.g. the back part of the body when facing the camera).

While depth cameras facilitate the problem, they are less ubiquitous than RGB cameras. Hence, we have also explored methods to recover non-rigid motion from RGB videos alone. We have shown real-time non-rigid tracking of general objects [5], and humans in clothing [4].

References


36.10 Combinatorial Image Analysis

Coordinators: Evgeny Levinkov and Bjoern Andres

The combinatorial image analysis group studies mathematical abstractions of image analysis tasks in the form of combinatorial optimization problems, and algorithms for solving these problems exactly, partially or locally. During the reporting period, the group’s research has been driven by the realization that seemingly unrelated computer vision tasks, including instance-separating semantic image segmentation, multiple object tracking and human body pose estimation, are fundamentally linked by a single combinatorial optimization problem whose feasible solutions define both a decomposition and a node labeling of a graph. The analysis of this problem, a generalization of the well-known correlation clustering problem, as well as the design of practical algorithms for computing feasible solutions have been the primary focus of the group (Section 36.10.1).

A secondary focus of the group has been on algorithms for biomedical image and data analysis (Section 36.10.2). In particular, we have continued to work on the problem of lineage tracing, i.e., the tracing of living cells as they move and divide, in sequences of light microscopy images. Previously, the group had proposed a mathematical abstraction of this task called the moral lineage tracing problem. During the reporting period, the group has improved its algorithm for solving this problem and applied the improved algorithm to sequences of images acquired by the technique of lens-free microscopy.

In addition to these research foci, the group has contributed to work with TU Munich on a challenging mixed-integer non-linear problem in the field of semi-supervised machine learning (Section 36.10.3).

36.10.1 Correlation Clustering

Investigators: Bjoern Andres, Eldar Insafutdinov, Andrea Hornakova, Andreas Karrenbauer, Jan-Hendrik Lange, Evgeny Levinkov, Mohamed Omran, Bernt Schiele, Paul Swoboda, Siyu Tang in cooperation with Thomas Brox, Margret Keuper (University of Freiburg), Alexander Kirillov, Carsten Rother, Bogdan Savchynskyy (TU Dresden), Jonas Uhrig (TU Darmstadt),

The research summarized in this section is concerned with a combinatorial optimization problem known as weighted correlation clustering. Among the problems whose feasible
solutions are decompositions (clusterings) of a graph, weighted correlation clustering has the special property that the number of clusters is not penalized or constrained in the problem statement and is instead defined by the (any) solution. This property makes the problem a suitable mathematical abstraction of computer vision tasks in which the number of clusters is not known a priori and is instead to be estimated from data.

**Generalizations of Weighted Correlation Clustering**

The weighted correlation clustering problem can penalize solutions in which neighboring nodes are put in distinct clusters. It cannot penalize solutions in which non-neighboring nodes are put in distinct clusters. In [1], we propose and study a generalization of the weighted correlation clustering problem that can. More specifically, we study the set of all decompositions of a graph through its characterization as a set of lifted multicuts. This leads us to practically relevant insights related to the definition of classes of decompositions by must-join and must-cut constraints and related to the comparison of clusterings by metrics.

To find optimal decompositions defined by minimum cost lifted multicuts, we establish some properties of some facets of lifted multicut polytopes, define efficient separation procedures and apply these in a branch-and-cut algorithm. In the special case of paths, the generalization of the correlation clustering problem we define can be solved efficiently, e.g., by dynamic programming. In [4], we offer a complete description of its feasible set, the lifted multicut polytope for paths.

While weighted correlation clustering has many applications already, more applications exist where the cost of assigning two nodes to distinct clusters depends on decision at the nodes. In [8], we state a combinatorial optimization problem whose feasible solutions define both a decomposition and a node labeling of a given graph. This problem offers a common mathematical abstraction of seemingly unrelated computer vision tasks, including instance-separating semantic segmentation, articulated human body pose estimation and multiple object tracking. Conceptually, it generalizes the unconstrained integer quadratic program and the minimum cost lifted multicut problem, both of which are NP-hard. In order to find feasible solutions efficiently, we define two local search algorithms that converge monotonously to a local optimum, offering a feasible solution at any time. To demonstrate the effectiveness of these algorithms in tackling computer vision tasks, we apply them to instances of the problem that we construct from published data, using published algorithms. We report state-of-the-art application-specific accuracy in the three above-mentioned applications.

**Partial Optimality in Correlation Clustering**

Weighted correlation clustering is hard to solve and hard to approximate for general graphs. Its applications in computer vision call for efficient algorithms. To this end, our work [6] makes three contributions. Firstly, it establishes partial optimality conditions that can be checked efficiently, and doing so recursively solves the problem for series-parallel graphs to optimality, in linear time. Secondly, it exploits the packing dual of the problem to compute a heuristic, but non-trivial lower bound faster than that of a canonical linear program relaxation. Thirdly, it introduces a re-weighting with the dual solution by which efficient local search algorithms converge to better feasible solutions. The effectiveness of these
methods is demonstrated empirically on a number of benchmark instances.

In [5], we propose additional partial optimality criteria for the multicut and max-cut problem as well as fast combinatorial routines to verify them. The criteria that we derive are based on mappings that improve feasible multicut, respectively cuts. Our elementary criteria can be checked enumeratively. The more advanced ones rely on fast algorithms for upper and lower bounds for the respective cut problems and max-flow techniques for auxiliary min-cut problems. Our methods can be used as a preprocessing technique for reducing problem sizes or for computing partial optimality guarantees for solutions output by heuristic solvers. We show the efficacy of our methods on instances of both problems from computer vision, biomedical image analysis and statistical physics.

**Algorithms for Correlation Clustering**

In [9], we propose a dual decomposition and linear program relaxation of the weighted correlation clustering problem. Unlike other polyhedral relaxations of the multicut polytope, it is amenable to efficient optimization by message passing. Like other polyhedral relaxations, it can be tightened efficiently by cutting planes. We define an algorithm that alternates between message passing and efficient separation of cycle- and odd-wheel inequalities. This algorithm is more efficient than state-of-the-art algorithms based on linear programming, including algorithms written in the framework of leading commercial software, as we show in experiments with large instances of the problem from applications in computer vision, biomedical image analysis and data mining. In the small empirical study [7], we compare local search algorithms for the weighted correlation clustering problem.

**Applications of Correlation Clustering**

We have applied the correlation clustering problem as a mathematical abstraction in the field of computer vision. In [3], we address the task of instance-aware semantic segmentation. Our key motivation is to design a simple method with a new modelling-paradigm, which therefore has a different trade-off between advantages and disadvantages compared to known approaches. Our approach, we term InstanceCut, represents the problem by two output modalities: (i) an instance-agnostic semantic segmentation and (ii) all instance-boundaries. The former is computed from a standard convolutional neural network for semantic segmentation, and the latter is derived from a new instance-aware edge detection model. To reason globally about the optimal partitioning of an image into instances, we combine these two modalities into a weighted correlation clustering problem. We evaluate our approach on the challenging Cityscapes dataset. Despite the conceptual simplicity of our approach, we achieve the best result among all published methods, and perform particularly well for rare object classes.

In [2], we address the task of jointly segmenting a video and tracking objects that appear in the video. Models for either task are commonly defined either w.r.t. low-level concepts such as pixels that are to be grouped, or w.r.t. high-level concepts such as semantic objects that are to be detected and tracked. Combining bottom-up grouping with top-down detection and tracking, although highly desirable, is a challenging problem. We state this joint problem as a co-clustering problem that is principled and tractable by existing algorithms.
demonstrate the effectiveness of this approach by combining bottom-up motion segmentation by grouping of point trajectories with high-level multiple object tracking by clustering of bounding boxes. We show that solving the joint problem is beneficial at the low-level, in terms of the FBMS59 motion segmentation benchmark, and at the high-level, in terms of the Multiple Object Tracking benchmarks MOT15, MOT16 and the MOT17 challenge, and is state-of-the-art in some metrics.

References


36.10.2 Biomedical Image and Data Analysis

Investigators: Bjoern Andres, Andrea Hornakova, Jan-Hendrik Lange, Markus List, Marcel H. Schulz, Jilles Vreeken in cooperation with Florian Jug, Corinna Blasse, Eugene W. Myers (MPI for Molecular Cell Biology and Genetics) Sanjeev Kumar, Bjoern H. Menze, Markus
The research summarized in this section is concerned with combinatorial optimization problems in the field of biomedical image and data analysis. In particular, we report on our work on lineage tracing, the problem of tracking living cells as they move and divide in a sequence of microscopy images. Lineage tracing is a powerful tool in developmental biology, e.g., for determining how the structure of living tissue emerges.

**Biomedical Image Analysis**

Lineage tracing, the joint segmentation and tracking of living cells as they move and divide in a sequence of light microscopy images, is a challenging task. A mathematical abstraction of this task, the moral lineage tracing problem (MLTP), has feasible solutions that define both a segmentation of every image and a lineage forest of cells. Branch-and-cut algorithms, however, are prone to many cuts and slow convergence for large instances. To address this problem, our work [3] makes three contributions: Firstly, we devise the first efficient primal feasible local search algorithms for the MLTP. Secondly, we improve the branch-and-cut algorithm by separating tighter cutting planes and by incorporating our primal algorithms. Thirdly, we show in experiments that our algorithms find accurate solutions on published instances and scale to larger instances, leveraging moral lineage tracing to practical significance.

In [2, 4], we apply moral lineage tracing to sequences of images acquired by the technique of lens-free light microscopy: In vitro experiments with cultured cells are essential for studying their growth and migration pattern and thus, for gaining a better understanding of cancer progression and its treatment. Recent progress in lens-free microscopy (LFM) has rendered it an inexpensive tool for label-free, continuous live cell imaging, yet there is only little work on analysing such time-lapse image sequences. Our work proposes a cell detector for LFM images based on fully convolutional networks and residual learning, and a probabilistic model based on moral lineage tracing that explicitly handles multiple detections and temporal successor hypotheses by clustering and tracking simultaneously. We benchmark our method in terms of detection and tracking scores on a dataset of three annotated sequences of several hours of LFM, where we demonstrate our method to produce high quality lineages. We evaluate its performance on a somewhat more challenging problem: estimating cell lineages from the LFM sequence as would be possible from a corresponding fluorescence microscopy sequence. We present experiments on 16 LFM sequences for which we acquired fluorescence microscopy in parallel and generated annotations from them. Finally, we showcase our methods effectiveness for quantifying cell dynamics in an experiment with skin cancer cells.

**Biomedical Data Analysis**

Genome-wide measurements of paired miRNA and gene expression data have enabled the prediction of competing endogenous RNAs (ceRNAs). It has been shown that the sponge effect mediated by protein-coding as well as non-coding ceRNAs can play an important regulatory role in the cell in health and disease. Therefore, many computational methods for the computational identification of ceRNAs have been suggested. In particular, methods based on Conditional Mutual Information (CMI) have shown promising results. However, the
currently available implementation is slow and cannot be used to perform computations on a large scale. Our work [1] presents JAMI, a Java tool that uses a non-parametric estimator for CMI values from gene and miRNA expression data. We show that JAMI speeds up the computation of ceRNA networks by a factor of approximately 70 compared to currently available implementations. Further, JAMI supports multi-threading to make use of common multi-core architectures for further performance gain.

References


36.10.3 Semi-supervised Learning

Investigators: Bjoern Andres, Jan-Hendrik Lange in cooperation with Daniel Cremers, Csaba Domokos, Emanuel Laude, Laura Leal-Taixé, Frank R. Schmidt, Jonas Schüpfer (TU Munich)

The need for annotated training data is threatening to become a bottleneck for machine learning research and applications. Semi-supervised learning methods have the potential to overcome this bottleneck. In principle, they offer mathematical abstractions for learning from structured knowledge instead of from examples. In practice, the mixed-integer non-linear optimization problems they give rise to pose a formidable challenge.

Our work [1] introduces a novel algorithm for transductive inference in higher-order Markov Random Fields, where unary energies are parameterized by a variable classifier. The considered task is posed as a joint optimization problem in the continuous classifier parameters and the discrete label variables. In contrast to prior approaches such as convex relaxations, we propose an advantageous decoupling of the objective function into discrete and continuous subproblems and a novel, efficient optimization method related to ADMM. This approach preserves integrality of the discrete label variables and guarantees global convergence to a critical point. We demonstrate the advantages of our approach in several experiments including video object segmentation on the DAVIS data set and interactive image segmentation.

References

36.11 Vision and Language

Coordinator: Zeynep Akata

Data and knowledge can be observed from multiple sources in different modalities with a given uncertainty. We can reduce the uncertainty of our data by combining multiple information sources which enables us to take more confident and accurate decisions. Vision and Language pertains complementary information, e.g. while language ambiguities can be recovered with perceptual information, language provides contextual information for various perception tasks.

Our research in vision and language consists of four main topics, i.e. visual description and question answering, explainable artificial intelligence (XAI), generative models with vision and language, and datasets. In the following subsections we detail each of the sub topic and summarize our relevant publications.

In visual description and question answering, is a new family of models and an exciting task that combines natural language processing and computer vision techniques. Here, we differentiate between visual description generation, i.e. automatically describing the content of an image with natural language, and visual question answering, i.e. automatically answering any question that may be asked about an image with natural language. In explainable artificial intelligence we take this one step further and define the task as automatically generating natural language or visualizations that explain a classifier’s decision or the answer of a question. In generative models with vision and language, we turn this idea around and generate realistic images of natural language descriptions. In datasets, we detail the datasets that we proposed in the past two years.

36.11.1 Datasets

Investigators: Anna Rohrbach, Niket Tandon, Bernt Schiele in cooperation with Marcus Rohrbach (UC Berkeley, Facebook AI Research), Atousa Torabi (Disney Research), Tegan Maharaj, Nicolas Ballas, Aaron Courville (Université de Montréal), Hugo Larochelle (Université de Sherbrooke, Google Brain), Christopher Pal (École Polytechnique de Montréal), Nelson Mukuze, Vera Demberg (Saarland University)

Datasets are indispensable to making progress in machine learning and artificial intelligence. At the same time, they are non-trivial and often costly to collect. In this section, we present our efforts on Vision and Language datasets, where we typically require aligned visual and language modalities. Specifically, we focus on video-text and image-text domains. First, we discuss our datasets for Movie Description [5] and Movie Fill-in-the-Blank [2] tasks. Then we present our dataset for predicting typical locations for verbs [3], which relies on images with captions.
Movie Description Datasets

Audio descriptions (ADs) make movies accessible to millions of blind or visually impaired people. AD provides an audio narrative of the most important aspects of the visual information, namely actions, gestures, scenes, and character appearance as can be seen in Figure 36.24. AD is prepared by trained describers and read by professional narrators. They are by design mainly visual and thus naturally form an interesting data source for computer vision and computational linguistics. In [5] we have introduced the Large Scale Movie Description Challenge (LSMDC) and the dataset which contains a parallel corpus of 128,118 sentences aligned to video clips from 200 movies (around 150 hours of video in total), extending our prior work [4]. The goal of the challenge is to automatically generate descriptions for the movie clips. First we characterize the dataset by benchmarking different approaches for generating video descriptions. Furthermore, we present and compare the results of several teams who participated in the challenges organized in the context of two workshops at ICCV 2015 and ECCV 2016.

Having models which can learn to understand video is of interest for many applications, but many domains lack sufficient data to explore and perfect video models. In order to address the need for a simple, quantitative benchmark for developing and understanding video, in [2] we present MovieFIB, a fill-in-the-blank question-answering dataset with over 300,000 examples, based on our LSMDC dataset. We perform a detailed analysis of 5 different models, and compare these with human performance. We also investigate the relative importance of language, static (2D) visual features, and moving (3D) visual features. We illustrate that: this task is not solvable by a language model alone; our model combining 2D and 3D visual information provides the best result; all models perform significantly worse than human-level. Our human evaluation for responses given by different models shows that accuracy on the MovieFIB evaluation corresponds well with human judgement.

Figure 36.24: Audio description (AD) and movie script samples from the movie “Ugly Truth”.

AD: Abby gets in the basket. Mike leans over and sees how high they are. Abby claps her hands around his face and kisses him passionately.

Script: After a moment a frazzled Abby pops up in his place. Mike looks down to see – they are now fifteen feet above the ground. For the first time in her life, she stops thinking and grabs Mike and kisses the hell out of him.
A Vision-grounded Dataset for Predicting Typical Locations for Verbs

In this work we are looking at images as opposed to video, and our goal is to capture the thematic fit for pairs of verbs and locations [3]. Specifically, we are interested in identifying typical and unusual combinations of verbs/locations for the task of selectional preference estimation (e.g. usually “eat” occurs in a “kitchen” but not in a “bathroom”). While information about location of an action is often absent in text, it may be visually present in images. It is hence a logical step to use multimodal data, and in particular, data from vision. We present a new dataset with the total of 20,000 human judgments for 2,000 verb/location pairs. We use MS COCO [1] images and captions to identify the presence of a verb or a scene type. Our dataset allows for evaluating text-based, vision-based or multimodal inference systems for the typicality of an event’s location. We also investigate three thematic fit baselines for our dataset: a state-of-the-art neural networks based thematic fit model learned from linguistic data, a model estimating typical locations based on the MSCOCO dataset and a simple combination of the two systems. Our experiments show that the visual scene probabilities are a useful cue for typical location prediction, and are complementary to the language estimates.

References


36.11.2 Explainable AI

Investigators: Zeynep Akata, Anna Rohrbach, Rakshith Shetty, Bernt Schiele in cooperation with Lisa-Anne Hendricks (UC Berkeley), Trevor Darrell (UC Berkeley), Dong Huk Park (UC Berkeley), Ronghang Hu (UC Berkeley), Jinkyu Kim (UC Berkeley), John Canny (UC Berkeley), Marcus Rohrbach (FAIR), Hamed R. Tavakoli (Aalto University), Ali Borji (UCF), Jorma Laaksonen (Aalto University)

Explaining decisions is an integral part of human communication, understanding, and learning, and humans naturally provide both deictic (pointing) and textual modalities in
Figure 36.25: A For a given question and an image, our Pointing and Justification Explanation (PJ-X) model predicts the answer and multimodal explanations which both point to the visual evidence for a decision and provide textual justifications. We show that considering multimodal explanations results in better explanations as visual and textual components complement each other. B In comparison to descriptions, our VQA-X explanations focus on the evidence that pertains to the question and answer instead of generally describing the scene. For ACT-X, our explanations are task specific whereas descriptions are more generic.

Multimodal Explanations: Justifying Decisions and Pointing to the Evidence

Deep models that are both effective and explainable are desirable in many settings; prior explainable models have been unimodal, offering either image-based visualization of attention weights or text-based generation of post-hoc justifications. We propose a multimodal approach [3] to explanation, and argue that the two modalities provide complementary explanatory strengths. We collect two new datasets to define and evaluate this task, and propose a novel model which can provide joint textual rationale generation and attention visualization. Our datasets shown in Figure 36.25 (left) define visual and textual justifications of a classification decision for activity recognition tasks (ACT-X) and for visual question answering tasks (VQA-X). We quantitatively show that training with the textual explanations not only yields better textual justification models, but also better localizes the evidence that supports the decision. We also qualitatively show cases where visual explanation is more insightful than textual explanation, and vice versa, supporting our thesis that multimodal explanation models offer significant benefits over unimodal approaches.

To illustrate the utility of multimodal explanations, consider Figure 36.25 (right). In both examples, the question “Is this a healthy meal?” is asked, and the PJ-X model correctly
What type of bird is this? It is a Cardinal because it is a red bird with a red beak and a black face.

Why not a Vermilion Flycatcher? It is not a Vermilion Flycatcher because it does not have black wings.

36 D2: Computer Vision and Multimodal Computing

Figure 36.26: Our phrase-critic considers grounded visual evidence to determine if candidate explanations are image relevant. For example, as many cardinals are red and have a black patch on their faces, mentioning and grounding those properties constitutes an effective factual explanation, i.e. rationalization. Furthermore, in our framework, informing the user of why an image does not belong to another class via the absence of certain attributes constitutes a counterfactual explanation.

answers either “no” or “yes” depending on the visual input. To justify why the image is not healthy, the generated textual justification mentions the kinds of unhealthy food in the image (“hot dog” and “toppings”). In addition to mentioning the unhealthy food, our model is able to point to the hot dog in the image. Likewise, to justify why the image on the right is healthy, the textual explanation mentions “vegetables”. The PJ-X model then points to the vegetables, which are mentioned in the textual explanation, but not other items in the image, such as the bread.

**Grounding Visual Explanations**

Existing visual explanation generating agents learn to fluently justify a class prediction. However, they may mention visual attributes which reflect a strong class prior, although the evidence may not actually be in the image. This is particularly concerning as ultimately such agents fail in building trust with human users. To overcome this limitation, we propose a phrase-critic model to refine generated candidate explanations augmented with flipped phrases which we use as negative examples while training. At inference time, our phrase-critic model takes an image and a candidate explanation as input and outputs a score indicating how well the candidate explanation is grounded in the image. Our explainable AI agent is capable of providing counter arguments for an alternative prediction, i.e. counterfactuals, along with explanations that justify the correct classification decisions. Our model [1] improves the textual explanation quality of fine-grained classification decisions on the CUB dataset by mentioning phrases that are grounded in the image. Moreover, on the FOIL tasks, our agent detects when there is a mistake in the sentence, grounds the incorrect phrase and corrects it significantly better than other models.

Ideally, an agent that accurately explains a classifier’s decision via natural language, as depicted in Figure 36.26, is expected to generate explanations such as “This is a Cardinal
because it is a red bird with a red beak and a black face” where the phrases should be both class discriminative, i.e. a red beak is discriminative for cardinals, and image relevant, i.e. the image indeed contains a red beak. Moreover, an explainable AI agent should be capable of arguing why the image was not classified as another class such as Vermilion Flycatcher by mentioning a class-specific property such as “black wings” that differentiates a cardinal from a vermilion flycatcher. Contrasting two concepts via class-specific attributes provides an additional means of model interpretation.

One way to design such an agent is to use densely labeled data with ground truth part annotations. However, it can be time consuming to collect densely labeled data for every task. On the other hand, large, diverse datasets such as the Visual Genome with densely labeled out-of-domain data do exist. Detecting visual evidence in a sentence via off-the-shelf grounding models can be unreliable, especially when applied to new domains. Nevertheless, integrating a natural language grounding model trained on auxiliary data and an LSTM-based explanation model via our proposed phrase-critic effectively grounds discriminative phrases in generated explanations.

Textual Explanations for Self-Driving Vehicles

Deep neural perception and control networks have become key components of self-driving vehicles. User acceptance is likely to benefit from easy-to-interpret textual explanations which allow end-users to understand what triggered a particular behavior. Explanations may be triggered by the neural controller, namely introspective explanations, or informed by the neural controller’s output, namely rationalizations. We propose a new approach [2] to introspective explanations which consists of two parts. First, we use a visual (spatial) attention model to train a convolutional network end-to-end from images to the vehicle control commands, i.e., acceleration and change of course. The controller’s attention identifies image regions that potentially influence the network’s output. Second, we use an attention-based video-to-text model to produce textual explanations of model actions. The attention maps of controller and explanation model are aligned so that explanations are grounded in the parts of the scene that mattered to the controller. We explore two approaches to attention alignment, strong- and weak-alignment. Finally, we explore a version of our model that generates rationalizations, and compare with introspective explanations on the same video segments. We evaluate these models on a novel driving dataset with ground-truth human explanations, the Berkeley DeepDrive eXplanation (BDD-X) dataset.

In this work, we focus on generating textual descriptions and explanations, such as the pair: “vehicle slows down” and “because it is approaching an intersection and the light is red” as in Figure 36.27. Natural language has an advantage of being inherently understandable and does not require familiarity with the design of an intelligent system in order to provide useful information. In order to train such a model, we collect explanations from human annotators. Our explanation dataset is built on top of another large-scale driving dataset [26] collected from dashboard cameras in human driven vehicles. Annotators view the video dataset, compose descriptions of the vehicle’s activity and explanations for the actions that the vehicle driver performed.

Our contributions are as follows. (1) We propose an introspective textual explanation model for self-driving cars to provide easy-to-interpret explanations for the behavior of a
deep vehicle control network. (2) We integrate our explanation generator with the vehicle controller by aligning their attentions to ground the explanation, and compare two approaches: attention-aligned explanations and non-aligned rationalizations. (3) We generated a large-scale Berkeley DeepDrive eXplanation (BDD-X) dataset with over 6,984 video clips annotated with driving descriptions, e.g., “The car is driving forward because there are no other cars in its lane”, and a visual explanation in the form of attention, i.e. attended regions directly influence the textual explanation generation process.

**Paying Attention to Descriptions Generated by Image Captioning Models**

To bridge the gap between humans and machines in image understanding and describing, we need further insight into how people describe a perceived scene. In this paper, we study the agreement between bottom-up saliency-based visual attention and object referrals in scene description constructs. We investigate the properties of human-written descriptions and machine-generated ones. We then propose a saliency-boosted image captioning model in order to investigate benefits from low-level cues in language models. We learn that (1) humans mention more salient objects earlier than less salient ones in their descriptions, (2) the better a captioning model performs, the better attention agreement it has with human descriptions, (3) the proposed saliency-boosted model, compared to its baseline form, does not improve significantly on the MS COCO database, indicating explicit bottom-up boosting does not help when the task is well learnt and tuned on a data,
(4) a better generalization is, however, observed for the saliency-boosted model on un- seen data.

In [4], we address two intriguing questions: (1) How well do image descriptions, by humans or models, on a scene agree with saliency, (2) Can saliency benefit image captioning by machine? Answering the questions, we learn not only about the role of attention in describing images, but also about the quality of human-written descriptions and machine-generated ones. We first study the textual statistics of the sentences by human and machine. Then we investigate the attention correlation in the structure of human-written and machine-generated descriptions. To further evaluate the contribution of low-level cues, we propose a saliency-boosted captioning model and compare it against a set of baseline captioning models.

References


36.11.3 Generative Models with Vision and Language

Investigators: Zeynep Akata, Qianru Sun, in cooperation with Levent Karacan (Hacettepe University), Aykut Erdem (Hacettepe University), Erkut Erdem (Hacettepe University), Tat-Seng Chua (Tianjin University)

In this subsection, we provide the details of two papers that are under revision. Our first work [1] focuses on manipulating the attributes of natural scenes such that when the user changes the value of an attribute, e.g. snow, the semantic meaning of that change gets reflected in the generated images.

Manipulating Attributes of Natural Scenes via Hallucination

The visual world we live in constantly changes its appearance depending on time and seasons. For example, at sunset, the sun gets close to the horizon gives the sky a pleasant red tint, with the advent of warm summer, the green tones on the grass leave its place in bright yellowish tones and autumn brings a variety of shades of brown and yellow to the trees. Such visual changes in the nature continues in various forms at almost any moment with
Our method utilizes a single generator network to imagine the scene with respect to its semantic layout and the desired set of attributes. It directly transfers the look from the hallucinated output to the input image, without a need to have access to a reference style image. Given a natural image, we hallucinate different versions of the same scene in a wide range of conditions, e.g. night, sunset, winter, spring, rain, fog or even a combination of those.

the effect of time, weather and season. Such high-level changes are referred to as transient scene attributes, e.g. cloudy, foggy, night, sunset, winter, summer, to name a few.

In this study [1], we build a two-stage framework for enabling users to directly manipulate high-level attributes of a natural scene. The key to our approach is a deep generative network which can hallucinate images of a scene as if they were taken at a different season (e.g. during winter), weather condition (e.g. in a cloudy day) or time of the day (e.g. at sunset) as shown in Figure 36.28. Once the scene is hallucinated with the given attributes, the corresponding look is then transferred to the input image while preserving the semantic details intact, giving a photo-realistic manipulation result. As the proposed framework hallucinates what the scene will look like, it does not require any reference style image as commonly utilized in most of the appearance or style transfer approaches. Moreover, it allows to simultaneously manipulate a given scene according to a diverse set of transient attributes within a single model, eliminating the need of training multiple networks per each translation task. Our comprehensive set of qualitative and quantitative results demonstrate the effectiveness of our approach against the competing methods.

Our contributions are summarized as follows: (1) We propose a new two-stage visual attribute manipulation framework for changing high-level attributes of a given outdoor image. (2) We develop a conditional GAN variant for generating natural scenes faithful to given semantic layouts and transient attributes. (3) We build up an outdoor scene dataset annotated with layout and transient attribute labels by combining and annotating images from Transient Attributes [2] and ADE20K.

References


36.11.4 Visual Description and Question Answering

Investigators: Anna Rohrbach, Rakshith Shetty, Siyu Tang, Seong Joon Oh, Wenbin Li, Mario Fritz, Bernt Schiele in cooperation with Marcus Rohrbach (UC Berkley, Facebook AI Research), Mateusz Malinowski (Google Deepmind), Lisa Anne Hendricks (UC Berkley), Jorma Laaksonen, Hamed Tavakoli (Aalto University), Misha Wagner, Hector Basevi, Ales Leonardis (University of Birmingham)

Two popular tasks to measure the ability of computers to relate visual and language modalities is the task of visual description and question answering. While visual description involves generating a summary natural language description of a scene or a video, question answering task requires answering a natural language question about the visual scene. These tasks also have real-world applications in assistive technology for the blind to understand visual media online, and thus has spawned lot of research interest recently. We explore architectural choices to improve image and video description models [6], and an alternate training algorithm to improve the diversity of the generated descriptions [5]. We also propose a joint model to perform video description and provide visual grounding (localization) of the people referred to in the generated description [4]. Finally, we present a scalable end-to-end model for the visual question answering task with a thorough study of design choices which leads to improved performance on the task [1]. By combining such deep learning models with a physics engine we develop a hybrid system which can answer questions about future states of a scene [7].

Image and Video Captioning with Augmented Neural Architectures

Standard approach to generate image and video captions consists of a feature extraction stage where activations from a pre-trained convolutional network is used to represent the image and decoder stage where a recurrent network is used to generate the description based on the visual features. In [6] we propose to augment the dense CNN features with explicit object and scene features which represent the presence or absence of particular object or scene type. By using these explicit features to initialize the decoder, we show improved performance in both image and video captioning tasks. Additionally we propose a method to ensemble multiple video caption generators by training an additional discriminator model which picks the best candidate caption from the pool of captions generated by the ensemble. This combination of augmented features and ensembling enabled us to obtain best rating in human evaluation in the MSR-VTT video description challenge, 2016.

Improving the Diversity in Image Captioning

When tasked with describing images, humans tend to use diverse language which vary in both syntactic structure and the semantic concepts. In contrast, automatic models produce generic descriptions which overuse the most common phrases from the training data. In [5] we address this lack of diversity by training the caption generator in a generative adversarial network (GAN) framework. The generator learns by fooling a discriminator which tries to distinguish generated captions from the human written ones. We design the discriminator take into account both the correctness and the diversity of the captions to
Sophia gags as she pushes past him and walks out. Our: She and Jacob walk down the corridor

Visual Labels [34]: Someone strides to the window.

Figure 36.29: Qualitative results of our approach on the grounded movie description task. Given a previous grounding we predict a sentence, grounding and co-reference. Comparision to our prior work Visual Labels [2].

make this distinction. Our quantitative and qualitative evaluation shows that with the proposed adversarial training, the generated captions are significantly more diverse than the baseline models (e.g. doubling the vocabulary size). The generated captions by our model also better match the n-gram distribution in the human written captions.

Generating Descriptions with Grounded and Co-Referenced People

While we discussed in the previous sections how to generate descriptions, we are also working on jointly performing description and visual grounding. Although a few works have proposed to learn a grounding during the generation process in an unsupervised way (via an attention mechanism), it remains unclear how good the quality of the grounding is and whether it benefits the description quality. In this work we propose a movie description model which learns to generate description and jointly ground (localize) the mentioned characters as well as do visual co-reference resolution between pairs of consecutive sentences/clips. We also propose to use weak localization supervision through character mentions provided in movie descriptions to learn the character grounding. At training time, we first learn how to localize characters by relating their visual appearance to mentions in the descriptions via a semi-supervised approach. We then provide this (noisy) supervision into our description model which greatly improves its performance. Our proposed description model improves over prior work w.r.t. generated description quality and additionally provides grounding and local co-reference resolution. We evaluate it on the MPII Movie Description dataset [3] using automatic and human evaluation measures and using our newly collected grounding and co-reference data for characters, an example result can be seen in Figure 36.29.

Deep Learning and Hybrid Approaches to Visual Question Answering

To solve the task of automatic visual question answering, we require models to understand the visual content, parse the natural language question and generate a natural language
response. Prior work to address this task combined symbolic approaches to parse the question with deep learning based feature extractors to summarize the visual content. In [1], we propose a scalable end-to-end model which combines CNN feature extractors with LSTM based question encoder and answer generator. We also conduct extensive experiments on VQA and DAQUAR datasets to study the design choices on how to encode, combine and decode information in our model. Our experiments show that with end-to-end training this model can achieve double the performance compared to a symbolic approach. Additionally, we report various baselines including “question-only” baselines which measure how well one can answer the question without looking at the image. We show that this is a very strong baseline and a significant portion of questions can be answered without looking at the image by exploiting the biases in the dataset.

The visual question answering task discussed above primarily focuses on describing only the current state of the scenes under consideration. However another important aspect of scene understanding is the ability to forecast the future states of a scene under an action. In [7], we posit this as a visual “what-if” question answering task, where an answer has to be given about a future scene state, given observations of the current scene, and a question that includes a hypothetical action. We develop a hybrid model integrating a physics engine into a deep learning based question answering architecture to solve this task. The physics engine helps the system anticipate future scene states by simulating the object-object interactions caused by the action. We demonstrate first results on this challenging new problem and show that the proposed hybrid approach outperforms data-driven end-to-end learning approaches.

References


36.12 Perceptual User Interfaces

Coordinator: Andreas Bulling

Developing human-computer interfaces that fully exploit the information content available in non-verbal human behaviour is challenging, particularly in unconstrained daily life settings. Key challenges are 1) to develop sensing systems that robustly and accurately capture non-verbal human behaviour in ever-changing conditions, 2) to develop computational methods for automatic analysis and modelling that are able to cope with the large variability in human behaviour, and 3) to use the information extracted from such behaviour to develop novel human-computer interfaces that are highly interactive, multimodal and modelled after natural human-to-human interactions.

Our group works at the interface of human-computer interaction, computer vision, wearable computing, and eye tracking. We develop novel sensing systems and computational methods to analyse non-verbal human behaviour automatically using ambient and on-body sensors. We specifically focus on visual and physical behaviour as we believe that these modalities are most promising for developing interfaces that offer human-like perceptual and interaction capabilities. We study the systems and methods that we develop in the context of specific application domains, most importantly pervasive eye-based human-computer interfaces and computational behaviour analysis.

Specifically, one main line of research is on gaze estimation as a core research challenge towards sensing and understanding human visual behaviour in both mobile and stationary daily-life settings (Sect. 36.12.3 and 36.12.7). Building on these methods, a second main line of research is on gaze-based human-computer interaction (Sect. 36.12.4). Here, we specifically focus on personal calibration methods (Sect. 36.12.1), calibration-free gaze interaction with mobile and ambient displays (Sect. 36.12.6 and 36.12.9) as well as eye contact detection (Sect. 36.12.2). The third main line of research is on computational modelling of interactive behaviour in human-human and human-computer interactions (Sect. 36.12.8 and 36.12.5). Finally, we are active in symbiotic human-machine vision systems (Sect. 36.12.10) and investigate applications in usable security and privacy (Sect. 36.12.11).

36.12.1 Calibration Methods

Investigators: Andreas Bulling, Michael Xuelin Huang, Yusuke Sugano, Xucong Zhang in cooperation with Jiajia Li (South China University of Technology), Grace Ngai, Hong Va Leong (Hong Kong Polytechnic University), Markus Löchtfeld (Aalborg University), and Christian Lander, Sven Gehring, Antonio Krüger (DFKI)

Understanding human visual attention is essential for understanding human cognition, which in turn benefits human-computer interaction. Recent work has demonstrated a Personalized, Auto-Calibrating Eye-tracking (PACE) system, which makes it possible to achieve accurate gaze estimation using only an off-the-shelf webcam by identifying and collecting data implicitly from user interaction events. However, this method is constrained by the need for large amounts of well-annotated data. We thus present fast-PACE in [1], an adaptation to PACE that exploits knowledge from existing data from different users to accelerate the learning speed of the personalized model. The result is an adaptive, data-driven approach
that continuously “learns” its user and recalibrates, adapts, and improves with additional usage by a user. Experimental evaluations of fast-PACE demonstrate its competitive accuracy in iris localization, validity of alignment identification between gaze and interactions, and effectiveness of gaze transfer. In general, fast-PACE achieves an initial visual error of 3.98 degrees and then steadily improves to 2.52 degrees given incremental interaction-informed data. Our performance is comparable to state-of-the-art, but without the need for explicit training or calibration. Our technique addresses the data quality and quantity problems. It therefore has the potential to enable comprehensive gaze-aware applications in the wild.

Learning-based gaze estimation has significant potential to enable attentive user interfaces and gaze-based interaction on the billions of camera-equipped handheld devices and ambient displays. While training accurate person- and device-independent gaze estimators remains challenging, person-specific training is feasible but requires tedious data collection for each target device. To address these limitations, we present the first method to train person-specific gaze estimators across multiple devices in our work [3]. As shown in Figure 36.30, at the core of our method is a single convolutional neural network with shared feature extraction layers and device-specific branches that we train from face images and corresponding on-screen gaze locations. Detailed evaluations on a new dataset of interactions with five common devices (mobile phone, tablet, laptop, desktop computer, smart TV) and three common applications (mobile game, text editing, media center) demonstrate the significant potential
of cross-device training. We further explore training with gaze locations derived from natural interactions, such as mouse or touch input.

Gaze estimation usually requires extensive hardware and depends on a calibration that has to be renewed regularly. We present EyeMirror [2], a mobile device for calibration-free gaze approximation on surfaces (e.g., displays). It consists of a head-mounted camera, connected to a wearable mini-computer, capturing the environment reflected on the human cornea. The corneal images are analyzed using natural feature tracking for gaze estimation on surfaces. In two lab studies we compared variations of EyeMirror against established methods for gaze estimation in a display scenario, and investigated the effect of display content (i.e. number of features). EyeMirror achieved 4.03 degrees gaze estimation error, while we found no significant effect of display content.

References


### 36.12.2 Eye Contact Detection

*Investigators: Andreas Bulling, Xucong Zhang, Julian Steil, Yusuke Sugano, Philipp Müller and Michael Xuelin Huang*

Eye contact is an important non-verbal cue in social signal processing and promising as a measure of overt attention in human-object interactions and attentive user interfaces. Eye contact detection is different to the general gaze estimation problem in that it only requires a categorial decision of whether a user is looking at a certain object or person. In our research, we developed eye contact detection methods which exploit this specific problem setting in order to surpass the performance of general purpose gaze estimation methods.

**Detecting eye contact in the present moment**

In [3], we address the challenge of robust detection of eye contact across different users, gaze targets, camera positions, and illumination conditions. We present a novel method for eye contact detection that combines a state-of-the-art appearance-based gaze estimator with a novel approach for unsupervised gaze target discovery, i.e. without the need for tedious and time-consuming manual data annotation. We evaluate our method in two real-world scenarios: detecting eye contact at the workplace, including on the main work display, from cameras mounted to target objects, as well as during everyday social interactions with the wearer of a head-mounted egocentric camera. We empirically evaluate the performance
of our method in both scenarios and demonstrate its effectiveness for detecting eye contact independent of target object type and size, camera position, and user and recording environment.

Subsequently, we adapted this approach to enable robust eye contact detection between multiple people in real-world settings [1]. Our novel method exploits that, during conversations, people tend to look at the person who is currently speaking. Harnessing the correlation between people’s gaze and speaking behaviour therefore allows our method to automatically acquire training data during deployment and adaptively train eye contact detectors for each target user. We empirically evaluate the performance of our method on a recent dataset of natural group interactions and demonstrate that it achieves a relative improvement over our previous state-of-the-art method [3] of more than 60%.

**Forecasting eye contact**

While detecting eye contact in the present moment is important for many applications, it limits attentive user interfaces to adapt after the fact, i.e. after a shift of overt visual attention has already happened. In [2] we instead study attention forecasting – the challenging task of predicting users’ gaze behavior (overt visual attention) in the near future. We present a novel long-term dataset of everyday mobile phone interactions, continuously recorded from 20 participants engaged in common activities on a university campus over 4.5 hours each (more than 90 hours in total). We propose a proof-of-concept method that uses device-integrated sensors and body-worn cameras to encode rich information on device usage and users’ visual scene. We demonstrate that our method can forecast bidirectional attention shifts and whether the primary attentional focus is on the handheld mobile device. We study the impact of different feature sets on performance and discuss the significant potential but also remaining challenges of forecasting user attention during mobile interactions.

**References**


**36.12.3 Gaze Estimation**

*Investigators: Andreas Bulling, Mario Fritz, Yusuke Sugano, Xucong Zhang in cooperation with Seonwook Park, Otmar Hilliges (ETH Zurich), Tadas Baltrušaitis, Louis-Philippe*
Morency (Carnegie Mellon University), and Erroll Wood, Peter Robinson (University of Cambridge)

Appearance-based Gaze Estimation

Learning-based methods promise to work well for the task of domain-independent gaze estimation, i.e. gaze estimation from a monocular RGB camera without any assumptions regarding the user, environment, or camera itself. However, current gaze estimation datasets were collected in the laboratory and methods were not evaluated across multiple datasets. Our work [6] makes three contributions towards addressing these limitations. First, we present the MPIIGaze dataset that contains 213,659 full face images and corresponding ground-truth gaze positions that we collected from 15 users during everyday laptop use over several months using an experience sampling approach (see Figure 36.31). This approach ensured continuous gaze and head poses and realistic variation in eye appearance and illumination. To facilitate cross-dataset evaluations, 37,667 images were manually annotated with six facial landmarks (eye and mouth corners) and pupil centres. Second, we present an extensive evaluation of state-of-the-art gaze estimation methods on three current datasets, including MPIIGaze. We study key challenges for these methods including target gaze range, illumination conditions, and facial appearance variation. Finally, we propose an appearance-based method based on a VGG deep convolutional neural network that outperforms the previous best method by 22% on MPIIGaze for the most challenging cross-dataset evaluation. The full dataset is available at http://www.mpi-inf.mpg.de/MPIIGaze

Appearance-based gaze estimation is promising for unconstrained real-world settings, but the significant variability in head pose and user-camera distance poses significant challenges for training generic gaze estimators. Data normalization was proposed to cancel out this geometric variability by mapping input images and gaze labels to a normalized space. Although used successfully in prior works, the role and importance of data normalization remains unclear. To fill this gap, we study data normalization for the first time using principled evaluations on both simulated and real data in our work [3]. We propose a modification to the current data normalization formulation by removing the scaling factor and show that our new formulation performs significantly better (between 9.5% and 32.7%) in the different evaluation settings. Using images synthesized from a 3D face model, we demonstrate the benefit of data normalization for the efficiency of the model training. Experiments on real-world images confirm the advantages of data normalization in terms of gaze estimation performance.

Recent gaze estimation work indicated that information from the full face region can benefit performance. Pushing this idea further, we propose an appearance-based method that, in contrast to a long-standing line of work in computer vision, only takes the full face image as input [5]. Our method encodes the face image using a convolutional neural network with spatial weights applied on the feature maps to flexibly suppress or enhance information in different facial regions. Through extensive evaluation, we show that our full-face method significantly outperforms the state of the art for both 2D and 3D gaze estimation, achieving improvements of up to 14.3% on MPIIGaze and 27.7% on EYEDIAP for person-independent 3D gaze estimation. We further show that this improvement is consistent across different illumination conditions and gaze directions and particularly pronounced for the most challenging extreme head poses.
Evaluation of Appearance-Based Methods for Gaze-Based Applications

Appearance-based gaze estimation methods that only require an off-the-shelf camera have significantly improved and promise a wide range of new applications in gaze-based interaction and attentive user interfaces. However, these methods are not yet widely used in the human-computer interaction (HCI) community, partly because it remains unclear how they perform compared to dominant, special-purpose eye tracking equipment. To address this limitation, we evaluate the performance of state-of-the-art appearance-based gaze estimation for interaction scenarios with and without personal calibration, indoor and outdoor, for different sensing distances, as well as for users with and without glasses [4]. We discuss the obtained findings and their implications for current dominant gaze-based applications, specifically explicit eye input, attentive user interfaces, gaze-based user modelling, and passive eye monitoring. Finally, to democratise the use of appearance-based gaze estimation and inter-action in HCI, we present OpenGaze (www.opengaze.org), the first software toolkit that is specifically developed for gaze interface designers.

GazeDirector: Fully Articulated Eye Gaze Redirection in Video

We present GazeDirector in [2], a new approach for eye gaze redirection that uses model-fitting. Our method first tracks the eyes by fitting a multi-part eye region model to video frames using analysis-by-synthesis, thereby recovering eye region shape, texture, pose, and gaze simultaneously. It then redirects gaze by 1) warping the eyelids from the original image using a model-derived flow field, and 2) rendering and compositing synthesized 3D eyeballs onto the output image in a photorealistic manner. GazeDirector allows us to change where people are looking without person-specific training data, and with full articulation, i.e. we can precisely specify new gaze directions in 3D. Quantitatively, we evaluate both model-fitting and gaze synthesis, with experiments for gaze estimation and redirection on the Columbia gaze dataset. Qualitatively, we compare GazeDirector against recent work on gaze redirection, showing better results especially for large redirection angles. Finally, we demonstrate gaze redirection on YouTube videos by introducing new 3D gaze targets and by manipulating visual behavior.
Learning to Find Eye Region Landmarks for Remote Gaze Estimation in Unconstrained Settings

Conventional feature-based and model-based gaze estimation methods have proven to perform well in settings with controlled illumination and specialized cameras. In unconstrained real-world settings, however, such methods are surpassed by recent appearance-based methods due to difficulties in modeling factors such as illumination changes and other visual artifacts. We present a novel learning based method [1] for eye region landmark localization that enables conventional methods to be competitive to latest appearance-based methods. Despite having been trained exclusively on synthetic data, our method exceeds the state of the art for iris localization and eye shape registration on real-world imagery. We then use the detected landmarks as input to iterative model-fitting and lightweight learning-based gaze estimation methods. Our approach outperforms existing model-fitting and appearance-based methods in the context of person-independent and personalized gaze estimation.

References


36.12.4 Gaze Interaction

*Investigators: Andreas Bulling, Yusuke Sugano, Michaela Klauck in cooperation with Mohamed Khamis, Carl Oechsner, Thomas Mattusch, Anna Kienle, Florian Alt (LMU Munich), Mahsa Mirzamohammad (University of Jyvaskyla)*

**VRPursuits: Interaction in Virtual Reality using Smooth Pursuit Eye Movements**

Gaze-based interaction using smooth pursuit eye movements (Pursuits) is attractive given that it is intuitive and overcomes the Midas touch problem. At the same time, eye tracking is becoming increasingly popular for VR applications. While Pursuits was shown to be effective in several interaction contexts, it was never explored in-depth for VR before. In [2], in a user
study (N=26), we investigated how parameters that are specific to VR settings influence the performance of Pursuits. We found that Pursuits is robust against different sizes of virtual 3D targets and sizes to them. However, Pursuits’ performance improves when the trajectory size is larger, particularly if the user is walking while interacting. While walking, selecting moving targets via Pursuits is generally feasible albeit less accurate than when stationary. Finally, we discuss the implications of these findings and the potential of smooth pursuits for interaction in VR by demonstrating two sample use cases: 1) gaze-based authentication in VR, and 2) a space meteors shooting game.

GazeDrone: Mobile Eye-Based Interaction in Public Space Without Augmenting the User

Gaze interaction holds a lot of promise for seamless human-computer interaction. At the same time, current wearable mobile eye trackers require user augmentation that negatively impacts natural user behavior while remote trackers require users to position themselves within a confined tracking range. In [1], we presented GazeDrone, the first system that combines a camera-equipped aerial drone with a computational method to detect sidelong glances for spontaneous (calibration-free) gaze-based interaction with surrounding pervasive systems (e.g., public displays). GazeDrone does not require augmenting each user with on-body sensors and allows interaction from arbitrary positions, even while moving. We demonstrated that drone-supported gaze interaction is feasible and accurate for certain movement types. It was well-perceived by users, in particular while interacting from a fixed position as well as while moving orthogonally or diagonally to a display. We presented design implications and discussed opportunities and challenges for drone-supported gaze interaction in public.

Hidden Pursuits: Evaluating Gaze-selection via Pursuits when the Stimulus Trajectory is Partially Hidden

The idea behind gaze interaction using Pursuits is to leverage the human’s smooth pursuit eye movements performed when following moving targets. However, humans can also anticipate where a moving target would reappear if it temporarily hides from their view. In [4], we investigated how well users can select targets using Pursuits in cases where the target’s trajectory is partially invisible (HiddenPursuits): e.g., can users select a moving target that temporarily hides behind another object? Although HiddenPursuits was not studied in the context of interaction before, understanding how well users can perform HiddenPursuits presents numerous opportunities, particularly for small interfaces where a target’s trajectory can cover area outside of the screen. We found that users can still select targets quickly via Pursuits even if their trajectory is up to 50% hidden, and at the expense of longer selection times when the hidden portion is larger. We discussed how gaze-based interfaces can leverage HiddenPursuits for an improved user experience.
Noticeable or Distractive? A Design Space for Gaze-Contingent User Interface Notifications

Users are interrupted by an ever-increasing number of notifications, ranging from error messages, over new email or chat alerts, to advertisement pop-ups. In [3], we explored gaze-contingent user interfaces notifications that are shown depending on users’ current gaze location. Specifically, we evaluated how different design properties influence notification noticeability and distractiveness. We measured noticeability quantitatively by analyzing participants’ performance in confirming notifications and distractiveness using a questionnaire. Based on a 12-participant user study on a public display, we showed that each of these properties affects noticeability and distractiveness differently and that the properties, in turn, allow for fine-grained optimization of notification display. These findings inform the design of future attentive user interfaces that could optimize the trade-off between, for example, the notification importance and the cost of interruption.

References


36.12.5 Interaction Behaviour Understanding

Investigators: Michael Xuelin Huang, Andreas Bulling in cooperation with Jun Wang, Grace Ngai, Hong-Va Leong, Tiffany C.K. Kwok, Eugene Yujun Fu, Erin You Wu (Hong Kong Polytechnic University)

Subtle interaction behaviour can reflect users’ mental state, such as mental stress and user intention. We therefore conducted a series of studies that explored and analysed interaction signals, including gaze and mouse behaviors as well as their coordination to infer users’ mental states and intention in different interaction tasks. More specifically, we firstly investigated the behaviour consistency of both gaze and mouse movements within a specific task for mental stress inference [4]. We then studied the coordination between gaze and mouse for stress detection [3]. We are also interested in the understanding of user behaviours in crowdsourcing tasks and the design of intelligent interfaces that can alleviate human workload and increase efficiency. As such, we performed studies to collect a corpus of a manual segmentation task [1] and we also made attempt to predict user intention in a web-search task [2].
Infer Mental Stress from Gaze and Mouse Behaviours

Stress can deteriorate human’s attention and memory, which, when a user is engaged in interactive applications, will negatively affect the user experience and downgrade the delivered performance. Traditional stress inference is mainly based on user physical features like Blood Volume Pulse, Galvanic Skin Response, often captured via devices that intrude on the user space. In contrast, in [4] we propose a non-intrusive approach that exploits the consistency of users’ behavioral patterns when interacting with a user interface, specifically, in terms of eye gaze and mouse movement. The relationship between the stress experienced by the user and his/her eye gaze and gaze-mouse coordination patterns are investigated. We show that both eye gaze and gaze-mouse coordination patterns can be exploited to distinguish whether a user is under stress. We also discover that a user’s eye gaze behavior patterns are more consistent when he/she is under stress.

In addition to the consistency between gaze and mouse movement patterns, mouse-gaze coordination is also indicative of stress level. Despite prior success in non-intrusive stress detection through mouse and gaze dynamics, the dependency on layout information reduces the method generalizability in real usage, where most of the tasks have dynamic graphical user interfaces (GUIs). We therefore present a GUI-agnostic stress detection system: MoGa [3], to infer user stress level in tasks with dynamic GUIs. MoGa adopts an innovative mouse-gaze attraction model, which leverages the coordination between mouse and eye movement without considering GUI information. We evaluated our model in tasks with dynamic and fixed GUIs. Our experimental results demonstrate the effectiveness of MoGa and the underlying mouse-gaze attraction model, which can achieve 74.3% CCR of detecting stress in dynamic GUIs tasks, beat the state of the art approach by around 15%. It opens up a new avenue for cognitive-aware intelligent user interface and numerous advanced HCI studies.

Understand User Behaviour and Intention in Interaction Tasks

Given the mounting demand of data annotation and the need of automatic segmentation, we studied users’ gaze and interaction behaviours (on-screen content, mouse and keystroke dynamics) in a popular annotation task, i.e. phonetic segmentation. Phonetic segmentation is the process of splitting speech into distinct phonetic units. Human experts routinely perform this task manually by analyzing auditory and visual cues using analysis software, which is an extremely time-consuming process. Methods exist for automatic segmentation, but these are not always accurate enough. In order to improve automatic segmentation, we need to model it as close to the manual segmentation as possible. We therefore collected a corpus [1] that capture the human segmentation behavior by recording experts performing a segmentation task. We believe that this data will enable us to highlight the important aspects of manual segmentation, which can be used in automatic segmentation to improve its accuracy.

Besides, user intention is also of great interest to intelligent user interface. User experience could be enhanced if the computer could understand human interaction intention. In [2], we present an approach to predicting users’ intention in interaction tasks based on past mouse movements. We adopt a long short-term memory (LSTM) model to predict the users’ intention via their next mouse click interaction, upon being trained with past mouse
interaction behaviors. To evaluate, we consider two scenarios in daily computer usage: a more structured crowdsourcing annotation task and a more free-form, open-ended web search task. Our results indicate that we could predict the next interaction event with reasonable accuracy. We also conducted a pilot study to investigate the possibility of applying our model for non-intentional mouse click detection. We believe that our findings would be beneficial towards the development of better intelligent agents.

References


### 36.12.6 Interaction with Mobile Devices

**Investigators:** Andreas Bulling, Michael Xuelin Huang in cooperation with Mohamed Khamis, Anita Baier, Niels Henze, Florian Alt, Linda Bandelow, Stina Schick, Dario Casadevall (LMU Munich), Jiajia Li, Grace Ngai, Hong Va Leong (The Hong Kong Polytechnic University)

Gaze-based interaction with mobile devices has attracted quite a lot of research attention. However, most of the current techniques are devised for desktop applications. Only a few attempts have been made to explore mobile platforms, such as tablets. Prior studies on mobile gaze estimation generally produce lower accuracy than their counterparts in the desktop settings, probably due to user motion, severe head pose variations, or varying lighting conditions. Therefore, we analysed face and eye visibility in front-facing cameras of smartphones in the wild [2] and surveyed the past, present, and future of eye tracking on handheld mobile devices [1].

**Understanding Face and Eye Visibility in Front-Facing Cameras of Smartphones used in the Wild**

Commodity mobile devices are now equipped with high-resolution front-facing cameras, allowing applications in biometrics (e.g., FaceID in the iPhone X), facial expression analysis, or gaze interaction. However, it is unknown how often users hold devices in a way that allows capturing their face or eyes, and how this impacts detection accuracy. In [2], we collected...
Figure 36.32: Eye tracking and eye-based interaction on handheld mobile devices required external hardware in the past. Recent advancements in processing power of off-the-shelf devices, and their front-facing cameras made eye tracking and eye-based interaction feasible on unmodified handhelds. At this stage, the next step is pervasive eye tracking on handheld mobile devices in the wild.

25,726 in-the-wild photos, taken from the front-facing camera of smartphones as well as associated application usage logs. We found that the full face is visible about 29% of the time, and that in most cases the face is only partially visible. Furthermore, we identified an influence of users’ current activity; for example, when watching videos, the eyes but not the entire face are visible 75% of the time in our dataset. We found that a state-of-the-art face detection algorithm performs poorly against photos taken from front-facing cameras. We discuss how these findings impact mobile applications that leverage face and eye detection, and derive practical implications to address state-of-the-art’s limitations.

The Past, Present, and Future of Gaze-enabled Handheld Mobile Devices: Survey and Lessons Learned

While first-generation mobile gaze interfaces required special-purpose hardware, recent advances in computational gaze estimation and the availability of sensor-rich and powerful devices is finally fulfilling the promise of pervasive eye tracking and eye-based interaction on off-the-shelf mobile devices. This work [1] provides the first holistic view on the past, present, and future of eye tracking on handheld mobile devices. To this end, we discuss how research developed from building hardware prototypes, to accurate gaze estimation on unmodified smartphones and tablets. We then discuss implications by laying out 1) novel opportunities, including pervasive advertising and conducting in-the-wild eye tracking studies on handelds, and 2) new challenges that require further research, such as visibility of the user’s eyes, lighting conditions, and privacy implications. We discuss how these developments shape
MobileHCI research in the future, possibly the next 20 years. A history timeline of eye tracking and gaze-based interaction on handheld mobile devices is displayed in Figure 36.32.

References


36.12.7 Mobile Eye Tracking

Investigators: Julian Steil, Marc Tonsen, Yusuke Sugano, Michael Xuelin Huang, Andreas Bulling in cooperation with Marion Koelle, Susanne Boll (University of Oldenburg), Wilko Heuten (OFFIS – Institute for IT)

Research concerning mobile eye tracking can be distinguished into three main areas.

The first area considers research in sensing like common eye tracking data or multi-modal sensing hardware. Classical approaches for gaze estimation require high-resolution imaging sensors, resulting in rather bulky and obtrusive headsets as well as hand-optimised algorithms for eye landmark detection and geometric gaze mapping. InvisibleEye is a novel approach for mobile eye tracking that uses millimetre-size RGB cameras that can be fully embedded into normal glasses frames [3].

The second research area builds on the provided data from the used sensing modalities and analyse this data, e.g., how to identify eye movement like fixations. A key property of state-of-the-art methods is that they rely solely on gaze data, i.e. they typically do not take any other information into account, such as the target being looked at. This approach works well for remote eye trackers used in stationary settings in which the estimated gaze is analysed within a fixed frame of reference, i.e. the screen coordinate system. In contrast, fixation detection for head-mounted eye trackers and mobile settings is significantly more challenging. Gaze estimates are typically given in the eye tracker’s scene camera coordinate system but this frame of reference changes constantly with respect to the world coordinate system as the wearer moves around or turns his head while looking at a target. As a result, gaze estimates during a fixation seem to shift within the scene camera coordinate system, resulting in failures of fixation detection methods that rely solely on gaze information. Maintaining gaze on a particular real-world target consequently involves a complex combination of fixations, smooth pursuit, and vestibulo-ocular reflex movements. In [1], we are the first to address the challenging task of fixation detection for head-mounted eye tracking combining scene and gaze information.

In the third research area, useful, real-world applications are implemented. Eyewear devices, such as head-mounted displays or augmented reality glasses, have recently emerged as a new research platform in fields such as human-computer interaction, computer vision, or the behavioural and social sciences. An ever-increasing number of these devices integrate eye
tracking capabilities given their significant potential for analysing and better understanding users’ attention allocation, for computational user modeling, and as a versatile means for interaction. Head-mounted mobile eye tracking typically requires two cameras: An eye camera that records a close-up video of the eye and a high-resolution first-person (scene) camera to map gaze estimates to the real-world scene. The scene camera poses a serious privacy risk given that it may record highly sensitive personal information, such as login credentials, banking information, or personal text messages, and it may potentially allow individuals to be identified, thus also infringing on the privacy of bystanders. This privacy risk will become even more pervasive and severe with the unnoticeable integration of eye tracking in ordinary glasses frames. Therefore, we propose PrivacEye [2], the first method for privacy-preserving head-mounted eye tracking. The key idea and core novelty of our method is to detect users’ transitions into and out of privacy-sensitive everyday situations by leveraging both cameras available on these trackers.

InvisibleEye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning-Based Gaze Estimation

Analysis of everyday human gaze behaviour has significant potential for ubiquitous computing, as evidenced by a large body of work in gaze-based human-computer interaction, attentive user interfaces, and eye-based user modelling. However, current mobile eye trackers are still obtrusive, which not only makes them uncomfortable to wear and socially unacceptable in daily life, but also prevents them from being widely adopted in the social and behavioural sciences. To address these challenges, in [3], we present InvisibleEye, a novel approach for mobile eye tracking that uses millimetre-size RGB cameras that can be fully embedded into normal glasses frames. To compensate for the cameras’ low image resolution of only a few pixels, our approach uses multiple cameras to capture different views of the eye, as well as learning-based gaze estimation to directly regress from eye images to gaze directions. We prototypically implement our system and characterise its performance on three large-scale, increasingly realistic, and thus challenging datasets: 1) eye images synthesised using a recent computer graphics eye region model, 2) real eye images recorded of 17 participants under controlled lighting, and 3) eye images recorded of four participants over the course of four recording sessions in a mobile setting. We show that InvisibleEye achieves a top person-specific gaze estimation accuracy of 1.79° using four cameras with a resolution of only 5×5 pixels. Our evaluations not only demonstrate the feasibility of this novel approach but, more importantly, underline its significant potential for finally realising the vision of invisible mobile eye tracking and pervasive attentive user interfaces.

Fixation Detection for Head-Mounted Eye Tracking Based on Visual Similarity of Gaze Targets

Fixations are widely analysed in human vision, gaze-based interaction, and experimental psychology research. However, robust fixation detection in mobile settings is profoundly challenging given the prevalence of user and gaze target motion. These movements feign a shift in gaze estimates in the frame of reference defined by the eye tracker’s scene camera. To address this challenge, in [1], we present a novel fixation detection method for head-mounted
Figure 36.33: Overview of our method. Inputs to our method are scene camera frames with corresponding gaze estimates. First, our method (a) extracts gaze patches around the gaze estimates and (b) then computes similarity values with a state-of-the-art deep convolutional image patch similarity network. (c) In the next step, the similarity values are thresholded to classify patch pairs into fixation candidates. (d) Finally, fixation candidates are checked for a minimum length.

Our method exploits that, independent of user or gaze target motion, target appearance remains about the same during a fixation. It extracts image information from small regions around the current gaze position and analyses the appearance similarity of these gaze patches across video frames to detect fixations. We evaluate our method using fine-grained fixation annotations on a five-participant indoor dataset (MPIIEgoFixation) with more than 2,300 fixations in total. Our method outperforms commonly used velocity- and dispersion-based algorithms, which highlights its significant potential to analyse scene image information for eye movement detection.

**PrivacEye: Privacy-Preserving Head-Mounted Eye Tracking Using Egocentric Scene Image and Eye Movement Features**

Eyewear devices, such as augmented reality displays, increasingly integrate eye tracking but the first-person camera required to map a user’s gaze to the visual scene can pose a significant threat to user and bystander privacy. In [2], we present **PrivacEye**, a method to detect privacy-sensitive everyday situations and automatically enable and disable the eye tracker’s first-person camera using a mechanical shutter. To close the shutter in privacy-sensitive situations, the method uses a deep representation of the first-person video combined with rich features that encode users’ eye movements. To open the shutter without visual input, PrivacEye detects changes in users’ eye movements alone to gauge changes in the “privacy level” of the current situation. We evaluate our method on a first-person video dataset recorded in daily life situations of 17 participants, annotated by themselves for privacy sensitivity, and show that our method is effective in preserving privacy in this challenging setting.

**References**


36.12.8 Non-Verbal Behaviour Analysis

Investigators: Andreas Bulling, Michael Xuelin Huang, Philipp Müller, Sabrina Hoppe, in cooperation with Tobias Loetscher (University of South Australia, Australia), Stephanie A. Morey (Flinders University, Australia), Jiajia Li, Grace Ngai, Hong Va Leong, Eugene Yujun Fu (The Hong Kong Polytechnic University, Hong Kong) and Kien A. Hua (University of Central Florida, USA)

Nonverbal Behaviour in Social Interactions

Analysing the nonverbal behaviour of people can help machines to infer higher-level aspects of social situations. One particularly important aspect of social interactions is rapport, the close and harmonious relationship in which interaction partners are “in sync” with each other. Rapport was shown to result in smoother social interactions, improved collaboration, and improved interpersonal outcomes. In [4], we are first to investigate automatic prediction of low rapport during natural interactions within small groups. This task is challenging given that rapport only manifests in subtle non-verbal signals that are, in addition, subject to influences of group dynamics as well as inter-personal idiosyncrasies. We record videos of unscripted discussions of three to four people using a multi-view camera system and microphones. We analyse a rich set of non-verbal signals for rapport detection, namely facial expressions, hand motion, gaze, speaker turns, and speech prosody. Using facial features, we can detect low rapport with an average precision of 0.7 (chance level at 0.25), while incorporating prior knowledge of participants’ personalities can even achieve early prediction without a drop in performance. We further provide a detailed analysis of different feature sets and the amount of information contained in different temporal segments of the interactions.

Another aspect of social interactions, which is especially important in the context of surveillance, is the presence of human fight behaviour. Detecting such behaviour from videos is challenging, as data of real human fight events is scarce, hampering the performance of data-driven approaches. In [1] we address this challenge by presenting a novel cross-species learning method with a set of low-computational cost motion features for fight detection. It effectively circumvents the problem of limited human fight data for data-demanding approaches. Our method exploits the intrinsic commonality between human and animal fights, such as the physical acceleration of moving body parts. It also leverages an ensemble learning mechanism to adapt useful knowledge from similar source subsets across species. Our evaluation results demonstrate the effectiveness of the proposed feature representation for cross-species adaptation. We believe that cross-species learning is not only a promising solution to the data constraint issue, but it also sheds lights on the studies of other human mental and social behaviors in cross-disciplinary research.
User Modelling

In order for machines to naturally interact with users, knowledge about the personality of users can be helpful. In [2] we show that eye movements during an everyday task can be utilized to predict aspects of our personality. We tracked eye movements of 42 participants while they ran an errand on a university campus and subsequently assessed their personality traits using well-established questionnaires. Using a state-of-the-art machine learning method and a rich set of features encoding different eye movement characteristics, we were able to reliably predict four of the Big Five personality traits (neuroticism, extraversion, agreeableness, conscientiousness) as well as perceptual curiosity only from eye movements. Further analysis revealed new relations between previously neglected eye movement characteristics and personality. Our findings demonstrate a considerable influence of personality on everyday eye movement control, thereby complementing earlier studies in laboratory settings. Improving automatic recognition and interpretation of human social signals is an important endeavor, enabling innovative design of human-computer systems capable of sensing spontaneous natural user behavior to facilitate efficient interaction and personalization.

On a lower level, user modelling can be used to improve tasks like facial affect recognition. Existing solutions to user-specific facial affect recognition, however, have usability issues since the annotation can be long and tedious. In [3], we address this issue by proposing a novel user-adaptive model, which we have called fast-Personal Affect Detection with Minimal Annotation (Fast-PADMA). Fast-PADMA integrates data from multiple source subjects with a small amount of data from the target subject. Collecting this target subject data is feasible since fast-PADMA requires only one self-reported affect annotation per facial video segment. To alleviate overfitting in this context of limited individual training data, we propose an efficient bootstrapping technique, which strengthens the contribution of multiple similar source subjects. Specifically, we employ an ensemble classifier to construct pretrained weak generic classifiers from data of multiple source subjects, which is weighted according to the available data from the target user. The result is a model that does not require expensive computation, such as distribution dissimilarity calculation or model retraining. We evaluate our method with in-depth experimental evaluations on five publicly available facial datasets, with results that compare favorably with the state-of-the-art performance on classifying pain, arousal, and valence. Our findings show that fast-PADMA is effective at rapidly constructing a user-adaptive model that outperforms both its generic and user-specific counterparts. This efficient technique has the potential to significantly improve user-adaptive facial affect recognition for personal use and, therefore, enable comprehensive affect-aware applications.

References


36.12.9 Public Display Interaction

Investigators: Andreas Bulling in cooperation with Florian Alt, Mohamed Khamis, Christian Becker, Daniel Buschek, Tobias Thieron, Axel Hoesl, Alexander Klimeczak, Martin Reiss (LMU Munich), Michael Barz, Florian Daiber, Daniel Sonntag (German Research Center for Artificial Intelligence), Yanxia Zhang (Delft University of Technology), Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Hans Gellersen (Lancaster University)

As the advance of eye tracking techniques, increasing attention has been paid to gaze based interaction with public displays. Our recent studies investigated important yet previously unexplored issues, including multi-person interaction with large-size public displays [4, 2, 5] and error correction [3, 1] in such scenarios. These studies tackled the most practical issues relevant to gaze-based public display interaction. Specifically, EyeScout [4] uses a movable eye tracker that follows a user to address the constrained tracking scope of conventional eye trackers. We also studied the multi-person use cases for public display interaction, including the self-perception enhancement in a group of users [2] as well as the awareness enhancement of gaze location of the remote collaborator [5]. However, novel gaze-based interactions come with new sources of error. We thus conducted studies to handle the error caused by parallax effect [3] and to explore error-aware interfaces [1] for public display interaction.

Multi-Person Interaction with Large-Size Public Displays

Interacting with large-size public displays is one of the most interesting and popular scenarios in the field of public display interaction. While gaze holds much promise for hands-free interaction with public displays, remote eye trackers with their confined tracking box restrict users to a single stationary position in front of the display. We therefore propose EyeScout [4], an active eye tracking system that combines an eye tracker mounted on a rail system with a computational method to automatically detect and align the tracker with the user’s lateral movement. EyeScout addresses key limitations of current gaze-enabled large public displays by offering two novel gaze-interaction modes for a single user: In “Walk then-Interact” the user can walk up to an arbitrary position in front of the display and interact, while in “Walk and Interact” the user can interact even while on the move. Our user study shows that EyeScout is well perceived by users, extends a public display’s sweet spot into a sweet line, and reduces gaze interaction kick-off time to 3.5 seconds – a 62% improvement over state of the art solutions.

Multi-person interaction with public display is another common scenario with numerous interesting research issues, such as user perception and collaboration. While user representations are extensively used on public displays, it remains unclear how well users can recognize their own representation among those of surrounding users. In [2] we studied the most widely used representations: abstract objects, skeletons, silhouettes and mirrors. We identify five strategies that users follow to recognize themselves on public displays and we quantify the users’ recognition time and accuracy with respect to each representation type. Our findings
suggest that there is a significant effect of (1) the representation type, (2) the strategies performed by users, and (3) the combination of both on recognition time and accuracy.

Besides the enhancement of user perception, eye tracking technique can also be of great use to enhance user collaboration in public display interaction. Gaze information provides indication of users focus which complements remote collaboration tasks, as distant users can see their partner’s focus. We apply gaze for co-located collaboration [5], where users’ gaze locations are presented on the same display, to help collaboration between partners. We integrated various types of gaze indicators on the user interface of a collaborative search system, and we conducted two user studies to understand how gaze enhances coordination and communication between co-located users. Our results show that gaze indeed enhances co-located collaboration, but with a trade-off between visibility of gaze indicators and user distraction. Users acknowledged that seeing gaze indicators eases communication, because it let them be aware of their partner’s interests and attention.

Gaze Error Correction in Public Display Interaction

Eye gaze information can be used to improve touch accuracy in public display by exploiting parallax effect. The parallax effect describes the displacement between the perceived and detected touch locations on a touch-enabled surface. Parallax is a key usability challenge for interactive displays, particularly for those that require thick layers of glass between the screen and the touch surface to protect them from vandalism. To address this challenge, we present EyePACT [3], a method that compensates for input error caused by parallax on public displays. Our method uses a display-mounted depth camera to detect the user’s 3D eye position in front of the display and the detected touch location to predict the perceived touch location on the surface. Our evaluations demonstrate that EyePACT (1) significantly improves accuracy even with varying gap distances between the touch surface and the display, (2) adapts to different levels of parallax by resulting in significantly larger corrections with larger gap distances, and (3) maintains a significantly large distance between two users’ fingers when interacting with the same object. These findings are promising for the development of future parallax-free interactive displays.

Accurate gaze information can facilitate public display interaction, however, gaze estimation error may still occur in practice for different reasons. Gaze estimation error can severely hamper usability and performance of mobile gaze-based interfaces given that the error varies constantly for different interaction positions. We explore error-aware gaze-based interfaces that estimate and adapt to gaze estimation error on-the-fly. We implement a sample error-aware user interface for gaze-based selection and different error compensation methods: a na¨ıve approach that increases component size directly proportional to the absolute error, a recent model based on the two-dimensional error distribution, and a novel predictive model that shifts gaze by a directional error estimate. We evaluate these models in a 12-participant user study and show that our predictive model significantly outperforms the others in terms of selection rate, particularly for small gaze targets. These results underline both the feasibility and potential of next generation error-aware gaze-based user interfaces.
36 D2: Computer Vision and Multimodal Computing

References


36.12.10 Symbiotic Human-Machine Vision

Investigators: Hosnieh Sattar, Andreas Bulling, Mario Fritz

Human-machine symbiosis describes a vision of deeply integrated and cooperative interactions between men and machines. The goal is to facilitate computers with formulative thinking, and let men and computers take cooperative decisions in complex situations instead of being dependent on predetermined programs. Such cooperation can be facilitated using explicit human input or implicitly by monitoring human behaviour. A rich source of implicit information is in human gaze, as it reflects processes of cognition and perception of the visual scene. With eye tracking becoming a commodity, this opens up exciting new opportunities in particular for symbiotic human-machine vision systems in which part of the visual processing is carried out by the machine while another part is performed by a human and conveyed to the machine via gaze information.

We have explored such symbiotic human-machine vision systems in several projects ranging from inferring the category and attributes of visual search targets [2], decoding the visual search targets from fixation data [3], and Gaze Embeddings for Zero-Shot Image Classification [1].

Predicting the Category and Attributes of Mental Pictures Using Deep Gaze Pooling

Previous work focused on predicting visual search targets from human fixations but, in the real world, a specific target is often not known, e.g. when searching for a present for a friend. In this work we instead study the problem of predicting the mental picture, i.e. only an abstract idea instead of a specific target. This task is significantly more challenging given that mental pictures of the same target category can vary widely depending on personal biases, and given that characteristic target attributes can often not be verbalised explicitly. In [2] we instead propose to use gaze information as implicit information on users’ mental...
Figure 36.34: We propose a method to predict the mental picture – consisting of abstract ideas – from users’ gaze. Our method uses gaze information as implicit information and a novel gaze pooling layer to seamlessly integrate localized fixation information on images (as attention masks) into a deep image representation.

picture and present a novel gaze pooling layer to seamlessly integrate semantic and localized fixation information into a deep image representation. We show that we can robustly predict both the mental picture’s category as well as attributes on a novel dataset containing fixation data of 14 users searching for targets on a subset of the DeepFashion dataset. Our results have important implications for future search interfaces and suggest deep gaze pooling as a general-purpose approach for gaze-supported computer vision systems.

Visual Decoding of Targets During Visual Search From Human Eye Fixations

What does human gaze reveal about a users’ intents and to which extend can these intends be inferred or even visualized? Gaze was proposed as an implicit source of information to predict the target of visual search and, more recently [2], to predict the object class and attributes of the search target. In this work [3], we go one step further and investigate the feasibility of combining recent advances in encoding human gaze information using deep convolutional neural networks with the power of generative image models to visually decode, i.e. create a visual representation of, the search target. Such visual decoding is challenging for two reasons: 1) the search target only resides in the user’s mind as a subjective visual pattern, and can most often not even be described verbally by the person, and 2) it is, as of yet, unclear if gaze fixations contain sufficient information for this task at all. We show, for the first time, that visual representations of search targets can indeed be decoded only from human gaze fixations. We propose to first encode fixations into a semantic representation and then decode this representation into an image. We evaluate our method on a recent gaze dataset of 14 participants searching for clothing in image collages and validate the model’s predictions using two human studies. Our results show that 62% (Chance level = 10%) of the time users were able to select the categories of the decoded image right. In our second studies we show the importance of a local gaze encoding for decoding visual search targets of users.
Gaze Embeddings for Zero-Shot Image Classification

Zero-shot image classification using auxiliary information, such as attributes describing discriminative object properties, requires time-consuming annotation by domain experts. We instead propose a method [1] that relies on human gaze as auxiliary information, exploiting that even non-expert users have a natural ability to judge class membership. We present a data collection paradigm that involves a discrimination task to increase the information content obtained from gaze data. Our method extracts discriminative descriptors from the data and learns a compatibility function between image and gaze using three novel gaze embeddings shown in Fig. 36.35: Gaze Histograms (GH), Gaze Features with Grid (GFG) and Gaze Features with Sequence (GFS). We introduce two new gaze-annotated datasets for fine-grained image classification and show that human gaze data is indeed class discriminative, provides a competitive alternative to expert-annotated attributes, and outperforms other baselines for zero-shot image classification.

References


36.12.11 Usable Security

Investigators: Andreas Bulling, Michael Xuelin Huang, Julian Steil in cooperation with Mohamed Khamis, Mariam Hassib, Regina Hasholzner, Emanuel von Zezschwitz, Linda Bandelow, Stina Schick, Dario Casadevall, Florian Alt (LMU Munich), Inken Hagestedt (CISPA Helmholtz Center for Information Security)

In this section, we discuss our research into methods which improve users’ privacy and security, such as by using gaze as an authentication factor.

They are all after you: Investigating the Viability of a Threat Model that involves Multiple Shoulder Surfers

Many of the authentication schemes for mobile devices that were proposed lately complicate shoulder surfing by splitting the attacker’s attention into two or more entities. For example, multimodal authentication schemes such as GazeTouchPIN and GazeTouchPass require attackers to observe the user’s gaze input and the touch input performed on the phone’s screen. These schemes have always been evaluated against single observers, while multiple observers could potentially attack these schemes with greater ease, since each of them can focus exclusively on one part of the password. In [1], we studied the effectiveness of a novel threat model against authentication schemes that split the attacker’s attention. As a case study, we reported on a security evaluation of two state of the art authentication schemes in the case of a team of two observers. Our results showed that although multiple observers perform better against these schemes than single observers, multimodal schemes are significantly more secure against multiple observers compared to schemes that employ a single modality. We discussed how this threat model impacts the design of authentication schemes.

GazeTouchPIN: Protecting Sensitive Data on Mobile Devices using Secure Multimodal Authentication

Although mobile devices provide access to a plethora of sensitive data, most users still only protect them with PINs or patterns, which are vulnerable to side-channel attacks (e.g., shoulder surfing). However, prior research has shown that privacy-aware users are willing to take further steps to protect their private data. In [3], we proposed GazeTouchPIN, a novel secure authentication scheme for mobile devices that combines gaze and touch input. Our multimodal approach complicates shoulder-surfing attacks by requiring attackers to observe the screen as well as the user’s eyes to find the password. We evaluated the security and usability of GazeTouchPIN in two user studies (N=30). We found that while GazeTouchPIN requires longer entry times, privacy aware users would use it on-demand when feeling observed or when accessing sensitive data. The results showed that successful shoulder surfing attack rate drops from 68% to 10.4% when using GazeTouchPIN.
GTmoPass: Two-factor Authentication on Public Displays Using GazeTouch passwords and Personal Mobile Devices

As public displays continue to deliver increasingly private and personalized content, there is a need to ensure that only the legitimate users can access private information in sensitive contexts. While public displays can adopt similar authentication concepts like those used on public terminals (e.g., ATMs), authentication in public is subject to a number of risks. Namely, adversaries can uncover a user’s password through (1) surfing users, (2) thermal attacks, or (3) smudge attacks. To address this problem, in [2] we proposed GTmoPass, an authentication architecture that enables Multi-factor user authentication on public displays. The first factor is a knowledge-factor: we employed a shoulder-surfing resilient multimodal scheme that combined gaze and touch input for password entry. The second factor is a possession-factor: users utilize their personal mobile devices, on which they enter the password. Credentials are securely transmitted to a server via Bluetooth beacons. We described the implementation of GTmoPass and reported on an evaluation of its usability and security, which showed that although authentication using GTmoPass is slightly slower than traditional methods, it protected against the three aforementioned threats.

Privacy-Aware Eye Tracking Using Differential Privacy

With eye tracking being increasingly integrated into virtual and augmented reality (VR/AR) head-mounted displays, preserving users’ privacy is an ever more important, yet under-explored, topic in the eye tracking community. In [4] we reported a large-scale online survey (N=124) on privacy aspects of eye tracking that provides the first comprehensive account of with whom, for which services, and to which extent users are willing to share their gaze data. Using these insights, we designed a privacy-aware VR interface that uses differential privacy, which we evaluated on a new 20-participant dataset for two privacy sensitive tasks: We showed that our method could prevent user re-identification and protected gender information while maintaining high performance for gaze-based document type classification. Our results highlighted the privacy challenges particular to gaze data and demonstrated that differential privacy is a potential means to address them. Thus, this approach layed important foundations for future research on privacy-aware gaze interfaces.

References


342
36.13 Image and Video Segmentation

*Coordinators: Anna Khoreva and Bernt Schiele*

Segmentation of images and videos is a very challenging task in computer vision and one of the most crucial steps toward image and video understanding. A variety of applications, such as activity recognition, automatic driving and image/video indexing, utilize at some point a segmentation algorithm and their performance depends highly on the quality of the latter.

Employing convolutional networks for image and video segmentation tasks requires a large amount of training data with expensive pixel-level annotations. To reduce the annotation effort and thus make the training for new domains more affordable, in Sect. 36.13.1 we propose to explore weaker forms of supervision for training image segmentation networks, such as image-level labels and bounding boxes instead of expensive object mask annotations. In Sect. 36.13.2 we broaden the scope to video object segmentation and investigate ways to reduce the necessity for large volumes of densely annotated video data for training deep networks by leveraging available static images, sophisticated data augmentation techniques and exploiting language referring expressions.

One important direction is semantic scene labeling, where the task is to assign a unique label (or category) to every single pixel in the image, which can be considered as a dense classification problem. This field has achieved prominent progress mainly due to deep convolutional neural networks, which boost the ability of dynamic object understanding. However, the problem is far from being solved. In Sect. 36.13.3 we explore ways for improving semantic segmentation approaches by leveraging unlabeled and synthetic data for training and designing new compatible modules to existing network architectures.

### 36.13.1 Image Segmentation with Weak Labels

*Investigators: Anna Khoreva, Rodrigo Benenson, Seong Joon Oh, Jan Hosang, Zeynep Akata, Mario Fritz and Bernt Schiele in cooperation with Matthias Hein (Saarland University)*

Training convolutional networks for applications such as semantic labelling and instance segmentation requires expensive pixel-level annotations, and thus significant cost is involved in creating large enough training sets. In order to make both the training more affordable and to extend the amount of training data, there is a need to relax the requirement of high-quality pixel-level image annotations. In this section we explore weaker forms of supervision for training image segmentation networks, such as image label and bounding box annotations, which are cheaper and easier to define.

**Exploiting Saliency for Object Segmentation from Image Level Labels**

There have been remarkable improvements in the semantic labelling task in the recent years. However, the state of the art methods rely on large-scale pixel-level annotations. [2] studies
the problem of training a pixel-wise semantic labeller network from image-level annotations of the present object classes. Recently, it has been shown that high quality seeds indicating discriminative object regions can be obtained from image-level labels. Without additional information, obtaining the full extent of the object is an inherently ill-posed problem due to co-occurrences. We propose using a saliency model as additional information and hereby exploit prior knowledge on the object extent and image statistics. We show how to combine both information sources in order to recover 80% of the fully supervised performance – which is the new state of the art in weakly supervised training for pixel-wise semantic labelling.

![High level Guided Segmentation architecture.](image)

**Simple Does It: Weakly Supervised Instance and Semantic Segmentation**

Semantic labelling and instance segmentation are two tasks that require pixel-level mask annotations. Starting from weak supervision in the form of bounding box detection annotations, in [1] we propose a new approach that does not require modification of the segmentation training procedure. We view the problem of weak supervision as an issue of input label noise. We explore recursive training as a de-noising strategy, where convnet predictions of the previous training round are used as supervision for the next round. We show that when carefully designing the input labels from given bounding boxes, even a single round of training is enough to improve over previously reported weakly supervised results and when properly used, “classic computer vision techniques” for box-guided instance segmentation are a source of surprisingly effective supervision for convnet training. Overall, our weak supervision approach reaches 95% of the quality of the fully supervised model, both for semantic labelling and instance segmentation.

![Training sample, with box annotations](image) ![Test image, fully supervised result](image) ![Test image, weakly supervised result](image)

**Figure 36.37:** We propose a technique to train semantic labelling from bounding boxes, and reach 95% of the quality obtained when training from pixel-wise annotations.
36.13.2 Video Object Segmentation with Different Levels of Supervision

Investigators: Anna Khoreva, Anna Rohrbach, Rodrigo Benenson, and Bernt Schiele in cooperation with Federico Perazzi (Disney Research, ETH Zurich), Alexander Sorkine-Hornung (Disney Research), Eddy Ilg (University of Freiburg) and Thomas Brox (University of Freiburg)

Like most deep learning techniques, convolutional networks for video object segmentation benefit from large amounts of training data. Current state-of-the-art methods mostly rely on pixel accurate foreground/background annotations of video frames. Labelling videos at the pixel level is a laborious task (compared e.g. to drawing bounding boxes or providing class labels), and creating a large training set requires significant annotation effort. In this section we investigate ways to reduce the necessity for such large volumes of video data by employing static images for training, sophisticated data augmentation techniques and exploiting language referring expressions.

Learning Video Object Segmentation from Static Images

Inspired by recent advances of deep learning in instance segmentation and object tracking, in [3] we introduce video object segmentation problem as a concept of guided instance segmentation. Our model proceeds on a per-frame basis, guided by the output of the previous frame towards the object of interest in the next frame (see Fig. 36.38). We demonstrate that highly accurate object segmentation in videos can be enabled by using a convnet trained with static images only, relaxing the constraint of lack of densely annotated video data. The key ingredient of our approach is a combination of offline and online learning strategies, where the former serves to produce a refined mask from the previous frame estimate and the latter allows to capture the appearance of the specific object instance. Our method can handle different types of input annotations: bounding boxes and segments, as well as incorporate multiple annotated frames, making the system suitable for diverse applications. We obtain competitive results on three different datasets, independently from the type of input annotation.

Lucid Data Dreaming for Video Object Segmentation

Convolutional networks reach top quality in pixel-level video object segmentation but require a large amount of training data (1k~10k) to deliver such results. In [2] we propose a new training strategy which achieves state-of-the-art results across three evaluation datasets.
For each new video frame the network is guided towards the object of interest by feeding in the previous frame mask estimate. We therefore refer to our approach as guided instance segmentation.

while using $20 \times \sim 100 \times$ less annotated data than competing methods. Our approach is suitable for both single and multiple object segmentation. With the results for multiple object segmentation we took the second place in the 2017 DAVIS Challenge on Video Object Segmentation [1].

Instead of using large training sets hoping to generalize across domains, we generate in-domain training data using the provided annotation on the first frame of each video to synthesize (“lucid dream”) plausible future video frames (see Fig. 36.39). In-domain per-video training data allows us to train high quality appearance- and motion-based models, as well as tune the post-processing stage. This approach allows to reach competitive results even when training from only a single annotated frame. Our results indicate that using a larger training set is not automatically better, and that for the video object segmentation task a smaller training set that is closer to the target domain is more effective. This changes the mindset regarding how many training samples and general objectness knowledge are required for the video object segmentation task.

**Video Object Segmentation with Language Referring Expressions**

Most state-of-the-art semi-supervised video object segmentation methods rely on a pixel-accurate mask of a target object provided for the first frame of a video. However, obtaining a detailed segmentation mask is expensive and time-consuming. In [4] we explore an alternative way of identifying a target object, namely by employing language referring expressions, see Fig. 36.40. Besides being a more practical and natural way of pointing out a target object, using language specifications can help to avoid drift as well as make the system more robust to complex dynamics and appearance variations. Leveraging recent advances of language grounding models designed for images, we propose an approach to extend them to video data, ensuring temporally coherent predictions. To evaluate our approach we augment the popular video object segmentation benchmarks, DAVIS$_{16}$ and DAVIS$_{17}$ with language descriptions of target objects. We show that our language-supervised approach performs on par with the methods which have access to a pixel-level mask of the target object on DAVIS$_{16}$ and is
Figure 36.39: Starting from scarce annotations we synthesize in-domain data to train a specialized pixel-level video object segmenter for each dataset or even each video sequence.

competitive to methods using scribbles on the challenging DAVIS17 dataset.

Figure 36.40: Examples of the proposed approach. Classical semi-supervised video object segmentation relies on an expensive pixel-level mask annotation of a target object in the first frame of a video. We explore a more natural and more practical way of pointing out a target object by providing a language referring expression.

References


36.13.3 Semantic Segmentation

Investigators: Yang He, Rakshith Shetty, Wei-Chen Chiu, Seyed M. Azimi, Mario Fritz, and Bernt Schiele in cooperation with Dominik Britz, Michael Engstler and Frank Mücklich (Saarland University), Magret Keuper (University of Mannheim)

Semantic segmentation is a fundamental task in computer vision, aiming to assign a semantic label to each pixel of an image. It has broad application scenarios such as traffic scenes [3], indoor scenes [2], and material science [1]. Modern semantic segmentation systems are built on the fully convolutional networks, which rely on large amount of training data. Our work are consist of leveraging unlabeled and synthetic data for improving semantic segmentation approaches, designing new compatible modules to existing network architectures, analysing semantic segmentation systems. Finally, we cooperated with researchers in material science, developing a system which is able to segment and recognize the type of materials on the surface of an object.

Advanced Steel Microstructure Classification by Deep Learning Methods

Microstructure stores the genesis of a material and determines all its physical and chemical properties. A microstructural recognition and classification system is proposed in [1]. Inspired of the success of deep neural networks in computer vision, the system was created via a fully convolutional neural networks accompanied by a max-voting scheme, as shown in Fig 36.41. The designed system drastically outperforms previous state-of-the-art method. Beyond the strong performance of this system, this line of research offers a more robust and first of all objective way for the difficult task of steel quality appreciation.

Figure 36.41: Semantic image segmentation from multiple views using spatio-temporal data-driven pooling.

STD2P: RGBD Semantic Segmentation Using Spatio-Temporal Data-Driven Pooling

Beyond the success in classification, neural networks have recently shown strong results on pixel-wise prediction tasks like image semantic segmentation on RGBD data. However, the commonly used deconvolutional layers for upsampling intermediate representations to the full-resolution output still shows different failure modes, like imprecise segmentation boundaries and label mistakes particular on large, weakly textured objects (e. g. fridge, whiteboard, door). We attribute these errors in part to the rigid way, current network aggregate information, that can be either too local (missing context) or too global (inaccurate
boundaries). Therefore we propose a data-driven pooling layer [2] that integrates with fully convolutional architectures and utilizes boundary detection from RGBD image segmentation approaches. We extend our approach to leverage region-level correspondence across images with an additional temporal pooling stage as shown in Fig 36.42. We evaluate our approach in various datasets extensively, showing the effectiveness of proposed method.

![Image Sequence](Image Sequence)

**Figure 36.42:** Semantic image segmentation from multiple views using spatio-temporal data-driven pooling.

**Learning Dilation Factors for Semantic Segmentation of Street Scenes**

Dilated convolutions are widely used operations in neural networks, which shows effective in various architectures. Conventional convolutions utilize manually set dilation factors, while we propose an approach to learn dilation factor automatically in [3], which is compatible with existing network architectures and learning pipeline. Our proposed dilated convolutions can be used in any locations of a network. We demonstrate its effectiveness on street scene semantic segmentation tasks on different datasets across different cities and capturing environments and conditions.

![Illustration of our learnable dilated convolutions](Illustration of our learnable dilated convolutions)

**Figure 36.43:** Illustration of our learnable dilated convolutions.
Not Using the Car to See the Sidewalk: Quantifying and Controlling the Effects of Context in Classification and Segmentation

Importance of visual context in scene understanding tasks is well recognized in the computer vision community. To understand the role of context, we propose a method [4] to quantify the sensitivity of black-box vision models to visual context by editing images to remove selected objects and measuring the response of the target models. We apply this methodology on two tasks, image classification and semantic segmentation, and discover undesirable dependency between objects and context. Besides, we propose an object removal based data augmentation solution to mitigate this dependency and increase the robustness of classification and segmentation models to contextual variations, which also maintains performance on regular data.

References


36.14 Combinatorial Computer Vision

*Coordinator: Paul Swoboda*

In the combinatorial computer vision group we investigate efficient algorithms for solving discrete optimization problems (that is involving variables taking only e.g. values \{0, 1\}) for computer vision applications. The classical example are Markov Random Fields (MRF) which have been widely applied to diverse tasks such as segmentation, depth estimation from stereo, optical flow and many other tasks. However, especially in recent years, a large number of diverse discrete optimization problems have been proposed for various computer vision tasks.

In order for computer vision applications to benefit from such optimization problems, efficient solvers need to be available that return high-quality solutions. To this end we study generally applicable techniques from which one can build efficient solvers that come with some desirable theoretical properties such as connections to LP-relaxations and monotonical convergence. The benefit of taking such a general approach to algorithm design becomes apparent when, instead of relying on ad-hoc algorithms for any new discrete optimization problem, we can rely on a rich algorithmic tool-box and a proven set of algorithmic design
principles that enable us to write new optimization algorithms fast. To facilitate writing new solvers, we also work on a high-performance library that encapsulates the basic algorithmic techniques we develop.

Another line of research in the group concerns partial optimality on discrete optimization problems. Since the discrete optimization problems we solve are very large, it is often hard to solve them exactly in their original form. However, many parts of the optimization problems are structurally easy in the sense that many heuristic algorithms can already return solutions that usually agree with globally optimal ones on a large subset of variables. Identifying such variables as part of a preprocessing step enables us to shrink the optimization problem. Such an approach accelerates the overall solution process and improves solution quality. A general framework for partial optimality has been developed in the last few years among others by members of the group for the case of MRFs. We want to further improve applicability of fast partial optimality methods and extend their applicability to a large range of discrete optimization problems.

### 36.14.1 Graph Matching

**Investigators:** Paul Swoboda, Florian Bernard, Christian Theobalt, in cooperation with Dagmar Kainmüller, Ashkan Mokarian (Berliner Institut für Gesundheitsforschung, Germany) and Johan Thunberg (Halmstad University, Sweden)

The graph matching problem aims to establish correspondences between two sets of points. The correspondence one seeks can be either bijective or, as is more useful in settings with outliers, bijective on a subset of points. The subset on which to establish bijective correspondences is found as part of the optimization process. The graph matching generalizes the classical linear assignment problem that is polynomially solvable by allowing for a more complicated cost structure expressed in terms of costs that factorize according to a graphical structure. It is equivalent to the well-known quadratic assignment problem (QAP) that is notorious for being one of the empirically hardest NP-hard optimization problems.

Besides the graph matching problem, which is also referred to as the Lawler-formulation of the QAP, we also study the following specializations and extensions: The Koopmans-Beckman formulation, hypergraph matching and multi-graph matching.

**Koopmans-Beckman formulation**

The Koopmans-Beckman formulation of the QAP assumes a special factorization of the quadratic cost. In particular, the quadratic cost is given by computing for each pair of points in each point set a score, resulting in two matrices $A$ and $B$. The pairwise costs are given by $\sum_{ij} A_{ij} B_{\pi(i),\pi(j)}$, where $\pi$ is the mapping that established the correspondence between the point sets. In other words, the quadratic cost is the correlation of matrices $A$ and $B$ after permuting to account for the correspondence. Its advantages are a compact cost formulation in contrast to the general QAP and natural interpretation for a wide range of correspondence problems.
Hypergraph matching

The hypergraph matching problem extends the quadratic cost structure by higher order costs based. It can be used either to obtain more expressive models or better LP-relaxations for the plain graph matching problem.

Multi-graph matching

The multi-graph matching problem aims to simultaneously establish correspondences between an arbitrary number of point sets. For multiple matchings, the notion of cycle consistency arises: assume that $X_{pq}$ is the assignment matrix between point sets indexed by $p$ and $q$. The condition $X_{pr}X_{rq} = X_{pq} \forall p, q, r$ is called cycle consistency and guarantees that concatenations of correspondences lead to consistent results. The reason for studying multi-graph matchings is that spurious correspondences introduced by noise in the data can be corrected by jointly optimizing over multiple correspondences, since each correspondence between two graphs depends on other matches via cycle consistency.

Message passing for (multi-)graph matching

In [2] we proposed an efficient message passing solver for the graph matching problem. The algorithm builds upon efficient convergent message passing techniques that allow it to solve tight relaxations for large-scale problems occurring in computer vision. We are the first to propose an LP relaxation for the multi-graph matching together with an efficient algorithm for solving it. It is based on novel linear constraints for cycle consistency, that are added via a cutting plane procedure. With our new algorithm, novel applications in biological image analysis could be approached.

Higher-order Projected Power Iterations for Scalable Multi-Matching

We propose in [1] a heuristic method to obtain efficiently multi-graph matchings for Koopman-Becklams-QAPs. The algorithm works by defining steps that transform one feasible multi-graph matching into another one and applies this basic step until no improvement in the objective can be observed. The transformation is inspired by recent power iteration methods. Its advantages are its speed and empirical success on very large-scale multi-graph matching problems from e.g. shape analysis.

References


36.14.2 Convex Optimization

Investigators: Paul Swoboda in cooperation with Vladimir Kolmogorov (Institute of Science and Technology, Austria)

Convex optimization is the basic mathematical principle allowing for powerful algorithms that find global minima irrespective of initialization and that allow to exploit dual formulations of optimization problems which are often more efficient to solve than the original primal ones. Specifically, Lagrangean decomposition of discrete optimization problems can be understood as convex non-smooth optimization problems which are optimized in the dual form. A large range of algorithms was proposed for convex optimizatino problems in image analysis and computer vision, including message passing, subgradient and proximal algorithms.

Proximal bundle methods

One of the most basic algorithms for Lagrangean decomposition is subgradient ascent. It can be applied in the most basic settings where only subgradient is available. However, this method is known to lead to slow convergence. To obtain better convergence speed in practice, bundle methods are used. Bundle methods approximate the convex function by a piecewise linear local surrogate estimated from subgradients. Efficacy of bundle methods has been observed in practice, however not theoretically better convergence rates have been obtained. We propose a proximal bundle method based on the Frank-Wolfe algorithm [1]. Sequentially, local approximations are constructed that include a trust-region term. The trust region term makes the local approximations strongly convex. Such objectives can then be solved efficiently with the Frank-Wolfe method. We show that our method outperforms established proximal bundle methods on a variety of computer vision tasks, including Markov Random Fields inference, graph matching and discrete tomography.

References


36.15 Academic Activities

36.15.1 Journal Positions

Bernt Schiele:

- IEEE Transactions on Pattern Analysis and Machine Intelligence (Associate Editor-in-chief)
- International Journal of Computer Vision (Associate Editor)
- IEEE Pervasive Computing (Associate Editor)
- International Journal of Pervasive Computing and Communications (Associate Editor)
- Springer Series Advances in Computer Vision and Pattern Recognition (Advisory Board Member)
Andreas Bulling:

- Proceedings of the ACM on Interactive, Mobile, Wearable, and Ubiquitous Technologies (Associate Editor)
- ACM Transactions on Interactive Intelligent Systems (Associate Editor)
- Journal of Eye Movement Research (Associate Editor)

Zeynep Akata:

- Pattern Recognition Journal Elsevier (Associate Editor)

Mario Fritz:

- IEEE Transactions on Pattern Analysis and Machine Intelligence (Associate Editor)

36.15.2 Conference and Workshop Positions

Membership in program and organization committees

Bernt Schiele

- European Conference on Computer Vision, ECCV, Munich, Germany, 2018 (General Co-Chair),
- Neural Information Processing Systems, NeurIPS, Montreal, Canada, 2018 (Senior Area Chair),
- Neural Information Processing Systems, NeurIPS, Vancouver, Canada, 2019 (Senior Area Chair),
- International Conference on Computer Vision, ICCV, Venice, Italy, 2017 (Area Chair),
- International Conference on Computer Vision, ICCV, Seoul, South Korea, 2019 (Area Chair),
- IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017-19 (PC member),
- International Conference on Learning Representations, ICLR, 2017 (PC member),
- German Conference on Pattern Recognition, GCPR, 2017-19 (PC member),
- International Conference on Computer Vision Systems ICVS, 2017 (PC member),
- Co-Organizer and Program Committee member for various workshops

Mario Fritz:

- German Conference on Pattern Recognition GCPR 2018 (PC Chair)
- ACM Computer Science in Cars Symposium – Future Challenges in Artificial Intelligence & Security for Autonomous Vehicles (CSCS 2018) (Co-Organizer and PC Chair)
- Workshop on Interactive and Adaptive Learning in an Open World at ECCV 2018 (Co-Organizer)
Andreas Bulling:

- **Mensch und Computer, MuC 2019** (PC Chair)
- **ACM SIGCHI Conference on Human Factors in Computing Systems, CHI 2018-19** (Associate Chair)
- **ACM Symposium on Eye Tracking Research and Applications, ETRA 2018** (Associate Chair)
- **International Conference on Multimodal Interaction, ICMI 2018** (PC Member)
- **Augmented Human International Conference, AH 2017** (PC Member)
- **ACM Symposium on User Interface Software and Technology, UIST 2017** (PC Member)

Zeynep Akata:

- **IEEE Conference on Computer Vision and Pattern Recognition CVPR 2019** (Area Chair)
- **International Conference of Computer Vision, ICCV 2019** (Area Chair)
- **European Conference of Computer Vision, ECCV 2018** (Area Chair)
- **International Joint Conference on Artificial Intelligence IJCAI 2018** (Senior PC Member)
- **International Conference of Machine Learning ICML 2018, 2019** (PC Member)
- **International Conference on Learning Representations ICLR 2018** (PC Member)
- **IEEE Conference on Computer Vision and Pattern Recognition CVPR 2018** (PC Member)
- **Neural Information Processing Systems NeurIPS 2018** (PC Member)
Anna Khoreva:
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2018-19 (PC Member)
- Neural Information Processing Systems NeurIPS 2018 (PC Member)
- Asian Conference of Computer Vision ACCV 2018 (PC Member)
- International Conference of Machine Learning ICML 2019 (PC Member)

Seong Joon Oh:
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2018-19 (PC Member)
- Neural Information Processing Systems NeurIPS 2018 (PC Member)
- Asian Conference of Computer Vision ACCV 2018 (PC Member)
- International Conference of Machine Learning ICML 2019 (PC Member)

Gerard Pons-Moll:
- European Conference of Computer Vision, ECCV 2018 (Area Chair)
- International Conference on 3DV, 3DV 2019 (Area Chair)
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2018-19 (PC Member)
- International Conference on Computer Vision ICCV 2017 (PC Member)
- SIGGRAPH 2018 (Reviewer)
- SIGGRAPH-Asia 2018 (Reviewer)
- Eurographics 2018-19 (Reviewer)
- Symposium on Computer Animation 2017, (International Program Committee)
- International Conference of Machine Learning ICML 2018 (PC Member)
- Neural Information Processing Systems NeurIPS 2018 (PC Member)
- Computers and Graphics, 2017-18 (Reviewer)
- Transactions on Pattern Analysis and Machine Intelligence, TPAMI 2017-18 (Reviewer)
- International Journal on Computer Vision, (IJCV), 2017-18 (Reviewer)

Shanshan Zhang:
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2017-19 (PC Member)
- European Conference on Computer Vision ECCV 2018 (PC Member)
- Association for the Advancement of Artificial Intelligence AAAI 2019 (PC Member)
- International Joint Conference on Artificial Intelligence IJCAI 2017 (PC Member)
- International Conference on Computer Vision ICCV 2017 (PC Member)
- Vision and Learning Seminar VALSE 2019 (Workshop Chair)
Yongqin Xian:
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2018-19 (PC Member)
- Neural Information Processing Systems NeurIPS 2018 (PC Member)
- Asian Conference of Computer Vision ACCV 2018 (PC Member)
- International Conference of Machine Learning ICML 2019 (PC Member)
- International Conference on Learning Representation ICLR 2018 (PC Member)
- Transactions on Pattern Analysis and Machine Intelligence, TPAMI 2017-18 (Reviewer)
- Transactions on Image Processing, TIP 2018 (Reviewer)

Hosnieh Sattar:
- Transactions on Pattern Analysis and Machine Intelligence, TPAMI 2018 (Reviewer)
- Neurocomputing, 2018-19 (Reviewer)
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2019 (PC Member)
- International Conference on Computer Vision ICCV 2019 (PC Member)
- Women in Computer Vision Workshop ECCV 2018 (PC Member)

Tribhuvanesh Orekondy:
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2019 (PC Member)

Eldar Insafutdinov:
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2017-19 (PC Member)
- European Conference on Computer Vision ECCV 2018 (PC Member)
- International Conference on Computer Vision ICCV 2017 (PC Member)

Apratim Bhattacharyya:
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2019 (PC Member)
- Transactions on Pattern Analysis and Machine Intelligence TPAMI 2019 (Reviewer)
- ACM Computer Science in Cars Symposium – Future Challenges in Artificial Intelligence & Security for Autonomous Vehicles (CSCS18) (Reviewer)

Xucong Zhang:
- ACM SIGCHI Conference on Human Factors in Computing Systems, CHI 2019 (Reviewer)
– ACM Symposium on Eye Tracking Research and Applications, ETRA 2018 (Reviewer)
– ACM Symposium on User Interface Software and Technology, UIST 2016-17 (Reviewer)
– Transactions on Pattern Analysis and Machine Intelligence, TPAMI 2017-18 (Reviewer)

Julian Steil:
– ACM Journal on Interactive, Mobile, Wearable and Ubiquitous Technologies, IMWUT 2018 (Reviewer)
– ACM SIGCHI Conference on Human-Computer-Interaction with Mobile Devices and Services, MobileHCI 2018 (Reviewer)
– ACM Symposium on Eye Tracking Research and Applications, ETRA 2018 (Reviewer)
– ETRA Workshop on Pervasive Eye Tracking and Mobile Eye-based Interaction, PET-MEI 2018 (Reviewer)
– ACM Augmented Human International Conference, Augmented Human 2017 (Reviewer)
– ACM Symposium on User Interface Software and Technology, UIST 2017 (Reviewer)
– ACM International Symposium on Wearable Computers, ISWC 2017 (Reviewer)

Philipp Müller:
– ACM Symposium on Eye Tracking Research and Applications, ETRA 2018-19 (PC Member)
– ACM SIGCHI Conference on Human-Computer-Interaction with Mobile Devices and Services, MobileHCI 2018 (Reviewer)
– ACM Augmented Human International Conference, Augmented Human 2017 (Reviewer)
– Journal of Ambient Intelligence and Humanized Computing, AIHC 2018 (Reviewer)

Bjoern Andres:
– Neural Information Processing Systems NeurIPS 2017 (PC Member)
– International Conference of Machine Learning ICML 2017 (PC Member)
– IEEE Conference on Computer Vision and Pattern Recognition CVPR 2018 (PC Member)

Anna Rohrbach:
– International Conference on Computer Vision ICCV 2017 (PC Member)
– Conference on Empirical Methods in Natural Language Processing EMNLP 2017 (PC Member)
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2018 (PC Member)
- IEEE Transactions on Multimedia, TMM 2017 (Reviewer)
- IEEE Transactions on Image Processing TIP, 2017 (Reviewer)

David Stutz:
- SIGGRAPH, 2019 (Reviewer)
- IEEE Transactions on Pattern Analysis and Machine Intelligence TPAMI, 2018 (Reviewer)
- IEEE Transactions on Image Processing TIP, 2018 (Reviewer)

Rakshith Shetty:
- IEEE Conference on Computer Vision and Pattern Recognition CVPR 2018 (PC Member)

Membership in steering and other committees

Bernt Schiele:
- IEEE Social Computing (Steering Committee Member)
- IEEE International Symposium on Wearable Computing (Steering Committee Member)
- International Conference on Computer Vision Systems (Steering Committee Member)

Mario Fritz:
- ACM Technical Policy Committee Europe

Andreas Bulling:
- ACM International Joint Conference on Pervasive and Ubiquitous Computing

Zeynep Akata:
- BNVKI Special Interest Group on Artificial Intelligence (SIG-AI)

36.15.3 Invited Talks and Tutorials

Bernt Schiele:
- Symposium on Visual Computing and Perception, Braunschweig, Germany, July 2017 (Plenary Speaker)
- Intel NIS Network on Intelligent Systems Workshop 2017, Munich, Germany, August 2017 (Invited Talk)
- CS Department, TU Darmstadt, October 2017 (Invited Talk)
- ADAS Advanced Driver Assistance Systems – Computer Vision Round Table, Lindau, Germany, November 2017 (Invited Talk)
Mario Fritz:
- Center for Art and Media, Karlsruhe, 2018; “The Bright and Dark Side of Computer Vision: Latest Advances and Implications on Privacy”
- First International Workshop on The Bright and Dark Sides of Computer Vision: Challenges and Opportunities for Privacy and Security (CV-COPS 2017) at CVPR 2017
- ACM chapters computer science in cars Symposium (CSCS 2017), Munich
- Tagung Identifikation nach Bildern (Symposium on Forensics and Identification), Saarbrücken, 2017

Andreas Bulling:
- ACM Symposium on Eye Tracking Research and Applications, Poland, June 2018 (Keynote)
- Max Planck Institute for Biological Cybernetics, Germany, March 2017

Zeynep Akata:
- “Interpretable Machine Learning Tutorial”, CVPR 2019 (Keynote)
- “Women in Computer Vision Workshop”, WiCV@CVPR 2019 (Keynote)
- Data Science Statistics and Visualization (DSSV), Vienna, Austria, 2018 (Keynote)
- ISC- High Performance Computing Conference (ISC-HPC), Frankfurt, Germany 2017 (Keynote)
- British Machine Vision Conference BMVC, New Castle, UK, 2019 (Invited lecture)
- Eastern European Machine Learning Summer School, Bucharest, Romania, 2019 (Invited lecture)
Anna Khoreva:
- U Stanford, USA, July 2017
- 41st Pattern Recognition and Computer Vision Colloquium, Center for Machine Perception, CTU Prague, October, 2017

Eldar Insafutdinov:
- U Stanford, USA, July 2017
- 4th Christmas Colloquium on Computer Vision, Skoltech, Moscow, Russia, December, 2018

Gerard Pons-Moll:
- ASLLA symposium: Human Understanding through AI: Days of future present. Gangeung, Korea, November 2018
- The GAME (Graphics And Mixed Environment Webinar, https://games-cn.org/), China, November 2018 (Webinar tutorial)
- VR days, Amsterdam, Netherlands, October 2018
- 3D Humans Workshop. USA, CVPR 2018
- Stanford University, Palo Alto, CA, USA, June 2018.
- CMU, Pittsburg, USA, June 2018.
- Saarland Campus Lecture Series. Saarland Informatics Campus. Germany, May 2018.
- University of Surrey. Guilford, UK, April 2018.
- University College London (UCL). London, April 2018.
- EPFL Lausanne, March 2018.
- DFKI, Kaiserslautern. February 2018
- Max Planck for Intelligent Systems. Tuebingen, December 2017

Anna Khoreva:
- UC Berkeley, USA, May 2018
- The 1st Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Guangzhou, China, November, 2018

Bjoern Andres:
- MPI for Intelligent Systems, Tübingen, Germany, May 2017
- University of Tübingen, Tübingen, Germany, May 2017
- CTU Prague, Prague, Czech Republic, January 2017
Jan-Hendrik Lange:
- Imaging and Vision from Theory to Practice Workshop, Siegen, Germany, March 2018

Hosnieh Sattar:
- “Women in Computer Vision Workshop”, WiCV@ECCV 2018

36.15.4 Other Academic Activities

Bernt Schiele
- Vice President, Board of European Computer Vision Association ECVA (since 2017)
- Elected Member, DFG Fachkollegium Computer Science (DFG, German Research Foundation) (since 2012)
- Vice Chairman, DAGM, German Association for Pattern Recognition, Germany (since 2015)
- Member of Board of Trustees, IGD Fraunhofer Institute for Computer Graphics Research, Darmstadt, Germany (since 2007)
- Best Paper Award Committee, Asian Conference of Computer Vision ACCV, 2018
- Member of Scientific Advisory Board, VRVis, Center for Virtual Reality and Visualization, Vienna, Austria (2010 – 2017)
- Member of Scientific Advisory Board, Bernstein Focus: Neurotechnology Frankfurt, Germany (since 2008)
- Member of Scientific Advisory Board, Idiap Research Institute, Martigny, Switzerland (since 2013)
- Member of Audit Committee, Idiap Research Institute, Martigny, Switzerland (2014 – 2018)
- External Reviewer of PhD thesis: Marius Cordts, TU Darmstadt, Germany, 2017
- External Reviewer of PhD thesis: Martin Kiefel, ETH Zurich, Switzerland, 2017
- External Reviewer of PhD thesis: Ahmad Humayun, Georgia Institute of Technology, USA, 2018
- External Reviewer of PhD thesis: Marcel Simon, Friedirch-Schiller-University Jena, Germany, 2018

Andreas Bulling:
- Conference Workshop Egocentric Perception, Interaction, and Computing, ICCV 2017 (Organiser)

Anna Rohrbach:
- Conference Workshop The Joint Video and Language Understanding Workshop: MovieQA and The LSMDC, ICCV 2017 (Organiser)
Zeynep Akata:
- Conference Workshop *Uncertainty and Robustness in Deep Visual Learning*, CVPR 2019 (Organiser)
- Conference Workshop *Women in Computer Vision*, ECCV 2018 (Organiser)

Gerard Pons-Moll:
- Conference Workshop *PeopleCap’18: capturing and modeling human bodies, faces and hands*, ECCV 2018 (Organiser)
- Conference Workshop *PeopleCap’17: capturing and modeling human bodies, faces and hands*, ICCV 2017 (Organiser)
- Reviewer for grant proposals from DFG (German Research Council).
- Reviewer for grant proposals from Israel Science Foundation.

Bjoern Andres:

### 36.16 Teaching Activities

**Summer Semester 2017**
- High Level Computer Vision (Bernt Schiele, Mario Fritz)

**Winter Semester 2017/2018**
- Introduction to Machine Learning, University of Amsterdam (Zeynep Akata)
- Probabilistic Graphical Models and their Applications (Bernt Schiele, Gerard Pons-Moll)

**Summer Semester 2018**
- Introduction to Brain Computer Interaction (Andreas Bulling)
- High Level Computer Vision (Bernt Schiele, Mario Fritz)

**Winter Semester 2018/2019**
- Introduction to Machine Learning, University of Amsterdam (Zeynep Akata)
- Machine Learning (Bernt Schiele, Paul Swoboda, Gerard Pons-Moll)
- Probabilistic Graphical Models and their Applications (Paul Swoboda, Gerard Pons-Moll)
Master and Bachelor Theses

Guan Wang: Unconstrained Appearance-based Gaze Estimation from a Freely Moving Camera, Saarland University, 2017
Pascal Bies: Detecting Memorable Real-World Experiences in Egocentric Video, Saarland University, 2017
Patrick Ferber: Using Neural Networks for Distance Estimation in Planning, Saarland University, 2017
Mihai Fieraru: Learning to Track Humans in Videos, Saarland University, 2017
Assulan Nurkas: Computational Modelling of Visual Attention during Reading, Saarland University, 2018
Henry Lin: Credit Card Fraud Detection with Graph Convolutional Models, University of Amsterdam, 2018
Haitam Ben Yahiya: Graph Convolutional Generators Towards Learning to Generate Larger Graphs, University of Amsterdam, 2018
Edgar Schoenfeld: Generalized Zero-and Few-Shot Learning via Aligned Variational Autoencoders, University of Amsterdam, 2018
Govert Verkes: Explaining Uncertainty in Visual Question Answering, University of Amsterdam, 2018
Laurens Weitkamp: Visual Rationalizations in Deep Reinforcement Learning for Atari Games, University of Amsterdam, 2018
Ahmed Abbas: Bottleneck Potentials in Markov Random Fields, Saarland University, 2019

36.17 Dissertations, Habilitations, Awards

36.17.1 Dissertations

- Jan Hosang, Analysis and Improvement of the Visual Object Detection Pipeline, May 2017
- Anna Rohrbach, Generation and Grounding of Natural Language Descriptions for Visual Data, May 2017 (with distinction)
- Maksim Lapin, Image Classification with Limited Training Data and Class Ambiguity, May 2017
- Mateusz Malinowski, Towards Holistic Machines: From Visual Recognition To Question Answering About Real-World Images, June 2017 (with distinction)
- Siyu Tang, People Detection and Tracking in Crowded Scenes, September 2017 (with distinction)
- Anna Khoreva, Learning to Segment in Images and Videos with Different Forms of Supervision, December 2017
- Wenbin Li, From Perception over Anticipation to Manipulation, April 2018
- Seong Joon Oh, Image Manipulation against Learned Models: Privacy and Security Implications, August 2018 (with distinction)
- Xucong Zhang, Gaze Estimation and Interaction in Real-World Environments, September 2018 (with distinction)
36.17.2 Offers for Faculty Positions

- Zeynep Akata: Full Professor (W3) at University of Würzburg, Germany
- Andreas Bulling: Full Professor (W3) at University of Freiburg, Germany
- Andreas Bulling: Full Professor (W3) at University of Stuttgart, Germany
- Mario Fritz: Full Professor (W3) at University of Marburg, Germany
- Mario Fritz: Tenured Faculty Position at CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
- Björn Andres: Honorary Professor at University of Tübingen, Germany

36.17.3 Awards

- Gerard Pons-Moll: Awardee of Emmy Noether Programme 2018
- Anna Rohrbach: Otto Hahn Medal for the year 2017 (for her PhD)
- Siyu Tang: DAGM MVTec dissertation award for her PhD
- Mateusz Malinowski: DAGM MVTec dissertation award for his PhD
- Mateusz Malinowski: Dr.-Eduard-Martin award for his PhD
- Gerard Pons-Moll: Google Faculty Research Award 2019
- Bernt Schiele: Fellow, Institute of Electrical and Electronics Engineers IEEE (2017)
- Margret Keuper, Siyu Tang, Björn Andres, Thomas Brox, and Bernt Schiele: Winner of the CVPR 2017 Multi-Object Tracking Challenge (MOT17)
- Mohamed Omran, Gerard Pons-Moll, and Bernt Schiele: Best Student Paper Award at 3DV 2018 for Neural Body Fitting: Unifying Deep Learning and Model Based Human Pose and Shape Estimation
- Julian Steil, Philipp Mueller, Yusuke Sugano, and Andreas Bulling: Best Paper Award at MobileHCI 2018 for Forecasting User Attention During Everyday Mobile Interactions Using Device-Integrated and Wearable Sensors
- Mohamed Khamis, Florian Alt, and Andreas Bulling: Best Paper Honourable Mention Award at MobileHCI 2018 for The Past, Present, and Future of Gaze-enabled Handheld Mobile Devices: Survey and Lessons Learned
- Michael Barz, Florian Daiber, Daniel Sonntag, and Andreas Bulling: Best Paper Award at ETRA 2018 for Error-Aware Gaze-Based Interfaces for Robust Mobile Gaze Interaction
- Seonwook Park, Xucong Zhang, Andreas Bulling, and Otmar Hilliges: Best Presentation Award at ETRA 2018 for Learning to Find Eye Region Landmarks for Remote Gaze Estimation in Unconstrained Settings
- Erroll Wood, Tadas Baltrusaitis, Louis-Philippe Morency, Peter Robinson, and Andreas Bulling: Best Paper Honourable Mention Award at Eurographics 2018 for GazeDirector: Fully Articulated Eye Gaze Redirection in Video
– Mohamed Khamis, Christian Becker, Andreas Bulling, and Florian Alt: Best Paper Honourable Mention Award at ACM CHI 2018 for Which one is me? Identifying Oneself on Public Displays
– Marc Tonsen, Julian Steil, Yusuke Sugano, and Andreas Bulling: Distinguished Paper Award in Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) 2017 for InvisibleEye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning-Based Gaze Estimation
– Xucong Zhang, Yusuke Sugano, and Andreas Bulling: Best Paper Honourable Mention Award at UIST 2017 for Everyday Eye Contact Detection Using Unsupervised Gaze Target Discovery
– Mohamed Khamis, Linda Bandelow, Stina Schick, Dario Casadevall, Andreas Bulling, and Florian Alt: Best Paper Honourable Mention Award at MUM 2017 for They are all after you: Investigating the Viability of a Threat Model that involves Multiple Shoulder Surfers
– Mohamed Omran, and Bernt Schiele with co-authors, Best Paper Award of Fraunhofer IGD and the Visual Computing Groups of TU Darmstadt, for CVPR 2017 paper: The Cityscapes Dataset for Semantic Urban Scene Understanding
– Anna Rohrbach: Outstanding Reviewer at ICCV 2017, CVPR 2018
– Bernt Schiele: Outstanding Reviewer at CVPR 2018
– Mohamed Omran: Outstanding Reviewer at CVPR 2018
– Rakshith Shetty: Qualcomm Innovation Fellowship in Europe 2018
– Yongqin Xian: Qualcomm Innovation Fellowship Finalists in Europe, 2018
– Yongqin Xian: Chinese Government Award for Outstanding Students Abroad, 2018
– David Stutz: STEM Award IT 2018

36.18 Grants and Cooperations

– Bernt Schiele, Principal Investigator, Cluster of Excellence “Multimodal Computing and Interaction”
– Bernt Schiele and Mario Fritz, Principal Investigator, DFG CRC 1223: “Methods and Tools for Understanding and Controlling Privacy”
– Bernt Schiele, Principal Investigator, Industry cooperation with Toyota Europe on “Learning to Categorize Objects in Real-Word Scenes”
– Bernt Schiele, Principal Investigator, Intel Network on Intelligent Systems

Andreas Bulling: ANTICIPATE – ERC Starting Grant

ANTICIPATE aims to establish the scientific foundations for a new generation of user interfaces that pro-actively adapt to users’ future input actions by monitoring their attention and predicting their interaction intentions – thereby significantly improving the naturalness, efficiency, and user experience of the interactions. Realising this vision of anticipatory
human-computer interaction requires groundbreaking advances in everyday sensing of user attention from eye and brain activity. We will further pioneer methods to predict entangled user intentions and forecast interactive behaviour with fine temporal granularity during interactions in everyday stationary and mobile settings. Finally, we will develop fundamental interaction paradigms that enable anticipatory UIs to pro-actively adapt to users’ attention and intentions in a mindful way. The new capabilities will be demonstrated in four challenging cases: 1) mobile information retrieval, 2) intelligent notification management, 3) Autism diagnosis and monitoring, and 4) computer-based training.

- Starting date: February 2019.
- Duration: 5 years.
- Funding: 5 positions.
- MPI-INF investigators: Andreas Bulling

Gerard Pons-Moll: Real Virtual Humans (RVHu) – Emmy Noether Starting Grant

RVHu: Virtual humans are at the centre of diverse application areas such as medicine and psychology, virtual and augmented reality, and special effects in movies. Virtual humans that look and move indistinguishably from real humans have the potential to enable natural communication between humans and “machines”. Machines also need detailed internal mental models of real people in order to perceive (from sensory data) all subtleties that make us real. To achieve realism, all details matter: the facial expressions, the geometry of the body and its movements, the soft-tissue motion, the appearance and dynamics of clothing or the light reflecting on our body. All these components need to be modelled and perceived to the highest precision. Our hypothesis is that easier to control and more realistic models of humans and clothing can be learned by capturing real people using 3D/4D scans, videos and inertial sensors. The aim of this project is to build compact and rich representations of people and develop inference techniques to extract such representations from visual data. The two key research questions we want to address are: 1) How do we efficiently digitize humans without losing the detail that make us real? 2) Given a compact digitization, how can we train machines to perceive such rich representations from visual data?

- Starting date: January/February 2019.
- Duration: 6 years.
- Funding: 1 Principal Investigator + 2 Doctoral Students (DFG)
- MPI-INF investigators: Gerard Pons-Moll, Bharat Lal Bathnagar, Verica Lazova

Gerard Pons-Moll: Facebook Reality Labs Cooperation Agreement

The aim of the project is to build 3D dynamic model of clothing on humans. The basic idea is to learn models of clothing by “looking” at how real clothing deforms on top of the body. Specifically, we will compile lots of 3D sequences of real people in clothing, register them over time, and learn a model of “how clothing deforms on top of the body”. The initial goal should be to learn garment specific models that adapt to different body shapes and motions. The long-term goal is to build a model of clothing dynamics that generalizes across different garment topologies and styles, allowing to animate any garment on top of a virtual human.
The aim of the project is to build explainable artificial intelligence systems. The basic idea is to learn models in the intersection between vision and language where language is used as a means to explain the perceptual input and the decision based on the perceptual input. The funding is granted a consortium of principal investigators at University of California Berkeley, Boston University and Kitware Inc. University of Amsterdam as a sub-PI of UC Berkeley is working on building models that explain fine-grained visual classification decision in the form of language and bounding boxes, driving videos in the form of language and attention maps, combining classic machine learning models with deep learning in order to make deep decision makers more transparent and emergent language protocols in the context of multi-agent reinforcement learning.

- Starting date: March 2017.
- Duration: 4 years.
- Funding: 1-2 positions
- MPI-INF investigators: Zeynep Akata

Gerard Pons-Moll: Google Faculty Research Award

The goal of this proposal is to train a model capable of reconstructing the 3D geometry of a person including clothing layered on top of the body from one or several frames of a video.

- Starting date: 2019.
- Duration: 1 years.
- Funding: 1.5 position
- MPI-INF investigators: Gerard Pons-Moll

36.19 Publications

Books and proceedings


Journal articles and book chapters


**Conference articles**


[34] M. Khamis, F. Alt, and A. Bulling. The past, present, and future of gaze-enabled handheld mobile devices: Survey and lessons learned. In MobileHCI 2018, 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, Barcelona, Spain, 2018, Article 38. ACM.


Theses


Technical Reports


37 D3: Internet Architecture

37.1 Personnel

Head of Group
Prof. Anja Feldmann, Ph.D.

Researchers
Balakrishnan Chandrasekaran Ph.D. (March 2018—)
Volker Stocker (01.10.2018—)

Ph.D. Students
Corinna Coupette (01.10.2018—)
Mohamad Hoseini (01.06.2018—)
Thorben Krüger
Franziska Lichtblau
Aniss Maghsoudlou (01.11.2018—)
Mirko R. Palmer
Lars Prehn (01.11.2018—)
Jawad Said Saidi
Florian Streibelt

Secretary/Technical Staff
Birgit Hohmeier-Toure
Rainer May

37.2 Visitors

From January 2018 to February 2019, the following researchers visited our group:

Bruce MacDowell Maggs 21.06.2018–25.06.2018 Duke University & Akamai Technologies
Randy Bush 18.09.2018–23.09.2018 Arrcus & Internet Initiative Japan
Cristel Pelsser 18.09.2018–23.09.2018 Université de Strasbourg
Georgios Smaragdakis 06.05.2018–10.05.2018, 10.06.2018–13.06.2018, 27.01.2018–01.02.2019 Technische Universität Berlin
Vaibhav Bajpai 05.11.2018–09.11.2018 Technische Universität München
37.3 Group Organization

At the moment the department has a flat organizational hierarchy and consists of the director, two postdoctoral researchers, 9 Ph.D. students, a secretary and a technical staff. The department meets regularly every week (usually Tuesdays between 16:00 and 17:00) for discussions on research progress, activities, and other meta topics. The department also meets at least once a month for research talks by visiting scientists, collaborators from MPI or MPI-SWS, or fellow researchers in the group. The group maintains various resources ranging from source code repositories (e.g., a github server) to Web portals (e.g., wiki) to collect information relevant for the department including infrastructure details, project descriptions, data sets, seminar schedules, conference deadlines and the like as well as sharing code or paper repositories.

The department currently has two postdoctoral researchers: Balakrishnan Chandrasekaran and Volker Stocker. Balakrishnan was a PhD student in the computer science department at Duke University in the US. He did a one year stint as a postdoctoral researcher at Technische Universität Berlin before joining the MPI in March 2018. His interest spans network measurements, design and architecture. Volker joined the MPI in October 2018, and prior to that pursued his doctoral studies as a research assistant at the University of Freiburg. He has been a lecturer in economics at the Baden-Wuerttemberg Cooperative State University (DHBW) in Loerrach, Germany and spent time as a visiting research student at the TU Berlin, Germany, the University of Northumbria in Newcastle, UK, and the Massachusetts Institute of Technology (MIT) in Cambridge, US. His major research interests are in the fields of network economics of the Internet and Internet policy.

The department’s research focuses on three broad areas, namely Internet Measurement, Future-proofing the Internet, and Wide-Area Data Analytics. Within each of these areas we have multiple teams working on different subtopics. For Internet measurements these range from BGP-Communities to characterizing the network traffic asymmetry. Topics under the area of future-proofing the Internet range from alternative transport protocols for video streaming to Socket Intents. Lastly, research efforts in the area of wide-area data analytics cover topics from Flowstream—a system for enabling network traffic analytics—to an architecture for handling distributed mega-datasets.

37.4 BGP Communities

*Investigators: Anja Feldmann, Franziska Lichtblau, and Florian Streibelt in cooperation with Robert Beverly (Naval Postgraduate School), Randy Bush (Internet Initiative Japan), Christoph Dietzel (DE-CIX), Cristel Pelsser (University of Strassbourg), and Georgios*
The Border Gateway Protocol (BGP) communicates reachability information between neighbors in the Internet. As the network evolved, the complexity of connections, policies, and economics drove the need for similarly complex and fine-grained routing policies. As a result BGP, the de facto inter-domain routing protocol, has been extended to help support such policies, and provide value-added services. This effort focuses on one such extension, BGP communities, and the implications of its real-world implementation and deployment [3, 1, 2].

BGP communities are an optional transitive BGP attribute used to “tag” advertisements. Operators frequently configure their infrastructure to take different actions depending on community tags. So, communities provide not only a common label for groups of prefixes, but also the ability to signal semantics between ASes and between routers within an AS.

BGP communities are increasingly popular and are used to encode an ever-wider variety of information. Within the last year the number of observable communities increased by roughly 20%. As we describe in Section 2, communities are used to realize routing policies, bias path or peer selection, steer traffic, etc. ASes also use communities to offer value-added services for customers of ISPs and members of IXPs including tagging of route ingress points and origins, selective advertisement, traffic engineering, and remotely triggered blackholing (RTBH), i.e., dropping of traffic to a target destination to mitigate denial-of-service (DoS) attacks. Some providers even use communities to encode latency information.

Our research around BGP communities focuses on different aspects. On the one hand, we show that BGP communities are a seemingly innocuous feature and that they can be used to influence routing in unintended ways, while on the other hand we highlight how to use BGP communities to better respond to distributed denial-of-service (DDoS) attacks.

37.4.1 BGP Community-based Attacks

Although the BGP community-based attacks we consider require certain conditions for success, we show that these conditions hold sufficiently widely to warrant operational attention. Importantly, since our extensive measurements show that communities are widely propagated, an attacker exploiting the BGP communities of a particular AS need not be a directly connected peer. Further, we demonstrate the feasibility of attacks both with and without address space hijacking, suggesting that existing hijack detection methods are insufficient to detect community-based attacks.

The attacks are the result of weaknesses in the current use and implementation of BGP communities and community-based services. Services enabled by communities are typically relevant only between directly connected ASes—for instance, an AS tagging a backup route with a community to indicate that the remote AS should use a lower local preference. Intuitively then, one might expect communities to not propagate through multiple ASes, or beyond their intended destination AS. However, via large-scale analysis of passive BGP data sets, we find that more than 50% of the BGP communities traverse more than four ASes and we see 10% with a hop count of more than six. To better assess the potential vulnerabilities enabled by BGP communities, we design multiple scenarios that highlight intentional, unintentional, and malicious community use. These include the ability to remotely signal...
blackholing of a prefix for which the attacking AS is not responsible, traffic steering, and route manipulation of another AS's prefixes.

The key contributions of this effort are as follows. (i) We analyze BGP community propagation showing that 2.2K networks forward received BGP communities onward. We show that the majority of communities are propagated through the entire Internet. (ii) We highlight this routing system can of worms and identify sufficient conditions for community-based attacks on the routing system. (iii) We sketch three scenarios of how BGP communities can be misused. (iv) We show that these attacks are possible in lab experiments and in the wild. (v) We highlight traffic dropping due to remotely triggered blackholing, as well as remote steering of traffic and route manipulation, possibly through an interceptor, i.e., a rogue traffic monitor. Lastly, we provide recommendations on the use of communities.

37.4.2 Mitigating DDoS Attacks Using Advanced Blackholing

The revolution of the digital age fueled by the Internet has attracted the good but the evil alike. While the threats executed over the Internet are multifaceted from a criminalistics perspective, e.g., fraud, data and identity theft, espionage, or cyber terrorism, the dominant network threat is Denial-of-Service (DoS) attacks. The goal of DDoS attacks is to force a service or system to become unavailable by consuming crucial resources. These resources can be computing power at the servers or exploitation of application-layer vulnerabilities, i.e., semantic attacks, or network bandwidth, i.e., volumetric attacks. To conduct such volumetric attacks, adversaries often use DDoS. Traffic from numerous distributed sources is generated and steered towards a target service to make it unavailable. Once the network links to the target are congested due to the DDoS attack, legitimate traffic that traverses the same links is also affected.

DDoS threats are continuously increasing in terms of volume, frequency, and complexity. While the largest observed and publicly reported attacks were between 50 to 200 Gbps before 2015, current peaks are an order of magnitude higher and exceeded 1 Tbps in 2016, and 1.7 Tbps in early 2018. We also observe a massive rise in the number of DDoS attacks. Prior work report that a third of all active /24 networks were targeted by DDoS attacks between 2016 and 2017. Similar observations are also reported by researchers in the security industry. A particularly prominent DDoS attack type is amplification attacks. They take advantage of protocol design flaws, whereby a relatively small request triggers a significantly larger response. With a spoofed source IP address the response traffic is amplified and reflected to the target. Vulnerable protocols include classical protocols such as NTP, DNS, and/or SNMP, as well as relatively new protocols, e.g., DNSSEC and memcached. Amplification factors of up to 50,000-times have been witnessed in the wild. To exemplify, a request of 15 bytes can trigger a 750 Kbytes response.

This alarming increase in DDoS attacks and their sophistication and severity demands scalable yet cost-effective countermeasures. At this point, we are, however, left with various mitigation techniques and tools that can partially counteract the impact of the attacks. These include (i) Traffic Scrubbing Services (TSS), (ii) Router Access Control List Filters (ACL), (iii) Remotely Triggered Black Hole (RTBH), and (iv) BGP Flowspec.

In this research effort, we propose another approach for attack mitigation, called Advanced Blackholing (Advanced BH). Advanced Blackholing does not require trust, cooperation,
and sharing of resources among networks. It builds upon the excellent scalability of RTBH (to aggressively drop volumetric attack traffic) while incorporating the good properties of ACLs, Flowspec, and TSS (fine-granular filtering) in a lightweight fashion. Thus, Advanced Blackholing offers a new service in between RTBH and TSS and, as we will show, it can be deployed at scale, e.g., at IXPs.

IXPs offer an ideal deployment location for DDoS traffic mitigation as many ISPs use them to exchange traffic, e.g., more than 800 networks and more than 6 Tbps at DE-CIX in Frankfurt or AMS-IX in Amsterdam. Notice that by enabling such a service in one of these large IXPs, hundreds of member networks (as well as their customers and peer networks) will immediately benefit without the need to change anything else in the Internet protocols and the operation of the member networks, and without cooperation and coordination between two member networks. IXPs can also easily absorb the largest attacks seen to date, as they have Tbps of capacity before the attack reaches the ports (Gbps of capacity) of their members. Moreover, IXPs have existing routing infrastructure via the route servers, they have experienced network management teams, and they are increasingly hosting critical infrastructure, such as root DNS and NTP servers.

On the data plane, Advanced Blackholing combines on demand fine-grained filtering based on layer 2-4 header information with rate limiting. This can be done via vendor specific filters or SDN OpenFlow rules. IXP members can trigger Advanced Blackholing filters either via BGP attributes or SDN on their ports to drop or shape attack traffic. Thus, Advanced Blackholing achieves scalable scrubbing while giving feedback about the state and volume of the attack (telemetry) to the Advanced Blackholing users.

Our prototype, Stellar, relies on filtering and rate limiting of traffic and on BGP communities for signaling. We focus on the latter for the prototype to enable fast adoption in practice. Moreover, we show that Stellar requires no configuration by the IXP members, light configuration by the IXP operator, and the attack traffic is dropped at the IXP.

In addition to describing the design and implementation of the Stellar system, which realizes Advanced Blackholing at a major European IXP, this study reports on our initial experience with the system to mitigate attacks. Our evaluations indicate that Stellar scales well even if Advanced Blackholing requests and attack traffic increases at very high levels. Stellar also allows fast responses and is highly configurable, e.g., it provides traffic shaping to give telemetry feedback on the attack to its users. We are currently deploying Stellar as a service at a large IXP and we plan to install it at other IXPs in the near future.

References


Traffic asymmetry typically refers to an imbalance (or skew) in the contribution of ingress and egress traffic to the total traffic. Figure 37.1(a) reveals the prevalence of this traffic asymmetry, while Figure 37.1(b) shows its temporal characteristics. We calculate, for each hour of traffic observed by a large European Tier-1 ISP, the total ingress and egress traffic volumes. We plot, in Figure 37.1(a), the CDF of the percentage of ingress traffic in the total traffic, computed once for each hour of observed traffic over the one-week study period. The share of egress traffic is simply the complement of this CDF, and our choice of plotting the ingress share simply reflects the dominant characteristic of the data set: traffic observed by the ISP is mostly ingress (with a median ingress traffic of roughly 78%) and the percentage of ingress traffic varies from (a very high minimum of) 70% to a maximum of 82%. The 12% of variation in ingress traffic share, although appearing smaller, contributes to a substantial volume of traffic.

Figure 37.1(b) shows the temporal characteristics of traffic asymmetry, with time along the X-axis and hourly ingress traffic share along the Y-axis. The plot asserts that the asymmetry follows a diurnal cycle (consistently over the entire week) similar to that exhibited by end-user Internet traffic. Most of the end-user applications (e.g., Web browsing and video streaming) are highly asymmetric, i.e., size of requests are at least an order of magnitude smaller than that of responses. With such applications accounting for the majority of the network traffic, the traffic asymmetry we observe (along with its temporal characteristics) is not surprising. This high-level characterization of traffic asymmetry, nevertheless, leads to several follow-up questions. We enumerate a small subset of these questions that have huge implications for network planning and operations.
- Who contributes most to the traffic asymmetry: Is the traffic originating from and destined for ASes that directly peer with the ISP more asymmetric than that compared to others?
- What is the interplay between routing and traffic asymmetry: With more than one path between two ASes, how does traffic flow over these paths? From the viewpoint of either endpoint, for instance, is traffic always ingressing and egressing over the same network path?
- Are hypergiants the biggest contributor of traffic asymmetry? Are their traffic steering policies and mechanisms the primary reason behind ingress and egress traffic flowing over different paths?

We are currently investigating these questions using data sets obtained from diverse vantage points, e.g., a large tier-1 ISP and a large European IXP [1].

References

37.6 Optimal Transport for Video Streaming

Investigators: Balakrishnan Chandrasekaran, Mirko Palmer, and Anja Feldmann in cooperation with Ramesh Sitaraman (University of Massachusetts Amherst & Akamai Technologies) and Kevin Spiteri (University of Massachusetts Amherst)

Today, video traffic dominates the overall Internet traffic. Several forecasts further affirm that the volume of video traffic will continue to increase significantly: Cisco estimates, for instance, that video traffic will account for 82% of all IP traffic by 2021, up from 72% in 2016; Akamai, one of the largest content delivery networks (CDNs) in the world, projects that video will drive a seven-fold traffic increase by 2020. Further, with the addition of live streaming and broadcasts over the Internet, the need to deliver video traffic in an efficient manner (e.g., by imposing less load on the network) while ensuring that end-users’ experiences are unaffected (e.g., by avoiding video rebuffering or stalls) is clearer than ever before.

Streaming video over the Internet is, however, a challenging problem, exacerbated by the fact that video is a “real-time” service with either a “hard” real-time or rate-adaptive utility curve. The inherent challenges in streaming real-time video traffic over varying, and sometimes less than ideal, network conditions are only exacerbated by the choice of a reliable transport—so far, TCP. It is well known that TCP is not suited for video streaming: the rich body of prior work on optimizing and extending TCP, and adaptive bitrate (ABR) selection attest to this observation. TCP’s retransmissions of lost packets in a video stream, inadvertently lead to stalls in the video stream. TCP also performs poorly when it encounters packet losses that are not due to congestion. While prior efforts explored a wide range of the solution space, the simple question of “Why should we use only a reliable transport like TCP for streaming video in the Internet?” has, surprisingly, not received much attention. TCP,
unfortunately, still remains the de facto choice for video streaming due to the widespread use of adaptive streaming over HTTP.

Google’s Quick UDP Internet Connections (QUIC) protocol, recently, took a positive, albeit small, step towards changing this status quo. QUIC enables the design of sophisticated protocols on top of UDP, and allows multiple streams to be multiplexed over a single connection. The design, however, takes a crucial misstep: the streams support only reliable transport. By eschewing unreliable delivery, QUIC, thus, falls trap to most, if not all, of TCP’s problems for video delivery: needless retransmissions and rebuffering when network conditions are less than ideal. Indeed, in some instances, QUIC has been shown to perform poorer than TCP for video streaming. Lastly, QUIC is optimized for minimizing latency, which is often not the major cause for video streaming issues.

We propose the use of a hybrid approach—using both reliable and unreliable transport as needed—motivated by a simple observation: not all frames in a (widely-used) video encoding scheme are equally “important”; some frames (e.g., I-Frames or key frames) are more “important” than others (e.g., B- and P-Frames). By “importance” we refer to the implications of a loss of a frame, contained in a video stream, for the quality of experience (QoE) that an end user attributes when watching that video. Our aim is to enrich the interface between the transport protocol and the application to optimize video streaming, build support for selectively enabling reliability in the underlying transport, and investigate the design of an end-to-end optimal video streaming solution.

As part of this ongoing project, we have already implemented support for unreliable streams in QUIC, while ensuring that our implementation does not hamper the evolution of the protocol. We also solicited feedback from the Internet standardization bodies by submitted our proposal as an Internet draft [2]. We are also actively engaged in collaboration with a CDN to determine how we could test our solution in the wild. To engage the networking community, we also published a manuscript in the QUIC workshop detailing our implementation, preliminary findings and benefits of our approach [1].

In collaboration with researchers at the University of Massachusetts Amherst and Akamai Technologies, we are currently evaluating our hybrid transport for video streaming under diverse network conditions both in controlled “in-lab” environments and real “in-the-wild” experiments. Our findings and efforts should significantly inform the debate on video streaming in the Internet.

References


37.7 Socket Intents

Investigators: Anja Feldmann in cooperation with Theresa Enghardt and Philipp S. Tiesel (Technische Universität Berlin)

Today, most mobile devices can connect to the Internet via multiple access networks, e.g., WiFi/DSL and LTE. To take advantage of the networks’ diverse path characteristics (e.g., delay, bandwidth, and reliability) and aggregate bandwidth, we need smart strategies for choosing which interface(s) to use for what traffic. In this research effort, we present an approach on how to tackle this challenge: With the concept of Socket Intents, applications can express what they know about their communication pattern and their preferences [7, 2].

Using our Socket Intents Prototype, this information is used to choose the most appropriate path or path combination on a per message or per connection basis. Furthermore to make an informed choice of which network to use, a host requires accurate and up-to-date performance metrics. So far such network characteristics are, however, typically not readily available and can be highly volatile. We explore what performance metrics are available on a host by monitoring and aggregating them within the Socket Intents prototype. These metrics then feed into our access network selection policies to improve page load time in a testbed emulating various network characteristics. For Web browsing, the Socket Intents prototype can distribute resources across multiple interfaces, and, thus, multiple access networks. With the application telling us about the expected size of a resource, we can load small resources on an interface with a low observed latency, and large resources on an interface with a high observed download capacity. We evaluate the possible benefits for Web browsing in a testbed and “in the wild”. In a network scenario with asymmetric network characteristics, we find that the Socket Intents prototype improves Page Load Time (PLT) in most cases compared to using a single interface or even compared to MPTCP.

In addition to the design, development, and evaluation of the socket intents prototype in various scenarios, we published several Internet drafts [6, 4, 5] to the IETF and actively sought feedback from the standardization committees in an effort to standardize and deploy the socket intents framework. Our work on the architecture and development of the socket intents prototype also influenced several design discussions concerning the TAPS system [3, 8, 1].

References


### 37.8 Distributed and Timely Flow Summarization at Scale

**Investigators:** Anja Feldmann, Aniss Maghsoudlou, and Jawad Said Saidi in cooperation with Damien Foucard and Georgios Smaragdakis (Technische Universität Berlin)

Network operators have to continuously keep track of the activity in their network to identify a wide range of unusual events, e.g., attacks or disruptions, as well as to derive statistics over long time windows to provision their network or make informed peering decisions. Typically, they rely on flow-level or packet-level data captured at different routers in their network. Over the last few years, all major router vendors have made flow and packet capture utilities available in their products. A flow capture maintains information about the five-tuple (protocol, source IP, destination IP, source port, destination port, traffic exchanged) while a packet capture stores the header of, typically, sampled packets that traverse a router. Among the most popular flow captures are Netflow and IPFIX while the libcap library is the most common packet-level capture utility.

The increasing availability of large-scale flow and packet captures make it possible for the operators to get insights about the state and the health of their network. This benefits, however, comes at certain costs. The massive amount of data is increasingly difficult to store, transfer, or analyze promptly, as line speeds and the number of online users and applications steadily increase. Moreover, stored flow or packet captures do not provide an obvious way to answer typical operator questions, e.g., to find the start or end of an attack event. Indeed, most queries require scanning multiple captures—an inefficient, often manual, and slow process which restricts ad-hoc queries. This requirement stresses the need for online indexing of flows on top of existing captures.

Indeed, for most network operators the storage of flow and packet captures is a pressing issue. Since it is prohibitively expensive to store all the captured data, they typically have to delete past captures, e.g., using expiration dates. This practice further limits the range of investigation. To overcome this problem we need a way to maintain succinct (i.e., space-efficient) flow summaries which enable accurate answers for a large range of queries.

This problem is further exacerbated by the fact that flow captures are typically collected at multiple border and backbone routers in a network. Transferring the raw traces to a centralized processing location is (a) increasingly expensive due to the data volume or (b) may be forbidden due to regulatory restrictions or due to different jurisdictions. Yet, if we
had succinct mergeable data summaries we could tackle these problems by only transferring these or even the differences between consecutive summaries. For this, we need a data structure that supports the operators merge and diff to track network-wide activity across time and sites. To the best of our knowledge, a distributed flow processing system that can answer complex queries on-the-fly at scale is not available today.

To address this issue, we propose an efficient summarization and sharing system, called Flowstream. The Flowstream system (i) operates on top of existing flow capture utilities, (ii) guarantees prompt and accurate execution of queries for popular generalized flows, and (iii) does so across sites and/or across time. We envision that each router exports its data to a close-by Flowtree daemon using an existing API of e.g., Netflow or IPFIX. It continuously constructs and maintains a summary of the active flows using a novel self-adjusting data structure, called Flowtree. Flowtree maintains a hierarchical tree and keeps an accurate replica for the most popular generalized flows (the leafs of the tree) and aggregated statistics for less popular flows (interior nodes of the tree—upwards in the flow hierarchy). This data structure enables accurate and timely answer for queries across multiple dimensions. In effect, Flowtree creates an index of and sorts the active generalized flows based on their popularity. Thus, it is also easy to identify not only flows that are popular but also those that are not. Our preliminary evaluations of Flowtree show that this data structure can yield accurate yet scalable flow summaries with substantial memory savings compared to prior work on similar summarization or aggregation data structures [1].

References


37.9 The Need for New Computing Primitives

Investigators: Anja Feldmann in cooperation with Niklas Semmler and Georgios Smaragdakis (Technische Universität Berlin)

A few decades ago, digitization entered many processes of modern society, from public administration to e-commerce, from transportation to industrial automation, ultimately permeating our everyday lives. We are now in the middle of a digital revolution of unprecedented intensity where digitization will force the future convergence of the physical and digital worlds with ubiquitous, novel, and disruptive applications. A prerequisite for this digital future is timely and dependable information from the physical world that allows the physical processes and their digital representations to interact through multiple real-time control loops at different levels of time, spatial scale, and detail.

Maintaining a digital representation of the real world in a consistent and timely manner is challenging. The challenges arise when handling data at different scales of time (from sub-millisecond to hours, and days, spanning easily six to seven orders of magnitude) and space. Other challenges include handling heterogeneous, non-stationary, and non-uniform, distributed data streams. Furthermore, sensing must efficiently produce summaries of the physical world, while achieving specific (task-defined) approximation guarantees and respecting privacy during data integration and analysis.
Our vision in this realm is driven by the increasing deployment of sensors and their increasing resolution, particularly at the edge of the Internet. These sensors produce data at a rate that outpaces the capacity growth of wide-area networks. This trend has led to the creation of datasets that originate from many different sources. These datasets can no longer be fully stored and/or processed within a single computer system. We refer to them as “mega-datasets.” A mega-dataset can only be handled by a distributed, yet local system, e.g., a cluster of compute nodes. We define a distributed mega-dataset as a collection of physically distributed mega-datasets. Processing the distributed mega-datasets, in a coordinated, yet distributed fashion in real time requires novel scalable computing primitives. Such primitives need to summarize data into a form that can be transferred with low latency (while using minimal resources and addressing data lineage, quality criteria, and time constraints). Our objective is to explore possible architectures that will allow us to retrieve and store these mega-datasets and simple, novel computing primitives that will help in handling these datasets.

37.10 Academic Activities

37.10.1 Conference and Workshop Positions

Membership in program and organization committees

Prof. Anja Feldmann, Ph.D.:

- *IEEE International Conference on Computer Communications (INFOCOM) 2020*, Beijing, China, April 2020 (PC member).
- *IFIP Networking, 2019*, Warsaw, Poland, May 2019 (PC member).
- *The International Teletraffic Congress (ITC) 30*, Vienna, Austria, September 2018 (PC member).
- *IEEE International Conference on Computer Communications (INFOCOM) 2018*, Honolulu, Hawaii, USA, April 2018 (PC member).
Balakrishnan Chandrasekaran, Ph.D.:
- *ACM Symposium on SDN Research (SOSR) 2019*, San Jose, California (USA), April 2019 (PC member).

Franziska Lichtblau:
- *RIPE75*, Dubai, UAE, October 2017 (PC member).
- *RIPE76*, Marseilles, France, May 2018 (PC member).
- *RIPE77*, Amsterdam, The Netherlands, October 2018 (PC member).
- *RIPE79*, Rotterdam, The Netherlands, October 2019 (PC member).

Membership in steering and other committees

Prof. Anja Feldmann, Ph.D.:
- *ACM Internet Measurement Conference*, (Steering Committee),
- *IEEE/ACM Transactions on Networking (ToN)*, (Steering Committee),
- *Unesco Fachausschuss Kommunikation und Information*, (Member),
- *Dagstuhl Wissenschaftlicher Rat*, (Member).

### 37.10.2 Invited Talks and Tutorials

Prof. Anja Feldmann, Ph.D.:
- *IEEE LANMAN 2019: Keynote*, 01.07.2019, Paris, France
- *PAM 2019: Keynote*, 28.03.2019, Puerto Varas, Chile
- *Digitising Europe Summit*, 19.02.2019, Berlin, Germany
- *COMSNETS 2019: Keynote*, 09.01.2019, Bangalore, India
- *Distinguished Lecture Series*, 16.11.2018, RWTH, Aachen, Germany
- *MPG: Tag mit Wissenschaft*, 19.10.2018, Schloss Ringberg, Germany
- *Podiumsdiskussion Gutachtertagung des interdisziplinären Forschungsprojekts Assessing Big Data (ABIDA)*, 18.10.2018, BMBF Berlin, Germany
- *25 Jahre TKN “Ein Blick zurück, ein Blick nach vorn”*, 12.10.2018
- *BBAW: Wissenschaftliche Herausforderungen*, 10.09.2018
- *Sigcomm SecSoN: Keynote*, 24.08.2018, Budapest, Hungary
- *TMA Experts Summit*, 26.06.2018, Wien, Austria
- *MPI Lecture Series*, 06.06.2018, Saarbrücken, Germany
37.11 Teaching Activities

Winter Semester 2018/2019

Data Networks (A. Feldmann and B. Chandrasekaran)

37.12 Dissertations, Habilitations, Awards

37.12.1 Dissertations


37.12.2 Awards

- Anja Feldmann, Vodafone Stiftung: Vodafone Innovationspreis, 2018 for her work on innovation, development, and optimization of Internet architectures.
- Anja Feldmann, Bayerische Akademie der Wissenschaften: Friedrich Wilhelm Joseph von Schelling Prize, 2018, for her outstanding achievements in the field of computer science, especially regarding Internet protocols and architectures.
- TU-Berlin: Honorarprofessor, 2018
37.13 Publications

Books and proceedings


Journal articles


Conference articles


Workshop articles and posters


Preprints


38 D4: Computer Graphics

38.1 Personnel

Director
Prof. Dr. Hans-Peter Seidel

Senior Researchers and Group Leaders
Dr. Vahid Babaei (from September 2018)
Dr. Renjie Chen
Dr. Piotr Didyk (until March 2018)
Prof. Dr. Karol Myszkowski
Gurprit Singh, Ph.D. (from September 2017)
Prof. Dr. Christian Theobalt
Dr. Rhaleb Zayer

Researchers
Vamsi Adhikarla, Docter (until June 2017)
Dr. Shida Beigpour (until December 2017)
Dr. Florian Bernard (from June 2017)
Avishek Chatterjee, Ph.D. (until August 2017)
Mohamed Elgharib, Ph.D. (from July 2018)
Dr. Vladislav Golyanik (from August 2018)
Dr. Caigui Jiang (until January 2018)
Dr. Srinath Sridhar (until June 2017)
Dr. Okan Tursun
Weipeng Xu, Ph.D.
Hyeonseung Yu, Ph.D. (from May 2017)
Dr. Michael Zollhöfer (until November 2017)

PhD Students
Elena Arabadzhiyska
Mojtaba Bemana (from March 2018)
Mallikarjun Byrasandra Ramalinga Reddy (from January 2019)
Pablo Garrido (until November 2017)
Björn Golla (from December 2018)
Marc Habermann (from September 2017)
Ikhsanul Habibie (from February 2018)
Jozef Hladký
Hyeongwoo Kim
Thomas Leimkühler
Dushyant Mehta
Abhimitra Meka
Franziska Müller
Oliver Nalbach (until September 2017)
Michal Piovarči
Nadia Robertini (until August 2017)
Denis Sumin
Ayush Tewari
Edgar Tretschk (from October 2018)
Jiayi Wang (from May 2018)
Krzysztof Wolski (from January 2018)

Research Engineer
Oleksandr Sotnychenko

Project Coordination
Dr. Bertram Somieski

Secretaries
Sabine Budde
Ellen Fries

38.2 Visitors

From March 2017 to February 2019, the following researchers visited our group:

Christoph Lassner 23.04.2017–25.04.2017 MPI for Intelligent Systems, Tübingen, Germany
Markus Steinberger 22.05.2017–26.05.2017 Graz University of Technology, Austria
Mark Dokter 22.05.2017–26.05.2017 Graz University of Technology, Austria
Mohamed Elgharib 27.05.2017–31.05.2017 Qatar Computing Research Institute, Qatar
Adrien Bousseau 01.06.2017–02.06.2017 INRIA, Sophia Antipolis, France
Yulia Gryaditskaya 01.06.2017–02.06.2017 INRIA, Le Chesnay, France
<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
<th>Institution and Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligang Liu</td>
<td>03.06.2017–11.06.2017</td>
<td>University of Science and Technology, Hefei, P.R.China</td>
</tr>
<tr>
<td>Gerard Pons-Moll</td>
<td>12.06.2017–13.06.2017</td>
<td>MPI for Intelligent Systems, Tübingen, Germany</td>
</tr>
<tr>
<td>Vlastimil Havran</td>
<td>14.06.2017–16.06.2017</td>
<td>Czech Technical University, Prague, Czech Republic</td>
</tr>
<tr>
<td>Björn Golla</td>
<td>19.06.2017–30.11.2018</td>
<td>University of Magdeburg, Germany</td>
</tr>
<tr>
<td>Patrick Perez</td>
<td>25.06.2017–26.06.2017</td>
<td>Technicolor, Rennes, France</td>
</tr>
<tr>
<td>Gurprit Singh</td>
<td>23.06.2017</td>
<td>Dartmouth College, Hanover, USA</td>
</tr>
<tr>
<td>Peter Wonka</td>
<td>27.06.2017–29.06.2017</td>
<td>King Abdullah University of Science and Technology, Saudi-Arabia</td>
</tr>
<tr>
<td>Dawid Palen</td>
<td>01.08.2017–30.09.2017</td>
<td>West Pomeranian University of Technology, Szczecin, Poland</td>
</tr>
<tr>
<td>Michal Chwesiuk</td>
<td>01.08.2017–16.02.2018</td>
<td>West Pomeranian University of Technology, Szczecin, Poland</td>
</tr>
<tr>
<td>Wentao Liao</td>
<td>21.08.2017–31.01.2018</td>
<td>University of Science and Technology, Hefei, P.R.China</td>
</tr>
<tr>
<td>Daniel Mlakar</td>
<td>01.09.2017–31.10.2018</td>
<td>Graz University of Technology, Austria</td>
</tr>
<tr>
<td>Lucas Thies</td>
<td>06.09.2017–08.10.2017</td>
<td>University of Erlangen, Germany</td>
</tr>
<tr>
<td>Shuhang Gu</td>
<td>07.09.2017–10.09.2017</td>
<td>Polytechnic University of Hong Kong, P.R.China</td>
</tr>
<tr>
<td>Martin Winter</td>
<td>07.09.2017–11.09.2017</td>
<td>Graz University of Technology, Austria</td>
</tr>
<tr>
<td>Vahid Babaei</td>
<td>21.09.2017–31.08.2018</td>
<td>Massachusetts Institute of Technology, USA</td>
</tr>
<tr>
<td>Mojtaba Bemana</td>
<td>22.09.2017–21.03.2018</td>
<td>Sharif University of Technology, Tehran, I.R.Iran</td>
</tr>
<tr>
<td>Tianteng Bi</td>
<td>18.10.2017–21.06.2018</td>
<td>Beijing Institute of Technology, P.R.China</td>
</tr>
<tr>
<td>Mario Botsch</td>
<td>03.11.2017–04.11.2017</td>
<td>University of Bielefeld, Germany</td>
</tr>
<tr>
<td>Tobias Ritschel</td>
<td>17.11.2017</td>
<td>University College London, U.K.</td>
</tr>
<tr>
<td>Angela Dai</td>
<td>24.11.2017–25.11.2017</td>
<td>Stanford University, USA</td>
</tr>
<tr>
<td>Thabo Beeler</td>
<td>27.11.2017</td>
<td>ETH Zürich, Switzerland</td>
</tr>
<tr>
<td>Vladislav Golyanik</td>
<td>08.12.2017</td>
<td>University of Kaiserlautern, Germany</td>
</tr>
<tr>
<td>Tobias Rittig</td>
<td>05.01.2018–25.01.2018</td>
<td>Charles University Prague, Czech Republic</td>
</tr>
<tr>
<td>Johan Thunberg</td>
<td>23.01.2018</td>
<td>LCSB Luxembourg, Luxembourg</td>
</tr>
<tr>
<td>Michael Möller</td>
<td>05.02.2018–06.02.2018</td>
<td>University of Siegen, Germany</td>
</tr>
</tbody>
</table>
Rao Fu 25.02.2018–26.02.2018 Tsinghua University, Beijing, P.R.China
Mallikarjun BR 24.02.2018–27.02.2018 Mercedes Benz, Bangalore, India
Lingjie Liu 01.03.2018–20.01.2019 University of Hong Kong, P.R.China
Nitish Padmanaan 19.03.2018 Stanford Computational Imaging Lab, USA
Vincent Sitzmann 19.03.2018 Stanford Computational Imaging Lab, USA
Markus Steinberger 01.04.2018–08.04.2018 Graz University of Technology, Austria
Wenping Wang 03.04.2018–07.04.2018 University of Hong Kong, P.R.China
David Dunn 26.04.2018–27.04.2018 UNC Chapel Hill, USA
Marek Wernikowski 01.06.2018–31.12.2018 West Pomeranian University of Technology, Szczecin, Poland
Eakta Jain 21.06.2018 University of Florida, USA
Krzysztof Krejtz 21.06.2018 University of Warsaw, Poland
Markus Steinberger 26.06.2018–30.06.2018 Graz University of Technology, Austria
Sumit Shekhar 01.07.2018–26.09.2018 Saarbrücken, Germany
Seyedeharezou Fatemi 05.07.2018–21.09.2018 Sharif University, Tehran, I.R.Iran
Stephan Stiboller 09.07.2018–08.10.2018 Graz University of Technology, Austria
Markus Prettner 09.07.2018–31.08.2018 Graz University of Technology, Austria
Wolfgang Tatzgern 09.07.2018–31.08.2018 Graz University of Technology, Austria
Sepideh Sarajian Maralan 10.07.2018–10.09.2018 Sharif University, Tehran, I.R.Iran
Linjie Lyu 15.07.2018–09.09.2018 Tsinghua University, Beijing, P.R.China
Alexander Weinrauch 01.08.2018–15.11.2018 Graz University of Technology, Austria
Mallikarjun BR 01.08.2018–31.12.2018 Mercedes Benz, Bangalore, India
Markus Oberweger 29.08.2018–31.08.2018 Graz University of Technology, Austria
Ana Serrano 17.09.2018–13.11.2018 University of Zaragoza, Spain
Lan Xu 01.10.2018–01.04.2019 Hong Kong University of Science and Technology, P.R.China
Yue Zhang 18.10.2018–19.10.2018 Oregon State University, USA
Michael Neff 31.10.2018 University of California, Davis, USA
Yao Feng 19.11.2018–21.11.2018 Shanghai Jiao Tong University, P.R.China
38.3 Group Organization

Our research is currently organized into the following six research areas, each having its own small group of coordinators:

- Digital Geometry Processing (R. Chen, R. Zayer, and H.-P. Seidel)
- Computational Videography (C. Theobalt)
- Reconstructing the Static and Dynamic Real World (C. Theobalt)
- Realistic and Real-time Rendering (G. Singh)
- Perception and Advanced Displays (K. Myszkowski)
- Computational Fabrication (V. Babaei)

The coordinators coordinate the work in their areas together with Hans-Peter Seidel and they form the D4 steering committee. The steering committee meets on a weekly basis (Tuesday, 11 am) and discusses all group related issues. In particular, it addresses topics such as recruiting, guests and seminars, teaching, project acquisition, mid-term and long-term strategic planning.

The whole group meets thrice a week for the

- D4 lab meeting (Tuesday, 12:30 pm), where organizational issues are discussed and information is distributed by the members of the steering committee,
- D4 graphics colloquium (Tuesday, 1 pm), where visitors present their ongoing work to the group, the computer graphics group at Saarland University and to other interested people, and the
- D4 graphics lunch (Thursday, 11:30 am), where people from within D4 present their work in progress to the group. The main goal of this meeting is to keep the group informed on the ongoing projects, collect the group feedback and influence further project development at relatively early stages.

Apart from these formal meetings, there are several meetings and discussion groups that also take place frequently, but not on a totally regular basis, such as paper discussion groups that discuss papers of special interest, especially immediately preceding major conference events; technical meetings in special areas that are of particular interest to a specific subset of researchers (often in cooperation with people from the graphics group at Saarland University); internship and practical course meetings where all people involved in internships or FoPras meet and discuss; and last but not least meetings dedicated to single projects.
Digital Geometry Processing is concerned with the representation, analysis, manipulation, and optimization of digital shapes. Due to the advances in 3D acquisition and manufacturing technologies, the usage of geometric data is continuously increasing and the efficient processing of digital shapes plays an important role for a variety of applications in areas such as computer graphics, computer-aided design and manufacturing, medical imaging and surgery planning, architecture, entertainment, etc. We develop theoretical models, numerical schemes, and algorithms for a variety of problems in digital geometry processing.

38.4.1 Shape Deformation and Mapping Optimization

Investigators: Björn Golla and Renjie Chen

Shape deformation is a classical problem in computer graphics and animation that comes in a variety of forms and settings. Essential to all variants is the requirement to compute a map between different spaces with some desirable properties. Some requirements such as smoothness are easily achieved while other, such as local injectivity or strict bound on the induced metric distortion are hard (sometimes impossible) to guarantee. In [3], we proposed an interactive deformation system that is driven by user specified positional constraints. Central to our deformation system is a highly efficient planar meshless shape deformation algorithm. Our method is based on an unconstrained minimization of isometric energies, and is guaranteed to produce $C^\infty$ locally injective maps by operating within a reduced dimensional subspace of harmonic maps. We extend the harmonic subspace of [2] to support multiply-connected domains, and further provide a generalization of the bounded distortion theorem that appeared in that paper. Our harmonic map, as well as the gradient and the Hessian of our isometric energies possess closed-form expressions. A key result is a simple-and-fast analytic modification of the Hessian of the energy such that it is positive definite, which is crucial for the successful operation of a Newton solver. The method is relatively simple to implement and is specifically designed to harness the processing power of modern graphics hardware. Our modified Newton iterations are shown to be extremely effective, leading to fast convergence after a handful of iterations, while each iteration is fast due to a combination of a number of factors, such as the smoothness and the low dimensionality of the subspace, the closed-form expressions for the differentials, and the avoidance of expensive strategies to ensure positive definiteness. The entire pipeline is carried out on the GPU, leading to deformations that are significantly faster to compute than the state-of-the-art. Figure 38.1 shows the deformation result of a doubly-connected domain.

The effectiveness of the GPU-accelerated harmonic shape deformation method motivated us extended the Newton-method to triangle mesh discretizations of the surface in [4]. This enables the optimization of maps from surfaces in 3D space to the plane, which is the basis for parameterization. Based on a complex formulation of the piecewise linear mapping, we show that the Hessian of isometric energies has similar simple and compact analytic expressions. This allows analytical projection of the per-element Hessians to positive semidefinite matrices for efficient Newton iteration. Further, we inspect the spectra of the
Figure 38.1: Locally injective isometric deformation of Rex, a doubly-connected planar domain (left). Note the hole between the legs. Our solver converged after 6 iterations to the result depicted, which was computed in 0.07s on the GPU. The Newton-Eigen solver took 2.47s which is $\times 35$ times slower.

per triangle energy Hessians and show that given an initial mapping, simple global scaling can shift the energy towards a more convex state. This allows our method to converge faster than starting from the given initial state. Our method outperforms state-of-the-art methods on 2D deformation and parameterization, as shown in Figure 38.2. Additionally, our formulations support adding an energy smoothness term to the optimization with little additional effort, which improves the mapping results such that concentrated distortions are reduced.

Figure 38.2: Comparison of different optimization methods for shape deformation and parameterization.

Continuous conformal maps are typically approximated numerically using a triangle mesh which discretizes the plane, due to its popularity and hardware friendliness. Computing a conformal map subject to user-provided constraints then reduces to a sparse linear system, minimizing a quadratic “conformal energy”. In [1], we address the more general case of non-triangular elements, and provide a complete analysis of the case where the plane is discretized using a mesh of regular polygons, e.g. equilateral triangles, squares and hexagons, whose interiors are mapped using barycentric coordinate functions. We demonstrate experimentally that faster convergence to continuous conformal maps may be obtained this way. For example, relative to a regular triangle mesh of 217 vertices, a regular hex mesh with a comparable number of vertices provides a 7$\times$ reduction in conformal energy at the expense of only
25% more computation time, while a regular quad mesh with a comparable number of vertices provides a 2× reduction in conformal energy while requiring 40% less computation time. We provide a formulation of the problem and its solution using complex number algebra, significantly simplifying the notation. We examine a number of common barycentric coordinate functions and demonstrate that superior approximation to harmonic coordinates of a polygon are achieved by the Moving Least Squares coordinates. We also provide a simple iterative algorithm to invert barycentric maps of regular polygons, allowing to apply them in practical applications, e.g. for texture mapping.

References


38.4.2 Path Planning with Divergence-Based Distance Functions

Investigator: Renjie Chen

Path generation is an important problem in many fields, especially robotics. One way to create a path between a source point $z$ and a target point $y$ inside a complex planar domain $\Omega$ is to define a non-negative distance function $d(y, z)$, which is used to automatically plan a gradient-descent path towards $y$, avoiding obstacles that may be present. A key requirement from such distance functions is the absence of spurious local minima, and this has led to the common use of harmonic potential functions. This choice guarantees the absence of spurious minima, but is well known to be slow to numerically compute and prone to numerical precision issues. To alleviate the first of these problems, we propose a family of novel divergence distances in [2]. These are based on $f$-divergence of the Poisson kernel of the domain. Using the concept of conformal invariance, we show that divergence distances are equivalent to the harmonic potential function on simply-connected domains, namely generate paths which are identical to those generated by the potential function. We then discuss how to compute two special cases of divergence distances, one based on the Kullback-Leibler, the other on the total variation divergence, in practice by discretizing the domain with a triangle mesh and using Finite Elements (FEM) computation. We show that using divergence distances instead of the potential function and other distances has a significant computational advantage, as, following a pre-processing stage, they may be computed online in a multi-query scenario up to an order of magnitude faster than the others when taking advantage of certain sparsity properties of the Poisson kernel. Furthermore, the computation is “embarrassingly parallel”, so may be implemented on a GPU with up to three orders of magnitude speedup.
To further improve the computational efficiency and storage complexity, we show how to assign a set of reduced coordinates to each point in \( \Omega \) and to define a family of distance functions based on these coordinates using the concepts of harmonic measure and \( f \)-divergence. Figure 38.3 shows the paths generated using reduced coordinates. Since in practice, especially in robotics applications, the path is often restricted to follow the edges of a discrete network defined on a finite set of sites sampled from \( \Omega \), any method that works well in the continuous setting must be discretized appropriately preserve the important properties of the continuous case. We show how to define a network connecting a finite set of sites, such that a greedy routing algorithm, which is the discrete equivalent of continuous gradient descent, based on our reduced coordinates is guaranteed to generate a path in the network between any two sites. In many cases, this network is close to a planar graph, especially if the set of sites is dense. Guaranteeing the existence of a greedy route between any two points in the graph is a significant advantage in practical applications, avoiding the complexity of other path-planning methods, such as the shortest-path and A* algorithms. While the paths generated by our algorithm are not the shortest possible, in practice we found that they are close to that.

Figure 38.3: Path planning with reduced coordinates. While the \( L^2 \)-distance (left) works well for simple shapes, for complex domains it has local minima where the gradient-descent path terminates. In contrast, the \( f \)-divergence (right) is guaranteed to be void of local minima and can be implemented efficiently using reduced coordinates.

References


38.4.3 Design and Optimization of Space Structures

*Investigators: Caigui Jiang and Renjie Chen*

Space structures, also called space frames or space frame structures, are elegant and materially efficient truss-like structures consisting of beams (two-force members) connected at nodes.
Space structures are desirable and often necessary in industrial design and architectural construction. The design and optimization of space structures present many challenges, especially for designs with complex geometries.

In [4], we propose a systematic framework to design and optimize statically sound space structures that approximate freeform surfaces and satisfy the following goals. (Goal 1) First, the structure should be statically sound. This means that the structure is in force equilibrium with axial forces along the beams without bending moments. (Goal 2) Second, the structure should be constructed with regularly arranged beams and nodes to be aesthetically pleasing. (Goal 3) Third, the structure should approximate a given designed shape, e.g., a freeform architectural surface. (Goal 4) Fourth, the cost of the structure should be minimized. The most important factor associated with cost is material usage. A cost-effective space structure consists of beams with variable cross sections, given that most material is used for the beams. Because the beams are usually manufactured by extrusion (aluminium) or bending and welding (steel), followed by cutting, many beams should share the same cross section. To achieve the goals stated above, our framework allows a user to explore different configurations, optimize the node positions and connectivity, and adjust the beam cross sections. We first jointly optimize node positions and connectivity through a nonlinear continuous optimization algorithm. Next, with fixed nodes and connectivity, we formulate the assignment of beam cross sections as a mixed-integer programming problem with a bilinear objective function and quadratic constraints(4,8),(993,990). We solve this problem with a novel and practical alternating direction method based on linear programming relaxation. The capability and efficiency of the algorithms and the computational framework are validated by a variety of examples and comparisons. Figure 38.4 left shows a space structure design and optimized with our framework motivated by a real architectural project.

Designing optimal trusses that satisfy functional specifications with a minimal amount of material has interested both theoreticians and practitioners for more than a century. In [3], we introduce two main ideas to improve upon the state of the art. First, we formulate an alternating linear programming problem for geometry optimization. Second, we introduce two sets of complementary topological operations, including a novel subdivision scheme for global topology refinement inspired by Michell’s famed theoretical study. Based on these two ideas, we build an efficient computational framework for the design of lightweight trusses. We illustrate our framework with a variety of functional specifications and extensions. We show that our method achieves trusses with smaller volumes and is over two orders of magnitude faster compared with recent state-of-the-art approaches. Figure 38.4 right shows a bridge design started from a given input specification of supporting points (red) and two sets of external forces.

Figure 38.4: Examples of space structure design and optimization.
Besides under layer space frames or trusses, we also pay attention to the outer layer surfaces. In [2], we study the design and optimization of polygonal meshes with concave planar faces. To discretize freeform surfaces into polyhedral meshes, we propose a novel class of regularizers for mesh aesthetics based on symmetries. They are useful to generate concave polygons on negative Gaussian curvature region, and provide the necessary flexibility to create smooth transformation of planar faces across the region where Gaussian curvature alternated between positive and negative. In [1], we explore optimizing for centroidal Voronoi tessellations with rotational symmetry penalties in order to align mesh edges with the underlying curvature directions. The results indicate a high quality isotropic and field aligned remeshing with only few singularities.

References


38.4.4 High Performance Surface Tessellations

**Investigator: Rhaleb Zayer**

Probably the prevailing numerical model for spatial partitioning is the *Voronoi diagrams* (VD) model. A representation where seed points (generators) are separated by hyper-planes equidistant to their closest neighboring points. When generators are restricted to coincide with the center of mass of their respective cells, the resulting *Centroidal Voronoi tessellations* (CVT) exhibits more balanced cells. Both VD and CVT arise across various fields of study ranging from meteorology, geography, crystallography, and chemistry and remain a central theme in computer graphics and computational geometry. Their aesthetic appeal makes them a favorite choice in modern design and architecture as they evoke natural tessellations such as the patterns observed on drying earth or a giraffe.

While the creation of such tessellations in the planar or volumetric settings is facilitated by inexpensive distance evaluations, the surface setting remains a challenge due to expensive geodesic estimation and the toil of maintaining valid connectivity representations. Most existing formulations are inherently serial, lag in performance, and severely underuse modern parallel computing capabilities.

We re-examine the notion of Voronoi diagrams and propose an alternative model which both simplifies the creation of natural tessellations on surface meshes and brings unprecedented performance throughout a fully parallel algorithmic formulation suitable for fine grained parallelism [1]. Inspired by natural growth phenomena such as crystallization, where different cell fronts grow simultaneously till they collide and form cell walls, without requiring knowledge of the location of other seeds or the geodesic paths towards them. Unlike the
Figure 38.5: Illustration of the growth of natural tessellations on Amelia Earhart’s flight suit (10M facets mesh, courtesy of the Smithsonian Institution) initialized with 10K random seeds. Small scale geometric details (pocket fold, button fold) are processed successfully with our fully parallel GPU implantation.

VD model these cells are seldom separated by razor sharp lines. In order to capture this behavior, we regard cell boundaries as narrow bands and we model the partition as a set of smooth functions layered over the surface. Starting from an initial set of seeds or regions, the partition merges as the solution of a time dependent set of partial differential equations describing concurrently evolving fronts, see figure 38.5. The numerical solution is carried out by a simple Euler time-stepping scheme and a lean sparse matrix representation. Within this formalism, the cost per time-iteration is dominated by the multiplication of two sparse matrices and as such it lends itself to efficient implementation on modern graphics hardware. Although our formulation is intended for producing partitions separated by smooth narrow bands, sharp boundaries as well as the Delaunay-like dual mesh can be obtained and updated throughout CVT iterations without compromising performance.

References


### 38.4.5 High Performance Mesh Subdivision

**Investigator: Rhaleb Zayer**

Mesh subdivision is a ubiquitous method for generating free-form surfaces from coarser control meshes. Although subdivision surfaces have a long and rich history in research and
industry, their efficient evaluation still poses a challenge on modern parallel architectures. During subdivision, the control mesh undergoes a series of averaging, splitting and relaxation operations, which complicates efficient parallel implementation and data management. Traditionally, serial subdivision implementations rely on linked lists based mesh data structures, e.g., winged-edge and half-edge. Changes to the topology in such a data structure requires careful pointer updates to preserve consistency. Computations in the local neighborhood of mesh vertices? which are essential in subdivision? require pointer chasing. While those operations are efficient on the CPU, modern parallel devices, like the graphics processing unit (GPU), are faced with unbalanced workloads, synchronization issues, and scattered memory accesses?all of which significantly hurt performance.

With Algebra Subdivision (AlSub)\[1\], we provide a full fledged treatment mesh subdivision which recasts the problem into sparse linear algebra operations and allows for efficient parallel implementation. Our approach is modular and extensible, suitable for different subdivision schemes and handles specialized tasks, like mesh boundary treatment, creases, and selective subdivision. Our implementations not only ensures consistent results throughout all application scenarios, but also show significant performance improvements for all of them. While a direct implementation of this formulation \[4\] readily indicates high throughput (AlSub pure in Figure 38.6-left), we introduce a series of optimizations (AlSub opt.) that increase performance by another order of magnitude throughout all test cases, and dramatically reduces memory footprint, Figure 38.6-middle and largely outperforms the industry standard (OpenSubdiv \[3\]). Our approach significantly outperforms other implementations in scenarios where an input mesh must be subdivided once (2–3x faster than the patch-based GPU implementation in [2]) as shown in Figure 38.6-right. To the best of our knowledge AlSub and is the first fully GPU-enabled, universally applicable subdivision implementation.

References


38.4.6 Linear Algebra Kernels and Graph Algorithms

Investigator: Rhaleb Zayer

Linear algebra kernels and graph algorithms are the back bone of modern high performance computing. Being part of the critical path of many applications, every performance improvement at their level translates directly into gains for the entire application. Furthermore, high performance geometry processing can be achieved by channeling vectorization effort through these kernels as shown in Sections 38.4.4 and 38.4.5. We have therefore dedicated considerable efforts to improving the performance of these key elements as summarized in the following paragraphs.

Sparse matrix-vector product (SpMV) Conceptually, the sparse matrix-vector product (SpMV): \( y = A \cdot x \), where \( A \) is a sparse matrix and \( x \) and \( y \) are dense vectors, is a rather straightforward operation in linear algebra. Nonetheless, it is of utter importance, as it is often evaluated multiple times, be it in solving linear systems, performing eigenanalysis, or querying graph structures. Traditionally, the standard compressed sparse rows (CSR) representation of sparse matrices lends itself to a simple, yet efficient sequential SpMV implementation, as all non-zeros that contribute to one output element are placed next to another in memory. However, the advent of affordable, parallel architectures, poses new challenges for SpMV. Despite great strides over the last decade, most existing GPU-SpMV approaches trade off one aspect of performance against another. They either require preprocessing, exhibit inconsistent behavior, lead to execution divergence, suffer load imbalance or induce detrimental memory access patterns.

Figure 38.7: Trend line of the SpMV performance for all tested methods against our Ho-\( \lambda \)aSpMV over the entire University of Florida Sparse Matrix collection. Additional line thickness indicates variance. * CSR5 requires 10 – 30× of additional preprocessing, † yaSpMV requires up to 150 000× of additional preprocessing time.

410
In [3], we presented HolaSpMV, an uncompromising approach for SpMV on the GPU. It requires no separate preprocessing or knowledge of the matrix structure and works directly on the standard CSR format. From a global perspective, our approach exhibits a homogeneous behavior reflected in efficient memory access patterns and steady per-thread workload. From a local perspective, it avoids heterogeneous execution paths by adapting its behavior to the work load at hand, it uses an efficient encoding to keep temporary data requirements for on-chip memory low, and leads to divergence-free execution. We evaluated our approach on more than 2500 matrices comparing to vendor provided, and state-of-the-art SpMV implementations. Our approach not only significantly outperforms approaches directly operating on the CSR format (20% average performance increase), but also outperforms approaches that preprocess the matrix even when preprocessing time is discarded as can be seen in Figure 38.7. Additionally, the same strategies lead to significant performance increase when adapted for transpose SpMV.

Our efforts in [1] focused on exploring the potential of hierarchical matrix representations for SpMV for the GPU. Such representations are not new as their potential has been recognized in early work on the finite element method. Nonetheless their lack of popularity can be attributed to the burden of accommodating linear algebra primitives to the hierarchical representation. It has been observed in earlier work on vector processors that the use of hierarchical formats can lead to significant memory savings [2]. This storage cost reduction is a key motivation behind our effort to accommodate a hierarchical format on the GPU, where memory resources are much more limited. Furthermore, a hierarchical format offers several additional advantages: it scales well for large matrices, allows for implementing algorithms in a divide and conquer manner, and permits sharing/duplication of sub-matrices between different matrices and within a single matrix. The evaluation of our implementation of basic linear algebra routines, suggests that our hierarchical format is competitive to highly optimized standard libraries and significantly outperforms them in the case of transpose matrix operations. The results point towards the viability of hierarchical matrix formats on massively parallel devices such as the GPU.

**General sparse matrix-matrix multiplication (SpGEMM)** In the ongoing efforts targeting the vectorization of linear algebra primitives, (SpGEMM) has received considerably less attention than (SpMV). While both are equally important, this disparity can be attributed mainly to the additional formidable challenges raised by SpGEMM.

In [5], we presented AC-SpGEMM, a dynamic approach for addressing SpGEMM on the GPU. Our approach works directly on the standard compressed sparse rows (CSR) data format. In comparison to previous SpGEMM implementations, our approach guarantees a homogeneous, load-balanced access pattern to the first input matrix and improves memory access to the second input matrix. It adaptively re-purposes GPU threads during execution and maximizes the time efficient on-chip scratchpad memory can be used. Adhering to a completely deterministic scheduling pattern guarantees bit-stable results during repetitive execution, a property missing from other approaches. Evaluation on an extensive sparse matrix benchmark, see figure 38.8, suggests our approach being the fastest SpGEMM implementation for highly sparse matrices (80% of the set). When bit-stable results are sought, our approach is the fastest across the entire test set.
Sparse matrix assembly We revisited the problem of sparse matrix assembly in the context of the finite element method and we showed that the assembly of the arising system matrices can be concisely encoded as an augmented SpGEMM [8]. Using the lean unstructured mesh matrix representation [7] We demonstrate how the global graph connectivity of the assembled matrix can be captured through basic linear algebra operations and show how local interactions between nodes/degrees of freedom within an element can be encoded by means of concise representation, action maps. These ideas not only reduce the memory storage requirements but also cut down on the bulk of data that needs to be moved from global storage to the compute units, which is crucial on parallel computing hardware, and in particular on the GPU. The performance gains are shown in figure 38.9. Our approach can be easily incorporated within existing sparse matrix libraries without requiring substantial code modifications. To the best of our knowledge, our approach is the first to perform the whole assembly on the GPU without requiring any additional CPU preprocessing or postprocessing.

Figure 38.9: Assembly of the stiffness matrix for the 4-nodes (top), and 10-nodes (left) tetrahedron, 3DoF per node, for the bracket with hole shown to the left. Relative speedup of all approaches is shown w.r.t Sparse2 timings (in seconds).
Dynamic graph data structures In today's interconnected world, large, ever-changing data structures are in common use and dynamic graphs are favored in many application domains, ranging from communication networks, to social media networks, to intelligence data. As graphics processing units (GPUs) become ever more ubiquitous and comparatively inexpensive, massively parallel compute devices are available to deal with problems posed in such large scale domains. Currently, there are many static graph libraries and algorithms available that take advantage of this massive parallelism. However, most deal only with the static use case, as performance on the GPU is predicated on being able to managed thread divergence, memory locality and optimal work distribution, which becomes increasingly difficult in a dynamic setting, where the memory layout is constantly changing.

In [6] (Best paper award at HPEC), we presented a new, dynamic graph data structure, built to deliver high update rates while keeping a low memory footprint using autonomous memory management directly on the GPU. By transferring the memory management to the GPU, efficient updating of the graph structure and fast initialization times are enabled as no additional memory allocation calls or reallocation procedures are necessary since they are handled directly on the device.

We have further optimized the ideas above in[4] to allow for a fully-dynamic graph data structure on the Graphics Processing Unit (GPU) allowing not only for edge but also vertex updates. Performing the memory management on the GPU allows for fast initialization times and efficient update procedures without additional intervention or reallocation procedures from the host. Our optimized approach performs initialization completely in parallel; up to 300x faster compared to previous work. It achieves up to 200 million edge updates per second for sorted and unsorted update batches; up to 30x faster than state of the art. Furthermore, it can perform more than 300 million adjacency queries and millions of vertex updates per second. On account of efficient memory management techniques like a queuing approach, currently unused memory is reused later on by the framework, permitting the storage of tens of millions of vertices and hundreds of millions of edges in GPU memory. We evaluate algorithmic performance using a PageRank and a Static Triangle Counting (STC) implementation, demonstrating the suitability of the framework even for memory access intensive algorithms.

References


### 38.5 Computational Videography

**Coordinator: Christian Theobalt**

The research group *Graphics, Vision, & Video* ([gvv.mpi-inf.mpg.de](http://gvv.mpi-inf.mpg.de)) headed by Prof. Dr. Christian Theobalt explores systems and algorithms for advanced computational videography, which aim to extract real-world information from and provide realistic editing effects for live videos. The group’s research in this domain has been supported by two ERC grants.

Video editing is a very important application in several industries such as post-processing in movies and media, augmented/virtual reality, camera processing and related applications etc., particularly for videos of human body and faces.

A common requirement in all these cases is simple, ideally automated editing of the video at a high level of photorealism, which involves an accurate understanding of the material, lighting and geometry in the scene. The challenges in this field involve the algorithmic problem of performing photorealistic edits that cannot be distinguished from real video capture and system-level challenges such as performing the editing in real-time or live contexts. Although professional tools already exist for performing such video editing, they are mostly limited by the requirement of a trained technical artist to perform the editing at a slow off-line speed, often even requiring manual touch-up of individual frames on a per-pixel basis. In the reporting period, we made significant contributions in simplifying and automating such tools with minimal-to-no need of human interaction, and at a real-time framerate in live contexts using a single sensor.

Notably, novel advancements were made in photorealistic editing of videos consisting of human full-body and facial performances, enabling for the first time video editing tools in which performance of one actor can be replaced with another, while preserving their pose, expression or lighting, while only replacing the identity. It was shown that using the combination of generative model-based reconstruction with neural networks, models for a desired human body and face identity can be reconstructed which provide explicit control over geometric parameters such as pose and expression, which can be either controlled through human interaction, or driven by another performance.

We also explored the challenging problem of estimating materials and lighting from live videos in uncontrolled environments so as to virtually augment the scenes with photorealistic virtual object insertion or perform re-lighting of the real objects. This is achieved through
decomposing the video into the components of light transport based on parametric models of material and lighting, and driven by translative neural networks or fast GPU based optimization. These advances in videography provide insights into development of model representations for real-world physical parameters that can be used for fast and efficient image or video editing applications, hence enabling for the first time automated tools for live video augmentation in general uncontrolled scenes.

### 38.5.1 Inverse and Video-based Rendering

 Investigators: Abhimitra Meka, Michael Zollhoefer, Avishek Chatterjee, and Christian Theobalt

 Decomposing the intrinsic components of image formation in a video is essential for producing photorealistic, illumination-aware video editing effects. We developed a novel method to process RGB-D streams of combined color video and depth in real-time to decompose it into the reflectance (material) and shading (lighting) layers, while also reconstructing the geometry of the scene, using a fast GPU solver [1]. Estimated reflectance is stored directly into the voxel-geometry, hence providing a fused model of the scene that provides full and accurate appearance information, independent of the incident lighting. Potential inaccuracies in the estimation can be corrected by user-interaction directly in the scene in 3D. This enables photorealistic user-interactive augmented reality applications such as object relighting, recoloring and material-editing, all in real-time, Fig. 38.10.

 Estimating the material reflectance properties of an object surface is a key requirement for photorealistic computer graphics and augmented reality. We developed LIME [2, 3], the first method to estimate the reflectance of a material of a general shape under uncontrolled indoor illumination from a single color image. Our method is based on a set of jointly trained neural networks that decouple the problem of material estimation from geometry inference using a re-rendering pipeline based on intrinsic image decomposition of the image. A novel reconstruction loss resulting from a linear decomposition of light-transport intrinsics enables both high-frequency material and lighting estimation in real-time from a live video. This provides new capabilities in augmented reality, such as rendering virtual objects with real materials under consistent high-frequency real-world lighting.

 Related to these works, we developed a new method to render novel viewpoints of a (static) real world scene in an image-based manner. Traditional image-based rendering...
used hand-crafted multi-view blending schemes to view-dependently texture approximate geometry for view-dependent novel viewpoint rendering. However, this strategy cannot entirely prevent disturbing rendering artifacts, such as ghosting or popping due to blending of sparsely sampled images on an approximate shape. We therefore developed a new CNN-based approach that estimates a coherent diffuse texture on an approximate scene geometry from training images and explicitly decouples and renders view-dependent appearance effects, such as moving specular highlights, in a smooth and highly realistic way [4]. The final renderings of our method are more coherent and realistic than with established image-based rendering methods from the literature when using a comparable number of input images.

Structuring and editing large community media collections according to user-defined criteria is another essential and timely task in computational videography, but also in visual computing in general, where structuring large data sets of graphical assets is of utmost importance. In [5], we propose an interactive approach to arrange visual data along user-defined criteria. We learn low-dimensional continuous criteria via interactive ranking, so that a novice user needs only describe the relative ordering of examples. This is formed as semi-supervised label propagation in which we maximize the information gained from a limited number of examples. Further, we actively suggest data points to the user to rank in a more informative way than existing work. Our efficient approach allows users to interactively organize thousands of data points along 1D and 2D continuous sliders. Experiment with datasets of imagery and geometry demonstrate that our tool is useful for quickly assessing and organizing the content of large databases.

References


38.5.2 Advanced Synthesis and Editing of Human Actor Videos

*Investigators: Hyeongwoo Kim, Lingjie Liu, Florian Bernard, Weipeng Xu, Marc Habermann, and Christian Theobalt*

We develop new approaches that demonstrate an effectiveness of our image-based dynamic scene reconstruction techniques (Sec. 38.6) for enabling novel ways of editing videos of humans and general dynamic scenes with meaningful control parameters. Notably, we
Figure 38.11: Our Deep Video Portraits method [1] enables it, for the first time, to video-realistically edit the full 3D head pose and face expression of a person in a portrait video. It also enables us to transfer the rigid head pose, facial expression and eye motion with a high level of photorealism from a source video to a target video. The approach opens up new applications in photorealistic video editing for visual effects, AR/VR, as well as high quality visual dubbing.

develop new ways to combine traditional model-based capture and rendering with new neural network-based approaches for video synthesis. In this context, synthesizing and editing video portraits, i.e., videos framed to show a person’s head and upper body, is an important problem in computer graphics, with numerous applications in video editing and movie post-production, to name a few. We developed Deep Video Portraits [1], the first approach that enables photo-realistic re-animation of the entire posture and face expression of a human in a portrait videos using only an input color video, Fig. 38.11. The core of the approach is a generative adversarial neural network with a space-time architecture that converts computer graphics renderings of a (reconstructed) face model into photo-realistic and temporally-coherent video. This opens up a new level of capabilities in many applications such as face reenactment, visual dubbing of foreign language films, interactive face editing, movie post-production and video conferencing. In contrast to previous face editing and reenactment approaches we can, at interactive frame rates, edit the full face expression, including – for the first time – the full 3D head pose of an actor in a video. The algorithm then automatically synthesizes both plausible and highly realistic videos of foreground and (static) scene background.

We also demonstrated that video-based face capture technology can be used for video-based rendering and editing challenges that are specific to virtual reality telepresence. In our FaceVR approach we demonstrated the gaze-aware reenactment of stereo video content for VR telepresence applications [5, 6]. With this approach it is possible to computationally remove the VR headset from stereo face video of a person, while reproducing the full face expression and eye gaze “behind” the display. The core of the algorithm is an image-based face reconstruction and stereo video reenactment approach that uses a combination of external and in-display cameras to create the modified content. In this way, believable VR-based telepresence applications at a higher level of quality and immersion are feasible.

In the work [2] a novel learning-based approach for the animation and reenactment of human actor videos is presented, see Fig. 38.12. In contrast to conventional human character rendering, we do not require the availability of a production-quality photo-realistic 3D
Figure 38.12: Neural animation of human characters [2]: The top row shows frames from which a 3D motion is extracted. The bottom row shows the corresponding target person imagery synthesized by our network which shows the target person reenacting the source motion.

model of the human, but instead synthesize novel videorealistic animations using a human actor video in conjunction with a (medium-quality) controllable 3D template model of the person. The approach uses a neural network, specifically a new space-time conditional GAN, to synthesize plausible target video given the body template in a pose controlled by semantically meaningful parameters. The motion of this template is captured in the training data using one of our state-of-the-art human performance capture methods. Our approach significantly reduces production cost for video-realistic human body animation compared to conventional rendering approaches using production-quality 3D models, and can also be used to realistically edit existing videos.

The development of new efficient methods for high-quality editing and synthesis of videos also raises concerns that such technologies could be used to modify videos with malicious intent. Importantly, the same basic understanding of the performance and working range of synthesis methods is key to also develop approaches that can potentially detect or prevent such unwanted modifications with high reliability. We also make this an important part of our research agenda. In collaboration with the group of Bernt Schiele we also showed how our techniques that combine monocular face capture and video-realistic rendering can be used to effectively protect the privacy of people in community image and video databases [3, 4]. Today, many image sand videos are uploaded to community platforms, often times without the knowing of people that may be visible in the background of such scenes. However, automated face recognition systems could be used at scale to identify people in such images. By using one of neural face rendering approaches we were able to obfuscate the (face) identity of people in general real world imagery, such that: (1) the final image still looks like a plausible person image without other types of explicit obfuscation mechanisms (such as blurring or blacking out the face) that would render the image unpleasant to look at; and (2) such that the altered identity is hard to recognize by either a human or a machine-based identification algorithm.
References


38.6 Reconstructing the Static and Dynamic Real World

Coordinator: Christian Theobalt

The research group *Graphics, Vision, & Video (gvv.mpi-inf.mpg.de)* headed by Prof. Dr. Christian Theobalt works on methods advancing the state of the art in marker-less reconstruction of static and dynamic scenes. The group’s research in this domain has also been supported by two ERC grants, an ERC starting grant and an ERC consolidator grant. In static scene reconstruction, the ultimate goal is to develop methods that capture high-quality models of shape, appearance and lighting of large scale real world scenes, ideally with only a single lightweight color or depth camera. In 4D capture, the goal is to reconstruct highly detailed 4D (3D+time) models of geometry, motion, appearance and illumination of real world scenes. Of particular interest is the reconstruction of humans. But the reconstruction of all deformable shapes outdoors, in the real world, at high quality, ideally with only a handful or even just one camera, is the long term goal.

Newly emerging application domains have dramatically increased the need for high-quality marker-less 3D and 4D scene reconstruction methods in the recent past. For example, they do not only simplify but are often an essential precondition of content capture for visual effects, or 3D and free-viewpoint TV. Nowadays, their relevance goes far beyond these applications. High fidelity reconstruction is essential for believable virtual and augmented realities, as well as telepresence systems. It is also an important technology supporting the learning of believable models of humans from real world observations in order to create highly realistic virtual actors, avatars and agents. It is also an essential precondition for advanced real world perception abilities needed by future intelligent computing and autonomous systems that need to reconstruct the human world to understand it and to safely interact in it.
Despite great progress recent years, many problems in high quality 3D and 4D capture remain unsolved. In the reporting period, we advanced the state-of-the-art of both static and dynamic scene reconstruction methods along several important dimensions: the generality of scenes that can be handled, the accuracy and quality of the reconstruction, the efficiency and robustness of reconstruction, and the simplicity of sensors needed for capturing. Notably, we researched entirely new ways of fusing and deeply integrating model-based and deep learning-based scene reconstruction. This paves the way for a new generation of future real world reconstruction, perception and understanding methods that can learn and continuously refine their internal algorithmic concepts and model representations on even sparsely labeled or unlabeled real world examples. We also maintain one of the largest repositories of reference data sets for different sub-tasks in general static and dynamic scene reconstruction, as well as marker-less motion and performance capture, GVVPerfCapEva (see Sect. 38.10.2).

38.6.1 High-quality Static Scene Reconstruction with Lightweight Sensors

Investigators: Michael Zollhoefer, Weipeng Xu, Florian Bernard, and Christian Theobalt

Large-scale high-quality reconstruction of shape and appearance of static scenes with a moving camera, such as an RGB-D camera [5], is an important enabling technology in mixed and augmented reality reality, in robotics, but also in architecture, digital heritage, or real world navigation, to name a few. We developed BundleFusion, the first approach for globally-consistent online reconstruction of large scale scenes with an RGB-D camera [1]. It leads to reconstruction results that are even on par with high-quality offline approaches. Global consistency in the captured geometry is achieved by real-time bundle adjustment on truncated signed distance field that is dynamically updated and effectively stored in space and time.

Depth-camera based scanning is further challenged by the limited resolution and stark noise of contemporary camera hardware. Therefore, very thin structures and features are very difficult to reconstruct with hand-held scanning and merging approaches like the aforementioned one. We therefore developed CurveFusion [3], a new approach that uses curve features as fusion primitive, Fig. 38.13. It is one of the first methods to scan very
thin and tubular structures, in addition to the larger scale scene content, with a hand-held commodity depth camera.

Scanning results are also impaired by sensor noise. In [4], a new algorithm for point cloud denoising is presented, where it is assumed that a given point cloud comprises (noisy) points that were sampled from an underlying surface that is to be denoised. The point cloud denoising problem is phrased in terms of a dictionary learning framework that operates on local patches that are extracted from the point cloud. Experimentally, it was demonstrated that the method outperforms existing denoising approaches in various noise scenarios.

While the previous approaches aim at surface capture without a semantic decomposition into individual objects and entities. We also proposed a novel convolutional neural network approach to jointly perform 3D object retrieval and pose estimation from monocular images (see Fig. 38.14) [2]. In order to alleviate both pose bias for 3D shape retrieval and categorical bias for pose estimation, we learn an embedding space from 3D data that is explicitly disentangled into a shape vector and a pose vector. Our approach outperforms the previous state-of-the-art methods on both 3D object retrieval and pose estimation tasks.

References

38.6.2 Deformable Surface and Performance Capture in Controlled and Uncontrolled Scenes

Investigators: Marc Habermann, Edgar Tretschk, Vladislav Golyanik, Weipeng Xu, Michael Zollhoefer, Nadia Robertini, Florian Bernard, Avishek Chatterjee, and Christian Theobalt

In the past, we made pioneering contributions to dense 4D scene capture with cameras, notable marker-less performance capture. In the reporting period, we greatly advanced the capabilities of marker-less human performance capture methods, as well as methods capturing general deformable objects with cameras. In particular, the reconstruction accuracy and performance were drastically increased, and robustness to general scene conditions was greatly enhanced. Further on, new approaches empowered by new ways of deeply integrating template-based and machine learning-based reconstruction have enabled some of the first methods for performance and deformable shape capture from a single color camera.

For example, we developed a new efficient method for template-based non-rigid surface tracking from monocular RGB videos [3]. It uses an analysis-by-synthesis approach which involves a novel texture term specifically tailored towards tracking objects with uniform texture but fine-scale structure, such as the regular micro-structural patterns of fabric. Our texture term exploits the orientation information in the micro-structures of the objects, e.g., the yarn patterns of fabrics. This enables for the first time accurate tracking of uniformly colored materials that have these high frequency micro-structures, for which traditional photometric terms are usually less effective.

In the area of human performance capture, we developed new tools to register and align large corpora of laser scans of humans and to build a simple and effective model, the MPI Human Shape Model, that combines skinning-based pose deformations with a PCA space of human shape variations [5]. Alignment tools, aligned scans and the learned model developed in this collaboration with the group of Bernt Schiele are made publicly available.

We further proposed, in collaboration with the team of Gerard Pons-Moll, the first method for reconstructing accurate static 3D body models and surface textures of people in general clothing from a single, monocular video in which a person is rotating on spot [2]. Our main contribution is a method to non-rigidly deform the silhouette cones corresponding to the dynamic human silhouettes, resulting in a visual hull in a common, articulation-normalized reference frame that enables surface reconstruction. Requiring only a smartphone or webcam, our method enables everyone to create their own fully animatable digital double, e.g., for social VR applications or virtual try-on for online fashion shopping. A further improvement texture stitching strategy enabled model reconstruction with even higher texture quality [1].

In addition, we developed entirely new template-based methods to capture the performance of humans in motion from a single color camera. Our MonoPerfcap algorithm is the first marker-less approach for temporally coherent 3D performance capture of a human with general clothing from monocular RGB video [7]. We tackle this challenging problem by using a novel approach that employs sparse 2D and 3D human pose detections from a convolutional neural network using a batch-based pose estimation strategy and then refining the surface geometry based on fully automatically extracted silhouettes to enable medium-scale non-rigid alignment. We demonstrate state-of-the-art performance capture results that enable exciting applications such as video editing and free viewpoint video, previously infeasible from monocular video.
Figure 38.15: We developed LiveCap, the first real-time human performance capture approach that reconstructs dense, space-time coherent deforming geometry of entire humans in general clothing from just a single RGB video. [4]

While MonoPerfcap is an off-line method, LiveCap [4] is the first template-based real-time human performance capture approach that reconstructs dense, space-time coherent deforming geometry of entire humans in general clothing from just a single RGB video, Fig. 38.15. Our efficient skeleton pose optimizer fits the skinned template to regressed 2D and 3D skeleton joint positions, to sparse detected facial landmarks, and to the foreground silhouette. Our novel monocular real-time capable algorithm for non-rigid analysis-by-synthesis tracking reconstructs 3D deformations of even loose apparel. The new method is orders of magnitude faster than the state of the art and the first to demonstrate dense capture of motion and non-rigid surface deformation at real-time frame rates. It achieves an accuracy on par with with off-line performance capture techniques.

We also further advanced the accuracy and performance of performance capture from multi-view video. A new analysis-by-synthesis formulation using a dense set of Gaussian functions on the template surface, as well as a new implicit alignment formulation between model and images enables accurate dense capturing of deformable human shapes in more general outdoor scenes outside of controlled studios [6].

References

38.6.3 Marker-less 3D Human Pose Estimation and Skeletal Motion Capture in General Scenes and with Lightweight Sensors


In the reporting period, we developed new methods for marker-less 3D skeletal motion capture, without surface reconstruction, from video that improve over previous work in several ways. Marker-less motion capture is an important enabling technology in many application areas, ranging from computer animation and special effects, to medicine and biophysics, to human-computer interaction, and to immersive virtual and augmented reality. We improved multi-view motion capture to be more robust in scenes with difficult lighting conditions. We also developed the first approaches for real-time full 3D human motion capture with a single color camera, as well as even a body-worn fisheye camera. We further developed new concepts to capture 3D human motion of multiple people under strong occlusions by objects and other individuals in the scene.

To improve multi-view motion capture, we developed, in collaboration with the team of Bernt Schiele, a new approach that combines generative skeletal motion estimation based on an implicit Gaussian image and scene representation with CNN-based 2D joint detections [1]. It enables marker-less motion capture a state-of-the-art accuracy even in outdoor scenes with multiple people and notable (self-occlusions), also with a low number of cameras.

In [7] the problem of multi-view marker-less motion capture under changing and non-uniform real world lighting conditions is addressed. In order to account for changing lighting conditions an intermediate image representation that is invariant to the scene lighting is used. Model-based pose fitting and CNN-based pose detection are synergistically combined with a lighting robust segmentation and material classification approach. We experimentally show that method is able to handle difficult real world scene and lighting conditions better than related multi-view approaches.

We also pushed the envelope of marker-less motion capture by developing new methods working with just monocular color video as input.

We developed VNect [6], the first method for accurate and real time 3D skeletal motion capture from monocular RGB video. It combines a learning based approach with generative fitting of a kinematic skeleton model. The system uses a novel fully-convolutional 3D pose estimation approach which strongly couples 3D pose inference of each body joint to its image evidence. A kinematic skeleton fitting term further reconciles the 2D and 2D pose predictions, and provides temporally smooth joint angle estimates, which can be readily employed for controlling virtual characters as shown in Figure 38.16. We improved VNect to make it faster and more accurate [5], allowing it to run off a laptop.

We also proposed the first real-time approach for the egocentric estimation of 3D human body pose in a wide range of unconstrained everyday activities (see Fig. 38.17) [8, 9]. To this end, we built a novel lightweight setup that converts a standard baseball cap to a mobile motion capture device featuring a single fish-eye camera facing downwards. The
Figure 38.16: VNect [6] allows temporally smooth real-time 3D human body pose estimation from a single RGB camera, enabling various applications for entertainment, sports analysis, and health.

Figure 38.17: Our real-time egocentric 3D full body pose estimation algorithm uses a single monocular cap-mounted fisheye camera that is attached to a standard baseball cap[8, 9]. The setup is lightweight and enables 3D pose estimation in everyday situations, without being constrained to a constrained recording volume.
Figure 38.18: Our new multi-person 3D pose estimation approach [4] reconstructs the 3D pose of multiple humans in a scene from monocular images, even under notable occlusions. Also shown is the compositing process to generate multi-person MuCo-3DHP scenes from MPI-INF-3DHP for training our approach, as well as frames from our multi-person 3D pose benchmark MuPoTS-3D.

core of the method is a CNN-based hierarchical real-time monocular 3D pose estimation method that takes the unique properties of the egocentric viewpoint into account, such as strong distortions and difficult self-occlusions. The method enables motion capture outside a fixed recording volume, and opens up new applications in immersive virtual and augmented reality.

The publicly available training data for single person 3D body pose estimation from RGB input are starkly limited in appearance diversity, which limits the generalizability of the trained models on general scenes. Our work [2] addresses this issue on two fronts. We use a multi-view marker-less motion capture system to capture multiple subjects in front of a green screen, wearing street clothing. Further, each subject is also captured with uniformly colored clothing, which allows an increased scope of appearance augmentation. This data set, which we call MPI-INF-3DHP, allows us to make our trained models invariant to clothing appearance and scene setting. We also establish a new benchmark to test the generalization of the trained models to unseen scene settings. Further, we leverage in-the-wild 2D pose data sets, where annotations are easier to obtain at scale, to learn more robust features, allowing us to beat the state of the art on standard benchmarks at the time by over 25%.

We further build upon our work on single person 3D body pose estimation from monocular color images in un-occluded in-the-wild settings [2, 6] to enable true in-the-wild body pose estimation even under a substantial degree of occlusion. The learning based approach uses a novel occlusion robust pose formulation during inference with a CNN [4]. The ability to handle occlusions enables multi-person 3D body pose estimation in general scenes where inter-personal and person-object occlusions are common. To train our system, we create plausible multi-person composites at scale from our MPI-INF-3DHP [2] single person dataset. Further, we propose a new multi-person monocular RGB 3D pose estimation benchmark called MuPoTS-3D with real sequences, and ground truth 3D pose acquired using a marker-less motion capture system from the Captury (a spin-off of the GVV group). See Fig. 38.18 for examples from the proposed datasets, and results of our method. A live system [3] based on this work allows real-time multi-person pose estimation and interactive character control.
References


38.6.4 High-quality Lightweight Facial Performance Capture

Investigators: Aysh Tewari, Hyeongwoo Kim, Pablo Garrido, Mohamed Elgharib, Michael Zollhoefer, Florian Bernard, Weipeng Xu, and Christian Theobalt

Reconstruction of high-quality 3D models of moving human faces is essential for the creation of digital human actors in movies and games, for believable avatars in virtual/augmented reality and telepresence, as well as for advanced video editing. It is also an important component in perception algorithms that interpret human action and behavior, as they are used in man machine interaction, attention detection or biometry.

In the past, we contributed novel algorithms enabling the reconstruction of high quality dynamic shape, appearance and lighting models of human faces from sparse input, e.g., [8]. In addition, the computer graphics and vision communities in general have dedicated long...
standing efforts in building computerized tools for reconstructing, tracking, and analyzing human faces based on visual input. We presented a state-of-the-art report [10] which summarized recent trends in monocular facial performance capture and discussed its applications, which range from performance-based animation to real-time facial reenactment. In the reporting period we greatly improved over the state of the art of face performance capture with sparse, notably monocular camera input, in several ways. Notably, we presented new algorithms combining model-based and deep learning-based face reconstruction by means of new end-to-end trainable architectures that achieve state-of-the-art reconstruction accuracy from only a single color video at unprecedented runtime performance. We also show modern 3D face performance capture algorithms pave the way for new applications in immersive VR and telepresence.

We proposed MoFA [5], the first unsupervised approach for monocular reconstruction of the identity geometry, albedo texture, face expression and scene illumination from a single color image. This was made possible by a new type of algorithm, a model-based face autoencoder which tightly integrates in a new way a 3D parametric face model, differentiable image formation and layers of a convolutional neural network. The algorithm can be trained on unlabeled community image data and achieves state-of-the-art reconstruction accuracy at an unprecedented 4 ms on a modern PC. We extended [5] in [3], where we additionally demonstrated faster training of our networks, and moreover, we proposed and evaluated analysis-by-synthesis and shape-from-shading refinement approaches to achieve a high-fidelity reconstruction.

In [4] we profoundly extended the concepts and architecture introduce in MoFa. We proposed the first approach that jointly learns 1) a regressor for face shape, expression, reflectance and illumination on the basis of 2) a concurrently learned parametric face model, directly from community images (see Fig. 38.19) that only need very sparse landmark annotations. The multi-level face model includes a fixed base model and a learned corrective model. This paves a new way to go beyond earlier parametric models that were learned from hard-to-capture corpora of 3D laser scans. The approach further improves reconstruction accuracy and yields similar extremely fast runtime performance.

In follow-up work, we improved monocular reconstruction quality as well as the learned parametric face model even further. We proposed FML: Face Model Learning [2], which demonstrated multi-frame video-based self-supervised training of a deep network that 1) learns a face identity model both in shape and appearance while 2) jointly learning to
FaceVR [6, 7] uses image-based face performance capture and external and internal cameras of a head-mounted VR display to virtually remove the headset from the face of a user in a VR stereo teleconference.

reconstruct 3D faces. Our approach was the first face reconstruction method which did not rely on an existing 3D parametric face model. It directly learned such a face model from videos, and yields even further enhanced reconstruction quality and generalizability at extremely fast runtime performance.

A different approach to monocular reconstruction of 3D faces is presented in our Inverse-FaceNet algorithm [1]. It uses a deep convolutional inverse rendering framework that jointly estimates facial pose, shape, expression, reflectance and illumination from a single input image on the basis of a parametric face model and a spherical harmonics lighting basis. The algorithm employs a new self-supervised boosting scheme to close the domain gap to real images if only synthetic training images with ground truth annotation are available.

We also showed how latest face performance capture techniques enable new applications In FaceVR, we show the gaze-aware reenactment of stereo video content for VR telepresence applications [6, 7]. With this approach it becomes possible to computationally remove the VR headset from stereo face video of a person, while reproducing the full face expression and eye gaze “behind” the display, Fig. 38.20. In this way, entirely new believable VR-based telepresence applications are feasible.

In HeadOn [9], we showed how the combination of monocular face performance capture from a depth camera, 3D body capture from an RGB-D camera, and a new dynamic texture synthesis algorithm on only approximate 3D geometry can be used to control personalized avatar videos of humans, e.g., for the purpose of believable teleconferencing in VR.

References

Figure 38.21: We presented the first approach [2, 3] to capture the full 3D hand motion in real-time from a monocular RGB video. It can be applied to a wide range of data and scene conditions, including legacy video.


### 38.6.5 Marker-Less Motion Capture of Hands in Interaction

**Investigators: Franziska Mueller, Jiayi Wang, Florian Bernard, Dushyant Mehta, and Christian Theobalt**

Many timely interactive application settings of high practical relevance, such as gesture-based computer input, human-robot interaction, or interactions in augmented reality, require high quality capture of the 3D motion of the human hand, ideally with a low number of or a single camera, and without using complex instrumentation of the hand. Marker-less hand motion capture is a highly challenging problem. The human hand is tremendously dexterous and is capable of complex motions and interactions with objects. Also, hands can perform very fast motions with many occlusions, the hand has uniform appearance, and real world scenes in which humans interact exhibit substantial scene clutter. As a consequence of all
these factors, marker-less hand motion capture in real world scenes, in particular in real-time and with a single (color or depth) camera is a highly difficult problem with many unresolved challenges.

In the reporting period, we have developed state-of-the-art techniques to track hands, in real-time without markers or gloves, and using only a single RGB-D or color camera. The deep methodical integration of model-based and learning-based reconstruction also enable some of the first methods that handle non-trivial hand-object interactions in cluttered scenes.

Most previous hand tracking research focused on an isolated hand gesturing in mid-air, a scenario decoupled from many real world interactions. We therefore advanced the state of the art with our real-time method for hand tracking under occlusion from an egocentric RGB-D sensor [4]. To gather plausible example interactions with ground-truth annotations, we employed a mixed reality capture setup where users can interact with virtual objects.

We also developed the first method for reconstructing the full 3D hand pose in real time from a monocular RGB video [2, 3]. The method combines CNN-based pose regression with generative model-fitting for full 3D hand capture. The neural network component of our approach was trained using our enhanced GANerated data. To this end, we used a generative adversarial network to close the domain gap between synthetic and real hand images, without the need for paired examples. Hence, our method works accurately in various scenarios as shown in Fig. 38.21.

Many methods for depth-based hand pose estimation assume a segmented or cropped input image that only shows the hand segmented from the surrounding scene. For most real applications, this assumption breaks since humans use their hands frequently to interact with their environment. Developing a robust hand segmentation approach is therefore a challenging and highly relevant problem in itself. Thus, we introduced HandSeg, an automatically labeled dataset for hand segmentation from depth images [1], as well as a set of state-of-the-art learning-based hand segmentation methods trained on this data corpus.

In our research, we also investigate new ways to use advanced monocular hand tracking for human-computer interaction. With FingerInput [5], we proposed a consolidated design space for thumb-to-finger microgestures. Using a single body-mounted depth camera, our real-time system demonstrated accurate detection of continuous finger gestures that surpassed the expressiveness and subtlety of previously presented gesture-based interaction schemes.

We also developed WatchSense [6], a new wearable system for gesture-based input sensing from a single depth camera attached to the arm. Our method employs a random decision forest to reliably detect hand and finger gestures performed on and above the back of the hand. It thereby turns the back of the hand into a finger-based multi-touch surface. This significantly extends the interaction space for smart devices like watches and phones (see Fig. 38.22).

References

Figure 38.22: WatchSense [6] enables on- and above-skin multi-finger and multi-touch interactions on the user’s back of the hand, with a single wearable depth sensor.


38.6.6 Foundational Algorithms and Tools for Real World Reconstruction

Investigators: Florian Bernard, Dushyant Mehta, Michael Zollhoefer, and Christian Theobalt

The GVV group also investigates foundational algorithmic questions that are of cross-cutting relevance to key aspects of both expert designed and learned images-based reconstruction methods in particular, and visual computing or machine learning algorithms in general.

Designing neural networks that have a compact memory-footprint, have fast inference times and maintain high accuracy is key for many practical reconstruction applications. While evaluating convolutional neural network (CNN) acceleration approaches, we found that CNNs trained with adaptive stochastic gradient descent (SGD) approaches surprisingly exhibit implicit filter level sparsity, meaning higher compactness at stable accuracy, which is at par with explicit, often heuristically designed, filter level sparsification (i.e. network compression) approaches. We conducted an extensive investigation [8] of the conditions under which the sparsity manifests, and how various seemingly unrelated hyperparameters interact with the sparsity. We find that in addition to regularization, feature selectivity is also tied to feature sparsification, and thus we establish a previously unknown link between
Figure 38.23: A novel multi-matching method that is both \textit{scalable} and able to account for \textit{geometric consistency} is proposed [3].

between the two. Our findings not only allow filter level pruning for CNN acceleration without any modification to the traditional deep learning pipeline, but also provide novel insights into the practical performance gap between vanilla SGD and adaptive flavors. We believe our findings will inform future theoretical and practical investigations in the domain.

Further, the matching of multiple objects (e.g. shapes or images) is a fundamental problem in vision and graphics (Fig. 38.23). The main limitations of existing multi-matching methods are that they either ignore geometric consistency and thus have limited robustness, or they are restricted to small-scale problems due to their (relatively) high computational cost. We address these shortcomings by introducing a Higher-order Projected Power Iteration method, which is (i) efficient and scales to tens of thousands of points, (ii) straightforward to implement, (iii) able to incorporate geometric consistency, and (iv) guarantees cycle-consistent multi-matchings [3].

In [1] we study convex relaxations of quadratic optimization problems over permutation matrices which are at the core of many matching and reconstruction problems. While existing semidefinite programming approaches can achieve remarkably tight relaxations, they have the strong disadvantage that they lift the original $n \times n$-dimensional variable to an $n^2 \times n^2$-dimensional variable, which limits their practical applicability. In contrast, we present a lifting-free convex relaxation that is provably at least as tight as existing (lifting-free) convex relaxations.

In [2] we propose a method for the synchronization of partial multi-matchings based on non-negative matrix factorisation. With that, given a set of partial pairwise multi-matchings, our method is able to establish cycle consistency. Moreover, we also developed distributed approaches for synchronization, in particular for the case of Euclidean transformations [9], and for synchronization on the unit sphere [10].

Another common observation is that many inverse problems in image-based static and dynamic scene reconstruction require the efficient solution of high-dimensional non-convex optimization problems. The GVV group has contributed various custom GPU solvers to optimize such task-specific energy functions, e.g. [4, 7].

Programming such problem-specific solvers on the GPU has been a tedious and time-consuming task in the end. To make such real-time optimization techniques more easily accessible to users and to conveniently adapt optimization methods to new energy functions and problem settings, we have developed a domain specific optimization language (Opt) [6] that allows for fast prototyping of a wide range of challenging computer graphics and computer vision problems on grid and graph structured domains, see Fig. 38.24. Based
on a high level description of the objective function, \textit{Opt} fully automatically generates application-specific high performance GPU optimization code.

Figure 38.24: Our domain specific optimization language (\textit{Opt}) [5] allows for fast prototyping of solvers for optimization problems encountered in a wide range of challenging computer graphics and computer vision problems.

References

38.7 Realistic and Real-time Rendering

Coordinator: Gurprit Singh

Synthesizing realistic images in time bound manner demands effort on different frontiers from simulating distribution effects, faithful representation of digital content, fabricating realistic appearance to faster computation of high-dimensional light transport integrals. Important related problems are novel-view synthesis [2] and stylization [1]. Depth-of-field (DoF) or motion blur (MB) are quite commonly used distribution effects for realistic image synthesis. Reproducing these distribution effects in synthesized imagery is a typical and well-understood part of most photo-realistic rendering systems. Splatting the point-spread function (PSF) of every pixel is general and provides high quality distribution effects, but requires prohibitive compute time. In 38.7.1, we present our work on accelerating splatting of PSFs to obtain cinematic quality distribution effects based on a sparse representation of Laplacian of the PSFs. We also suggest a novel efficient framework to produce complex distribution effects by exploiting coherency among RGB-D images rendered via pin-hole cameras.

Developing a resolution-independent representation of digital content is quintessential. Vector graphics precede raster graphics as a representation of digital content, yet, remain relevant today, since a resolution-independent representation allows artifact-free display on everything from a tiny smartwatch to a huge wall-size display. Unfortunately, efficient rendering of vector graphics at high resolutions still forms a challenging task for computer graphics. In 38.7.2, we present a new curved primitive and derive a parallel, hierarchical rasterization approach to render these primitives in an efficient way on current GPU hardware. We further look into synthesizing appearance from given per-pixel attributes using a convolutional neural network.

With the advent of powerful compute units (CPUs, GPUs), real-time renderers are paving their way towards full global illumination pipelines (e.g., in game industry). However, much is left to understand regarding sample correlations that directly affects the quality of the images (with aliasing or noise) irrespective of the underlying application. In 38.7.3, we present recent developments on Monte Carlo (MC) based samples techniques that are critical in approximating high dimensional light integrals representing radiant light energy. These works establish sound theoretical formulations that can represent error in closed-form and propose the first deep learning framework that can synthesize samples without any special hand-crafting or mathematical intricacies.

References

38.7.1 Real-time Distribution Effects

Investigators: Thomas Leimkühler and Tobias Ritschel

Distribution effects like depth-of-field and motion blur are an important factor to cinematic quality in synthetic images. Unfortunately, computing these effects can take long to compute. A common method in the real-time/interactive realm is to first render plain pinhole images and, in a second step, apply post-processing operations to create the desired effects. Drawing point spread functions (e.g., circles to create simple defocus blur) for every pixel is a general algorithm for achieving high-quality distribution effects. While this usually is an order of magnitude faster than classical MC methods, it is still not suitable for applications requiring interactive feedback.

Our work [2] accelerates splatting of point spread functions in two steps: The first step is a pre-process. Here, we determine and generate the required point spread functions, while taking aliasing problems into account. Then we find sparse representations of the Laplacian of the point spread functions, which we call spreadlets. This is achieved using a custom multi-step optimization procedure. The second step happens at runtime. Here, the spreadlets are efficiently drawn to the Laplacian of an image. Integrating this image produces the final result. Due to the sparsity of our representation we achieve an additional order of magnitude in execution speed, while scaling faithfully to strong motion and large out-of-focus areas.

Our approach is general enough to be applicable to simple defocus and motion, physical lens models, light scattering in participating media, as well as arbitrarily stylized effects. Furthermore, the algorithm can both synthesize distribution effects from pinhole images as well as reconstruct from stochastic images.

Observing that many visual effects (depth-of-field, motion blur, soft shadows, spectral effects) and several sampling modalities (time, stereo or light fields) can be expressed as a sum of many pinhole camera images, we suggest a novel efficient image synthesis framework that exploits coherency among those images [1]. We introduce the notion of distribution flow that represents the 2D image deformation in response to changes in the high-dimensional time-,
lens-, area light-, spectral-, etc. coordinates. Our approach plans the optimal traversal of the distribution space of all required pinhole images, such that starting from one representative root image, which is incrementally changed (warped) in a minimal fashion, pixels move at most by one pixel, if at all. The incremental warping allows extremely simple warping code, typically requiring half a millisecond on an Nvidia GeForce GTX 980Ti GPU per pinhole image.

References


38.7.2 Vector Graphics, Appearance Synthesis and Simulations

*Investigators: Oliver Nalbach, Jozef Hladky, Tobias Ritschel, and Markus Steinberger*

Efficient rendering of vector graphics at high resolutions still forms a challenging task for computer graphics, as there is no parallel approach that comes close to the elegance and efficiency of triangle rasterization. Inspired by polygon rasterization, we introduce CPatch, a curved primitive that can be used to construct arbitrary vector graphics [1]. A CPatch is a generalization of a 2D polygon: Any number of curves up to a cubic degree bound a primitive. Each curve divides the space into a positive and a negative half-space. The union of all positive half-spaces defines the inside of a CPatch, similar to the use of edge equations in polygon rasterization. We show that a CPatch can be rasterized efficiently in a hierarchical manner on the GPU, locally discarding irrelevant portions of the curves. Our rasterizer is fast and scalable, works on all patches in parallel, naturally supports all kinds of color spaces, blending, super-sampling and does not require any approximations. Additionally, we show how vector graphics input can efficiently be converted to a CPatch representation, solving challenges like patch self-intersections and false inside-outside classification. Results indicate that our approach is faster than the state-of-the-art, more flexible and could potentially be implemented in hardware.

In computer vision, convolutional neural networks (CNNs) achieve unprecedented performance for inverse problems where RGB pixel appearance is mapped to attributes such as positions, normals or reflectance. In computer graphics, screen space shading has boosted the quality of real-time rendering, converting the same kind of attributes of a virtual scene back to appearance, enabling effects like ambient occlusion, indirect light, scattering and many more. We propose to consider the diagonal problem [4]: synthesizing appearance from given per-pixel attributes using a CNN. The resulting Deep Shading renders screen space effects at competitive quality and speed while not being programmed by human experts but learned from example images.

On the frontiers of physics simulations in computer graphics, we propose a system [3, 2] for modeling fire dynamics using stochastic models of flickering and buoyant diffusion. We
present extensive behavioral control system where flame spread and motion is computed using differential equations that take account of wind fields, buoyancy, diffusion and velocity of the burning surface.

References


38.7.3 Analysis and Synthesis of Sample Correlations

Investigators: Thomas Leimkühler, Karol Myszkowski, Tobias Ritschel, and Gurprit Singh

Error in MC rendering is analyzed using different spatial and Fourier tools [4, 3]. To simplify the error analyses, the closed-form MC error formulations obtained so far assumes stationarity (shift-invariance), an infinite sampling domain and no importance sampling (IS). Unfortunately, none of these simplifying assumptions are true in practical rendering scenarios, making the theory mispredict error and convergence rates in practice. We propose a generalized closed-form variance formulation that properly predicts error without any assumption and works for IS in conjunction with correlated samples [5]. Our analysis reveals that the finite sampling domain, stationarity and IS all interplay to introduce or inhibit $C_0$ discontinuities in the integrand, affecting error and convergence rate. We also show that multiple importance sampling (MIS) is not able to overcome these detrimental effects as it inherits the convergence properties of the worst of the constituent strategies. With these interactions in mind, we show that (M)IS convergence improves by 1) removing certain discontinuities (on the integration domain boundary thanks to a mirroring strategy or within the domain by integrand filtering) or 2) by pushing them to the boundary of the integration domain.

These theoretical analyses propose certain design principles as an end product to shape the samples’ expected power spectra for improved error convergence rate and low variance. These include low discrepancy, blue noise spectra with anisotropy and progressivity. However, designing sampling patterns with such properties can require substantial hand-crafting effort, both in coding, mathematical derivation and compute time. Tackling this issue, we suggest another level of abstraction: a toolkit to optimize over all sampling methods to find the one producing user-prescribed properties such as a spectrum or a point correlation function [2]. A user simply implements a loss and the sampling method is found automatically—without
coding or mathematical derivation—by making use of unstructured non-linear filters in conjunction with the back-propagation abilities of modern deep learning frameworks. In an orthogonal work [1], we leverage spectral graph-theoretic tools (Laplacian-Beltrami operator) to develop a framework for detecting, quantifying and visualizing changes between two snapshots of a dynamic network.

References


38.8 Perception and Advanced Displays

Coordinator: Karol Myszkowski

Modeling the dynamics of perceptual mechanisms in the human visual system (HVS) offers a great promise in designing novel and advanced display systems which can attain much higher visual quality and viewing comfort that is not usually possible with conventional displays. The main reason is the limited capability of existing display technologies in reproducing all visual cues that are present in natural viewing conditions when an observer sees the real world. In order to provide a fully immersive viewing experience, it is desirable to provide all perceptual cues for the HVS such as binocular disparity and accommodative depth cues [1]. To this end, there is a huge amount of ongoing research effort for addressing perceptual challenges in new display technologies such as stereoscopic and auto-multiscopic displays. In this research direction, recently emerging light-field representation of visual data is the most promising candidate as the future display format. However, for a smooth transition from existing technologies to highly capable multiscopic and light-field displays, a connection between the current and upcoming imaging systems is required. In Sect. 38.8.1, we present our work on the conversion of existing monoscopic videos to more advanced stereoscopic and multi-view visuals as a part of ongoing efforts for providing a smooth transition between these technologies. In Sect. 38.8.2, we introduce our work focused primarily on appearance editing of light-field data and our advanced display prototypes. As standard image editing methods are not fully compatible with advanced light-field format, we propose a new approach for this task based on intrinsic decomposition of light-fields. In addition, we address conventional
displays’ lack of perceptual support for accommodative depth cues by introducing our award-winning advanced Membrane AR display, which is a very promising candidate addressing this shortcoming of existing near-eye AR display designs. In Sect. 38.8.3, we investigate novel quality measurement methods which are tailored towards evaluating images and light-fields, which is very critical for benchmarking acquisition and processing quality of both existing and future methods and displays in this field. Finally, in Sect. 38.8.4, we address well-known problems and limitations encountered in recent efforts to extend gaze-contingent rendering paradigm to the advanced stereo and HMD display systems using eye tracking.

References

38.8.1 Mono2Stereo and Stereo2Multiview Image Conversion

Investigators: Petr Kellnhofer, Thomas Leimkühler, Piotr Didyk, and Karol Myszkowski

The adoption of 3D videos is mostly hindered by the lack of high-quality 3D content. While image and video formats which support stereoscopic content is getting more and more popular, most of the existing footage is captured and stored in 2D format. For the wide use of 3D displays, it is important to develop a method to convert the existing 2D content to stereoscopic content, or to even more advanced multi-view format for glass-free auto-multiscopic displays [1, 2, 3].

We propose two approaches for 2D-to-stereo 3D video conversion. Our first work is a novel, data-driven method for 2D-to-3D video conversion which transfers the depth gradients from a large database of 2D+Depth images [1]. Instead of capturing 2D+Depth databases, we address this problem by creating a synthetic database from computer games and showing that this synthetic database can effectively be used to convert real videos. We show that our method produces high-quality 3D videos that are almost indistinguishable from videos shot by stereo cameras. In addition, our method significantly outperforms the current state-of-the-art, showing up to 20% improvement in the perceived depth.

The second approach for 2D-to-3D conversion is to infer binocular disparity from a monocular video stream in real-time [3]. We compute perceptually plausible disparity, that is numerically inaccurate, but results in a very similar overall depth impression with a plausible
overall layout, sharp edges, fine details and agreement between luminance and disparity. We use several simple monocular cues to estimate disparity maps and confidence maps of low spatial and temporal resolution in real-time. These are complemented by spatially-varying, appearance-dependent and class-specific disparity prior maps, learned from sample stereo images. Scene classification selects this prior at run time. Fusion of prior and cues is done by means of robust MAP inference on a dense spatio-temporal conditional random field (CRF) with high spatial and temporal resolution. Using normal distribution allows this in constant-time, parallel per-pixel work.

We also develop a novel method for converting stereoscopic content to multi-view content [2]. Auto-multiscopic display technology is an appealing approach that provides a glasses-free 3D experience for multiple viewers. However, a technical challenge is the lack of native multi-view content that is required to deliver a proper view of the scene for every viewpoint. We propose a real-time system that can convert stereoscopic video to a high-quality, multi-view video that can be directly fed to auto-multiscopic displays as shown in Fig. 38.26. A key to our solution lies in combining Lagrangian and Eulerian approaches for both the disparity estimation and novel view synthesis, which leverages the complementary advantages of both techniques. The solution preserves all the features of Eulerian methods (e.g. sub-pixel accuracy, high performance, robustness to ambiguous depth cases, and easy integration of inter-view aliasing) while maintaining the advantages of Lagrangian approaches (e.g. robustness to large disparities and possibility of performing non-trivial disparity manipulations through both view extrapolation and interpolation). Our method operates locally, mostly on 1D scan-lines, which allows for an efficient implementation both using a GPU and an FPGA. Therefore, our hardware implementation demonstrates that Eulerian techniques and their combination with Lagrangian approaches are good alternatives to hardware solutions that are based on a Lagrangian approach.

References


38.8.2 Light-Fields and Advanced Displays

Investigators: Shida Kunz Beigpour, Petr Kellnhofer, Piotr Didyk, Elena Arabadzhieyska, Okan Tarhan Tursun, and Karol Myszkowski

Light-field technology offers clear advantages over conventional imaging techniques because it provides a more complete representation of visual information by encoding the direction of light rays traveling in 3D space in addition to their intensities. Its use has been recently growing in popularity especially in Virtual and Augmented Reality (VR/AR) applications
and it is a more convenient format for performing fundamental vision tasks such as depth estimation and occlusion detection.

Appearance editing is one of the fundamental research topics in vision and computer graphics. It is widely used where it is desirable to alter the appearance of objects and surfaces on the existing visual content and it has been extensively studied for conventional visual formats. Unfortunately, the existing methods which have been developed specifically for 2D formats are not applicable to light-fields. Therefore, we introduce a novel appearance editing approach which is based on intrinsic decomposition of light-fields into albedo, diffuse, shading and specularity layers [1]. Different from the existing approaches, our method eliminates the need for a full “inverse rendering”, it does not need the knowledge of geometrical data and it works on surfaces with different specularity and reflectivity characteristics. In our work, we formulate intrinsic decomposition as an energy minimization problem with Retinex, disparity and absolute shading scale terms in the cost function, which is defined on image color gradient. In order to obtain different appearance effects, we decompose each layer into spatial frequency bands using band-sifting method. Then, we are able to get different material appearance effects by modifying the amplitude or inverting the sign of each band. Our method also has a highly-optimized efficient GPU implementation. The implementation allows interactive appearance editing of light-fields where each edit takes as short as 50 – 60 ms.

Unfortunately, the benchmarking of light-field intrinsic decomposition techniques are limited to qualitative evaluations due to the lack of a ground truth dataset. We address this need for a proper ground truth in our work by providing a new dataset consisting of 3 real and 8 synthetic scenes [5]. We capture the real scenes using our hardware setup which consists of a high-quality camera and a stepper-motor both of which are controlled by an Arduino board. Here, the decomposition of real world into diffuse, specular, shading and albedo layers is a challenging task and for a proper decomposition, we use polarizing filters and 3D printing technology to prepare physical chromatic (colored) and achromatic (gray) copies of the objects in our scenes. As a part of this work, we also use our new dataset to evaluate the performance of 4 recent intrinsic decomposition algorithms and share the results. We make our dataset publicly available for other researchers’ use (please refer to Sect. 38.10.5 for details). The potential uses of our dataset are not limited to the evaluation of decomposition quality and we believe that future works in closely related research areas such as light-field reconstruction, depth-based image rendering, 3D reconstruction and specularity removal will also benefit from our dataset.

In addition to our work on light-fields, we improve the state-of-the-art in displaying high-quality visuals by introducing an advanced near-eye display (NED) design, which addresses the lack of accommodative cues in conventional displays. Enabling accommodative cues is especially critical in AR applications where computer-generated overlays must be properly combined with real world stimuli. Our display prototype, which is designed in collaboration with the University of North Carolina, essentially solves this problem with a novel deformable membrane half-mirror setup [2, 3]. The membranes are installed on air-tight chambers and the optical principle behind our design is approximating a spherical concave surface whose curvature is controlled by changing the air pressure inside the chamber. The curvature has an effect of changing the virtual depth level which is formed by reflecting the image on a display from our deformable mirror as shown in Fig. 38.27. This design has many
advantages over the existing solutions because it is easy to build with a simple single optical
element and it provides a wide field of view. We show that this approach successfully solves
the well-known Vergence-Accommodation conflict with a series of perceptual subjective
experiments. Our design received the Best Paper Award at IEEE VR 2017 and Emerging
Technologies DCEXPO Special Prize in SIGGRAPH 2017.

As a part of ongoing efforts to get high-fidelity visuals on HMD displays, the maximum
spatial frequencies that can be reproduced in VR/AR applications are still far below what
HVS can perceive despite recent improvements. In another work, we present a novel method
for improving the apparent resolution of HMD displays to overcome this limitation of HMDs
[4]. Our work focuses on using perceptual optimizations for real-time apparent resolution
enhancement based on a computationally efficient two-step filtering approach. Motivated by
the fact that HVS integrates visual signals over a short period of time, we first estimate a
set of velocity-specific box filters which model the retinal blur introduced by eye movement
between subsequent frames. We invert these filters and apply them to the high-resolution
image. In the second step, we estimate another set of filters which model the optimal
decomposition of high-resolution image into a set of sub-frames. We apply the proper set of
estimated filters by analyzing the magnitude and direction of optical-flow resulting from both
the scene and head movement of the observer. Due to the temporal integration performed
by the HVS on visual signals, the apparent resolution is improved, which is validated by a
series of subjective experiments.

References

  and D. Luebke. Membrane AR: Varifocal, wide field of view augmented reality display from
deformable membranes. In ACM SIGGRAPH 2017 Emerging Technologies, Los Angeles, CA,
USA, 2017, Article 15. ACM.
  and H. Fuchs. Wide field of view varifocal near-eye display using see-through deformable
38.8.3 Image and Light-Field Quality Evaluation

Investigators: Vamsi Kiran Adhikarla, Denis Sumin, Krzysztof Wolski, Piotr Didyk, and Karol Myszkowski

As light-fields are gaining popularity, there is a growing interest in optimizing different stages of their processing pipeline and designing novel approaches for other fundamental operations such as re-sampling, and compression. In most of these tasks, it is critical to control the trade-off between quality and potential gains in other aspects. Therefore, recently a need to properly measure and quantify the loss of quality has arisen, which would greatly help guiding the design and evaluation of different methods developed for light-fields.

In our first work, we test a number of existing objective quality metrics to determine how well they can predict the quality of light-fields [1]. In order to do that, we collect a new dataset of dense reference and distorted light-fields as well as the corresponding quality scores which are scaled in perceptual units. The scores are acquired in a subjective experiment using an interactive light-field viewing setup. The dataset contains typical artifacts that occur in light-field processing chain due to light-field reconstruction, multi-view compression, and limitations of auto-multiscopic displays. The outcome of our study shows that the existing image quality metrics provide good measures of light-field quality but they require dense reference light-fields for the optimal performance. For more complex tasks of comparing two distorted light-fields, their performance drops significantly, which reveals the need for new, light-field-specific metrics.

When running quality metrics, we observe that it is difficult to assess the particular role of a local image distortion in the overall quality score because most of the existing image quality metrics provide a single scalar value representing the quality of the whole input frame. This limitation makes their use impractical for evaluating and improving novel approaches because usually the exact position of distortions are required to precisely identify and track any deviation from ground truth. The visibility metrics which produce visual difference maps and are specifically designed for detecting just noticeable distortions, could be a solution, but their predictions are often inaccurate. We argue that the main reason behind this problem is the lack of large image collections with a good coverage of possible distortions that are observed in different applications. To address the problem, we collect an extensive dataset of reference and distorted image pairs together with user markings indicating whether distortions are visible or not (please refer to Sect. 38.10.4 for details) [2]. We propose a statistical model that is designed for the meaningful interpretation of such data, which is affected by visual search and imprecision of manual marking. We use our dataset for training existing metrics and we demonstrate that their performance significantly improves. We show that our dataset with the proposed statistical model can be used to train a new CNN-based metric, which outperforms the existing solutions. We demonstrate
Figure 38.28: Three sample images from our dataset for predicting local visible differences and results of different metrics. The result of our CNN-based metric is shown in the last column, which predicts the average detection probabilities shown in the second column more accurately than the existing approaches.

the utility of such a metric in visually lossless JPEG compression, super-resolution and watermarking. Results of the trained metrics are presented in Fig. 38.28. In [3] the metric was enhanced by supporting varying display brightness and viewing distance. In the future, we plan to check the performance of this metric in light-field evaluation.

References


38.8.4 Eye Tracking

Investigators: Elena Arabadzhiyska, Okan Tarhan Tursun, Piotr Didyk, Hyeonseung Yu, Mojtaba Bemana, Gurprit Singh, and Karol Myszkowski

In the recent years, there have been rapidly increasing rendering demands on high-quality HMD and light-field displays as well as standard desktop displays. Higher visual quality and realistic rendering is addressed by more accurate simulation of the physical propagation
and reflection properties of light but this trend is almost always in direct correlation with higher computational costs. As the computational budget is limited by the capabilities of the hardware, it is always desirable to use it as efficiently as possible by improving the perceived visual quality and not wasting the rendering resources on the features which are not visually perceivable. Motivated by this observation, the gaze-contingency paradigm and the use of eye trackers have emerged as a practical solution in computer graphics applications [2].

Unfortunately, wide adoption of eye tracking in gaze-contingent rendering is mainly hindered by the limited sampling and refresh rates of affordable eye trackers and displays. This limitation is observed as an increased latency in updating the visual content and it is visible as disturbing temporal artifacts due to a temporary mismatch between actual gaze position and the location reported by the eye tracker during rapid eye movements called saccades. Saccades are one of the most frequent type of eye movements and they naturally occur many times when humans move their eyes around to observe a scene. Fortunately, once a saccade is initiated by the HVS, its trajectory follows a pre-programmed path with predictable acceleration and deceleration patterns. Using this ballistic property of saccadic eye movements, we build a model to predict the visual target of observers when their visual system initiate saccades (see Fig. 38.29a) [1]. We show an application of our model in foveated rendering, where the region rendered in high-resolution is moved to the target position before the end of the saccade to counteract the effects of system latency observed in standard gaze-contingent rendering methods. This allows a perceptually seamless foveated rendering experience which is free of temporal artifacts even on low-end eye tracking and display devices by eliminating the need for more capable and expensive hardware.

We show another successful application of eye tracking on multi-layer accommodative displays by optimizing the decomposition of 3D content into individual display layers (see Fig. 38.29b) [3]. Our work is based on two popular decomposition strategies; namely, Linear Blending (LB) and Light-Field Synthesis (LFS). In order to improve apparent visual quality, we introduce a gaze-dependent viewpoint sampling for LFS and perform a set of perceptual experiments to develop a domain-specific Structural Similarity Index (SSIM) for measuring perceived visual differences between two decomposition strategies under various types of stimuli and viewing conditions. Then, we propose a unified hybrid decomposition framework.
based on LB and LFS strategies and a highly efficient GPU implementation for real-time applications. In our subjective experiments, the participants report a visual quality similar to high-quality LFS decomposition while total rendering time is reduced approximately by a factor of 4 under typical viewing conditions.

References


38.9 Computational Fabrication

Coordinator: Vahid Babaei

The recent, wide availability of advanced manufacturing devices, such as 3D printers, has triggered huge interest in academia and industry. Additive manufacturing is opening the eyes of our community to the immense but untapped potential of the manufacturing industry for computational methods. Most of the time, hardware devices offer capabilities which are never utilized, because the necessary software and algorithms are underdeveloped. In recent years, our group has paid a particular attention to this emerging field. In Sect. 38.9.1, we summarize our contributions in appearance fabrication using multi-material 3D printers where we are leading this research area by introducing new algorithms and standards. In Sect. 38.9.2, our efforts for design and fabrication of drawing tools that enable controlling the tactile properties are presented. In Sect. 38.9.3, we review directional screens – efficient screens that reflect light from projectors in the most important directions.

38.9.1 Appearance Fabrication

Investigators: Vahid Babaei, Piotr Didyk, Denis Sumin, and Karol Myszkowski

We have a particular interest in visual appearance of objects and strive for fabrication algorithms that help creating products with novel and useful appearance characteristics using advanced manufacturing tools. Design for manufacturing of objects with high-fidelity appearance specifications is a key engineering task. On the computational/algorithimic front, there is a huge opportunity for revolutionizing the computational appearance design and manufacturing, and significantly improving the appearance quality of many products. The results of this research will enable numerous applications in rapid prototyping and manufacturing of end-use products. This spans several application domains from medical devices and surgical training, to cultural heritage preservation and anti-counterfeiting.
Similarly to 2D printers, multimaterial 3D printers are also binary devices where only a single material with a fixed concentration can be deposited in every volume element. Therefore, there is a tendency to adapt 2D color reproduction methods, and above all halftoning, to 3D printing. However, the usual issues and artifacts of original 2D halftoning methods persist, and they significantly influence the quality of 3D prints. In a first work [2, 1], we were able to eliminate halftoning artifacts by utilizing the inherent ability of 3D printers that put layers on top of each other. Our proposed “contoning” technique prints layers of transparent inks around an opaque body made of a white, diffuse material. In contrast to halftoning methods, contoning mixes the colors within the volume of the object. Hence, regardless of the viewing distance, the decomposition of the color into multiple materials remains invisible to the eye of the observer. Our results (Figure 38.30) suggest that, compared to halftoning algorithms, contoning is free of spatial artifacts.

Capturing and fabricating the object’s spectral reflectance in the visible range of light is called spectral reproduction. The input to a spectral reproduction workflow is a multispectral image. In a followup to [2], we employed multi-material 3D printing for spectral reproduction of paintings [4]. We took advantage of the current 3D printers capabilities of combining highly concentrated inks with a large number of layers, to expand the spectral gamut of a set of inks. We used a deep neural network to both predict the spectrum of a printed ink stack and optimize for the stack layout that best matches a target spectrum. In addition, we were able to successfully combine the contoning method with error-diffusion halftoning, which simultaneously solves the layout discretization and color quantization problems, accurately and efficiently. Our workflow outperforms the state-of-the-art models for spectral prediction and layout optimization. We demonstrated reproduction of a number of real paintings painted by our artist collaborator using our 3D printers that uses 10 custom inks (Figure 38.31).

The current printing materials for inkjet 3D printing, the most viable technology for appearance printing, suffer from excessive, unwanted translucency. This results in volumetric light transport (cross-talk) between surface points on a 3D print and severe blurring of details. In another work [3], we addressed exactly this problem. Our works stands in striking contrast with other research efforts in color 3D printing which ignore volumetric light transport and treat each point on a surface independently from its neighborhood. We employed a general Monte-Carlo simulation of heterogeneous scattering that predicts the color at each location on the surface very accurately by accounting for the internal structure of the...
Figure 38.31: Our workflow generates accurate reproduction of paintings and provides faithful color reproduction under various light sources. Paintings ©Azadeh Asadi.

Figure 38.32: We find an optimal material arrangement which results in superior appearance, preserving the spatial details despite the highly translucent materials.

volume surrounding each surface point. Our workflow employs a numerical optimizer on top of this accurate forward simulation and for any input texture produces an optimal volumetric map of printing materials. We are able to counteract heterogeneous scattering to obtain the impression of a crisp albedo texture on top of the 3D print, by optimizing for a fully volumetric material distribution that preserves the target appearance. Despite the inherent translucency of the medium, we reproduce detailed surface textures on 3D prints (Figure 38.32).

References


Figure 38.33: Our system involves quantifying the mechanics and perception of drawing tools, thereby designing and fabricating digital drawing tools that mimic the feel of real tools.

38.9.2 Fabrication of Digital Drawing Tools

Investigator: Piotr Didyk

New digital fabrication techniques enable customization of objects’ surface finish which consequently affects tactile properties. In this area, our research focuses on fabricating tools and surfaces for reproducing different tactile feedback during human-object interaction. Currently, we focus on designing and fabricating digital drawing styli which mimic traditional drawing tools. The research is driven by the observation that while digital drawing is becoming a favorite technique for many artists, a significant disadvantage is poor haptic feedback. Artists are usually limited to one surface and a few different stylus nibs, and while they try to find a combination that suits their needs, this is typically challenging. In our recent work [1], we addressed this problem and proposed a method for designing, evaluating, and optimizing different stylus designs.

To this end, we first characterized a large set of traditional drawing tools (Figure 38.33). We then conducted a series of user experiments to investigate the perception of drawing tools. These led to a derivation of a perceptual space that encodes perceptually-relevant attributes of drawing materials expressed as a function of measurable physical properties of the tools. We also developed a new, data-driven simulation technique for characterizing stylus-surface interaction, which allows us to embed new styli designs into the space and compare them with other designs. We finally leveraged all the components and recent advancements in multi-material 3D printing to demonstrate the application of our system in the design of new digital drawing tools that mimic traditional drawing materials. Our experiments demonstrated that using our system we can customize and fabricate styli which exhibit more similar tactile properties to the traditional drawing materials than currently available solutions.

References

38.9.3 Display Design and Fabrication

Investigator: Piotr Didyk

The appearance fabrication has applications beyond reproducing a look of a particular object. We also exploited our expertise in this field in the context of optimizing front projector screens. Here, the main challenge lies in fabricating surfaces which reflect light uniformly in all the directions where viewers are expected. Additionally, efficient screens should not waste light by reflecting it towards areas such as walls or ceiling. The problem boils down to designing and fabricating large surfaces with spatially-varying reflectance properties. Consequently, in our work [1, 2], we propose a new type of front projection screens, so-called directional screens. They are composed of tiny, highly reflective surfaces which reflect the light coming from a projector only towards the audience. Additionally, they reflect the light uniformly, avoiding “hot-spotting”, and can support non-standard audience layouts. In our research, we proposed an efficient algorithm for generating such surfaces. We evaluated our designs and prototypes demonstrating a potential for significant quality gains and energy savings.

References


38.10 Software and Datasets

As part of the research process, several libraries, development tools, large corpora of reference data sets, and application frameworks have been developed by members of the group. Also, a spin-off company was founded. In this section we describe some of them that evolved to a level where it was appropriate to either distribute them as open source projects or let members of other research institutions benefit from software that had been developed in our group.

38.10.1 TheCaptury GmbH – A Spin-off Company Commercializing Marker-less Motion and Performance Capture

Investigators: Nils Hasler, Michal Richter, and Christian Theobalt

TheCaptury GmbH (www.thecaptury.com) is a spin-off company from the Graphics, Vision, & Video group of Prof. Dr. Christian Theobalt that was founded in 2013. It is based in Saarbruecken and was supported in its inception by the the joint IT Inkubator of Saarland University and the Max Planck Society.

TheCaptury commercializes state-of-the-art marker-less motion and performance capture technologies originally developed in the Graphics, Vision, & Video group (see Sect. 38.6) and further refined for productization. On the motion capture side, theCaptury provides state-of-the-art motion capture systems, that require neither markers, nor tight suits, and
work in real-time. The systems work with standard video cameras and succeed to capture people in regular everyday apparel in general indoor and outdoor scenes. The technology allows tracking humans, props, and animals, using commodity video cameras. In addition to traditional motion capture, theCaptury develops technology that allows computing spatially and temporally coherent 3D geometry of the performer for sequences. This allows the software to re-render the scene from novel, virtual view-points to generate camera sweeps with freeze-frame effects using only few and inexpensive video cameras.

Currently, the two main products are an off-line software version for marker-less motion capture with arbitrary video camera systems, called Captury Studio (Fig. 38.34). The company also sells a real-time system including video recording and real-time processing hardware, called Captury Live.

The founding of theCaptury was supported by an EXIST Forschungstransfer grant by the German Ministry of Economy (around 500k Euros). TheCaptury received several awards, including the Main Prize of the IKT Innovativ Competition at CeBIT 2013, as well as a main prize in the 123go business plan competition of the greater Saarland/Luxemburg/Lorraine and Belgium region.

TheCaptury’s solution offers unique features: it needs no markers, it works in real-time, works for multiple people in interaction, works in general scenes outside controlled studios, works for people in general everyday clothing, works even with cheap unsynchronized action cameras, can capture very fast motions, can capture humans and even animals, works even in very unconventional environments (e.g., underwater), and – importantly – measures at an accuracy comparable to marker-based methods.

TheCaptury’s technology is used by several major computer game and visual effects studios (customer names cannot be disclosed), where it provides a much simpler and faster way to capture data for virtual actors, now also directly onset and outside studios. Further, theCaptury enables new applications in medicine and the life sciences. In combination with a software that biomechanically analyze captured motions, theCaptury provides a new way to analyze the forces and strains in the body, the joints and the musculoskeletal system during motion can be analyzed. This solution is commercially used for large scale analysis of the musculoskeletal health of workers in companies, from which suggestions for therapy,
exercise, or workplace improvement can be derived, and from which some injuries in their forming can even be predicted. Several companies have used theCaptury to measure and improve workplace ergonomics. The same solution is also used in professional sports by several internationally known teams (names cannot be disclosed). Notably, theCaptury is used by the Chinese national track and field Olympic team (Fig. 38.35). Here the marker-less video-based motion capture analysis of live athletic motions, without any obstructing body instrumentation, is used to improve the athletes’ performance, to predict injuries that are building up, and to further reduce the risk of injuries by optimizing motions.

TheCaptury motion analysis is also used in academia to fine-tune brain implant stimulator settings for Parkinson patients. This implant reduces a patient’s tremor, and needs regular recalibration by patient observation. With this non-obstructing, easy-to-setup motion capture system this can be done faster and by means of quantitative motion evidence rather than mere qualitative observation. TheCaptury’s system further opens the door for new functionality in virtual reality. A challenge in VR is to capture the motion of a person wearing a head-mounted VR display such that a virtual avatar, which mimics her real-world motion, is visible in the virtual world. This is important for the impression of self-embodiment in VR, e.g., when the person is looking down on her virtual body. With theCaptury, the full body motion of the user can be capture in real-time. This new solution for controlling avatars in VR is already used by several industrial companies. Finally, theCaptury’s technology is relevant for the factory of the future (Industry 4.0), where hybrid teams of humans and robots will have to collaborate effectively and safely in the same workspace. To enable this, machines must understand both the large scale and fine-scale motion of humans in their environment in real-time, at high accuracy, and without disturbing on-body instrumentation. Only then, robots can safely work together with humans or even learn their tasks by observing them. Together with several industrial and research partners (e.g. together with Airbus, Kuka,
DFKI), the company currently develops new solutions for hybrid human/robot teams in
the factory of the future as part of the Hybr-it project (www.hybr-it-projekt.de) funded by
BMBF.

38.10.2 GVVPPerfCapEva – A Repository of Human Shape and Performance
Capture Datasets

Investigators: Christian Theobalt and Collaborators

GVVPPerfCapEva (gvvperfcapeva.mpi-inf.mpg.de) is a repository of validation and train-
ing data sets for various aspects of dynamic scene reconstruction in general, and human
motion and performance capture in particular, that were created by the Graphics, Vision, &
Video headed by Prof. Dr. Christian Theobalt, and partner research groups at MPI for Infor-
matics and elsewhere. The datasets capture a wide range of human shape and performance
including full body, faces, and hands (see Fig. 38.36). The datasets span different sensor
modalities such as depth cameras, multi-view video, optical markers, and inertial sensors.
Notable recent additions are new extensive datasets for training monocular 3D hand pose
estimation methods from color (GANerated hands) and depth cameras (SynthHands), also
in interaction with real world objects (EgoDexter). Also, we provide some of the largest
data sets of monocular human pose video with ground truth annotations from a marker-less
multi-view motion capture system, both showing individual humans (MPI-INF-3DHP) and
multiple humans (MuCo-3DHP). Large test sets of in-the-wild images with 3D ground truth
are also provided alongside these data sets. Statistical models of human shape are widely
used in pose estimation and computer graphics. The ScanDB dataset provides 3D registered
scans of human subjects, and a statistical shape model built using the scans. The repository
also features data sets for general deformable shape capture from cameras. For some datasets,
external results obtained with newly developed methods can be uploaded and compared to
previous approaches.

Figure 38.36 shows examples of some of the datasets available for download.

GVVPPerfCapEva is the largest such repository for data on human shape and performance
with more than 25 datasets available for download. The datasets enable researchers worldwide
to compare their work and evaluate their algorithms. Since its inception, the repository
webpage has been visited by thousands of visitors, and datasets are widely used in the
literature.

38.10.3 PFSTOOLS for Processing High Dynamic Range Images and Video

Investigators: Rafal Mantiuk, Grzegorz Krawczyk, Tunç O. Aydin, and Ivo Ihrke

The pfstools package is a set of command line programs for reading, writing, manipulating
and viewing high-dynamic range (HDR) images and video frames. All programs in the
package exchange data using a simple generic high dynamic range image format, pfs,
and they use unix pipes to pass data between programs and to construct complex image
processing operations.

pfstools come with a library for reading and writing pfs files. The library can be used
for writing custom applications that can integrate with the existing pfstools programs.
pfstools offer also a good integration with high-level mathematical programming languages, such as MATLAB or GNU Octave. pfstools can be used as the extension of MATLAB or Octave for reading and writing HDR images or simply to store effectively large matrices.

The pfstools package is an attempt to integrate the existing high dynamic range image formats by providing a simple data format that can be used to exchange data between applications.

The pfstools package is accompanied by the pfscalibration and pfstmo packages. The pfscalibration package provides an algorithm for the photometric calibration of cameras and for the recovery of high dynamic range (HDR) images from the set of low dynamic range (LDR) exposures. The pfstmo package contains the implementation of ten state-of-the-art tone mapping operators suitable for convenient processing of both static images and animations.

The pfstools, pfscalibration and pfstmo packages are licensed as an Open Source project under a General Public License (GPL). The project web pages can be found at:

http://www.mpi-inf.mpg.de/resources/pfstools/

The software received wider interest of Open Source community and third party contributors prepared installation packages which are now included in several Linux distributions including Debian, Fedora and Suse. The software was presented on the Electronic Imaging Conference 2007 and a general introduction to the package was published in the proceedings [1].

References

38.10.4 LocVis – Local Visibility Maps of Artifacts and Distortions in Images

Investigators: Krzysztof Wolski and Karol Myszkowski

Numerous applications require a robust metric that can predict whether image differences are visible or not. However, the accuracy of existing visibility metrics, such as HDR-VDP, is often not good enough. In our paper, we argue that the main reason behind this problem is the lack of sufficient image collection with a good coverage of possible distortions [4]. We address this problem by creating a new dataset in collaboration with University of Cambridge.

The LocVis dataset\(^1\) consists of 557 images with 170 unique scenes, where the distortions and artifacts were manually labeled by human observers. Many of the images are generated for up to 3 distortion levels of artifacts, for example, different quality settings of image compression. Each scene is marked by approximately 15-20 observers in a local manner – instead of a single score, i.e. observers were asked to paint the mask and depending on their responses, pixel-wise visibility maps of distortions are recorded by our software tool.

Many of the images used in our dataset are based on the data available in previous studies and they were selected to cover an extensive set of typical artifacts which are frequently observed in computer graphics research and applications (see Fig. 38.37). In order to provide a convenient access for other researchers, we group the images into several subsets depending on the type of artifacts. The subset \textit{mixed} is an extended LOCCG dataset including such distortions such as: high-frequency and structured noise, virtual point light (VPL) clamping, light leaking artifacts, local changes of brightness, aliasing and tone mapping artifacts [2, 3]. The subset \textit{perceptionpatterns} consists of artificial patterns designed to expose well known perceptual phenomena, such as luminance masking, contrast masking and contrast sensitivity. The subsets \textit{aliasing}, \textit{peterpanning}, \textit{shadowacne}, \textit{downsampling} and \textit{zfighting} contain real-time rendering artifacts. The subset \textit{compression} contains distortions due to experimental low-complexity image compression. The subset \textit{deghosting} contains artifacts due to HDR merging, which exposes shortcomings of popular deghosting methods. Images of \textit{ibr} and \textit{cgibr} subsets contain artifacts produced by view-interpolation and image-based rendering methods, and they are mainly adopted from the previous work of Vamsi et. al. [1].

References


\(^1\) https://www.repository.cam.ac.uk/handle/1810/274368
http://visibility-metrics.mpi-inf.mpg.de/files/marking_datasets.zip
Figure 38.37: The figure presents examples of stimuli from the dataset. The insets show the closeup of the artifacts.


### 38.10.5 MPI Light-Field Archive

*Investigators: Vamsi Kiran Adhikarla, Denis Sumin, Shida Kunz Beigpour, and Karol Myszkowski*

Light-fields become a popular representation of three dimensional scenes, and there is interest in their processing, resampling, and compression. As those operations often result in loss of quality, there is a need to quantify it. With a purpose of new Light-field metric design, we created the *Light-field Archive*[^1], which is a collection of dense reference and distorted light-fields as well as the corresponding quality scores which are scaled in perceptual units [1].

The archive consists of nine synthetic and captured five real-world scenes, spanning large variety of different conditions such as: specular and transparent surfaces, thin and irregular structures, multiple light sources, multiple diffuse objects, day and night lights, indoor and outdoor, high-contrast color and texture, highly reflecting surfaces and shadows, glossy surfaces and simple objects. Examples of the dataset are presented in Fig. 38.38. All the light-fields are of identical spatial and angular resolution (960 × 720 × 101). The angular resolution was chosen high enough to avoid visible angular aliasing. This was achieved by assuring that the maximum on-screen disparities between consecutive view are around 1 pixel. To guarantee comfortable viewing, we limit the total disparity range to 0.2 visual degree. In the dataset typical light-field distortions that are specific to transmission, reconstruction,

[^1]: [https://lightfields.mpi-inf.mpg.de/Dataset.html](https://lightfields.mpi-inf.mpg.de/Dataset.html)
and display were considered. For each distortion multiple light-fields were generated, by varying the distortion severity level.

The reason, why Light-field imaging became so popular, is that it has various advantages over the traditional 2D photography, such as depth estimation and occlusion detection. This features can aid intrinsic decomposition, which enable multiple applications, such as light-field appearance editing. However, the current light-field intrinsic decomposition techniques primarily resort to qualitative comparisons, due to lack of ground-truth data. The Light-Field Intrinsic Dataset \(^3\) addresses this problem providing number of real world and synthetic 4D and 3D (only horizontal parallax) light-fields \(^2\). The ground-truth intrinsic data comprises albedo, shading and specularity layers for all sub-aperture images. In case of synthetic data, we also provide ground-truth depth, normals, and further decomposition of shading into direct and indirect components. For real-world data acquisition, we make use of custom hardware and 3D printed objects, assuring precision during multi-pass capturing.

References


\(^3\)https://lfid.mpi-inf.mpg.de
38.11 Academic Activities

38.11.1 Journal Positions

Karol Myszkowski is on the editorial board of

- ACM Transactions on Graphics (since 2018),
- IEEE Transactions on Visualization and Computer Graphics (since 2015),
- Journal of Virtual Reality and Broadcasting (since 2004),
- ACM Transactions on Applied Perception (since 2002),

Hans-Peter Seidel is on the editorial board of

- Visual Informatics (chief editor) (since 2017),
- International Journal of Shape Modeling (IJSM) (since 2001),
- Computer Aided Geometric Design (CAGD) (since 1999),
- Graphical Models (GMOD) (since 1995).

Christian Theobalt is on the editorial board of

- ACM Transactions on Graphics (since 2016),
- Visual Informatics (since 2016).

38.11.2 Conference and Workshop Positions

Membership in Program Committees

Renjie Chen:

- Symposium on Geometry Processing, Milan, Italy, July 2019.

Karol Myszkowski:

- ACM SIGGRAPH Asia Technical Papers, Brisbane, Australia, November 2019.
- ACM SIGGRAPH Technical Papers, Los Angeles, USA, July 2019.
- ACM SIGGRAPH Technical Papers, Vancouver, Canada, August 2018.
- EUROGRAPHICS Full Technical Papers, Delft, Netherlands, April 2018.
- Human Vision and Electronic Imaging XXV (HVEI’18), San Francisco, USA, February 2018.
- ACM SIGGRAPH Asia Technical Papers, Bangkok, Thailand, December 2017.
Christian Theobalt:

- **Vision, Modeling and Visualization (VMV)**, Stuttgart, Germany, October 2018.
- **German Conference on Pattern Recognition (GCPR)**, Stuttgart, Germany, October 2018.
- **IEEE International Conference on 3D Vision (3DV)**, Verona, Italy, September 2018.
- **European Conference on Computer Vision (ECCV)**, Munich, Germany, September 2018.
- **High Performance Graphics**, Vancouver, Canada, August 2018.
- **IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)**, Salt Lake City, USA, June 2018.
- **EUROGRAPHICS STAR Reports (Program Chair)**, Delft, Netherlands, March 2018.
- **IEEE International Conference on Computer Vision (ICCV)**, Venice, Italy, October 2017.
- **IEEE International Conference on 3D Vision (3DV)**, Qingdao, China, October 2017.
- **German Conference on Pattern Recognition (GCPR)**, Basel, Switzerland, September 2017.
- **Vision, Modeling and Visualization (VMV)**, Bonn, Germany, September 2017.
- **IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)**, Hawaii, USA, June 2017.

Okan Tarhan Tursun:


**Membership in Organizing Committees**

Karol Myszkowski:


Hans-Peter Seidel:

– Conference Steering Committee – *Eurographics Annual Conference Series*, European Association for Computer Graphics (Eurographics)
– Conference Steering Committee – *Vision, Modeling and Visualization (VMV) Annual Conference Series*

**Christian Theobalt:**

### 38.11.3 Invited Talks and Tutorials

**Vahid Babaei**
– *Appearance Fabrication for Two and Three Dimensional Printing*, Invited Talk, Hubert Curien Laboratory, Saint-Etienne, France, October 2018.

**Karol Myszkowski:**
– *Perceptual Display: Apparent Enhancement of Scene Detail and Depth*, Keynote Speaker, Human Vision and Electronic Imaging XXV (HVEI’18), San Francisco, USA, February 2018.
– *Perceptual Modelling for Stereoscopic 3D*, Invited Talk, University of Konstanz, Germany, December 2017.

**Hans-Peter Seidel:**
– *3D Image Analysis and Synthesis*, Invited Talk, Stanford University, November 2017
Gurprit Singh:


Christian Theobalt:

- *Capturing and Editing Models of the Real World in Motion*, Invited Talk, Technical University of Graz, Austria, November 2018.
- *Capturing Human Motion in the Wild*, Invited Talk, PoseTrack Challenge: Articulated People Tracking in the Wild, ECCV 2018 Workshop, Munich, Germany, September 2018.
- *Capturing and Editing the Real World in Motion* Invited Talk, Technical University of Munich, Germany, February 2018.

Hyeonseung Yu:

Rhaleb Zayer:

- *Demystifying Algebraic Topology*, Invited Talk, Technical University of Graz, Austria, June 2018.

38.11.4 Other Academic Activities

Karol Myszkowski:

- Reviewer for funding proposals submitted to Austrian Science Fund (FWF).
- External reviewer for PhD theses (only for the period of 2017-2019): Vienna University of Technology, CTU Prague, Charles University in Prague, EPFL, DTU Copenhagen, Graz University of Technology, University of British Columbia (UBC), UNC Chapel Hill, Zaragoza University.

Hans-Peter Seidel:

- Scientific Coordinator, Cluster of Excellence on Multimodal Computing and Interaction (MMCI), Saarland University (since 2007).
- Co-Director, Max Planck Center for Visual Computing and Communication (MPC-VCC), Stanford/Saarbrücken (since 2003).
- ERC Starting Grant Evaluation Panel, European Research Council (ERC) (2013/15/17).
- Swiss SNSF Research Council (since 2018).
- ACM SIGGRAPH Academy Committee (since 2018).
- Senate Evaluation Committee (SAE), Gottfried Wilhelm Leibniz Association (Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e.V.) (since 2014).
- Advisory Board of IT Incubator of Max Planck Society/Saarland University (since 2018).
- Review Board, Bavarian Research Foundation (Bayerische Forschungsstiftung), (since 2015).
- Scientific Advisory Board, Bavaria California Technology Center (BaCaTec), (since 2009).
- Executive Committee, Eurographics Association (since 1992).

Christian Theobalt:

- Member of the Steering Committee of the Intel Visual Computing Institute, Germany (2009-2017).
- Advisory Board of IT Incubator of Max Planck Society/Saarland University (since 2016).
- Co-Founder of and Scientific Mentor to the Captury GmbH (www.thecaptury.com).

Graduate Student Selection Committees of the International Max Planck Research School for Computer Science (IMPRS-CS) and the Saarland Graduate School of CS (since 2009).

Reviewer for ERC Starting/Advanced Grants.

Reviewer for funding proposals submitted to German Ntl. Science Foundation (DFG), Austrian Science Fund (FWF), Canadian Leader Opportunity Fund, Swiss National Science Foundation (SNF), Vienna Science & Technology Fund (WWTF).

Reviewer for EU FP7 and Horizon 2020 projects.

External Reviewer for PhD theses (only for the period of 2017-2019): EPFL, TU Graz, ETH Zürich, Erlangen University, University Paris Seine, INRIA Sophia Antipolis

### 38.12 Teaching Activities

#### Summer Semester 2017

**Lectures:**
- Computational Fabrication (P. Didyk)
- Geometric Modelling (H.-P. Seidel, R. Zayer)
- Realistic Image Synthesis (P. Slusallek, K. Myszkowski)

#### Winter Semester 2017/2018

**Lectures:**
- Perception in Computer Graphics (K. Myszkowski, P. Didyk)

#### Summer Semester 2018

**Lectures:**
- Geometric Modelling (H.-P. Seidel, R. Zayer)
- Realistic Image Synthesis (P. Slusallek, K. Myszkowski, G. Singh)

**Seminars:**
- 3D Shape Analysis (F. Bernard, C. Theobalt)

#### Winter Semester 2018/2019

**Lectures:**
- Perception in Computer Graphics (K. Myszkowski, O. Tursun)
Seminars:
Computational Fabrication (H.-P. Seidel, V. Babaei)

Master and Bachelor Theses
Zeeshan Khan Suri: Low-dimensional Shape Matching, ongoing.

38.13 Dissertations, Habilitations, Awards

38.13.1 Dissertations

Completed and Defended
Oliver Nalbach: Smarter Screen Space Shading, 10.11.2017.

Completed, but not yet defended
- Thomas Leimkühler: Artificial Intelligence for Efficient Image-based View Synthesis.
- Hyeongwoo Kim: Learning-based Face Reconstruction and Editing.
- Nadia Robertini: Model-based Human Performance Capture in Outdoor Scenes.

38.13.2 Offers for Faculty Positions
- Dan Casas: Tenure-Track Assistant Professor, Universidad Rey Juan Carlos (URJC), Madrid, Spain, 2018.
- Avishek Chatterjee: Tenure-Track Assistant Professor, IIT Kharagpur, India, 2017
- Renjie Chen: Tenured Associate Professor, University of Science and Technology of China (USTC), 2019.
- Piotr Didyk: Tenure-Track Assistant Professor, USI Lugano, Switzerland, 2018
- Cigui Jiang: Tenured Associate Professor, Xi’an Jiatong University, China, 2019
- Helge Rhodin: Tenure-Track Assistant Professor, UBC, Vancouver, Canada, 2019

38.13.3 Awards

- Pablo Garrido: Eurographics PhD Dissertation Award, Eurographics, 2018
- Marc Habermann: Guenter-Hotz Medal, Universität des Saarlandes, 2017
- Petr Kellnhofer: Dr.-Eduard-Martin-Preis, Universität des Saarlandes, 2017 
  Eurographics PhD Dissertation Award, Eurographics, 2018
- Petr Kellnhofer, Piotr Didyk, Karol Myszkowski: Best Paper Award, IEEE VR, 2017
  Digital Content Assoc. of Japan Award, ACM SIGGRAPH Emerging Techn., 2017
- Karol Myszkowski: IEEE TVCG Best Associate Editor Award 2016, IEEE VGTC, announced December 2017
- Thomas Leimkühler: Best Paper Award, Eurographics Symposium on Rendering, 2017
- Franziska Müller, Christian Theobalt (co-authors together with M. Soliman, L. Hegemann, J. Sol Roo, and J. Steimle from Saarland University) Best Academic Paper Award, ACM International Conference on Interactive Surfaces and Spaces (ISS), 2018
- Franziska Müller: Google Europe PhD Fellowship, 2017
  Women STEM Award, 2017
- Hans-Peter Seidel: EG Gold Medal, Eurographics, 2017 
  HPG 2017 Test-of-Time Award (with J. Günther, S. Popov and Ph. Slusallek), 2017
- Christian Theobalt: ERC Consolidator Grant – 4DRepLy, European Research Council, 2017
  Karl Heinz Beckurts Preis, Karl Heinz Beckurts-Stiftung, 2017
- Rhaleb Zayer (co-author together with M. Winter and M. Steinberger), Best Paper Award, IEEE HPEC, 2017
38.14 Grants and Cooperations

- Vahid Babaei with Wojciech Matusik (CSAIL MIT) and Szymon Rusinkiewicz (Princeton): Computational Fine Art Reproduction
- Renjie Chen with Craig Gotsman (NJIT), Kai Hormann (University of Lugano): Distance Functions for Fastest Path Computation
- Renjie Chen with Ofir Weber (Bar Ilan University): Real-time Shape Deformation
- Renjie Chen with Markus Steinberger (TU Graz): Delaunay Triangulation Computation on GPU
- Karol Myszkowski with David Luebke (Nvidia) and Henry Fuchs (UNC at Chapel Hill): Varifocal See-through Near-eye Display
- Karol Myszkowski with Diego Gutierrez and Belen Masia (Zaragoza University): Image Quality in VR
- Karol Myszkowski, Vahid Babaei with Bernd Bickel (IST Austria), Tim Weyrich (UCL) and Jaroslav Krivanek (Charles University): Appearance Fabrication
- Karol Myszkowski, Gurprit Singh with Tobias Ritschel (UCL): Multi-dimensional Sampling
- Karol Myszkowski with Tobias Ritschel (UCL): Rendering Artifact Restoration
- Karol Myszkowski with Piotr Didyk (USI Lugano): Advanced Displays
- Karol Myszkowski with Rafał Mantiuk (Cambridge University): Data-driven: Image Quality Evaluation
- Karol Myszkowski, Horizon 2020, Innovative Training Network (ITN), DISTRO: Distributed 3D Object Design, grant agreement 642841
- Karol Myszkowski, Horizon 2020, Innovative Training Networks (ITN), RealVision: Hyperrealistic Imaging Experience, grant agreement 765911
- Karol Myszkowski, Hans-Peter Seidel, Joint FhG/MPG Research Grant, Perceptually-aware Light Field Capture, Processing and Display, funded by MPG and FhG
- Hans-Peter Seidel, Scientific Coordinator, Cluster of Excellence on Multimodal Computing and Interaction, funded by DFG
- Hans-Peter Seidel, Principal Investigator, Cluster of Excellence on Multimodal Computing and Interaction, funded by DFG
- Hans-Peter Seidel, Co-Director, Max Planck Center for Visual Computing and Communication, funded by BMBF, MPG and Stanford University
– Gurprit Singh with Wojciech Jarosz (Dartmouth College), Kartic Subr (University of Edinburgh), David Couerjolly, Victor Ostromoukhov (LIRIS/CNRS): Monte Carlo Importance Sampling Analysis

– Gurprit Singh with Cengiz Öztireli (Disney Research), Abdalla G.M. Ahmed (KAUST), Oliver Deussen (University of Konstanz), Ravi Ramamoorthi (UCSD): Analysis of Sample Correlations for Monte Carlo Rendering

– Gurprit Singh with Iliyan Georgiev (Autodesk Solid Angle): Error Distribution Analysis

– Gurprit Singh with Toshiya Hachisuka (University of Tokyo, Japan): Gradient Domain Analysis

– Christian Theobalt, ERC Starting Grant CapReal

– Christian Theobalt, ERC Consolidator Grant 4DRepLy

– Christian Theobalt, Steering Committee of Intel Visual Computing Institute (until 2017)

– Christian Theobalt, Principal Investigator, Intel Visual Computing Institute (until 2017)

– Christian Theobalt, Unconstrained Performance Capture, funded by Intel Visual Computing Institute

– Christian Theobalt, User-centric Video Processing, funded by Intel Visual Computing Institute

– Christian Theobalt, Advanced Video Processing and Face Reconstruction, with Patrick Perez, Technicolor Research and Development

– Christian Theobalt with Shahram Izadi (Google): 3D and 4D Reconstruction

– Christian Theobalt with Matthias Niessner (TUM): 3D/4D Scene Reconstruction, Face Capture, VR/AR

– Christian Theobalt with Pat Hanrahan (Stanford University): Programming Tools for Optimization

– Christian Theobalt with Kalyan Sunkavalli (Adobe Research): Face Appearance Capture

– Christian Theobalt with Jürgen Steimle (Saarland University): Gesture-based Human-Computer Interaction

– Christian Theobalt with Antti Oulasvirtta (Aalto University): Gesture-based Human-Computer Interaction

– Christian Theobalt with Marc Stamminger (University of Erlangen-Nuremberg): 3D and 4D Reconstruction, Inverse Rendering, Image-based Rendering
– Christian Theobalt with Michael Zollhöfer (Stanford University): 4D Reconstruction, Video-based Rendering and Editing

– Christian Theobalt with Ohad Fried, Maneesh Agrawala (Stanford), Adam Finkelstein (Princeton): Model-guided Video Editing

– Christian Theobalt with Wenping Wang (University of Hong Kong): Performance Capture and Neural Rendering

– Christian Theobalt with Niloy Mitra (University College London): Thin Structure Reconstruction

– Christian Theobalt with Christian Richardt (University of Bath): User-centric Video, Computational Videography

– Christian Theobalt with Kwang In Kim (University of Bath, UNIST Korea): Structuring Community Media Collections, Neural Network Design

– Christian Theobalt with Pascal Fua (EPFL): Marker-less Monocular and Egocentric Pose Estimation

– Christian Theobalt with Carsten Stoll and Ron Mallet (Facebook Reality Labs): Deformable Scene Capture with Sparse Sensors

– Christian Theobalt with Vincent Lepetit (U. Bordeaux), Andrea Tagliasacchi (Google), Kwang Moo Yi (U. Victoria): Hand Segmentation

– Christian Theobalt with Marcus Magnor (TU Braunschweig): Video-based Reconstruction of People Models

– Rhaleb Zayer with Markus Steinberger (TU Graz): Fine-grained Computational Kernels

### 38.15 Publications

#### Journal articles and book chapters


Conference articles


arXiv papers


39 D5: Databases and Information Systems

39.1 Personnel

Director
Prof. Dr. Gerhard Weikum

Researchers
Dr. Daniel Bär (– July 2017)
Dr. Oana Balalau (January 2018 –)
Dr. Klaus Berberich
Dr. Mario Boley (– September 2018, now Monash University, Australia)
Dr. Luciano Del Corro (– December 2018, now Goldman Sachs, Frankfurt)
Dr. Patrick Ernst (May 2018 –)
Dr. Johannes Hoffart (– December 2018, now Goldman Sachs, Frankfurt)
Dr. Pauli Miettinen (– July 2018, now University of Eastern Finland)
Dr. Koninika Pal (September 2018 –)
Dr. Paramita Mirza
Dr. Simon Razniewski (November 2017 –)
Dr. Rishiraj Saha Roy
Dr. Daria Stepanova (– August 2018, now Bosch Stuttgart)
Dr. Jannik Strötgen (– February 2018, now Bosch Stuttgart)
Dr. Erisa Terolli (May 2018 –)
Dr. Liqiang Wang (February 2019 –)
Dr. Andrew Yates

Research programmers
Dragan Milchevski (– December 2018)

PhD Students
Abdalghani Abujabal (– October 2018)
Hiba Arnaout (February 2018 –)
Asia J. Biega (– February 2019)
Natalia Boldyrev, née Prytkova (– June 2017)
Cuong Xuan Chu
Patrick Ernst (– April 2018)
Mohamed Gad-Elrab
Azin Ghazimatin
Anna Christina Guimaraes
39.2 Visitors

In the time period from April 2017 to April 2019, the following researchers visited our group:

<table>
<thead>
<tr>
<th>Name</th>
<th>Dates</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matei Oltean</td>
<td>27.03.2017–23.08.2017</td>
<td>École polytechnique, Paris-Saclay</td>
</tr>
<tr>
<td></td>
<td>(1 day per month)</td>
<td></td>
</tr>
<tr>
<td>Sean Macavaney</td>
<td>08.05.2017–02.09.2017</td>
<td>Georgetown University</td>
</tr>
<tr>
<td>Thomas Rebele</td>
<td>29.05.2017–02.06.2017</td>
<td>Télécom Paris Tech University</td>
</tr>
<tr>
<td>Manolis Koubarakis</td>
<td>01.06.2017–30.09.2017</td>
<td>National and Kapodistrian</td>
</tr>
<tr>
<td></td>
<td>04.06.2018–20.09.2018</td>
<td>University of Athens</td>
</tr>
<tr>
<td>Jeff Z. Pan</td>
<td>12.06.2017–17.06.2017</td>
<td>University of Aberdeen, UK</td>
</tr>
<tr>
<td></td>
<td>03.09.2018–07.09.2018</td>
<td></td>
</tr>
<tr>
<td>Dora Erdos</td>
<td>21.06.2017</td>
<td>Boston University</td>
</tr>
</tbody>
</table>
Stephan Günnewamm 05.07.2017–06.07.2017 TU München
Demetrios 01.08.2016–16.08.2017 University of Cyprus
Zeinalipour-Yazti
Marco Ponza 01.08.2017–15.02.2018 University of Pisa
Jianqiu Xu 24.08.2017–25.08.2017 Nanjing University of Aeronautics and Astronautics
Zhen Jia 25.08.2016–31.08.2017 SouthWest Jiaotong University
18.01.2018–28.02.2018
18.01.2019–28.02.2019
Jyotsna Singh 01.09.2017–30.06.2018 IIT Kharagpur
Marc Spaniol 15.10.2017–17.10.2017 Université de Caen Basse-Normandie
Rakesh Agrawal 13.11.2017 EPF Lausanne
Martin Theobald 01.12.2017 Université du Luxembourg
Lukas Faber 20.12.2017 Hasso-Plattner-Institut
Bernhard Häupler 13.02.2018–16.02.2018 CMU
Mahak Sarin 01.03.2018–30.06.2018 IIT Ropar
(1 day every second month)
Nicole Immorlica 02.05.2018–04.05.2018 Microsoft Research New England
Thomas Rothvoss 14.05.2018–17.05.2018 University of Washington
Niloy Ganguly 28.05.2018 IIT Kharagpur
Dennis Diefenbach 06.07.2018 University of Saint-Etienne
Tanmoy Chakraborty 20.07.2018 IIT Delhi
Liqiang Wang 05.10.2016–30.09.2018 Shandong University
Soumajit Pramanik 02.11.2018–31.10.2019 IIT Kharagpur
Tim Baldwin 05.11.2018 University of Melbourne
Rui Zhang 19.11.2018–30.11.2018 University of Melbourne
Pauli Miettinen 06.12.2018–07.12.2018 University of Eastern Finland
Archit Sakhadeo 27.08.2018–21.12.2018 Pune Institute of Computer Technology
Bayu Distiawan Trisedya 03.10.2018–21.12.2018 University of Melbourne, Australia
Siddhant Arora 02.01.2019–04.05.2019 IIT Delhi
Vaibhav Agrawal 29.01.2019–15.07.2019 BITS, India
Samarth Mehrotra 15.02.2019–15.07.2019 BITS, India
Yohan Chalier 01.03.2019–31.08.2019 Télécom Paris Tech


39.3 Group Organization

D5 was established in 2003. It is headed by Gerhard Weikum and, as of March 2019, 21 doctoral students 4 senior researchers (Klaus Berberich, Simon Razniewski, Jilles Vreeken and Andrew Yates) and 6 other post-doctoral researchers. In addition, 3 senior researchers (Pauli Miettinen, Daria Stepanova, Jannik Strötgen) contributed to the results during the timeframe 2017–2019, but have left the institute. Our research is currently organized into five technical areas:

- Knowledge Base Construction and Curation (coordinated by Simon Razniewski and Daria Stepanova (until August 2018))
- Data Mining and Exploratory Data Analysis (coordinated by Jilles Vreeken and Pauli Miettinen (until July 2018))
- Information Retrieval, Information Extraction, and Text Analysis (coordinated by Andrew Yates and Klaus Berberich)
- Natural Language Understanding (coordinated by Rishiraj Saha Roy, Paramita Mirza and Jannik Strötgen (until February 2018))
- Privacy, Trust and Fairness (coordinated by Gerhard Weikum)

39.4 Knowledge Base Construction and Curation

Coordinators: Simon Razniewski and Daria Stepanova

Enhancing computers with “machine knowledge” that can power intelligent applications is a long-standing goal of computer science [4]. Major advances on knowledge harvesting – methods for turning noisy Internet content into crisp knowledge structures on entities and relations – have made this formerly elusive vision practically viable today.

Automatically constructed knowledge bases (KBs) are a powerful asset for search, analytics, recommendations and data integration, with intensive use at big industrial stakeholders. Examples are the knowledge graphs for search engines (e.g., Google, Bing, Baidu) and social networks (e.g., Facebook), as well as domain-specific KB’s (e.g., Bloomberg, Walmart). These achievements are rooted in academic research and community projects. Our long-term research on the Yago KB has been a trendsetter for this strategic direction. Today, the largest general-purpose KB’s with publicly accessible contents are BabelNet (babelnet.org), DBpedia (dbpedia.org), Wikidata (wikidata.org) and Yago (yago-knowledge.org). They contain millions of entities, organized in hundreds to hundred thousands of semantic classes, and billions of relational facts on entities. These and other knowledge and data resources are interlinked at the entity level, forming the Web of Linked Open Data.

We continue maintaining and further improving the Yago knowledge base; see Sect. 39.4.1. However, the focus of our research on knowledge bases is shifting to novel topics. Despite their wealth of facts, none of the major KB’s can ever be complete. KB’s have been constructed and are maintained with focus on encyclopedic knowledge about prominent and business-relevant entities, and often with strong reliance on Wikipedia. This way of knowledge acquisition misses out on a number of important dimensions, posing open challenges for next-generation KB’s. Our recent and ongoing research on knowledge bases aims to close these gaps.
**Commonsense Knowledge:** Automatically constructed KBs have mostly focused on harvesting encyclopedic fact knowledge. However, for semantic search and other intelligent applications (e.g., conversational bots in social media), machines need a broader understanding of the world: properties of everyday objects, human activities, plausibility invariants and more. This calls for the goal of distilling commonsense knowledge from Internet sources: properties of objects like size, color, shape, parts or substance of which an object is made of, etc., and knowledge on which objects are used for which activities as well as when and where certain activities typically happen. For example, a rock concert involves musicians, instruments – almost always including drums and guitars, speakers, a microphone for the singer; the typical location is a stage, and so on. Research endeavors on acquiring commonsense include ConceptNet [5] and our own project WebChild [6]. However, it is still a long way to go for computers to learn what every child knows.

**Fictional Domain Knowledge:** Fiction and fantasy are a core part of human culture, spanning from traditional literature to movies, TV series and video games. Well known fictional domains are, for instance, the Greek mythology, the Mahabharata, Tolkien’s Middle-earth, the world of Harry Potter, or the Simpsons. These universes contain many hundreds or even thousands of entities and types, and are subject of search-engine queries – by fans as well as cultural analysts. For example, fans may query about Muggles who are students of the House of Gryffindor (within the Harry Potter universe). Analysts and researchers in humanities may be interested in understanding character relationships [3], learning story patterns or investigating gender bias in different cultures [1]. Thus, organizing entities and classes from fictional domains into clean taxonomies is of great value.

**Emerging and Dynamic Knowledge:** Attribute values of entities (e.g., city populations) and relationships between entities (e.g., the CEO of a company or a person’s spouse) change over time. New entities of interest are created all the time and need to be added to the KB (e.g., new songs, sports matches, babies of celebrities). Existing entities may be irrelevant for a KB at some point, but become prominent at a later point. So KBs must be continuously updated, and recompiled from novel and sparse text sources.

**Incomplete Knowledge:** KBs have been built in an opportunistic manner, mostly relying on Wikipedia. If Wikipedia does not have the information or if a fact is stated only in sophisticated form in the article’s text, all the KBs miss out on it. For example, what is notable about particular singers, such as Joan Baez or Patty Smith? In addition to details of their songs, albums, record companies, etc., there are salient facts such as Patty Smith singing on behalf of Bob Dylan at the Nobel Prize Ceremony in December 2016 or Joan Baez having written the song “Diamonds and Rust” about Bob Dylan. All the KB’s miss out on these. Part of the problem is the poor coverage of predicate types: KBs lack predicates like fillInFor and songIsAbout. Even for standard predicates, there are often incompleteness issues, though. For example, even the largest and seemingly comprehensive KB’s list some children or ex-souses of famous people, but not all of them [2].
Personal Knowledge: Internet users commonly have their personal data spread over several devices and services. This includes emails, messages, contact lists, calendars, location histories, and many other kinds of digital traces. However, commercial systems often function as data traps, where it is easy to check information in but difficult to query and control (e.g., for revocation). We aim to put the user back in control of her own data. We introduce a novel framework that extracts and enriches personal information from different sources into a single knowledge base (KB) that lives on the user’s machine, a machine she controls.

Rules from Knowledge Bases and Text: Rules over a Knowledge Graph (KG) capture interpretable patterns in data and various methods for rule learning have been proposed. Since KGs are inherently incomplete, rules can be used to deduce missing facts. Statistical measures for learned rules such as confidence reflect rule quality well when the KG is reasonably complete; however, these measures might be misleading otherwise. We propose a rule learning method that utilizes probabilistic representations of missing facts. In particular, we propose to iteratively extend rules induced from a KG by relying on feedback from a precomputed embedding model over the KG and external information sources including text corpora.

Explainable Fact Checking: Prior methods for automatic fact checking focus on producing final truthfulness scores, which are hard to interpret for humans. Even methods that provide some meta-evidence for these scores solely yield explicit occurrences of facts in text. We moved forward towards deriving more human-understandable evidence based on background knowledge in the form of rules. Our framework combines evidence from both knowledge graphs and text sources.

Visual Knowledge: The social media explosion has populated the Internet with a wealth of images. Modern computer vision has been greatly advanced on making sense of this visual content, based on deep learning and large amounts of labeled data (e.g., for classifying objects in images, generating captions, visual question answering, etc.). However, learning-based tagging still yields noisy labels and is largely restricted to concrete objects, often missing out on generalizations and abstractions. We propose to address these limitations by semantically refining and expanding the labels suggested by learning-based object detection, combining visual cues with commonsense knowledge.

The following subsections present our recent contributions on these research directions. Section 39.4.1 discusses YAGO. Section 39.4.2 discusses commonsense knowledge. Section 39.4.3 discusses KBs for fictional domains. Section 39.4.4 deals with KBs on emerging content. Section 39.4.5 discusses KB recall and completeness (see also Section 39.7.4 for related research on information extraction for counting quantifiers). Section presents a new line of work on personal knowledge. Sections 39.4.7 and 39.4.8 discuss the usage of rules for knowledge discovery and fact checking. Last but not least, Sect. 39.4.9 presents our research on knowledge-based visual tag refinement.
References


39.4.1 YAGO Knowledge Base

Investigators: Johannes Hoffart and Gerhard Weikum in cooperation with Thomas Rebele (Telecom ParisTech University) and Fabian Suchanek (Telecom ParisTech University)

Our strategic endeavor on knowledge harvesting has its roots in research on semantic search starting in 2004. Later it became the Yago-Naga project, and led to the first release of the Yago KB in February 2007. The salient strength of Yago is its rich type system with hundred thousands of fine-grained classes, based on carefully integrating WordNet synsets with Wikipedia categories [6]. When IBM Watson won the Jeopardy quiz show, it harnessed Yago’s taxonomy for semantic type checking [2]. It is the only publicly available KB with statistical quality assurance: at least 95% accuracy (i.e., correct triples) based on sampling and Wilson confidence intervals. It is interlinked, at the entity level, with other knowledge and data resources in the Web of Linked Open Data [1]. The influence and value of Yago has been recognized by the research community through the AIJ Influential Paper Award 2017 (for the 2013 Yago2 paper in the Artificial Intelligence Journal [3]) and the W3C Seoul Test of Time Award 2018 (for the original WWW 2007 paper on Yago [6]).

Recent Yago releases added temporal and spatial knowledge [3] as well as multilingual dimensions [4, 5]. In 2017, the source code of the knowledge extractors has been made available to researchers and developers under the GPL license. Yago is maintained as a joint project of the Max Planck Institute for Informatics and the Télécom ParisTech University. The knowledge base is publicly accessible at http://yago-knowledge.org. Yago is also a key asset for the ambiverseNLU software suite developed in our group (see Section 39.7.1).

1urlhttp://aij.ijcai.org/aij-awards-list-of-previous-winners
2https://www.iw3c2.org/updates/ToT/
39 D5: Databases and Information Systems

References


39.4.2 Commonsense Knowledge

Investigators: Cuong Xuan Chu, Koninika Pal, Simon Razniewski, Julien Romero, Archit Sakhadeo, Gerhard Weikum in cooperation with Gerard de Melo (Rutgers University), Niket Tandon (AI2 Lab in Seattle), Jeff Pan (University of Aberdeen)

With the continued advances in natural language processing and artificial intelligence, the general public is increasingly coming to expect that systems exhibit what may be considered intelligent behavior. While machine learning allows us to learn models exploiting increasingly subtle patterns in data, our systems still lack more abstract, generic forms of commonsense knowledge. Examples of such knowledge include the fact that fire causes heat, the property of ice being cold, as well as relationships such as that a bicycle is generally slower than a car. Previous work in this area has mostly relied on handcrafted or crowdsourced data, consisting of ambiguous assertions, and lacking multimodal data. The work on ConceptNet [3], for instance, relied on crowdsourcing to obtain an important collection of commonsense data. In the context of our WebChild project [4, 5], we developed automated methods for the acquisition of large-scale, semantically organized commonsense knowledge.

**Object Properties:** This work presents a method for automatically constructing a large commonsense knowledge base, called WebChild, from Web contents [5]. WebChild contains triples that connect nouns with adjectives via fine-grained relations like hasShape, hasTaste, evokesEmotion, etc. The arguments of these assertions, nouns and adjectives, are disambiguated by mapping them onto their proper WordNet senses. Our method is based on semi-supervised Label Propagation over graphs of noisy candidate assertions. We automatically derive seeds from WordNet and by pattern matching from Web text collections. The Label Propagation algorithm provides us with domain sets and range sets for 19 different relations, and with confidence-ranked assertions between WordNet senses.
Recently, we have revived this line of research with a new focus on acquiring typical and salient properties of objects and concepts: those that an ordinary human would immediately associate with a target concept (as opposed to collecting all kinds of possible properties that hold for some instances of an object class but are not necessarily typical and not necessarily salient).

**Procedural Knowledge:** This work devised a method for automatically constructing a formal knowledge base on tasks and task-solving steps, by tapping the contents of online communities such as WikiHow [1]. Our methodology first applies Open-IE techniques to WikiHow articles, in order to extract noisy and ambiguous candidates for task and sub-tasks. Subsequently, we use judiciously devised clustering techniques to clean and organize these candidates, and to infer attribute values. To canonicalize tasks and sub-tasks, we leverage word embeddings to distinguish different meanings of the same phrase (e.g., “use keyboard”). The resulting knowledge base, called HowToKB, contains 1.29 Million tasks and sub-tasks organized into a clean taxonomy. Each task or sub-task is represented by a frame with attributes for parent task, preceding sub-task, following sub-task, required tools or other items, and linkage to visual illustrations. The quality of the constructed HowToKB knowledge base has been evaluated by crowdsourcing. For all task attributes, we achieve precision above 80 percent, and well above 90 percent for some attributes. As an extrinsic use case, we developed a technique for searching relevant YouTube videos to tasks in HowToKB. HowToKB datasets are available at [http://www.mpi-inf.mpg.de/yago-naga/webchild/HowToKB](http://www.mpi-inf.mpg.de/yago-naga/webchild/HowToKB)

**Image Descriptions:** Audio description (AD) provides linguistic descriptions of movies and allows visually impaired people to follow a movie along with their peers. Such descriptions are by design mainly visual and thus naturally form an interesting data source for computer vision and computational linguistics. In this work we propose a novel dataset which contains transcribed ADs, which are temporally aligned to full length movies [2]. In addition we also collected and aligned movie scripts used in prior work and compare the two sources of descriptions. We introduce the Large Scale Movie Description Challenge (LSMDC) which contains a parallel corpus of 128,118 sentences aligned to video clips from 200 movies (around

<table>
<thead>
<tr>
<th>Task: paint a bedroom wall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location: bedroom, house</td>
</tr>
<tr>
<td>Time: day, weekend</td>
</tr>
<tr>
<td>Agents: man, husband</td>
</tr>
<tr>
<td>Prop: paint, wall, paint brush</td>
</tr>
<tr>
<td>Parent: decorate the house</td>
</tr>
<tr>
<td>Previous: clean the wall, cover floor</td>
</tr>
<tr>
<td>Next: dry wall, redo floor</td>
</tr>
<tr>
<td>Sub-tasks: hit the edges first with paint, dip the roller into the paint, roll the paint onto the wall</td>
</tr>
</tbody>
</table>

Figure 39.1: Example task frame from HowToKB.
150 h of video in total). The goal of the challenge is to automatically generate descriptions for the movie clips. First we characterize the dataset by benchmarking different approaches for generating video descriptions. Comparing ADs to scripts, we find that ADs are more visual and describe precisely what is shown rather than what should happen according to the scripts created prior to movie production.

The main contribution of this work is the Large Scale Movie Description Challenge (LSMDC) which provides transcribed and aligned AD and script data sentences. The LSMDC was first presented at the Workshop “Describing and Understanding Video & The Large Scale Movie Description Challenge (LSMDC)”, collocated with ICCV 2015. The second edition, LSMDC 2016, was presented at the “Joint Workshop on Storytelling with Images and Videos and Large Scale Movie Description and Understanding Challenge”, collocated with ECCV 2016. Both challenges include the same public and blind test sets with an evaluation server3 for automatic evaluation. This has been joint work with Bernt Schiele’s group (Department D2 at MPI-INF).

References


39.4.3 Fictional Domain Knowledge

Investigators: Cuong Xuan Chu, Simon Razniewski, Gerhard Weikum

Introduction: Fictionitious contents is richly covered by a variety of Internet sources, like community-based descriptions of movies and TV series (hosted by Wikia), summaries and plots or full-fledged scripts of movies and books (e.g., imdb, librarything, shelfari, ...), or entire book series. The underlying domains often form self-contained universes by themselves, examples being Game of Thrones, the Greek mythology, Harry Potter or the Simpsons. As these reflect interesting aspects of human sub-culture (or even historic epochs and cultural contexts), it is worthwhile organizing knowledge about fictitious domains in a formal knowledge base with types, entities and relationships. This could support analysts (e.g., for forecasting the popularity of new computer games) and researchers in the humanities and social sciences. We started this line of research by first tackling the sub-task of organizing entities and their types. Taxonomies, also known as type systems or class subsumption

488
hierarchies, are an important resource for a variety of tasks related to text comprehension, such as information extraction, entity search or question answering, and a core piece in knowledge graphs such as DBpedia, YAGO, Wikidata, and so on. Taxonomy induction is a common problem which has been explored in many works [4, 2, 3], and is a crucial step on information extraction systems. However, while taxonomy construction for encyclopedic knowledge about the real world has received considerable attention already, taxonomy construction for fictional domains is a new problem that comes with specific challenges: (1) prior methods for taxonomy induction intensively leveraged Wikipedia categories. However, the coverage of fiction and fantasy in Wikipedia is very limited, and their categories are fairly ad-hoc, (2) both Wikipedia and other content sources like fan-community forums cover an ad-hoc mixture of in-domain and out-of-domain entities and types. For example, they discuss both the fictional characters (e.g., Lord Voldemort) and the actors of movies (e.g., Ralph Fiennes) and other aspects of the film-making or book-writing, and (3) unsuitability of existing lexicographic methods for inducing type hierarchy, especially dealing with proper names specific to fictional domains.

**Methodology:** We developed the first taxonomy construction method specifically geared for fictional domains, called TiFi [1].

The TiFi architecture is depicted in Figure 39.2. The first step, category cleaning, aims to clean the original set of categories by identifying categories that truly represent classes within the domain of interest, and by removing categories that represent, for instance, meta-categories used for community coordination or concern topics outside of the fictional domain, like movie or video game adoptions, awards etc. We devised a supervised classification approach that combines rules with additional graph-based features. This way, taxonomy construction for a new domain only requires new training examples instead of new rules. Moreover, our experiments show that, to a reasonable extent, models can be transferred across domains.

The second step, edge cleaning, identifies the edges from the original category network that truly represent subcategory relationships. We again pursued a supervised approach, allowing us to combine existing lexical and embedding-based techniques with semantic and novel graph-based features.

For the third step, top-level construction, basic design choices were to construct the top levels of the domain taxonomies from input category networks [4], and to consider also
existing upper-level taxonomies such as WordNet [5].

**Results:** TiFi outperforms the state-of-the-art taxonomy induction system for the first two steps, HEAD [3], by 21-23% and 6-8% percentage points in F1-score, respectively. An extrinsic evaluation based on entity search shows the value that can be derived from our taxonomies, where, for different queries, our taxonomies return answers with 24% higher precision than the input category systems.

**References**


### 39.4.4 On-the-fly Knowledge Base Construction

**Investigators:** Dat Ba Nguyen and Gerhard Weikum in cooperation with Martin Theobald (University of Luxembourg)

Even the largest KBs today (i) are still limited in up-to-date coverage of what happens in the real world, and (ii) miss out on many relevant predicates that precisely capture the wide variety of relationships among entities. To overcome both of these limitations, we propose a novel approach to build on-the-fly knowledge bases in a query-driven manner.

KB construction generally faces an inherent trade-off between precision (i.e., fraction of correct tuples among the acquired ones) and recall (i.e., fraction of correct tuples among the ones that could possibly be acquired from the input). In traditional KB construction, the priority is usually precision, since large KBs (e.g., commercial knowledge graphs, DBpedia, Yago, Wikidata, Freebase, etc.) are an infrastructure asset meant to support a wide variety of applications. In contrast, on-the-fly KB building is intended to support analysts in ad-hoc exploration and querying. Therefore, recall is the primary priority, and good precision is a secondary goal within this regime.

In this work we have developed QKBfly, a novel system for constructing query-driven, on-the-fly KBs [2]. Based on our experience with various IE tasks, our focus in this work is to develop an end-to-end system for KB construction, which may be triggered by an ad-hoc user query (e.g., when an analyst or journalist becomes interested in a particular person, organization or event).

The goal of dynamic and broader construction of KBs has received substantial attention in the database research community recently. The DeepDive project [3] has developed a
highly versatile tool suite for information extraction (IE) and KB population (KBP), based on Markov Logic and further techniques, including a variety of optimizations. Another ground-breaking project in this space is SystemT [1], which uses declarative rules for IE in a wide range of applications, including enterprise content analytics. However, these prior works still require a specification of which predicates are of interest to the IE/KBP process. Unless predicates like has adopted child, filed divorce from or plays role in are made explicit by the application architect (or “knowledge engineer”), they will not be discovered automatically.

The system takes as input an entity-centric query or a natural language question, automatically retrieves relevant source documents (via Wikipedia and news sources), runs a novel form of knowledge extraction on the sources, and builds a high-coverage KB that is focused on the entities of interest. Compared to mainstream KBs, we acquire facts for a much larger set of predicates. Compared to Open IE methods, arguments of facts are canonicalized, thus referring to unique entities with semantically typed predicates which are derived from precomputed clusters of phrases. Besides supporting analytical queries, QKBfly thus also facilitates the application of current question-answering (QA) frameworks, which increasingly rely on structured knowledge backends, to currently popular events and queries. At the heart of QKBfly is a semantic-graph representation of sentences that captures per-sentence clauses, nounphrases, pronouns, as well as their syntactic and semantic dependencies. Based on this graph, we devise an efficient inference technique that performs three key IE tasks, namely named-entity disambiguation, entity co-reference resolution and relation extraction, in a light-weight and integrated manner. Because of the clause-based representation of sentences, QKBfly is not limited to binary predicates but can also extract ternary (or higher-arity) predicates.

References


39.4.5 Knowledge Base Completeness

Investigators: Hiba Arnout, Shrestha Ghosh, Paramita Mirza and Simon Razniewski in cooperation with Fariz Darari (University of Indonesia), Nitisha Jain (Hasso Plattner Institute Potsdam), Werner Nutt (Free University of Bozen-Bolzano)

Despite their usefulness, KBs have major gaps and limitations in what they cover, thus posing the challenge of detecting and resolving these “unknown unknowns”. In this research we use various approaches to mapping knowledge base recall, including 1) relative recall measures, 2) logical reasoning, and 3) the compilation of recall metadata during information extraction [4].

Relative Recall Measures: In this work [1], we present an approach towards measuring Relative Completeness in Wikidata by comparison with data present for similar entities. The term completeness ideally defines if the knowledge base captures all known information about an item in the form of statements. Intuitively, a boolean parameter should be sufficient to indicate if the item is complete or not. For a defined domain unlike Wikidata this may be feasible. Also, unlike for properties, for which it is largely possible to indicate whether they contain all values relevant to the item, representing entity completeness by a Boolean parameter would convey little information as to the quality of the entity. Labeling an entity that has a single property as incomplete and an entity with over 100 properties as complete does not provide much specific information on their quality. Also identifying the complete entities is a rather hard and infeasible approach because of the fact that certain properties in Wikidata capture information that is bound to change over time, e.g., the property medical condition for items of type human. To quantify completeness under these circumstances, we propose the use of a relative notion of completeness: capturing recall in comparison with other, similar subjects. For example, to assess the completeness of data about Trump, one should look at the KB contents for other US presidents, and for assessing the completeness of a city, one should look at the data for similar cities. Formally, relative completeness relies on two components: (1) a similarity function between subject pairs $\text{sim}(S_1, S_2)$ that can be used to compute a (weighted) set of similar subjects $S$ for a subject $S$; (2) a scoring function $\text{score}(S, S)$ that computes a score or rank for the completeness of $S$ with regard to a set of comparison subjects $S$.

This relative completeness approach is easily scalable with the introduction of new classes in the knowledge base, and has been implemented for all available entities in Wikidata. The results provide an intuition on the completeness of an entity comparing it with other similar entities.

Reasoning about Completeness: The Semantic Web is commonly interpreted under the open-world assumption meaning that information available (e.g., in a data source) only captures a subset of the reality. Therefore, there is no certainty about whether the available information provides a complete representation of the reality. The broad aim of this paper is to contribute a formal study of how to describe the completeness of parts of the Semantic Web stored in RDF data sources. We introduce a theoretical framework allowing to augment RDF data sources with statements, also expressed in RDF, about their completeness. One
immediate benefit of this framework is that now query answers can be complemented with information about their completeness. We study the impact of completeness statements on the complexity of query answering by considering different fragments of the SPARQL language, including the RDFS entailment regime, and the federated scenario. We implement an efficient method for reasoning about query completeness and provide an experimental evaluation in the presence of large sets of completeness statements [2].

Recall-aware Information Extraction: Information extraction from text, IE for short, is the backbone of automated knowledge base construction. IE usually comes with precision estimates; however, it lacks awareness of recall.

In this work we present RecallIE, a methodology for estimating the recall of textual information extraction [3]. RecallIE uses distant supervision to estimate from textual features whether a text segment contains all objects for a given subject-predicate pair, e.g., whether a given text mentions all children of Trump. For an experimental study, we seed RecallIE with fact counts for 5 Wikidata relations as ground truth. We approximate the ideal output of an IE method via the combination of open information extraction (Open IE) and object label matching. Note that this specific choice is not decisive for our approach and merely serves as a concrete instantiation of our framework.

Using the fact counts from Wikidata, we train and evaluate on Wikipedia-extracted and Web-extracted sentences and paragraphs. To build the classifiers, we leverage the state-of-the-art methods for text classification, namely feature-based Support Vector Machines (SVMs) as well as neural-based Long Short-Term Memory networks (LSTMs) Our preliminary findings indicate that recall estimation is generally feasible and yields informative assessments.

References

39.4.6 Personal Knowledge Bases

Investigators: Ghazaleh Haratinezhad Torbati, Johannes Hoffart, Dragan Milchevski, Paramita Mirza, Gerhard Weikum, Andrew Yates

The wide variety of online information sources opens up the need for personalizing search and recommendation tasks by finding the most relevant and interesting content for each user. On the other hand, personal information is scattered over various resources such as user’s desktop, emails, cloud storage, calendars, social media, user’s online behaviour like search and clicks as well as data from the user’s phone like location history. In order to make use of the user’s data to the fullest, our vision is to automatically create a Personal Knowledge Base (PKB) from the user’s data sources.

A Personal KB should contain a person’s long-term information such as age, profession, gender, family members, hobbies and interests, as well as, short-term information such as daily agenda and person’s mood and emotions. In addition to the personalized search and recommendation using a PKB, its creating will also allow a person to inspect and control their information over different sources.

Prior work in this direction has started with Personal Information Management (PIM) and Personal Dataspace Management systems. The core idea in these systems is to facilitate the means of controlling, searching and browsing the personal data. One of the recent works in this area is Thymeflow [3]. Thymeflow integrates personal information into a single whole. It represents the personal knowledge in SPO format with a static ontology. Although its knowledge does not go further than contacts, emails, events and some other explicit

---

Figure 39.5: Simplified pipeline of the RecallIE approach.

information, it can be a starting point to our research. Another work in this direction is DigitalSelf [2] which addresses the problem in a different manner. Instead of creating a fixed personal schema they designed a conceptual data model based on human’s episodic memory. They consider “w5h” questions of “What, Who, When, Where, Why and How” to model’s human’s retrieval of events in the memory. In their work, they create scripts to extract the events from the user’s data.

In addition to these prior works, there have been new efforts on creating a decentralized Web. The most notable ongoing project is Solid, led by Tim Berners-Lee. The vision here is to give users full control over their personal data, such that they can choose where to store their own data, who can access it, when and where it can be revoked etc. In comparison to the Solid world, our vision will be closer to reality in two directions. First, creating the PKB would be easier when people are the true owners of their data (rather than storing everything in a Solid mediator). Second, the real power of personal data can be leveraged for personalized services in a flexible way, without disclosing all data to a mediator or one of the big Internet stakeholders. Moreover, our project also empowers Solid’s point of view, by giving the user an overview of their personal data over different sources.

The knowledge representation for a PKB is one of the challenges on this theme. We pursue an approach based on a hybrid model consisting of explicit knowledge in the form of SPO (Subject, Predicate, Object) triples and implicit knowledge in the form of latent embeddings. This entails various challenges. One is the personal ontology which can be dynamic depending on the user, meaning that the extracted information may differ from one person to another. The other interesting challenge is finding a way to migrate from implicit to explicit knowledge, i.e. deriving SPOs from embeddings. Following our earlier work in concept and named entity identification and linking [1], we find the entities (including named entities and concepts) in the person’s textual data like posts, emails and conversations. Using the extracted entities, we can find the user’s interests or the domain of the user’s interests. In the figure 39.6 part (a), concepts such as “play video games” and “watch movies and series” and entities such as “Diablo III” and “Planet Coaster” capture the user’s leisure activities.

One of the research obstacles in this field is obtaining personal data from real users. To this end, we have conducted a user study. We called for participation among university students and recruited 10 participants. We asked for their personal information, including demographic information such as age, nationality and family status, their hobbies i.e. how they spend their free time, their interests like favourite books and book genres and also some more abstract questions to capture their implicit profiles for example “what they value in life”. The participants were given the option to only answer the questions they wanted to, which also ensure the trueness of their responses. In addition to the questionnaires, we gathered conversations between different pairs of participants in order to learn their language style and additional explicit and implicit knowledge about them from the utterances. Figure 39.6 shows a sample of user conversations. In (a) there are explicit cues about the hobbies of the user which can be extracted and used in advanced search services for games or movies. In (b) there is implicit descriptions of the user’s future and the previous job which extraction is more difficult than the information in (a). However, if extracted, can be useful in employment.

3https://solid.mit.edu/
what do you like to do when you are not studying or working??

I like to sometimes explore saarland, or browse the internet or play video games What about you?

I watch movies or series, and I also play videogames

Oh, what video games do you like?

from Blizzard I love Diablo III, from Steam I love How to survive, 7 days to died, Planet coster

Pretty cool, so after finishing, you work at a school? Or at the education ministry?...

actually I want to work creating apps for improve learning

I was working in something like that in Mexico

I created the content and the activities of e-books that were used in elementary school

That's pretty amazing

(a) (b)

Figure 39.6: Sample of chats from our user study.

recommendation and searches. We can observe that the conversational data is a valuable source for capturing different personal knowledge and it is a more natural way of acquiring personal interests instead of asking the user to give them explicitly and without any context. Of course, for the whole process to be transparent, the user should be able to control their extracted information.

Currently, we are pursuing our vision by running different experiments on the data that we acquired. We are measuring the accuracy of automatically extracted personal information from conversations. We also study the PKB benefit by investigating downstream tasks with and without personalization.

References


39.4.7 Rules from Knowledge Bases and Text

Investigators: Daria Stepanova, Mohamed Gad-Elrab, Vinh Thinh Ho, Simon Razniewski and Paramita Mirza in cooperation with Evgeny Kharlamov (Oxford University), Thomas Pellisier Tanon (Telecom ParisTech University)

Rules over a knowledge base are of the form head ← body, where head is a binary atom and body is a conjunction of, possibly negated, binary or unary atoms. For example, the rule livesIn(X, Y) ← marriedTo(X, Z), livesIn(Z, Y) states that married people typically live together. Rules capture interpretable patterns in data, and they can be used for deducing missing facts as well as correcting errors in KBs. Various methods for rule learning have been proposed (e.g., see [5] for an overview).

Statistical measures for learned rules such as confidence (i.e., fraction of facts predicted by the rule that are indeed true in the KB) are used for rule ranking, and they reflect the rule quality well when the KB is reasonably complete. However, as discussed in Section 39.4.5 the majority of modern KBs are inherently incomplete, and therefore confidence as well as other measures might be misleading when used for evaluating rules over them.

For example, a KB that stores a lot of information about families of popular scientists but lacks data in other domains, would yield a heavily biased rule r1 : hasChild(X, Y) ← worksAt(X, Z), educated(Y, Z), stating that workers of certain institutions often have children among the people educated there, as this is frequently the case for scientific families. This rule is obviously biased and when being applied to the data might result in incorrect rule predictions.

So it is difficult to learn high-quality rules from an incomplete KB alone, and scalability dictates that only a small set of candidate rules should be generated. Therefore, the ranking and pruning of candidate rules is a major problem, which we have studied in several works [4, 3, 2, 1, 5].

Completeness-aware rule learning from knowledge bases

Recently, efforts have been put into detecting the concrete numbers of facts of certain types that hold in the real world (e.g., “Einstein has 3 children”) by exploiting Web extraction and crowd-sourcing methods. Such meta-data provides a lot of hints about the structure of KBs, and reveals parts that should be especially targeted by rule learning methods.

In [4, 3] we have proposed to make use of such external meta-information. More specifically, we have defined the problem of learning rules from incomplete KBs enriched with the exact numbers of missing relations of certain types, and proposed several novel rule ranking measures that effectively exploit the meta-knowledge about complete and incomplete KB parts: completeness confidence, completeness precision, recall and the (weighted) directional metric. Our measures have been injected in the rule learning prototype CARL\(^4\) (Completeness-Aware Rule Learning) and evaluated on real-world and synthetic KBs, demonstrating significant improvements both with respect to the quality of mined rules and predictions they produce. To further facilitate our approach, we have also developed a method for learning relation count information by extracting rules like “If a person has more than 2 siblings, then his parents are likely to have more than 3 children”.

\(^4\)https://github.com/Tpt/CARL
Rule learning from knowledge bases guided by embedding models

The problem of KB incompleteness has been tackled by methods that (learn to) predict missing facts for KBs (or actually missing relations between existing entities). A prominent class of approaches is statistics-based and includes tensor factorization or neural-embedding-based models. Intuitively, these approaches turn a KB, possibly augmented with external sources such as text into a probabilistic representation of its entities and relations, known as embeddings, and then predict the likelihood of missing facts by reasoning over the embeddings.

These kinds of embeddings can complement the given KB and are a potential asset in overcoming the limitations that arise due to KB incompleteness. In [2, 1] we exploited this observation and have proposed a method for learning rules from KBs that dynamically exploits feedback from a precomputed embedding model. Our method has been implemented within the RuLES system prototype (Rule Learning with Embedding Support), and its overview is presented in Figure 39.7, where arrows depict information flow between blocks.

In the input RuLES obtains a KB $G$, possibly a text corpus, and a set of user specified parameters for terminating rule construction. The KB and text corpus are used to train the embedding model that in turn is used to produce the probabilistic function $f$. The Rule Learning block constructs rules over the input KB, Rule Evaluation supplies it with quality scores $\mu$ for rules $r$, using both $G$ and $f$.

We have evaluated our approach with various embedding models on real-world datasets and observed significant improvements over state-of-the-art rule learning systems. Importantly, our approach is general in that any embedding model can be utilized including text-enhanced ones, which indirectly allows us to harness unstructured web sources for rule learning.

In Table 39.1 we provide some examples of rules learned by our system from a fragment of the Wikidata knowledge base.

Table 39.1: Example rules generated by RuLES on Wikidata KB.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r^1$:</td>
<td>$\text{nationality}(X, Y) \leftarrow \text{graduated_from}(X, Z), \text{in_country}(Z, Y), \text{not_research_uni}(Z)$</td>
</tr>
<tr>
<td>$r^2$:</td>
<td>$\text{scriptwriter_of}(X, Y) \leftarrow \text{preceded_by}(X, Z), \text{scriptwriter_of}(Z, Y), \text{not_tv_series}(Z)$</td>
</tr>
<tr>
<td>$r^3$:</td>
<td>$\text{noble_family}(X, Y) \leftarrow \text{spouse}(X, Z), \text{noble_family}(Z, Y), \text{not_chinese_dynasties}(Y)$</td>
</tr>
</tbody>
</table>

Figure 39.7: RuLES system overview.

https://github.com/hovinhthinh/RuLES
39.4.8 Explainable Fact Checking

Investigators: Mohamed Gad-Elrab and Daria Stepanova in cooperation with Jacopo Urbani (Vrije Universiteit Amsterdam)

Validating Knowledge Graphs Facts: Knowledge graphs grow continuously; new fact are added via automatic information extraction, crowdsourcing, or methods for KG completion which can either utilize rules to infer missing facts [6, 3] or are based on statistical predictions over learned embeddings [4]. Therefore, KGs contain doubtful if not incorrect triples.

Traditionally, fact-checking has been performed manually by human reviewers which is time-consuming. Therefore, with the increase of false facts on the Web, the automation of fact-checking is gaining more attention. Methods for automatic fact-checking proceed in two steps. First, they perform fact-spotting by searching for occurrences of a fact candidate, such as \( \langle \text{Sadiq Khan citizenOf UK} \rangle \), and possible alternatives, such as \( \langle \text{Sadiq Khan citizenOf Pakistan} \rangle \), in the Web sources. This is done by expanding the predicate into paraphrases (e.g. “has nationality”, “has passport”) and searching for it jointly with the alias names of the subject and object arguments. Then, the extracted evidence (or counter-evidence) is used to infer the truth value of the candidate fact.

Numerical scores produced by fully automated methods are not adequate whenever the final decision is made by KG curators. For humans, such scores are hard to understand or justify without explanations. Some approaches attempt to show the sources used in computing the scores as an explanation. Yet, the collected syntactic clues using fact-spotting are often not sufficient since textual sources are incomplete and biased in what is stated explicitly. For instance, the citizenship of London’s mayor \textit{Sadiq Khan} would rarely be mentioned.
In addition, some predicates (e. g. influencedBy) are ambiguous, and their interpretation is domain-specific.

**ExFaKT:** In [1] we introduced a novel framework for finding semantically related evidence in Web sources and the underlying KG, and for computing human-comprehensible explanations for facts. Extracted explanations can be consumed later by KG curators to judge the correctness of candidate facts. We refer to our framework, as **ExFaKT** (Explaining Facts over KGs and Text resources).

The key for detecting semantic evidence is intensional background knowledge in the form of rules. For example, the rule

\[ \text{citizenOf}(X, Y) \leftarrow \text{mayorOf}(X, Z), \text{locatedIn}(Z, Y) \]

intuitively states that mayors of cities are normally citizens of countries where these cities are located. Such rules can be specified by humans or automatically extracted from KGs using rule mining methods [5]. As the latter may fall short of covering all interesting situations, hand-crafted rules proved to be a valuable asset.

As shown in Figure 39.8, given a set of rules and a query for a fact candidate, **ExFaKT** rewrites the query into a set of subqueries. Whenever we find evidence in the KG or text that the body of the rule holds, the credibility of the head increases. This process creates semantic traces that explain, in a human-readable format, why a fact is likely to be true (or potentially false).

**ExFaKT** utilizes rules to guide the decomposition of the query into more frequently stated and thus easier-to-spot related facts. This way, we counter the reporting sparseness and bias. Moreover, rules can encode domain-specific knowledge to better cope with ambiguous predicates. Finally, rules combine knowledge from both textual Web sources and the KG. For example, a rule could find the mayors of cities in news articles and look up the countries of cities in the KG.

**Usability of ExFAKT for Humans:** We conducted a user study over Amazon Mechanical Turk to assess the utility of **ExFaKT** for human KG curators. We asked participants to
In ExFaKT, we compared the participants' performance while using ExFaKT explanations with the performance of performing a traditional fact validation. We observed that our method doubled the performance of the participants and enhanced their judgment accuracy as well. Furthermore, in [2] we implemented a prototype for ExFaKT with a web interface to capture further usability challenges such as output representation and results ranking.

**References**


### 39.4.9 Knowledge for Visual Content Tagging

*Investigators: Sreyasi Nag Chowdhury, Simon Razniewski and Gerhard Weikum in cooperation with Niket Tandon (Allen Institute of AI), Hakan Ferhatosmanoglu (University of Warwick)*

**Motivation and Problem:** The enormous growth of social media has populated the Internet with a wealth of images. On one hand, this makes image search easier, as there is redundancy for many keywords with informative text surrounding the images. On the other hand, it makes search harder, as there is a huge amount of visual contents that is hardly understood by the search engine. There are two paradigms for searching images: content-based image retrieval (CBIR) and tag-based image retrieval (TBIR).

CBIR finds images similar to a query image based on visual features that are used to represent an image [3]. These features include color, shape, texture, SIFT descriptors etc. Recent advances in deep-learning-based object detection have lifted this approach to a higher level, by assigning object labels to bounding boxes (e.g., [2]). However, these labels are limited to concrete object classes (e.g., *truck, SUV, Toyota Yaris Hybrid 2016*, etc.), often
trained (only) on (subsets of) the ca. 20,000 classes of ImageNet [1]. Thus, they miss out on generalizations (e.g., vehicle) and abstractions (e.g., transportation, traffic jam, rush hour). Figure 39.9 (a) shows the top-confidence visual labels by LSDA [2] for an example case of incorrect labels.

TBIR retrieves images by textual matches between user query and manually assigned image tags (e.g., from collaborative communities such as Flickr). While some of the semantic gap in CBIR is reduced in TBIR, the performance of TBIR often suffers from incomplete and ambiguous tags. Figure 39.9 (b) illustrates this point: there is only a single tag happiness and none for the concrete objects in the image. For the big search engines, one way of overcoming this bottleneck is to exploit query-and-click logs. The query keyword(s) associated with a click can be treated as label(s) for the clicked image. However, this method crucially relies on the labels to appear in (sufficiently many) queries (or, traditionally, salient text surrounding the image). [4] gives a survey on TBIR and tag assignment and refinement.

This work addresses the above limitations by semantically refining and expanding the labels suggested by learning-based object detection. Experiments show that our method, called VISIR, improves the quality of the state-of-the-art visual labeling tools like LSDA [2].

**Approach and Contribution:** We leverage state-of-the-art CBIR by considering the visual tags of an existing object detection tool (LSDA [2] in our experiments) as a starting point. Note that there are multiple labels for each bounding box with varying confidence scores, and our goal is to compute the most informative labels for the entire image. We impose a constrained optimization on these initial labels, in order to enforce their semantic coherence. We also consider labels that are visually similar to the detected ones, to compensate for omissions. In addition, we utilize lexical and commonsense knowledge to generate candidate labels for generalizations (hypernyms from WordNet) and abstractions (concepts from ConceptNet). So we both refine and expand the initial labels. The joint inference on the entire label space is modeled as an optimization problem, solved by an integer linear program (using the Gurobi ILP solver).

Figure 39.10 shows examples for the input labels from the deep-learning-based visual object detection (left column) and the output labels that VISIR (our work) computes (right
column). The labels from LSDA illustrate a clear semantic incoherence for these specific examples. VISIR labels are coherent, adds generalizations (in blue) and abstractions (in green). Incorrect labels are marked red. Although our work aligns with the existing TBIR research on social tagging and tag refinement, there are key differences:

**Granularity:** Our starting point is tags for bounding boxes, whereas user-provided tags refer to an entire image.

**Cardinality:** The number of bounding boxes in an image can be quite large. Moreover, object detectors usually produce a long list of varying-confidence tags for each bounding box.

**Noise:** As a result, many of the visual candidate tags in our approach are of mixed quality, whereas traditional social tagging typically has few but trusted annotations per image.

For these factors, our notion of tag refinement is unlike the one in prior work. Therefore, we

<table>
<thead>
<tr>
<th>LSDA Labels</th>
<th>VISIR labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>allosaurus</td>
<td>person</td>
</tr>
<tr>
<td>loggerhead turtle</td>
<td>guitar</td>
</tr>
<tr>
<td>person</td>
<td><strong>stringed instrument</strong></td>
</tr>
<tr>
<td><strong>bird</strong></td>
<td><strong>self-expression</strong></td>
</tr>
<tr>
<td>bone china</td>
<td>food processor</td>
</tr>
<tr>
<td>stove</td>
<td>bowl</td>
</tr>
<tr>
<td>WC, loo</td>
<td>cup or mug</td>
</tr>
<tr>
<td>cup or mug</td>
<td><strong>utensil</strong></td>
</tr>
<tr>
<td>cucumber</td>
<td>snake</td>
</tr>
<tr>
<td>snake</td>
<td><strong>reptile</strong></td>
</tr>
<tr>
<td>green mamba</td>
<td><strong>slithery</strong></td>
</tr>
<tr>
<td></td>
<td><strong>poisonous</strong></td>
</tr>
<tr>
<td>racket</td>
<td>tennis bat</td>
</tr>
<tr>
<td>person</td>
<td><strong>individual</strong></td>
</tr>
<tr>
<td><strong>bathing cap</strong></td>
<td><strong>play tennis</strong></td>
</tr>
<tr>
<td>tennis ball</td>
<td>tennis ball</td>
</tr>
<tr>
<td>head cabbage</td>
<td></td>
</tr>
</tbody>
</table>

Figure 39.10: Images with labels from LSDA and VISIR
refer to our task as **Visual Tag Refinement**.

**Visual Tag Refinement** can be broken down into three sub-tasks, for which this work provides effective solutions:
1) elimination of incoherent tags among the initial visual labels,
2) expansion of the tag space by adding visually similar tags missed by object detectors, and adding candidate tags for generalization and abstraction,
3) joint inference on the enriched tag space, by integer linear program.

VISIR is a new method for refining and expanding visual labels for images. Its key strengths are cleaning out noisy labels from predictions by object detection tools and adding informative labels that capture generalizations and abstractions. Our model makes this feasible by considering the visual similarity of labels, the semantic coherence across concepts, and various kinds of background knowledge. The joint inference on an enriched label candidate space is performed by means of a judiciously designed Integer Linear Program. Our experiments show the viability of the approach, and also demonstrate significant improvements over two state-of-the-art object detection and tagging tools.

**References**


**39.5 Data Mining and Exploratory Data Analysis**

*Coordinators: Jilles Vreeken and Pauli Miettinen*

In this research area we focus on developing theory and methods that help to answer the question “This is my data, what can you tell me about it?” Our methods should hence give you insights on how your data is structured, such that you can make a more educated guess to what your next step will be, i.e., what type of analysis to run, whether to gather more or different data, or to adjust your hypothesis on how you think the data is distributed.

A key element of exploratory data analysis is that we aim to explore the data, not the assumptions you may have about it. This means we want all our methods to be as free of parameters as possible, as well as free of assumptions of how your data may be distributed. The following subsections give an overview of the research projects that we have been pursuing. We have grouped these projects thematically, going from statements about dependencies, causality, associations, re-descriptions, patterns, to matrix and graph
decompositions, and wrap up with improving the parallelization efficiency of a wide class of learning algorithms.

### 39.5.1 Dependency Discovery

The first theme we cover is that of discovering dependencies, or, in other words, correlation. We used information theory to study the definition of correlation, how to robustly measure and discover approximate functional dependencies, and how to robustly determine conditional independencies in data.

#### Correlation by Compression

*Investigators: Kailash Budhathoki and Jilles Vreeken*

We defined the notion of correlation from the ground up, without having to make any assumptions on the data generating model [1], using Algorithmic Information Theory. Loosely speaking, we measure how correlated two sets of observations \( x \) and \( y \) are by how many bits we save compressing the data together, exploiting the dependencies between their corresponding observations, instead of compressing the observations over \( x \) and \( y \) separately. Our framework is highly general in the sense that it can pick up *any* algorithmic dependency and it naturally extends to multivariate data of *any* kind. That is, we define our score over mathematical objects in general, regardless of whether these are discrete or continuous tabular data, for sequential data, or graphs. We depict the general intuition in Fig. 39.11.

Specifically, we propose two measures, both of which explicitly penalise redundancy to avoid free-rider patterns. That is, if \( x \) and \( y \) are correlated, but independent of \( z \), \( s(x, y) \) should be larger than \( s(x, y, z) \). We show how to instantiate these scores in practice through the Minimum Description Length (MDL) principle using existing lossless compressors. Thorough empirical evaluation on both synthetic and real-world data confirms these measures have very high statistical power, are robust with regard to redundancy, and identify interesting correlations in real-world data. In short, we formalise, for the first time, the connection between correlation and algorithmic information theory. It is important to note that our proposal is *orthogonal* to existing methods. Our framework can be used both through existing compressors, as well as suggests what compressors to study in the future such that we can measure correlation over non-standard data types such as complex multivariate sequences and graphs. By developing better compressors, we more closely approximate Kolmogorov complexity, and hence will be able to better detect richer classes of correlation.

#### Discovering Approximate Functional Dependencies

*Investigators: Panagiotis Mandros, Mario Boley, and Jilles Vreeken*

Given a data sample \( D_n = \{d_1, \ldots, d_n\} \) drawn from the joint distribution \( p \) of some input variables \( \mathcal{I} \) and an output variable \( Y \), it is a fundamental problem in data analysis to find variable subsets \( \mathcal{X} \subseteq \mathcal{I} \) that jointly influence or (approximately) determine \( Y \). The *functional dependency discovery* problem, i.e., to find

\[
\arg \max \{ Q(\mathcal{X}; Y) : \mathcal{X} \subseteq \mathcal{I} \}
\]  

(39.1)
Figure 39.11: **Algorithmic model of correlation.** Let $x$ be a data object representing a series of $n$ observations $a_1, a_2, \ldots, a_n$, and $y$ represents $b_1, b_2, \ldots, b_n$. (left) One way to generate the observations of $x$ and $y$ is by independent shortest programs $P_x$ and $P_y$. (right) The alternative way is by a single shortest program $P_{xy}$. The program $P_{xy}$, however, unlike $P_x$ and $P_y$, additionally uses the information from corresponding observation $b_i$ of $y$ when generating an observation $a_i$ of $x$, and vice versa. Therefore if $x$ and $y$ are correlated, $P_{xy}$ is the most succinct way to generate observations of $x$ and $y$.

Figure 39.12: **Dependency-by-chance.** Estimated fraction of information for variables $X$ of increasing domain size (4 to 2048) to independent $Y$ (domain size 4) for fixed sample size (1000). Estimated dependency increases for naive estimator $\hat{F}$, while the corrected-for-chance estimator $F_0$ accurately estimates population value $F(X; Y) = 0$.

for some real-valued measure $Q$ that assesses the dependence of $Y$ on $X$, is a classic topic in the database community, but also has many other applications including feature selection and knowledge discovery. For instance, finding such dependencies can help identify compact sets of descriptors that capture the underlying structure and actuating mechanisms of complex scientific domains. For categoric input and output variables, the measure $Q$ can be chosen to be the fraction of information defined as

$$F(X; Y) = (H(Y) - H(Y | X))/H(Y) ,$$

where $H(Y) = \sum_{y \in Y} p(y) \log p(y)$ denotes the Shannon entropy. This score represents the relative reduction of uncertainty about $Y$ given $X$. It takes on values between 0 and 1 corresponding to independence and exact functional dependency, respectively.

Estimating the score naively with empirical probabilities $\hat{p}$, however, leads to an overestimation of the actual dependence between $X$ and $Y$, a behavior known as dependency-by-chance. In particular, since the bias is increasing with the domain size of variables, it is unsuitable for dependence discovery where we have to soundly compare different variable sets of varying
dimensionality and consequently of widely varying domain sizes (see Fig. 39.12). The fraction of information can be corrected by subtracting its estimated expected value under the hypothesis of independence. This gives rise to the reliable fraction of information [2] defined as

\[
\hat{F}_0(\mathcal{X}; \mathcal{Y}) = \hat{F}(\mathcal{X}; \mathcal{Y}) - \hat{E}_0(\hat{F}(\mathcal{X}; \mathcal{Y})) ,
\]

(39.2)

where \( \hat{E}_0(\hat{F}(\mathcal{X}; \mathcal{Y})) = \sum_{\sigma \in S_n} \hat{f}(\mathcal{X}; \mathcal{Y}_\sigma) / n! \) is the expected value of \( \hat{F} \) under the permutation model, i.e., under the operation of permuting the empirical \( \mathcal{Y} \) values with a random permutation \( \sigma \in S_n \). This estimator can be computed efficiently in time \( O(nk) \) for \( \mathcal{X} \) with domain size \( k \). Moreover, the maximization problem can be solved effectively by a simple branch-and-bound scheme: the maximally attainable \( \hat{F}_0 \) for supersets of some partial solution \( \mathcal{X} \) can be bounded by the function \( f_{\text{mon}}(\mathcal{X}) = 1 - \hat{E}_0(\hat{F}(\mathcal{X}; \mathcal{Y})) \), which follows from the monotonicity of \( \hat{E}_0(\hat{F}(\cdot; \mathcal{Y})) \) [2]. This, however, turned out to be a rather simplistic bounding function that left room for substantial improvements. We provided [3] the following contributions:

1. We showed that the problem of maximizing the reliable fraction of information is NP-hard. This justifies the usage of worst-case exponential-time algorithms as well as heuristic search methods.

2. Motivated by this insight, we greatly improve the practical performance for both of these optimization styles by deriving a novel admissible bounding function \( f_{\text{spc}}(\mathcal{X}) \). This function is not only tighter than the previously proposed \( f_{\text{mon}}(\mathcal{X}) \) but in particular we have that the supremum of \( f_{\text{mon}}(\mathcal{X}) / f_{\text{spc}}(\mathcal{X}) \)—and thus the potential for additional pruning in search—is unbounded.

3. Finally, we reported extensive empirical results evaluating the proposed bounding function and the various algorithmic strategies. In particular, we consider the approximation ratio of the greedy algorithm and show that in fact, it produces highly competitive results in a fraction of time needed for complete branch-and-bound style search—motivating further investigation of this fact.

For this work, we received the IEEE ICDM 2018 Best Paper Award.

**Testing Algorithmic Conditional Independence by Stochastic Complexity**

*Investigators: Alexander Marx and Jilles Vreeken*

Testing for conditional independence plays a key role in the discovery of Bayesian, and causal networks. If the true probability distribution of the observed data is faithful to the underlying causal graph, conditional independence tests can be used to recover the undirected causal network. In essence, under the faithfulness assumption finding that two random variables \( X \) and \( Y \) are conditionally independent given a set of random variables \( Z \), denoted as \( X \perp \perp Y \mid Z \), implies that there is no direct causal dependency between \( X \) and \( Y \).

We proposed [5, 4] a new robust independence test for discrete data. Given access to the true distribution, conditional mutual information would be the perfect criterium for conditional independence. Estimating \( CMI \) purely from limited observational data leads, however, to discovering spurious dependencies—in fact, it is likely to find no independence
at all. To use CMI in practice, it is therefore necessary to set a threshold. This is not an easy task, as the threshold should depend on the domain sizes of the involved variables as well as the sample size.

The main problem of existing tests is that these struggle to find the right balance for limited data: either they are too restrictive and declare everything as independent or not restrictive enough and do not find any independence. To tackle this problem, we build upon algorithmic conditional independence, which has the advantage that we not only consider the statistical dependence, but also the complexity of the distribution. Although algorithmic independence is not computable, we can instantiate this ideal formulation with stochastic complexity. In essence, we compute stochastic complexity using either factorized or quotient normalized maximum likelihood (fNML and qNML), and formulate SCI, the Stochastic complexity based Conditional Independence criterium.

Importantly, we show that we can reformulate SCI to find a natural threshold for CMI that works very well given limited data and diminishes given enough data. In the limit, we prove that SCI is an asymptotically unbiased and $L_2$ consistent estimator of CMI. For limited data, we find that the qNML threshold considers the sample size as well as the dimensionality of the data. The fNML threshold, however, additionally considers the estimated probability mass functions of the conditioning variables. In practice, as we show in our experiments, this reduces the type II error. Moreover, when applying SCI based on fNML in constraint based causal discovery algorithms, we observe a higher precision and recall than related tests. In addition, in our empirical evaluation SCI shows a sub-linear sample complexity.

References


39.5.2 Causality

Causal inference, telling cause from effect, is perhaps one of the most important problems in science. To make absolute statements about cause and effect, carefully designed experiments
are necessary, in which we consider representative populations, instrument the cause, and control for everything else. In practice, setting up these experiments is very expensive, or even impossible. We hence consider causal inference from observational data. That is, the goal is to infer the most likely direction of causation from data that has not been obtained in a completely controlled manner but is simply available. In recent years large strides have been made in the theory and practice of discovering causal structure from observational data.

**Causal Inference in Discrete Data**

*Investigators: Kailash Budhathoki and Jilles Vreeken*

We proposed a general framework for causal inference on observational data [1, 5]. We based it on the solid foundations of Kolmogorov complexity, and develop a score for pairs of data objects that identifies not only the direction, but also quantifies the strength of causation, all the while being unbiased to the complexities of the individual objects, without making any assumptions on the distribution nor the type of causal relation between the data objects, and without requiring any parameters to be set.

Kolmogorov complexity is not computable, however, and hence we derived practical, computable versions based on Shannon Entropy and the Minimum Description Length (MDL) principle. Our first proposal, *Origo* [1, 5], considers pairs of multivariate binary attributes $X$ and $Y$, and uses decision trees to model dependencies, and a two-part MDL score to determine the causal direction. With *Cisc* [2] we focused further on the case where $X$ and $Y$ are univariate, which allowed us to use a more robust MDL score based on Stochastic Complexity. Finally, with *Acid* [3] we showed that by optimising conditional Shannon entropy through discrete regression we can not only achieve a highly accurate inference framework, but one that is guaranteed to infer the correct direction given sufficient data, no hidden confounders, and an additive noise model data generating process.

Extensive experiments on synthetic, benchmark, and real-world data confirm that these algorithms indeed perform very well in practice. They are highly robust to noise, offer high statistical power, and outperform existing proposals by a wide margin.

**Causal Inference in Event Sequences**

*Investigators: Kailash Budhathoki and Jilles Vreeken*

In many settings we collect discrete-valued time series—event sequences. This raises the question of whether we can determine that is more likely that $x^n$ caused $y^n$, or the other way around, that $y^n$ caused $x^n$. The perhaps most well-known framework for causal inference on time series is Granger causality, which postulates that $x^n$ is likely a cause of $y^n$ if the past values of $x^n$ help to significantly better sequentially predict the values of $y^n$ than we can do with just the data over the past of $y^n$. While this framework makes intuitive sense, to put it to practice it does require us to make assumptions on the generating process, as well as on the measure of predictability.

We took a related, but subtly different approach [4] following the information theoretic causal framework defined above. Simply put, we say that $X$ causes $Y$ if we save more bits
by compressing the data of \( Y \) with additionally the past of \( X \), than vice versa. To optimally compress the data, we would need to know its distribution. In practice, however, we only have observed data and a class of possible prediction strategies. We hence build our framework on the notion of sequential normalized maximum likelihood (SNML), which is a strategy that is guaranteed to give the minimum number of additional bits (regret) compared to the true distribution, regardless of input, and regardless of whether or not the true distribution is in the model class under consideration.

We gave the general theory for causal inference on event sequences using SNML, including a detailed exposition on how to derive our causal indicators for binary event sequences based on the class of bernoulli distributions—from which the extension to multinominal distributions is trivial. Importantly, for discrete data in general, Cute, which stands for causal inference on event sequences, has only a linear time worst case runtime complexity. Results on synthetic data show that unless effect is only time-shifted, it outperforms transfer entropy by a wide margin. Additionally, we consider two case studies on real world data, where we find that Cute with high accuracy reconstructs the ground truth in water elevation levels in two rivers, as well as in discovering excitatory connections in neural spike train data.

**Causal Inference in Continuous-Valued Data**

*Investigators: Alexander Marx and Jilles Vreeken*

While discrete data lends itself naturally to a compression-based approach, perhaps most data of interest is numeric. We hence also studied the problem of inferring the most likely direction between two univariate numeric random variables \( X \) and \( Y \). Traditional methods, that rely on conditional independence tests, cannot decide between the Markov equivalent classes of \( X \rightarrow Y \) and \( Y \rightarrow X \). Recently, it has been postulated however that if \( X \rightarrow Y \), there exists an independence between the marginal distribution of the cause, \( P(X) \), and the conditional distribution of the effect given the cause, \( P(Y \mid X) \). The state of the art exploits this asymmetry in various ways, and overall obtain up to 70% accuracy on a well-known benchmark of cause-effect pairs. We broke this barrier, and give an elegant score that is computable in linear-time and obtains over 82% accuracy on the same benchmark [7, 9].

We base our methods on the algorithmic Markov condition, which states that if \( X \) causes \( Y \), the factorization of the joint distribution \( P(X,Y) \) in the causal direction admits a simpler description—in terms of Kolmogorov complexity—than the anti-causal direction. Kolmogorov complexity is not computable, however, so we need a practical instantiation. In this paper we do so using the Minimum Description Length (MDL) principle. In practice, we propose to fit a regression model from \( X \) to \( Y \), and vice versa, measuring both the complexity of the function, as well as the error it makes in bits, and infer that causal direction by which we can describe the data most succinctly.

We showed that our score is robust against sample size and correlates strongly with accuracy. Moreover, it admits an elegant and effective analytical statistical tests on the difference in score between the two causal directions. A thorough evaluation of SLOPE on synthetic and real data shows our method works remarkably well in practice. To illustrate its strength, we show the results over the above-mentioned benchmark in Fig. 39.13. In particular, we show the accuracy over the top-\( k \) pairs ranked according to the score at hand, the so-called decision rate. The plot shows that SLOPE leads by large margin over
Figure 39.13: [Higher is better] Weighted accuracy of our method, SLOPE, versus the state of the art in causal inference for univariate numeric pairs as identified in a recent survey on the Tuebingen benchmark data set (98 pairs). The gray area indicates the 95% confidence interval of a fair coin toss.

its competitors; it is 100% accurate over the 32 pairs it is most certain about, and is 90% accurate over its top-74 out of 98 pairs. Unlike its competitors, its accuracies are strongly significant with regard to the 95% confidence interval of a fair coin flip. Moreover, our score comes with a natural significance test that allows us to weed out insignificant results; when we consider the 83 significant pairs only we find that SLOPE is even 85% accurate.

Causal Inference in Mixed Type Data

Investigators: Alexander Marx and Jilles Vreeken

In practice, X and Y do not have to be of the same type. The altitude of a location (real-valued), for example, determines whether it is a good habitat (binary) for a mountain hare. In fact, neither X nor Y have to be univariate. Whether or not a location is a good habitat for an animal is not just caused by a single aspect, but by a combination of conditions which are not necessarily of the same type. We are therefore interested in the general case where X and Y may be of any cardinality, and may be single or mixed-type. To the best of our knowledge there did not exist any method for this general setting: all existing methods that consider two variables are only defined for single-type pairs. Additive Noise Models (ANMs), for example, have only been proposed for univariate pairs of real-valued or discrete variables, and similarly so for methods based on the independence of $P(X)$ and $P(Y \mid X)$. Trace-based methods require both X and Y to be strictly multivariate real-valued, and whereas cumulative entropy-based methods also works for univariate pairs, these again have to be real-valued.

In this work [8], we define an MDL score for coding forests, a model class where a model consists of classification and regression trees as this allows us to consider both discrete and continuous-valued data with one unified model. By allowing dependencies from X to Y, or vice versa, we can measure the difference in complexity between $X \rightarrow Y$ and $Y \rightarrow X$. Discovering a single optimal decision tree is already NP-hard, and hence we cannot efficiently discover the coding forest that describes the data most succinctly. We therefore proposed
Crack, an efficient greedy algorithm for discovering good models directly from data. The inferences we make hence are all with respect to the class of coding trees, and the specific encoding we define. We discuss the implications with regard to identifiability, and through extensive empirical evaluation on synthetic, benchmark, and real-world data, we showed that Crack performs very well in practice—even under adversarial settings.

Our main contributions were as follows. We introduce the first framework for inferring the causal direction from univariate and multivariate single and mixed-type data—as opposed to existing methods that are only able to deal with either nominal or numeric data. We propose a new causal indicator based on the algorithmic Markov condition, instantiate it through MDL, and propose a fast algorithm to compute it. We provide extensive empirical evaluation of our method, in which we additionally introduce new multivariate cause-effect pairs with known ground truth.

**Telling Confounded from Causal**

*Investigators: David Kaltenpoth and Jilles Vreeken*

One of the main assumptions in causal inference is that of causal sufficiency. That is, to make sensible statements on the causal relationship between two statistically dependent random variables $X$ and $Y$, it is assumed that there exists no hidden confounder $Z$ that causes both $X$ and $Y$. In practice this assumption is often violated—we seldom know all factors that could be relevant, nor do we measure everything—and hence existing methods are prone to spurious inferences.

We hence studied the problem of inferring whether $X$ and $Y$ are causally related, or, are more likely jointly caused by an unobserved confounding variable $Z$. To do so, we build upon the algorithmic Markov condition (AMC). Simply put, this means that if $Z$ causes both $X$ and $Y$ the complexity of the factorization according to this model, $K(P(Z)) + K(P(X|Z)) + K(P(Y|Z))$, will be lower than the complexity corresponding to the model where $X$ causes $Y$, $K(P(Z)) + K(P(X)) + K(P(Y|X))$. As we obviously do not have access to $P(Z)$, we propose to estimate it using latent factor modelling. Second, as Kolmogorov complexity is not computable, we use the Minimum Description Length (MDL) principle as a well-founded approach to approximate it from above. This is the method that we developed in this paper [6].

In particular, we consider the setting where given a sample over the joint distribution $P(X,Y)$ of continuous-valued univariate or multivariate random variable $X = (X_1,\ldots, X_m)$, and a continuous-valued scalar $Y$. Although it has received little attention so far, we are not the first to study this problem. Recently, Janzing and Schölkopf showed how to measure the “structural strength of confounding” for linear models using resp. spectral analysis and ICA. Rather than implicitly measuring the significance, we explicitly model the hidden confounder $Z$ via probabilistic PCA. While this means our approach is also linear in nature, it gives us the advantage that we can fairly compare the scores for the models $X \rightarrow Y$ and $X \leftarrow Z \rightarrow Y$, allowing us to define a reliable confidence measure.

Through extensive empirical evaluation on synthetic and real-world data, we show that our method, CoCa, short for Confounded-or-Causal, performs well in practice. This includes settings where the modelling assumptions hold, but also in adversarial settings where they do not. We show that CoCa beats both baselines as well as the recent proposals mentioned.
above. Importantly, we observe that our confidence score strongly correlates with accuracy. That is, for those cases where we observe a large difference between the scores for causal resp. confounded, we can trust CoCa to provide highly accurate inferences.

References

- [6] D. Kaltenpoth and J. Vreeken. We are not your real parents: Telling causal from confounded by MDL. In *Proceedings of the 2019 SIAM International Conference on Data Mining (SDM 2019)*, Calgary, Canada, 2019. SIAM. Accepted.

### 39.5.3 Subgroup Discovery in Multivariate Data

Correlation and causal analysis can reveal important information about the data, such that the survival rate of some disease correlates, or is caused by with the sex of the patient. Any patient would rather want to know what this means for them, i.e. what is the association between their sex and survivability. Discovering interesting associations in data is the second an important research topic theme of our group, instantiated through our research on subgroup discovery, the discovery of descriptors (patterns, queries) that single out subgroups of the data (e.g. female patients) that stand out with regard to the overall distribution of the variable of interest (e.g. female patients have a higher chance of surviving a certain medical procedure).
Discovering Structure-Property Relationships of Materials by Subgroup Discovery

Investigators: Mario Boley and Jilles Vreeken in cooperation with Brian Goldsmith, Luca Ghiringhelli and Matthias Scheffler (FHI Berlin)

Big-data analytics applied to materials-science data often focuses on the inference of a global prediction model for some property of interest for a given class of materials. However, the underlying mechanism for some target property could differ for different materials within a large pool of materials-science data. Consequently, a global model fitted to the entire dataset may be difficult to interpret and may well hide or incorrectly describe the actuating physical mechanisms. In these situations, local models describing subgroups would be advantageous to global models. For illustration, a globally optimal regression model could predict a negative relationship between two material properties, whereas among subgroups there exists a positive relationship. As a more physical example, the transition metals of the periodic table are a subgroup, and the actinides, lanthanides, and halogens are other subgroups. Thus, identification of subgroups is useful to gain an understanding of similarities and differences between materials.

We demonstrated how through subgroup discovery we can identify and describe local patterns, correlations, and descriptors in materials-science data according to some desired target property (or properties). At first, we begin by formulating the subgroup discovery algorithm for materials-science applications. Next, we demonstrate that subgroup discovery can identify physically meaningful models that classify the crystal structures of 82 octet binary (OB) semiconductors as either rocksalt (RS) or zincblende (ZB) from only information of its chemical composition. The OB compounds have long been studied, and we consider it an exemplary dataset for the demonstration of subgroup discovery to find descriptors of materials. Notably, subgroup discovery helps us to find a two-dimensional model derived from only the atomic radii of valence s and p orbitals that properly classify the crystal structures for 79 of the 82 OB semiconductors. Subsequently, we applied subgroup discovery to 24,400 configurations of neutral gas-phase gold clusters with 5–14 atoms. Small gold clusters have different physical and chemical properties than their bulk counterpart, and they exhibit a diverse array of physicochemical properties depending on their size and shape. The aim of investigating gold clusters here is two-fold: (1) to search for general structure–property relationships holding across gold clusters of different sizes and vastly different configurations; and (2) to demonstrate the versatility of subgroup discovery on a large and heterogeneous dataset. It is established that subgroup discovery can help identify unexpected and general, size-independent, patterns within the dataset of gold cluster configurations.

This work is the first publication [3] following from an intense collaboration between our group and that of the Theory group of the Fritz Haber Institute of the Max Planck Society, in which we are applying data mining techniques to discover structure in material science data. Another direct result from this collaboration is the MaxNet project on Big-Data-Driven Material Discovery. Two further joint manuscripts are currently in the final stages of preparation.
Mining Dispersion-Corrected Subgroups

Investigators: Mario Boley and Jilles Vreeken in cooperation with Brian Goldsmith and Luca Ghiringhelli (FHI Berlin)

Subgroup discovery is a well-established KDD technique with applications, e.g., in Medicine, Social Science, and Materials Science. In contrast to global modeling, which is concerned with the complete characterization of some variable defined for a given population, subgroup discovery aims to detect intuitive descriptions or selectors of subpopulation in which, locally, the target variable takes on a useful distribution. In scientific domains, like the ones mentioned above, such local patterns are typically considered useful if they are not too specific (in terms of subpopulation size) and indicate insightful facts about the underlying physical process that governs the target variable. Such facts could for instance be: ‘patients of specific demographics experience a low response to some treatment’ or ‘materials with specific atomic composition exhibit a high conductivity’. For numeric (metric) variables, subgroups need to satisfy two things to truthfully represent such statements: the local distribution of the target variable must have a shifted central tendency (effect), and group members must be described well by that shift (consistency). The second requirement is captured by the group’s dispersion, which determines the average error of associating group members with the central tendency value.

Despite all three parameters—size, central tendency, and dispersion—being important, the only known approach for the efficient discovery of globally optimal subgroups, branch-and-bound search, is restricted to objective functions that only take into account size and central tendency. A problem with all such functions is that they inherently favor larger groups with scattered target values over smaller more focused groups with the same central tendency. That is, they favor the discovery of inconsistent statements over consistent ones—surprisingly often identifying groups with a local error that is almost as high or even higher than the global error. To alleviate this, we extended [1] branch-and-bound search to objective functions of the form

\[ f(Q) = g(|Q|, \text{med}(Q), d(Q)) \] (39.3)

where \( g \) is monotonically increasing in the subpopulation size, monotonically decreasing in any dispersion measure \( d \) around the median, and, besides that, depends only (but in arbitrary form) on the subpopulation median. This involves developing an efficient algorithm for computing the tight optimistic estimator given by the optimal value of the objective function among all possible subsets of target values, which has been shown to be a crucial ingredient for the practical applicability of branch-and-bound. Unlike the state of the art, we do not fix the size of subset \( R_i \) as in the previous approach but instead fix its median to target value \( i \). It turns out that this suffices to efficiently compute the tight optimistic estimator for all objective functions of the form of Eq. (39.3). Moreover, we end up with a linear time algorithm in the important special case where the dependence on size and dispersion is determined by the dispersion-corrected coverage defined by

\[ \text{dcc}(Q) = \frac{|Q|}{|P|} \max \left\{ 1 - \frac{\text{amd}(Q)}{\text{amd}(P)}, 0 \right\} \]

where \( \text{amd} \) denotes the average absolute deviation from the median. This is the same computational complexity as the objective function itself. Consequently, this new approach
can discover subgroups according to a more refined selection criterion without increasing the worst-case computational cost. Additionally, as demonstrated by empirical results on a wide range of datasets, it is also highly efficient and successfully reduces the error of result subgroups in practice.

**Mining Fair Subgroups**  *Investigators: Janis Kalofolias, Mario Boley and Jilles Vreeken*

An orthogonal problem with traditional subgroup discovery is its simplistic notion of generality: if a subpopulation is relatively sizeable it is considered general, even though it might show arbitrary statistical obscurities. This lack of representativeness is a key problem in many important scenarios.

In scientific discovery and theory development, we often seek to identify local factors that influence some variable, but want to control for the influence of other potential explanations. For instance, in materials science we may want to discover structural patterns that characterise the HOMO-LUMO energy gap in gold nanoclusters, independent of the parity of their atom count, which is already known to have a strong influence, while in policy we are not allowed to discriminate and while students with a high chance of obtaining a degree are reasonable candidates for defining the application criteria of a scholarship, we might still want to ensure that the eligible population is gender-balanced.

All of the above settings share the requirement of subgroups to not only be relatively sizeable, but also statistically representative w.r.t. some control variable. Specifically, this variable should have a similar distribution between the subgroup and the global population, exhibiting what is called statistical parity. In contrast, simply removing the control variable, to avoid it influencing the result, is infeasible, since it can usually be approximately recovered by the remaining variables (known as red-lining effect).

In this work [4] we extend branch-and-bound subgroup discovery to include a general representativeness term which is based on the statistical distance between the local and the global distribution of the control variable. Moreover, we show how the resulting representative subgroup discovery problem can be solved efficiently for the case of a binary control and a numeric target variable. In particular, we propose RaWR, an algorithm to compute the tight optimistic estimator for the representativeness-aware objective function, in $O(n \log n)$ time. Experiments show that, when employing this algorithm in the branch-and-bound framework, we can prune orders of magnitude of candidates in comparison to the state of the art, which, besides reducing memory consumption, leads to orders of magnitude gain in runtime; therewith, RaWR makes it possible to mine representative subgroups in otherwise computationally infeasible settings.

**Mining Causal Rules**  *Investigators: Kailash Budhathoki, Mario Boley and Jilles Vreeken*

Although both impressive and useful, stating neither stating that there exists a causal relationship from a set of variables $X$ towards a certain variable of interest $Y$, nor that under a set of conditions target $Y$ shows exceptional behaviour satisfies the curiosity of the domain expert — for a domain expert it is of particular interest to know those conditions under which a causal effect is visible, such as the specific combinations of drugs that lead to
severe side-effects. That is, we are interested in discovering causal rules from observational data that consistently maximize rule effect formalized as

\[ e(\sigma) = \mathbb{E}Y | \sigma = \text{true} - \mathbb{E}Y | \sigma = \text{false}. \]

Though simple to state, this task is not only computationally hard; algorithmic solutions also have to cope with an intricate combination of two semantic problems. The first is the well-known phenomenon of overfitting. This phenomenon results from the high variance of the naive empirical (or “plug-in”) estimator of \( e \) for rules with too small sample sizes for either of the two events, \( \sigma = \text{true} \) or \( \sigma = \text{false} \). Combined with the maximization task over a usually very large rule language, this variance turns into a strong positive bias that dominates the search and causes essentially random results of either extremely specific or extremely general rules. The second problem is often referred to as Simpson’s paradox: even strong and confidently measured effects of a rule might not actually reflect true domain mechanisms, but can be mere artifacts of the effect of other variables.

We presented [2] a theoretically sound approach to the discovery of causal large effect rules that remedies each of the aforementioned problems. To address the structural problem, we propose to control for the effect of a given set of potential confounder variables \( Z \). In particular, we identify the admissible data generation process under which it is possible to discover truly causal rules. To address the overfitting problem, we proposed to measure and optimize the reliable effect of a rule. In contrast to the plug-in estimator, we propose a conservative empirical estimate of the population effect, that is not prone to overfitting. Additionally, and in contrast to other known rule optimization criteria, it is also consistent, i.e., with increasing amounts of evidence (data), the measure converges to the actual population effect of a rule. We develop a practical algorithm for efficiently discovering the top-\( k \) reliable effect rules. In particular, we show how the optimization function can be cast into a branch-and-bound approach based on computationally efficient tight, selection unaware, optimistic estimators.

In sum, we proposed a practically applicable rule induction technique that is able to discover true causal rules, and therewith insights to the application domain underlying a dataset. Moreover, our approach lends itself naturally to an iterative data mining approach in which we can discover insights beyond the factors included in \( Z \). We supported our claims by experiments on real-world datasets as well as by reporting the required computation times on a large set of benchmark datasets. An extended version of this work is currently under review.

**References**

Figure 39.14: Map visualizing the redescription $L. lynx \lor L. canadensis \iff [-24.4 \leq t^+ < 3.4]$. Purple denotes the areas where both queries hold, red the areas where only the left-hand query holds, and blue the area where only the right-hand query holds.

39.5.4 Redescription Mining

Investigators: Megan Humble, Janis Kalofolias, and Pauli Miettinen in cooperation with Esther Gabrun (University of Eastern Finland)

Finding multiple ways to characterize the same entities is a problem that appears in many areas of science. In medical sciences, for example, one typically wants to find a subset of patients sharing similar symptoms and similar genes. In biology, the bioclimatic constraints that must be met for a certain species to survive constitute that species’ bioclimatic envelope, and finding such envelopes can help, e.g. to predict the results of global warming.

But this process is only semi-automatic. For instance, to find the bioclimatic envelopes, an expert first selects a species and then uses some method to find the envelope for this particular species. More complex combinations of species, or even any combinations at all, are rarely studied, as manually iterating over all possible combinations would be far too laborious.

It is here where redescription mining [2, 3] comes to help. In redescription mining the input contains entities with two sets of characterizing variables. The task is to find a pair of queries, one query for both sets of variables, such that both queries describe (almost) the same set of entities. In niche-finding, the entities would be spatial locations, one set of variables would be the fauna and the other set would contain the bioclimatic variables. A simple example of a redescription in this setting is presented in Fig. 39.14.

The redescription in Fig. 39.14 is $L. lynx \lor L. canadensis \iff [-24.4 \leq t^+ < 3.4]$. It describes some areas in the Earth on one hand as those areas where either the Eurasian or Canada lynx lives and on the other hand, as those areas where March’s maximum temperature is between $-24.4$ and $3.4$ degrees Celsius.
Our redescription mining tool Siren [4] provides a visual interface for mining and analysing the redescriptions. Individual redescriptions can be visualised in multiple linked visualisations, and the queries can be edited interactively (see Fig. 39.15).

**Selecting Surprising Redescriptions**

The redescription mining task is an example of exhaustive data mining: the goal is to find all redescriptions that satisfy the constraints. Many of these redescriptions can be redundant, however. For example, applying the ReReMi algorithm to data that contains land mammal habitants and climate yields in six different redescriptions describing the climate of polar bear’s habitat; in short, they are all different ways of expressing that the polar bear lives in cold environments. Yet, as soon as the user has seen one of them, they knows the essence of the rest of them.

To find only the interesting redescriptions, we use the framework of subjective interestingness [7, 6]. Our approach is based on maximum-entropy distributions, as proposed by De Bie [1]. We consider the likelihood of seeing a particular redescription given the earlier redescriptions we have seen. Initially, we have no background information, and the redescriptions are ranked by their accuracy. We add the first redescription as a constraint, and learn the maximum-entropy distribution of the data under that constraints. From that distribution we can calculate the likelihoods of all remaining redescriptions, and sort them in ascending likelihood (that is, descending in how surprising they are). The most surprising redescription is then added as a constraint to the model, new maximum-entropy distribution is calculated, and the remaining redescriptions’ likelihoods are updated.

Redescriptions present few important challenges to this general framework. First, the variables in the queries can be of mixed type (some Boolean, some categorical, some...
numerical), and second, the queries can, in principle, be arbitrary Boolean queries. Together, these two features mean that the maximum-entropy distribution can have an exponential number of terms to solve. Fortunately, we can engineer around this problem in the real-world scenarios, as most redescriptions are short, and the numerical variables only depend on intervals, not arbitrary sets. In order to model the data rows, we present two approaches. When calculating the likelihood of a redescription, one of them considers all rows of the data, while the other only considers the rows where either of the queries in the redescription hold.

Our experiments show that the maximum-entropy approach is efficient in removing redundant redescriptions, and that our implementation can do the ranking in a reasonable time, even if it cannot currently be used for interactive ranking.

### Analysing Financial Data

Standard redescription mining algorithms are not tailored to handle time series data. In [5], the basic ReReMi algorithm for redescription mining was extended to handle time series by allowing it to build constraints over the time; for instance, it could find a redescription that is only valid between certain dates, or starting from a certain date.

Being able to handle time series data can be useful in many contexts. In [5], the extended algorithm was used to analyse financial data, in particular, data containing attributes of exchange traded funds (ETFs). The analysis of the results found, for instance, redescriptions that showed that S&P 500, emerging markets, and German stock index at least were tightly correlated in the first half of 2012. On the other hand, the algorithms were not able to find any seasonal effects, such as the “January effect” (i.e. stocks are down on December and up on January). This might be because the ETFs are so large and diverse that they absorb such seasonal variations.

### References


39.5.5 Patterns, Anomalies, and Interestingness

When we consider to read a book we can read the synopsis we find on the back. When considering to investigate a database such synopses are not readily available. We study techniques for summarising data in easily understandable terms, e.g., a small and non-redundant set of the patterns. We do so for different data types and description languages, as well as how we can use these patterns to uncover anomalies, generate realistic discrete-valued population samples, and how to define scores to identify interesting and surprising nodes in graphs.

Summarising Event Sequences with Rich and Interleaving Patterns

Investigators: Apratim Bhattacharyya and Jilles Vreeken

Discovering the key patterns from a database is one of the main goals of data mining. Modern approaches do not ask for all patterns that satisfy a local interestingness constraint, such as frequency, but instead ask for that set of patterns that together describe the data best. How well we can describe the data depends on the description language we use. The richer this language, the more relevant structure we can identify. At the same time, a richer language means a larger search space, and hence requires more efficient search.

In this paper we consider databases of event sequences, and are after that set of sequential patterns that together describe the data best—as we did previously with Sqs. Like Sqs we describe a database with occurrences of patterns. Whereas Sqs requires these occurrences to be disjoint, however, we allow patterns to interleave. This leads to more succinct descriptions as well as better pattern recall. Moreover, we use a richer class of patterns. That is, we do not only allow for gaps in occurrences, but also allow patterns to emit one out of multiple events at a certain location. For example, the pattern ‘paper [proposes | presents] new’ discovered in the JMLR abstract database matches two common forms of expressing that a paper presents or proposes something new.

With this richer language, we can obtain much better compression rates with much fewer patterns. To discover good models we propose SQUISH [1], a highly efficient and versatile search algorithm. Its efficiency stems from re-use of information, partitioning the data, and in particular from considering only the currently relevant occurrences of patterns in the data. It is a natural any-time algorithm, and can be ran for any time budget that is opportune.

Extensive experimental evaluation shows that SQUISH performs very well in practice. It is orders of magnitude faster than the state of the art, results in better models, as well as discovers meaningful semantics in the form patterns that identify multiple choices of values. We give a number of examples of such patterns as discovered on the abstracts of the Journal of Machine Learning, and from a collection of addresses made by US presidents in Table 39.2. Importantly, SQUISH is highly extendable [2], allowing for richer pattern classes to be considered in the future.
**Presidential Addresses**

1. [coordin. | execut.] branch govern
2. fellow [citizen | american | countrymen]
3. [discharg | perform | commenc] duti
4. god [bless | help]
5. [exercise | grant | balanc] power
6. power [grant | vest]
7. [eighteenth | fifteenth | fourteenth] amendment
8. [guard | war] against

**JMLR Abstracts**

1. [high | curse | low] dimension.
2. [empirical | structural] risk minimisation
3. [independent | principle] component analysis
4. paper [proposes | presents] new
5. [Mahalanobis | edit | Euclidean | pairwise] distance
6. [data | train] set
7. [conditional | Markov] random field
8. [gradient | coordinat.] descent

<table>
<thead>
<tr>
<th>Table 39.2: Sample choicisodes discovered by SQUISH in the Addresses and JMLR datasets.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surprising Co-Occurrences</strong></td>
</tr>
</tbody>
</table>

*Investigators: Jilles Vreeken in cooperation with Arno Siebes (U. Utrecht), Roel Bertens (U. Utrecht)*

The recognition of anomalies provides useful application-specific insights. For transactional data anomaly detection usually boils down to pointing out those transactions that show unexpected behaviour. While existing methods all have their own advantages, however, none of them is able to detect an anomaly based on the presence of multiple items in a single transaction that are not expected to occur together. As an example consider a dataset containing people’s drinking habits where roughly half of the people drinks soft drink $C$ and the other half drinks soft drink $P$. Now each individual who drinks $C$ or $P$ is not surprising. Moreover, someone drinking both $C$ and $P$ also does not seem surprising as it can be compressed well, it contains multiple frequent patterns and there is nothing missing. However, in this dataset almost everyone drinks either $C$ or $P$, but not both. Therefore, someone drinking both $C$ and $P$ is an anomaly, as drinking both is unexpected.

We proposed [4] to score each transaction based on the most unlikely co-occurrence between patterns and therefore our method is able to find the described class of anomalies.

For this example, the score we introduce is based on the well-known interestingness measure *lift* of an association rule. More precisely, we take minus the log of the minimal lift of the two association rules $C \rightarrow P$ and $P \rightarrow C$. So, the difference is that we do not score a rule, but a transaction and do so by the minimal lift of all the rules that apply to this transaction. The higher the score, the more anomalous the transaction. So, if both $C$ and $P$ are frequent, a transaction $t$ containing both is anomalous if $\{C, P\}$ is infrequent, and the more infrequent it is, the more anomalous $t$ is. Note that this is the opposite of rare
patterns, or rare association rules. For rare rules either $C$, $P$, or both should be infrequent, while the confidence of, e.g., $C \rightarrow P$ should be high.

For rare association rules either $C$ or $P$ is expected to be infrequent, and multiple or adaptive minimal support thresholds can be used for efficient discovery. Since we assume both $C$ and $P$ to be frequent and only \{C, P\} to be infrequent such ideas cannot be used here. Rather, an exhaustive algorithm requires all frequent sets with a support equal or larger than 1, since any of these may be formed by an unexpected combination of patterns. Finding the most surprising transactions using this humongous set is infeasible on all but the most trivial data sets. For this reason we introduce a heuristic algorithm based on the pattern sets computed by algorithms such as Krimp or Slim.

In extensive experiments we firstly show that this heuristic algorithm finds all anomalies we hide in synthetic data. Secondly, we show that the transactions found to be anomalous in real world data sets are indeed strange. For example, in the well-known Adult data set, the top-ranked transaction contains the very unexpected co-occurrence of someone whose attribute sex is female yet whose relationship status is husband. It is highly probable that this is a mistake, but it is certainly an anomaly the data scientist should be aware of before analysing the data set.

**Generating Realistic Populations**

*Investigators: Jilles Vreeken in cooperation with Nikolaj Tatti (Aalto), Hao Wu (Google), Naren Ramakrishnan (Virginia Tech)*

Many research areas, e.g., epidemiology, public health, social science, study the behavior of large populations of individuals under natural scenarios as well as under human interventions. A key need across these domains is the ready availability of realistic synthetic datasets that can capture key attributes and activities of large populations. For instance, in epidemiology, synthetic populations are necessary to study disease propagation and intervention measures before implementation. Information from the US census is typically used to model such synthetic datasets.

As a case in point, in epidemiology, one important task is to simulate disease spread and potential outbreaks on the city- or nation-level, and provide useful information to public health officials to support policy and decision making. To make such simulations as accurate as possible, synthetic populations that have the same structural and behavioral properties as the real population are needed. In domains like health care, privacy is an additional issue motivating the design of synthetic populations. In these applications, the necessary datasets to be generated can be represented as tuples with categorical data attributes.

Motivated by these emerging needs, we focused our attention on constructing a generative model that captures given characteristics of categorical population attributes, and best estimates the underlying data generation distribution [5]. However, modeling multi-dimensional categorical data and estimating distributions can be quite challenging due to the exponential possibilities of data spaces in terms of the number of dimensions of categorical data tuples. To address these challenges and difficulties, we propose to model categorical data with statistical constraints, we apply the classical and statistically well-founded maximum entropy model. We construct a generative maximum entropy model, which takes the data schema
and a set of patterns. By sampling the categorical tuples from the maximum entropy model, synthetic population datasets can be generated.

Generally, solving maximum entropy models can be infeasible in practice. In this paper, we show that by leveraging the structure of the categorical data space in our setting, the maximum entropy model could be inferred quite efficiently. We also propose a heuristic together with the Bayesian information criterion (BIC) to select a simple as well as informative model. Using results on both synthetic datasets and real US census data, we demonstrate that the proposed maximum entropy model is capable of recovering the underlying categorical data distribution and generating relevant synthetic populations.

Interesting and Surprising Nodes in Graphs

Investigators: Jilles Vreeken in cooperation with Robert Pienta, Minsuk Kahng, Zhang Lin, Dueng Horn Chau (Georgia Tech), James Partha Talukdar (IIS), Abello (Rutgers U.), and Ganesh Parameswaran (Yahoo!)

Large graphs are ubiquitous. They are natural representations for many domains, and hence we find graph structured data everywhere. As data collection becomes increasingly simple, and many domains remain complex, real-world graphs are rapidly increasing in size and data richness. These graphs may have over millions and billions of nodes and edges and also have thousands or more of attributes. It is fair to say that many graphs are in fact too big; exploring such large graphs, where the goal of the user is to gain understanding, is a highly non-trivial task.

Visualization is perhaps the most natural approach to exploratory data analysis. Under the right visualization, finding patterns, deciding what is interesting, what is not, and what to investigate next becomes easy tasks — in a sense the answers “jump to us” as our brains are highly specialized for analyzing complex visual data. It is therefore no surprise that visualization has proven to be successful in many domains.

Visualizing large graphs in an intuitive and informative manner has proven to be difficult, however. Even with advanced layout techniques, plotting a graph can create a hard-to-read cluster of overlapping nodes and edges, from which little can be deduced. This is the case even for graphs with only thousands of nodes. Instead of plotting the whole graph (a), visualizing only part of the graph (b) seems more promising. However, as many real world graphs are scale-free (follow a power law degree distribution), selecting relevant subgraphs to visualize can be challenging. Moreover, because of high-degree nodes, even a single-hop neighborhood expansion from a node can be visually overwhelming.

We take a different approach, and proposed to adaptively explore large graphs from a local perspective. That is, starting from an initially selected node — e.g., explicitly queried by a user, or proposed by an outlier detection algorithm — we only show the most interesting neighbors as the user explores the graph from node to node [3]. We identify these nodes by their subjective interestingness based on how surprising their and their neighbors’ data distributions are (e.g., do these neighbors’ degree distributions follow a power law distribution like when considering all nodes?), as well as by how similar those distributions are compared to those of the nodes the user has explored so far. By only showing the parts of the graph that would be more interesting to the user, the visualization does not get too complex. By
being adaptive, it allows users to explore facets of the graph that are more subjectively interesting to them.

We call our adaptive approach FACETS, which steer users towards local regions of the graphs that match best with their current browsing interests, helping them better understand and visualize the graphs at the same time. FACETS ranks nodes based on how interesting and unexpected their neighborhoods are. To the best of our knowledge, our work is the very first that adopts these notions to help user explore and visualize large graphs. To illustrate how FACETS works in practice, we consider the results of FACETS for a movie database graph, and two users in Figure 39.16.

Figure 39.16: Visualizations of the (a) Blade Runner and (b) Toy Story case studies. The gray, circular nodes form a profile of films investigated by the user via graph exploration or text-search. The star nodes are the last-clicked node for which FACETS has produced the 5 most interesting (red), and surprising (blue) nodes. Nodes can be both surprising and interesting (purple).

References

39.5.6 Reductions and ASP Solvers for Frequency-Based Pattern Mining

Investigators: Daria Stepanova and Pauli Miettinen in cooperation with Sergey Paramonov (Katholieke Universiteit Leuven) and Stefan Neumann (University of Vienna)

Frequent itemset mining is one of the classical problems in data mining. Much of the research has focused on developing specialised algorithms for specific mining problems. These specialised algorithms are fast, but they are not easy to adapt even to slight variations of the problems. On the other hand, the concentration of specific problems has also blurred the “big picture”, leaving it unclear how the different problems are related to each other. To address the first problem, we are proposing a hybrid solution that uses specialised mining algorithms for fast mining, but general Answer Set Programming (ASP) solvers for applying constraints. To address the second problem, we show that under novel reductions, most maximal pattern mining problems can be reduced to each other.

ASP solvers. Availability of vast amounts of data from different domains has led to an increasing interest in the development of scalable and flexible methods for data analysis. A key feature of flexible data analysis methods is their ability to incorporate users’ background knowledge and different criteria of interest. They are often provided in the form of constraints to the valid set of answers, the most common of which is the frequency threshold.

In general, all constraints can be classified into local constraints (e.g., frequency or size), that can be validated by the pattern candidate alone, and global constraints (e.g., various condensed representations like maximal, closed, skyline), that can only be validated via an exhaustive comparison of the pattern candidate against all other candidates. Combining local and global constraints in a generic way is an important and challenging problem, which has been widely acknowledged in the constraint-based mining community.

Answer Set Programming [1] as a declarative rule-based problem solving paradigm oriented towards difficult search problems, seems well-suited for specifying constraints conveniently thanks to its expressive and intuitive modelling language. In [6, 7, 8] we have exploited these properties of ASP and presented a hybrid approach for combining local and global constraints in a generic and efficient way.

More specifically, our approach comprises of two steps. In the first step, optimized algorithms are applied to discover a set of frequent patterns, and in the second step, the patterns are post-processed using declarative ASP. The key advantage of our approach stems from the fact that it preserves the generality of purely declarative methods with respect to the frequent pattern mining problems, while providing an efficient system to develop prototypes which can run on real-world datasets where typically only specialized algorithms are deployed. This is especially beneficial in the setting where a user considers new variations of pattern mining problems.

Our hybrid method has been applied to itemset, sequence and graph mining tasks as well to a problem of approximately tiling a database. Experiments on real-world datasets show the effectiveness of the proposed method and computational gains compared to the state-of-the-art purely declarative pattern mining approaches.

Reductions between maximal pattern mining problems. Reductions are a standard way to study the computational complexity of problems. They allow us to impose a partial order
on the problems based on the hardness of solving them. Reductions are not commonly used for many classical frequency-based pattern mining problems, though. Often, the argument is that “the output is potentially exponential,” rendering standard polynomial reductions useless. But other models of complexity, such as counting complexity \[9\] or enumeration complexity \[2\], can still be applied.

To establish connections between maximal frequent pattern mining problems, we developed \[4\] two novel reduction types, maximality-preserving reductions and maximality-preserving reductions for feasible frequency-based problems. These allow us to study the complexity of maximal pattern mining problems in different domains, including labelled graphs, DAGs, trees, subsequences with no repeated symbols, and itemsets, to name a few. Extending the work of Kimelfeld and Kolaitis \[3\], we build a number of novel reductions between these problems (see Figure 39.17). Our results show that the complexity of these maximal pattern mining problems essentially collapses.

Furthermore, we can also show that these reductions can be used in practice to efficiently mine frequent subgraphs from labelled graphs using slightly modified Apriori algorithm \[5\].

References


39.5.7 Matrix Factorization

Investigators: Sanjar Karaev, Saskia Metzler, and Pauli Miettinen in cooperation with James Hook (University of Bath, UK)

In a classical matrix factorization problem, the goal is, given the input data in a form of a matrix \(A \in \mathbb{R}^{n \times m}\) and a positive integer \(k\), find factor matrices \(B \in \mathbb{R}^{n \times k}\) and \(C \in \mathbb{R}^{k \times m}\) such that the error

\[
\|A - BC\| \quad (39.4)
\]

is minimized. Low rank matrix factorization \((k << \min\{n, m\})\) finds extensive applications in data analysis because it helps to uncover hidden patterns in the data. By finding the factor matrices \(B \in \mathbb{R}^{n \times k}\) and \(C \in \mathbb{R}^{k \times m}\), we represent the input matrix \(A \in \mathbb{R}^{n \times m}\) as an elementwise sum of \(k\) rank-1 patterns and a noise matrix. To see this, let us rewrite the matrix product in an alternative form:

\[
BC = F_1 + F_2 + \cdots + F_k, \quad (39.5)
\]

where each summand \(F_i\) is the outer product of the \(i\)th column of \(B\) and the \(i\)th row of \(C\). Now it becomes clear that the rank-1 matrices \(F_i\) are the “simple patterns”, and minimizing \((39.5)\) is equivalent to finding \(k\) rank-1 patterns that minimize

\[
\|A - (F_1 + F_2 + \cdots + F_k)\|. \quad (39.6)
\]

Now each element \(A_{ij}\) of the input matrix can be seen as an elementwise sum of the corresponding elements \((F_s)_{ij}\), with each such element making a contribution towards the final value of \(A_{ij}\).

**Subtropical matrix factorization.** While the component interpretation in \((39.6)\) seems like a viable approach to finding patterns from the data, patterns \(F_i\) tend to interfere with each other, often making it difficult to separate them. This happens because at each point \(ij\) there are \(k\) patterns contributing simultaneously, and it is difficult to tell them apart. In order to address this problem, we used matrix factorization over the so called “subtropical algebra”. It differs from the standard one in that it uses the maximum operation instead of the addition, and its domain of definition is restricted to nonnegative real numbers. Formally, the subtropical matrix factorization is defined as follows. Given a matrix \(A \in \mathbb{R}_+^{n \times m}\) and an integer \(k > 0\), find factor matrices \(B \in \mathbb{R}_+^{n \times k}\) and \(C \in \mathbb{R}_+^{k \times m}\) minimizing

\[
E(A, B, C) = \|A - B \otimes C\|, \quad (39.7)
\]
where $B \otimes C$ denotes the subtropical matrix product of $B$ and $C$:

$$(B \otimes C)_{ij} = \max_{s=1}^{k} B_{is} C_{sj}. \quad (39.8)$$

Using the maximum operation instead of the addition allows patterns to directly determine the outcome for every element in the matrix product. To understand why this is the case, let us write the subtropical equivalent of (39.5):

$$(BC)_{ij} = \max\{(F_1)_{ij}, (F_2)_{ij}, \ldots, (F_k)_{ij}\}. \quad (39.9)$$

It is clear that for every element $ij$ there is a dominant factor $F_s$, such that $(F_s)_{ij} \geq (F_t)_{ij}$ for $t \neq s$, and hence $(F_s)_{ij} = (BC)_{ij}$. We call this property the “winner takes it all” since the dominant factor is the only one making any contribution to the final value in the subtropical product.

We previously proposed two algorithms for solving the subtropical matrix factorization problem – Capricorn [6] and Cancer [5]. While they solve essentially the same problem, Capricorn should be used if the data is expected to have so called flipping noise (when some elements get changed to random values), and for dealing with continuous (e.g. Gaussian) noise, Cancer becomes a better choice. We have recently developed a unifying framework, which we called Equator [7], that incorporates both Capricorn and Cancer. We also added a new form of Cancer that optimizes the Jensen-Shannon divergence instead of the Frobenius norm and ran more comprehensive experiments. One of the main objectives of this work was to compare the subtropical structure to the one produced by the nonnegative matrix factorization (NMF). In particular, we found that Cancer was slightly superior to NMF in predicting missing data. In terms of data interpretation, we observed that subtropical methods are generally better at identifying prominent features in the data, while NMF’s strength lies in recovering smooth transitions between high and low values. To illustrate this difference, let us have a look at the results of Cancer and an NMF algorithm, called WNMF [8], on the European climate data\(^6\) (Figure 39.18). The data was a matrix with rows corresponding to geographical locations and columns to weather observations, such as average temperature or precipitation. We decomposed it into two factors and then interpreted the columns of the left-hand matrix. Since each row of the data represents a location, these columns can be plotted on a map. The example results of Cancer and WNMF show these algorithms’ interpretation of the annual precipitation patterns in Europe. By comparing them to the ground truth data, one can see that Cancer extracts the wettest areas, while WNMF captures the the transition from high to low precipitation areas, but fails to identify some of the wettest regions, such as the west Norwegian coast. This clearly demonstrates the difference between the two types of structures – while NMF recovers the gradual change from one part of the data to another, the subtropical algebra makes the most prominent features stand out.

**Mixed logistic-tropical matrix factorization.** As we saw above, both NMF and the subtropical matrix factorization models can find interesting patterns in the data. Moreover, the found features are often complementary to each other, making it worthwhile to try both

methods on the same dataset. The problem is, however, that some data might exhibit both structures simultaneously. Indeed, the structure of real-world data depends on the process that generated it, and it is perfectly plausible that, while some parts of the data follow the NMF structure, the rest of it is better explained using the subtropical matrix factorization. We developed a method to automatically detect which parts of the data have more NMF-like structure, and which require the subtropical algebra. We also proposed an algorithm, called Latitude, for solving the resulting mixed matrix factorization problem [3]. The idea is to represent the input matrix as a convex sum of the NMF and subtropical matrix products

\[ A_{ij} = \alpha_{ij}(BC)_{ij} + (1 - \alpha_{ij})(B \uplus C)_{ij} . \]  

(39.10)

This way, not only can we answer the question of whether a particular part of the data has more of NMF or subtropical structures, but also by how much. One downside of the above approach is that there is too much freedom in fitting the data. In order to avoid overfitting, we place constraints on the mixing matrix \( \alpha \). We do this by requiring that the matrix \( \alpha \) be a sigmoid of a rank-1 “tropical” matrix. The tropical algebra is related to the subtropical algebra, and its addition is also defined as the maximum operation. The difference is that it also has multiplication replaced with addition, and is defined over the set of extended real numbers \( \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \). Let us denote the tropical matrix product by \( \diamond \). Now the \( n \)-by-\( m \) matrix \( \alpha \) can be defined as follows:

\[ \alpha = \sigma(\phi \diamond \theta) , \]  

(39.11)

where \( \phi \) is a column vector of length \( n \) and \( \theta \) is a row vector of length \( m \).

To see how this method works in practice, let us have a look at the distribution of the NMF and subtropical structures produced from a subset of the Extended Yale Face collection of face images [2] (Figure 39.19). It is easy to see that the more prominent features of the face, such as eyes, nose, and mouth, lean heavily towards the subtropical structure, while the rest of the face is almost entirely NMF-like.

**Logistic-tropical matrix factorization.** Dense subgraph mining is a common task in data analysis. The goal is to (approximately) express a given graph as a union of a relatively
small number of dense subgraphs. By considering the adjacency matrix of a graph, this problem can be equivalently expressed as covering a binary matrix with dense submatrices (or quasi-cliques). Perhaps the biggest practical interest in this problem comes from identifying communities in graphs.

Traditionally communities were thought to be (quasi-)cliques and hence assumed to be dense submatrices of the adjacency matrix. In recent years, however, many other models have been observed, such as stars, hyperbolic shapes, or the core-periphery model. The latter one was studied quite extensively in social sciences (see e.g. [1]).

While the above models seem quite different, they can all be classified under the same umbrella, the so called nested matrix (see e.g. [10]). In order to define it, let us first introduce the notion of a step function. Let $[n] = \{1, 2, \ldots, n\}$ and $[m] = \{1, 2, \ldots, m\}$. We say that function $s: [n] \to [m]$ is a step function if $s(i) \geq s(j)$ for all $i$ and all $j < i$. A binary matrix $A \in \{0, 1\}^{n \times m}$ is directly nested if there exists a step function $s$ such that on each row $i \in [n]$ of $A$, $a_{ij} = 1$ if $j \leq s(i)$ and $a_{ij} = 0$ if $j > s(i)$. $A$ is nested if there exists a way to permute its rows and columns such that the permuted matrix $A'$ is directly nested.

Since nested matrices describe all of the community types discussed above, it is tempting to use them instead of quasi-cliques in traditional graph mining methods. One of the more popular approaches is to represent communities as rank-1 submatrices and use matrix factorization to mine them. Unfortunately, this method does not work directly with nested matrices since they are not really rank-1 matrices. It is possible, however, to view them as generalized rank-1 matrices, thanks to the recent characterization using so called rounding ranks (see [9]). The problem is that this characterization falls apart in higher dimensions, preventing us from formulating the nested subgraph mining problem in terms of matrix factorization. It is possible to show, however, that if we use matrix factorization over the tropical algebra, then we can avoid this issue. In our recent work [4], we developed a tropical matrix factorization model for community mining and proposed a highly efficient and scalable algorithm to solve it. We used it on various real-world datasets to prove that the proposed model can indeed find a more diverse set of communities than the classical model.
In order to demonstrate what kind of structure can be found using this model, we show in Figure 39.20 some sample communities mined from real-world data. For each community we show its adjacency matrix with indices sorted according to the nested order. We can see that the found communities are quite diverse – in particular we find both convex and concave shapes, and also an example of the core-periphery model (the leftmost picture).

References


39.5.8 Large Entries of Matrix Products

*Investigators: Christina Teflioudi in cooperation with Rainer Gemulla (University of Mannheim)*

We study the problem of efficiently retrieving large entries in the product of two given factor matrices $B$ and $C$, which we refer to as “large-entry retrieval problem”. Large entries in
the matrix product $BC$ correspond to strong interactions of data objects and are often of particular interest in applications. We focus on the setting where the factor matrices are tall and skinny, each with millions of rows and tens to hundreds of columns.

Tall and skinny matrices commonly arise from low-rank matrix factorization methods, such as singular value decomposition (SVD), non-negative matrix factorization (NMF), or latent-factor models, each an important building block in data mining. In the context of recommender systems, for example, latent factor models are a popular and successful approach for predicting the preference of users for items. In such models, both users and items are represented by vectors of latent features arranged in a user matrix and an item matrix, respectively. The product of both matrices captures the preference of each user for each item; we are interested in large entries, which correspond to the items of highest preference (for the corresponding user) and thus to potentially good recommendations.

In most applications, the matrix product is significantly larger than the factor matrices themselves and its complete computation is generally infeasible in practice. To avoid the computation of the full matrix product, we propose LEMP, an efficient algorithm to retrieve only the Large Entries of the Matrix Product. LEMP takes as input two real-valued factor matrices $B$ and $C$ as well as a threshold $\theta > 0$. It then determines the set of large entries in $BC$, i.e., the set of entries that have a value of at least $\theta$.

LEMP in its core solves a Maximum Inner Product Search (MIPS) problem: it interprets both matrices as sets of vectors, formed by the rows of $B$ and $C^T$ and outputs all pairs of query and probe vectors with an inner product of at least $\theta$. We refer to $B$ as the query matrix and to $C$ as the probe matrix. Our LEMP [1] algorithm is inspired by Fagin’s Threshold Algorithm (TA) and techniques from cosine similarity search. It makes use of the simple observation that both the Euclidean norm and the direction of two vectors influence the value of their inner product. In particular, LEMP groups the input vectors into buckets of similar norms and subsequently solves a smaller cosine similarity search problem for each bucket. In this way, LEMP (i) exploits vector norms for early pruning, (ii) is able to choose a suitable search technique separately for each bucket (and query), and (iii) improves cache locality by fitting the small problem instances into cache. LEMP is able to leverage existing cosine-similarity search techniques to process its buckets. In addition, we devised a number of novel search algorithms for both exact (based on TA) and approximate (based on Locality Sensitive Hashing) search. Our methods are tailored to our setting of tall and skinny matrices, in which the dimensionality of the input vectors is intermediate (say, 10–500), i.e., higher than usual for TA and lower than usual for cosine similarity search.

Our experiments suggest that LEMP is often multiple orders of magnitude faster than performing the full matrix product and between 2x and 20x faster than the best-performing alternative exact methods such as the cover tree algorithm and offers better speed-accuracy trade-off than state-of-the-art techniques for approximate MIPS based on asymmetric transformations.

References

39.5.9 Community Models for Graphs

Investigators: Saskia Metzler and Pauli Miettinen in cooperation with Stephan Günneumann (Technical University of Munich)

Cliques are frequently used to model communities in graphs. A community is then regarded a set of nodes where each pair is equally likely to be connected. But observing real world communities reveals that they have more structure than that: Typically, a community has few densely connected members and many loosely associated ones. Hence, not every edge between a pair of nodes in a community is equally likely and we observe a hyperbola-shaped pattern in the degree-ordered adjacency matrix.

We describe a model for capturing non-uniform edge distributions inside communities [1]. Different parametrizations of this model allow to view communities from different perspectives. On the degree-ordered adjacency matrix of a community, our model may be expressed by means of a hyperbola equation for a geometric view, as a mixture of different index orderings for a probabilistic characterization, or through an intuitive description of the shape using parameters for the core size and the tail height (see Fig. 39.5.9).

We find such community models by optimizing for the log-likelihood. And we show that the these three models are equivalent: given one, we can always find parameters for the other two that define exactly the same community.

In our work entitled “Stability and Dynamics of Communities on Online Question–Answer Sites” [2], we present a large scale analysis of the shape of real world networks of large online question-answer sites and their time evolution. We are particularly interested in the patterns of volunteer efforts within the communities. To that end, we examine the community structure of several large online question-answer sites and how they evolve over time. To describe the user interaction patterns concisely, we employ the hyperbolic community model. Using this model, we obtain a summary of each community in each time step by means of intuitive parameters that reflect the connectivity pattern within the network. Our study of the temporal evolution of these parameters reveals an important characteristic: In contrast to what has been observed earlier in the analyses of growth behaviour of online communities, we observe that the user activity within a community is constant with respect to its size throughout its lifetime. Furthermore, the structural organisation of different communities across different question–answer sites follows a common scheme: There is a small group of users who is responsible for the majority of the social interactions.

The design of community detection and graph mining algorithms requires thorough testing on realistic data. Therefore, it is important to generate random graphs with realistic features. In particular to successfully test community detection algorithms, it is essential to have reliable ground-truth information of what the communities are. We present a random graph generator [3] that is capable of generating hyperbolic communities. Their sizes and shapes are drawn from distributions that are either learned from a given (real-world) graph, or defined by the user. The shape is determined through a combination of the size of the core $\gamma$ and the thickness of the tail $H$. Our model preserves properties of the original graph. For example, the degree distribution and local and global clustering coefficients are preserved up to the effects of the noise. The resulting graphs also naturally preserve the shapes of the communities, while still generating highly random graphs.
Figure 39.20: Example nested factors. The orange dots are the 1s in the matrix and the found community is the area left and down from the blue line.

Figure 39.21: Visualization of parameters for the hyperbolic model. The parameter $\gamma$ indicates the core size and $H$ indicates the tail height.
References


39.5.10 Effective Parallelization for Machine Learning

Investigators: Mario Boley in cooperation with Michael Kamp (U. Bonn), Olana Missura (Google), and Thomas Gärtner (U. Nottingham)

With this paper [1] we contributed a novel and provably effective parallelisation scheme for a broad class of learning algorithms. The significance of this result is to allow the confident application of machine learning algorithms with growing amounts of data. In critical application scenarios, i.e., when errors have almost prohibitively high cost, this confidence is essential. To this end, we consider the parallelisation of an algorithm to be effective if it achieves the same confidence and error bounds as the sequential execution of that algorithm in much shorter time. Indeed, our parallelisation scheme can reduce the runtime of learning algorithms from polynomial to polylogarithmic. For that, it consumes more data and is executed on a quasi-polynomial number of processing units. To formally describe and analyse our parallelisation scheme, we consider the regularised risk minimisation setting. For a fixed but unknown joint probability distribution \( D \) over an input space \( \mathcal{X} \) and an output space \( \mathcal{Y} \), a dataset \( D \subseteq \mathcal{X} \times \mathcal{Y} \) of size \( N \in \mathbb{N} \) drawn iid from \( D \), a convex hypothesis space \( \mathcal{F} \) of functions \( f: \mathcal{X} \to \mathcal{Y} \), a loss function \( l: \mathcal{F} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \) that is convex in \( \mathcal{F} \), and a convex regularisation term \( \Omega: \mathcal{F} \to \mathbb{R} \), regularised risk minimisation algorithms solve \( \mathcal{L}(D) = \arg \min_{f \in \mathcal{F}} \sum_{(x,y) \in D} l(f, x, y) + \Omega(f) \). The aim of this approach is to obtain a hypothesis \( f \in \mathcal{F} \) with small regret \( Q(f) = \mathbb{E}[l(f, x, y)] - \arg \min_{f' \in \mathcal{F}} \mathbb{E}[l(f', x, y)] \)

Regularised risk minimisation algorithms are typically designed to be consistent and efficient. They are consistent if there is a function \( N_0: \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+ \) such that for all \( \epsilon > 0, \Delta \in (0, 1], N \in \mathbb{N} \) with \( N \geq N_0(\epsilon, \Delta) \), and training data \( D \sim D^N \), the probability of generating an \( \epsilon \)-bad hypothesis is no greater than \( \Delta \), i.e., \( P(Q(\mathcal{L}(D)) > \epsilon) \leq \Delta \). They are efficient if the sample complexity \( N_0(\epsilon, \Delta) \) is polynomial in \( 1/\epsilon, \log 1/\Delta \) and the runtime complexity \( T_\mathcal{E} \) is polynomial in the sample complexity. We considered the parallelisation of such consistent and efficient learning algorithms, e.g., support vector machines, regularised least squares regression, and logistic regression.

Our main theoretical contribution was to show that algorithms satisfying the above conditions can be parallelised effectively. We consider a parallelisation to be effective if the \((\epsilon, \Delta)\)-guarantees are achieved in time polylogarithmic in \( N_0(\epsilon, \Delta) \). The cost for achieving this reduction in runtime comes in the form of an increased data size and in the number of processing units used. For the parallelisation scheme presented in this paper, we are able to
bound this cost by a quasi-polynomial in $1/\epsilon$ and log $1/\Delta$. The main practical contribution of this paper is an effective parallelisation scheme that treats the underlying learning algorithm as a black-box, i.e., it can be parallelised without further mathematical derivations and without writing dedicated code.

The empirical evaluation confirmed its potential in practical settings. Given the same amount of data as the underlying learning algorithm, we achieves a substantial reduction of computation time in realistic applications. Using 150 processors, we are between 80 and around 700-times faster than the underlying learning algorithm on a single processing unit. Compared with parallel learning algorithms from Spark’s MLlib, it achieves hypotheses of similar quality, while requiring only 15 – 85% of their runtime.

References


39.6 Information Retrieval, Information Extraction, and Text Analysis

*Coordinators: Andrew Yates and Klaus Berberich*

Finding relevant documents and extracting information from them are common NLP tasks, with downstream applications such as search, knowledge base construction, and question answering. Consequently, much prior work has considered the core problems of identifying documents relevant to a query, extracting structured information from documents, and analyzing the contents of documents. One might be tempted to assume that these are solved problems given the widespread usage of Information Retrieval through Web search, of Information Extraction through the use of methods that rely on structured data extracted from text (e.g., question answering over knowledge bases), and of Text Analysis through the use of technologies that rely on document-level analysis (e.g., sentiment analysis for monitoring sentiment about a product). These tasks have been extensively studied under specific assumptions and within limited domains, however, prompting us to develop approaches that relax these assumptions and apply these tasks to more complex domains.

Our work in this space covers a variety of research directions. In the context of Information Retrieval, sections 39.6.1, 39.6.4 and 39.6.7 describe work reducing the handcrafted features in query-document ranking methods and applying search to complex new domains, such as identifying biomedical literature relevant to a clinical case report and personalizing search over health forums. In section 39.6.8 we address the problem of searching knowledge bases when exact match semantics are insufficient, such as when a query returns so many results that ranking is required and when a query must be relaxed to return any results. Sections 39.6.2 and 39.6.3 bridge the gap between Information Retrieval and Text Analysis with approaches for going beyond keyword queries by identifying temporal events in documents and incorporating semantic annotations into queries. Similarly, in section 39.6.9 we address the
problem of identifying relationships between quantities in tables and in text with the goal of developing a next generation search engine capable of answering queries containing quantities. Finally, we describe our work applying Information Extraction to the difficult domain of dialogue in section 39.6.5 (e.g., inferring a speaker’s profession from their utterances) and extracting richer, higher-arity relations in section 39.6.6.

### 39.6.1 Neural Matching and Ranking

*Investigators: Klaus Berberich, Kai Hui and Andrew Yates in cooperation with Arman Cohan, Luca Soldaini, Sean MacAvaney (Georgetown University), Gerard de Melo (Rutgers University), Ben He, Canjia Li, Le Sun, Yingfei Sun, Le Wang and Jungang Xu (University of Chinese Academy of Sciences)*

Despite the widespread use of deep neural models across a range of linguistic tasks, the extent to which such models can improve information retrieval (IR) and what components an IR retrieval model should include remain open questions and hot research topics [9]. For ad-hoc IR tasks, the goal is to produce a ranking of relevant documents given an open-domain (“ad hoc”) query and a document collection. We study the problems of (i) designing neural models for ad-hoc IR and (ii) adapting neural models to perform other tasks, such as complex answer retrieval.

Given the success with traditional IR approaches of modeling positional information, such as term proximity and term dependencies [11, 8], we propose the PACRR model [3], which captures n-gram signals by applying convolutional neural network layers to an embedding similarity matrix [4] that represents a query-document pair. We find that PACRR outperforms previously proposed neural IR models on a standard Web search benchmark. Beyond positional information, there still exist multiple factors that are believed to improve the performance of a retrieval model, including disambiguation of query terms [12] and the density of relevant information in a document. [1] We improve upon PACRR’s performance by modeling these factors in a neural framework, along with a novel query regularization approach, in the Co-PACRR model. [5] Finally, we propose a framework for performing query expansion through pseudo-relevance feedback [10, 2] with neural IR models. [6] This is a non-trivial task because, unlike traditional IR models, neural models generally lack an explicit term weighting mechanism that can be used to reduce the impact of expansion terms. Our approach, called NPRF, treats a given neural model as a black box and uses it to consider feedback document-document similarities between feedback documents from an initial result set and other documents. While this approach improves over the base neural model’s performance on standard ad-hoc benchmarks, it is limited in that it can only be used with unigram models due to assumptions made to reduce its computational cost. We are investigating ways to remove this limitation in ongoing work.

We adapt the PACRR model to perform complex answer retrieval (CAR) [7], which is the task of retrieving answer paragraphs to queries that have multi-faceted or nuanced answers. For example, in the context of CAR, the query coffee should return a broad overview of the subject including paragraphs across facets such as its history, cultivation, processing, health effects, roles in various cultures, etc. The extent to which facet terms appear in relevant paragraphs varies greatly with the facet. For example, paragraphs discussing coffee’s history rarely use the term “history”, and are thus more difficult for a retrieval model to identify,
whereas paragraphs discussing its health effects are likely to include at least one of the two terms. We call this idea *facet utility* and provide PACRR with new signals to allow it to more weakly match low-utility facets (e.g., rather than requiring relevant paragraphs to include the term “history”, the model might consider the presence of “century” or “ancient” to be sufficient). We find that considering facet utility significantly improves PACRR’s performance on the CAR task.

References


39.6.2 Event-centric Search and Exploration

Investigators: Arunav Mishra, Dhruv Gupta and Klaus Berberich in cooperation with Avishek Anand (L3S Research Center), Vinay Setty (University of Stavanger), Mittul Singh (Aalto University) and Dietrich Klakow (Saarland University)

Human history can be described as a chain of important events, and thus they are the center of discussion in many kinds of document collections, e.g., news archives and social media. To leverage events for better search and exploration of such document collections, a better understanding of which events a document alludes to is required. One line of our research in the past two years has been to come up with better models of events (e.g., their importance and their associated temporal scope) based on their descriptions in documents, which can then be used to provide users with improved means of search and exploration on document collections.

Estimating Event Importance

Given an incoming stream of news articles from different news sources, an acute problem is to identify ongoing events, as described in a multitude of articles, and to estimate their importance. This has direct applications in news-aggregation services that seek to inform users about important ongoing events and refer them to recent news articles with details about them. In [7] we develop an approach to identify ongoing events in news articles that relies on a clustering approach inspired by but different from traditional density-based clustering. The approach views news articles as bags of shingles (i.e., short text segments) and named entities (i.e., persons, locations, and organizations). In the clustering step, the approach groups together news articles that refer to similar named entities using similar textual descriptions. In our experiments, we observed that the obtained clusters are able to capture ongoing events with high accuracy. Going beyond the mere detection of relatively short-span events (e.g., a series of attacks in Syria), our approach is also able to chain them together to capture long-span events (e.g., the Syrian Civil War), again making use of the representation of articles as bags of shingles and named entities. Since not all events are of equal importance, a last but crucial building block in our approach is its ability to rank events according to their importance. To this end, we consider different signals including the number of sources that report on a specific event, their geographic diversity, and their authority estimated based on the citations that they receive from Wikipedia. Experiments on two large-scale real-world collections of news articles demonstrate that our approach is able to detect events with high accuracy and that the determined ranking of events shows good agreement with the manually curated Wikipedia Current Events Portal.
Generating Textual Summaries of Events

In our earlier research [5] we had proposed an approach to automatically generate so-called event digests. Given an event query, consisting of a textual description, geographic locations, a time interval, and involved named entities, the approach would construct a summary of the event having a specific length. To this end, our approach relied on extractive summarization based on statistical language models for text, temporal expressions, geographic locations, and named entities. From a set of pseudo-relevant documents retrieved for the event query, a subset of sentences was extracted using an ILP, aiming at high relevance of individual sentences to the event and high diversity within the set of sentences. Within the reporting period, we built on this earlier research in two directions.

Commonly used effectiveness measures for automatic text summarization, such as the family of ROUGE measures, compare the generated summary against a gold-standard summary created by a human. One way to conduct this comparison is to measure the precision and recall of the generated summary in terms of \( n \)-grams that are contained in the gold-standard summary. What is thus completely ignored by these measures is the narrative flow of the generated summary, i.e., whether the order of sentences included in the generated summary is meaningful to users. In [6] we investigate through a crowdsourced study to what extent the order and proximity of sentences in a summary affect its readability as perceived by users. Starting from a set of gold-standard summaries, we create modified versions of these that differ in terms of sentence order from the original. This includes versions with randomized sentence order, reversed sentence order, and a sentence order placing adjacent sentences from the original summary at maximal distance. Users, recruited using a crowdsourcing platform, are asked to express their relative preference for different versions of a summary. Our findings include that sentence order and their proximity have a substantial effect on the user-perceived quality of summaries.

In [8] we extend our earlier approach [5] by integrating richer language models based on deep learning. More specifically, we propose two novel language models based on long short-term memory (LSTM) that can better capture context across sentence boundaries, resulting in a more informed selection of sentences for the final summary. Our experiments on the testbed from [5] show that integrating the novel language models results in small yet statistically significant improvements in terms of ROUGE-2 and ROUGE-SU4 over our earlier method and other state-of-the-art extractive summarization methods.

Exploring Events in the News

Events can provide an effective means to explore news collections. In [3] we develop a system coined BioNEx that is focused on the biomedical domain and uses biomedical news events (e.g., disease outbreaks) to let users explore a news collection. To detect such biomedical news events, the system relies on a clustering approach that views news articles as bags of shingles and disambiguated named entities including biomedical named entities (e.g., diseases), reusing some of the ideas described in [7]. The system then provides the user with a rich interface, shown in Figure 39.22, which visualizes clusters along different dimensions. Users can then interact with the shown identified events and, for instance, explore related historical events or look into the underlying news articles that report on an event.
Temporal Scoping of Events and Knowledge Graph Facts

Descriptions of events can be incomplete and, as an important constituent, lack information about the time when they happened. In the reporting period, we developed two approaches that can help with determining a temporal scope for an incompletely described event.

In [2] we make use of neural word embeddings to estimate the temporal scope of an event based on a short textual description of it. To this end, in a first step, we learn embeddings, relying on the well-established skip-gram model, both for words and temporal expressions (e.g., years, months, days) on large-scale document collections including Gigaword and TREC ClueWeb12. This first step thus embeds words and temporal expressions in a common vector space, allowing us to estimate the semantic similarity between two words, two temporal expressions, or a word and a temporal expression. Given a short textual description of an event, in a second step, we need to identify a semantically similar temporal expression, which is then reported as the temporal scope of the event. For this second step, we consider two options: (i) early fusion first aggregates the learned embeddings for the words from the textual description and picks the semantically most similar temporal expression; (ii) late fusion identifies semantically similar temporal expressions for each word from the textual description and aggregates the resulting rankings of semantically similar temporal expressions. Our experiments based on event descriptions from year pages in Wikipedia show that our methods typically return the correct year for an event at rank 2 or above, thereby outperforming state-of-the-art methods from the literature.

Many facts in knowledge graphs correspond to events. Consider, as a concrete example, the fact Albert Einstein was awarded Nobel Prize, which may lack a temporal scope (here: the year 1921) in a knowledge graph. In [4] we resort to document collections annotated with temporal expressions to estimate a temporal scope for a given fact from a knowledge graph. To this end, we first make use of the lexical knowledge contained in knowledge graphs (i.e., surface forms and relation paraphrases) to construct a keyword query that can be used
to identify documents containing sentences that refer to the fact. Such a sentence must include surface forms for the involved named entities, a paraphrase for the relation, and mention a temporal expressions. From the identified sentences, using a model for temporal expressions proposed in [1], we can determine the most likely temporal expression at a specific granularity (e.g., year, month, or day). Our experimental evaluation based on a temporal fact benchmark from the literature and using Gigaword as well as The New York Times Corpus as background document collections shows that our proposed method typically returns the correct year for a fact at rank 3 or above. For more than 25% of the facts from the benchmark, the correct year is returned at rank 1, which means that the method could pick the correct year automatically.

References


39.6.3 Querying Text with Semantic Annotations

Investigators: Dhruv Gupta, Klaus Berberich and Jannik Strötgen in cooperation with Demetris Zeinalipour (University of Cyprus)

It is nowadays feasible, thanks to progress in natural language processing methods and increased computing capacity, to adorn even large document collections with linguistic and semantic annotations. Such annotations include part-of-speech tags, named entities (e.g., persons, organizations, and locations), and temporal expressions (e.g., years, months, or days). Semantic annotations of this kind and the better understanding of document contents that they bring allow to improve search functionality for end users but also to speed up knowledge-centric downstream applications (e.g., information extraction). Making use of semantic annotations for an improved search experience and making them accessible more efficiently for downstream applications has been an active line of our research during the reporting period.

Generating Semantic Aspects

For ambiguous and underspecified queries it is often hard for users to sift through query results and identify groups of documents that relate to a common aspect. Earlier attempts to remedy this and help the user navigate through the result have included result clustering (e.g., based on topic models) and faceted search. While the former often requires careful parameter tuning to yield meaningful clusters, which might then be hard to interpret, the latter relies on manually curated facets which have to be assigned to documents. In [3] we propose an approach that makes use of semantic annotations in result documents to identify so-called semantic aspects. Assuming that a user has issued the underspecified query "olympic games," a relevant semantic aspect, according to our definition, combines different kinds of semantic annotations that are found to co-occur frequently. For the example at hand, a relevant semantic aspect could combine a temporal expression (e.g., the year 2012), with a set of person named entities (e.g., Usain Bolt and Michael Phelps), and a location (e.g., London) to represent documents related to the 2012 London Olympics.

To identify semantic aspects in a set of pseudo-relevant documents retrieved for the original query, we rely on an approach coined xFACTOR that is inspired by the Apriori algorithm for frequent itemset mining. Much like Apriori, our approach considers sets in increasing order of their cardinality. Different from Apriori, though, xFACTOR does not merely determine support, i.e., count how often a set of semantic annotations has been observed. Instead, it is informed about the inherent semantics of the different kinds of annotations and takes it into account when determining how often an annotation has been witnessed. Thus, for geographic locations as a concrete example, an occurrence of LONDON would count towards how often the geographic location UNITED KINGDOM has been observed. The approach determines a ranked list of semantic aspects, each of which is linked to the documents in which it has been observed, making it easy for users to explore the query result.

Our experimental evaluation on four document collections, including TREC ClueWeb 12 and Gigaword, showed that xFACTOR is able to identify meaningful semantic aspects for underspecified queries, substantially outperforming competitors based on topic models.
Indexing Semantically Annotated Document Collections

Semantic annotations allow for more powerful querying functionality, which can be useful for expert users and downstream applications (e.g., for information extraction). Thus, as a concrete example, if documents have been semantically annotated, a user could issue a query such as \((\text{paris hilton}) \oplus (\text{PERSON})\), asking for occurrences of the phrase “paris hilton” that coincide with a person named entity. When semantically annotated, a document can be viewed as consisting of layers including the layer of words and multiple annotation layers on-top of it. Figure 39.23 shows a concrete example of a semantically annotated document.

In [1] we propose a query language, inspired by regular expressions, to formulate semantic queries on annotated document collections. Moreover, we investigate how such document collections can be indexed, so that semantic queries can be processed efficiently. To index semantically annotated document collections, we stick to the well established combination of an inverted index and a direct index, which is common in many retrieval systems. However, two key challenges arise that need careful design decisions.

First, given that annotations can span multiple words, it is no longer sufficient to keep track of positions, but the index has to record position intervals at which a term (i.e., a word or an annotation) occurs. This also entails that algorithms for processing specific kinds of queries (e.g., a phrase or wildcard query) have to be modified to consider position intervals instead of simple positions.

Second, and more challenging, one has to decide which terms should be indexed in the inverted index. Simply adopting the standard approach from text retrieval to index
The GUI features a search bar in which the user can enter the query using the operators described in Section 2. The user can then select to formulate the following query:

(LOCATION) joined nato (DATE)

This query retrieves sentences matching the phrase “joined nato” and including a geographic location and a temporal expression.

Figure 39.25 shows a screenshot of GYANI’s GUI. The query “(LOCATION) joined nato (DATE)” retrieves sentences matching the phrase “joined nato” and including a geographic location and a temporal expression.

n-grams up to a specific length falls short for the semantic annotation layers. To illustrate this, consider a combination of part-of-speech tags (e.g., IN DT), which is likely to occur very often in any given document collection and is thus not helpful for query processing. Figure 39.24 illustrates the design space of indexing units that we consider for GYANI. The indexing units that we found to give the biggest benefits were k-Fragments and k-Stitches. The idea here is to have an indexing unit that spans layers of the document. A concrete 2-Fragment would be (last year)((2018, 2018)) providing efficient access to any document where “last year” has been annotated as referring to the year 2018. Similarly, the 2-Stitch (PERSON)((1990, 1990)) provides efficient access to any document where a person named entity is mentioned together with the year 1990 within the same sentence. While adding these indexing units comes at the cost of increased index size, we observe across multiple document collections that the additional cost is comparable to that of an n-gram index keeping track of unigrams, bigrams, and trigrams.

To make effective use of the additional indexing units, we design efficient algorithms that implement the operators (e.g., wildcard and stacking) of our query language. When provided with a semantic query issued by a user or a downstream application, GYANI relies on a greedy algorithm to compile a query plan consisting of operators to process the given query.

Our experimental evaluation studies the efficiency of GYANI for several knowledge-centric tasks, as found in downstream applications, on three large-scale document collections including Wikipedia and Gigaword. Knowledge-centric tasks that we consider include information extraction (e.g., finding all occurrences of (PERSON)(raised in)(LOCATION)) and fact spotting (e.g., finding sentences mentioning that a company has its headquarters in a particular city). We observe that GYANI typically accomplishes speed ups in query-time response time by at least an order of magnitude, when compared to simpler baselines that only instantiate a subset of the indexing units shown in Figure 39.24.

GYANI has been implemented as an end-to-end system [2] and been deployed on several document collections. Figure 39.25 shows a screen shot of its GUI. Users can issue a semantic query according to our query language and are presented with matching sentences from
the document collection, in which the relevant semantic annotations are highlighted. The example query shown in Figure 39.25 demonstrates how GYANI can be used for the simple information extraction task of identifying countries with the year in which they joined NATO. This query could be issued by an application, which could then further process the sentences retrieved by GYANI, for instance, by aggregating identified years based on the country.

References


39.6.4 Analyzing Health Forums and Clinical Texts

Investigators: Andrew Yates in cooperation with Arman Cohan (Allen Institute of AI and Georgetown University), Bart Desmet (Ghent University), Nazli Goharian, Sean MacAvaney, Luca Soldaini, Sydney Young (Georgetown University) and Ayah Zirikly (U.S. National Institutes of Health)

We consider the problems of supporting mental health diagnoses and medical doctors’ clinical decisions. In the context of mental health, the goal of this project is to develop tools for studying mental health conditions using social media. This encompasses both methods for identifying users suffering from a health condition (“diagnosed users”) and models for differentiating between diagnosed users and control users. This work is motivated by the observation that mental health has a unique connection to language. That is, mental health problems differ from many other medical conditions in that they cannot be diagnosed with methods like imaging techniques (e.g., x-rays) or laboratory analyses (e.g., blood tests). Thus studying patients’ language use has potential for achieving a better understanding of the underlying disease. We propose a pattern-based approach for identifying Reddit users who claim to be suffering from a mental health condition (e.g., PTSD, depression) and matching them with appropriate control users [1, 6, 3]. Using this data, we developed a neural approach for differentiating between depressed users and control users, finding that the model learned to look for signs that a user consistently expressed negative sentiments over time. [6] While this observation is not novel, it demonstrates that a model can differentiate between these users classes based on their general language use. This work received a Best Paper Award at EMNLP 2017. In the context of support forums in which users suffering from depression chat with each other and with trained moderators, we additionally propose neural and feature-based approaches for helping the moderators identify posts that suggest a risk of self-harm and thus require prompt attention [2, 6].
In the context of supporting doctors’ clinical decisions, we are motivated by the observation that it is difficult for doctors to keep current with the large and rapidly increase volume of biomedical literature. Clinical Decision Support (CDS) is an approach for reducing this problem by identifying biomedical documents to support clinical decisions (e.g., diagnosis of a patient’s condition, tests that can be conducted to inform a diagnosis, and treatment options for a condition). In the context of the TREC CDS shared task, the input for a CDS system is a medical case report, and the output is a ranked list of biomedical documents that can inform the doctor’s decision. In our work we focus on the problem of reformulating the medical case report into a query that effectively identifies relevant documents (i.e., we both expand the case report with informative terms and remove uninformative terms). We propose a neural reformulation approach that combines the agreement between terms and a learned case report representation with other features in order to predict which case report terms should be removed from the query and which candidate terms should be added to it [4, 5]. We find that this approach performs comparably with the best-performing methods from the TREC CDS benchmark.

References


39.6.5 Information Extraction from Dialogues

Investigators: Anna Tigunova, Andrew Yates, Paramita Mirza, and Gerhard Weikum

Recently dialogue agents have become an essential part of daily life, assisting various aspects of it via natural language interaction with users. With an increasing focus on the use of
statistical and machine learning based approaches, the last few years have seen some truly remarkable conversational agents appear on the market (e.g. Apple Siri, Microsoft Cortana, Google Assistant). These agents can perform simple tasks, answer factual questions, and sometimes also aimlessly chit-chat with the user. Still, creating agents capable of holding personalized conversations, while at the same time building and expanding their knowledge repository about users’ personal information, remains a challenge. In this work, we investigate the topic of capturing background knowledge about users from what they said, that would make intelligent agents capable of more friendly and effective interactions.

**Personal Knowledge Base from and for Dialogues**

As part of our research towards Personal Knowledge Bases (PKB, cf. Section 39.4.6) for individual users, we propose to extract such personal knowledge from the user’s utterances directly. The advantages of having an explicit PKB are as follows: (i) it can be easily reused and updated to support dialogue agents, especially when the current conversation span is short, and (ii) it is transferable for different applications such as personalized search and recommendation systems. Moreover, it provides a convenient way of explaining and justifying the agent’s statements to the human whenever requested.

There are three issues, which should be taken into consideration to construct a PKB from and for dialogues:

1. What is relevant and interesting personalized knowledge about the user?
2. How to extract or infer such knowledge from the user’s utterances?
3. How to incorporate the background knowledge into dialogue agents?

Addressing the first issue, we consider interesting facts about a user to be attributes (age, gender, occupation, etc), interests and hobbies, relationships to other people (family status, names of friends, etc) or sentiments towards certain topics or people. Many of those are subjective and mutable, however, hence in this work we focused on the basic ones, such as gender, age, profession and family status.

The second issue is concerned with Information Extraction (IE) from text, assuming we have transcribed users’ dialogues, textual chat logs or users’ posts on discussion forums. Prior works on IE for Knowledge Base Population mostly focused on long structured formal texts, such as Wikipedia pages or news articles. The conversation domain on the other hand is much more challenging: the utterances are short and non-explicit, noisy, and mostly in informal language. Therefore, it is an additional challenge to be able to distill valuable information from them, and this is the issue we focused on, which will be elaborated further in the next section.

The third issue on incorporating PKB into a dialogue system is a natural research direction once we have built the PKB. We aim to address this in the future and are actively working towards this goal.

**Hidden Attribute Model (HAM)**

We devised a neural architecture, called *Hidden Attribute Model (HAM)* [1], which is trained to predict personal attributes (e.g., profession or age) of a given user. More formally, given
a subject $S$ and a predicate $P$, our goal is to predict a probability distribution over object values $O$ for the predicate based on the subject’s utterances from a dialogue corpus (e.g., a movie script). Thus, the model is trained on subject-predicate pairs on a per-predicate basis. For example, given the movie script excerpt shown in Figure 39.26, the profession predicate, and the subject Edwards, policeman, should be ranked as the subject’s most likely profession.

The operation of HAM is the following. Each subject is associated with a sequence of utterances containing terms; each term is represented as an embedding vector. The model takes the subject’s utterances as input and creates the latent representation via either convolutional networks or attention mechanisms. The created representation is then used to perform classification of the object values for the subject with respect to a certain predicate. In our current work, we leverage gender, age, profession and family status as predicates. We trained and tested our models on various datasets. First, we create a unique dataset of film characters from a corpus of movie scripts, labeled with their profession, age and gender. Secondly, we construct a dataset containing profession, gender, age and family status labels for Reddit users, with the use of pattern matching. Finally, we use PersonaChat dataset [2] as existing benchmark. Due to the different nature of these datasets, we can stress-test our model in different scenarios, including the challenging setting of transfer learning.

Extensive experiments demonstrate the viability of our methods and their superior performance compared to state-of-the-art baselines, in terms of various evaluation metrics (MRR and AUROC). Moreover, our qualitative study has shown that the model is able to produce valuable terms, resulting from attention distribution or convolution operation, which are the cues for the model to learn the correct label for a given attribute. For example, our approach infers that a subject who often uses terms like theory, mathematical, and species is likely to be a scientist, while a subject who uses terms like senate, reporters, and president may be a politician.

References


39.6.6 Information Extraction for Higher-Arity Relations

Investigators: Patrick Ernst, Amy Siu, Gerhard Weikum

Knowledge bases, such as YAGO, DBpedia, KnowLife, etc., have proven their usefulness for many applications. They are key components for search engines and recommender systems as well as domain-specific use cases, such as health care (e.g., curation of biological databases, medical question answering, and guided search and exploration of biomedical literature). However, a major limitation of many knowledge bases is that the majority of their facts refer to binary relations only, i.e. relationships between two entities. For example, YAGO knows that Marie Curie has won a Nobel Prize in 1903 and another one in 1911, but it does not keep the fields (Physics and Chemistry) as explicit predicates. KnowLife includes facts about drug treatment of diseases, but no information about the appropriate dosages and target groups. Especially in such advanced domains, it is often crucial to consider higher-arity relations. An example is to capture which drug is used for which disease at which dosage (e.g. 2.5 mg/day) for which kinds of patients (e.g., children vs. adults).

To overcome this problem, we introduced *HighLife* [1], a text-based knowledge extraction method, which goes beyond binary facts and is able to populate knowledge bases with higher-arity facts. *HighLife*’s method is twofold. We use seed facts as distant supervision to learn patterns from natural language text, apply these patterns to extract fact candidates, and iterate these steps. A key difficulty addressed by *HighLife*, lies in the observation that higher-arity facts are often expressed only partially: with some but not all of their arguments. For example, the statement, that “Google bought Nest for a price of $3.2 billion in 2014” encoded by the higher-arity fact \(\text{acquired}(\text{Google}, \text{Nest}, 2014, \text{USD} 3.2\text{bbl})\), could be expressed as:

- “Google acquired Nest in 2014” without stating the price, or
- “Google bought Nest for 3.2 Billion” without giving a date.

We address this issue by extending the fact-pattern duality paradigm to higher-arity as well as partial patterns and facts. While achieving high recall, the approach is susceptible to noise and target drifts. Therefore, we use constraint reasoning to eliminate spurious fact candidates. For example, we can apply type constraints to identify when facts about winning the Pulitzer prize are for movies or songs (instead of books), and we can exploit value constraints when confusing the numbers for amount and year on a company acquisition. To incorporate such constraints we extend a Weighted MaxSat-based reasoner to higher-arity case, including reasoning over the composability of partial fact candidates into full facts.

By conducting extensive experiments we showed the validity and versatility of *HighLife*. The method is general and applicable to any domain and a wide range of text genres. Furthermore, the experiments demonstrated that *HighLife* is able to harvest facts with higher-arity as well as high precision.
39.6.7 Search on Health Corpora

Investigators: Erisa Terolli, Patrick Ernst, Gerhard Weikum

Health forums have become an important source of information for people with medical conditions. Users post questions, often along with specific descriptions of their personal case histories, and other users reply in a community question-answering (CQA) manner. These online communities comprise other patients as well as medical doctors who bring in their expertise (including the possible advice to visit a doctor). Typical use cases of posts include patients who are unhappy with their current treatment, because they have non-standard symptoms and the diagnosis is unclear (even after involving doctors and hospitals), or because they suffer from adverse side effects of their drugs or are otherwise dissatisfied with their therapies. In such cases, the user is looking for advice on additional hypotheses and tests or alternative treatments. Either way, the experience of patients with similar conditions could be helpful. Common to these cases is that the user’s information need is highly specific for the individual history; it is not about general information about certain diseases and drugs (which could be obtained from online portals such as mayoclinic.org).

All health forums have a search box, to support users beyond the CQA functionality. In practice, however, the retrieval results are often too broad and not really focused on the patient’s specific needs. As an example, consider a user posting the question “Does this sound like a thyroid / hormonal problem?”. If we use this literally as a query, the answers will be way too general. The post body contains a 50-line detailed description of the user’s symptoms, blood test results (so she visited doctors already), drug prescriptions and self-observations. We could ask the user to formulate a well-crafted keyword query from this long text, but most end-users would be unable (and unwilling) to do so. For this reason, our goal is to support end-users for search health forums, by automatically generating personalized queries from their questions and full post contents.

Our approach to tackle this problem is based on the method of query expansion as shown in Figure 39.27. We start with the user question, that is, the post title, as if it were a keyword query, and then expand and personalize it by judiciously selecting and weighting terms from the body of the user post. This method has been integrated into a prototype system that supports exploratory search over a large body of documents from health communities. We explore methods for query expansion. In contrast to the mainstream approach of expanding by semantically related terms using large background corpora or thesauri, or expanding by pseudo relevance feedback, our methods leverage individual user posts for personalization.

We devise various ways of identifying the most informative biomedical terms and entities in user posts, based on a collection of patient-oriented health-portal pages (ehealthforum.com/healthcenter/), using the UMLS thesaurus (www.nlm.nih.gov/research/umls/) as entity repository. The query weights of the expansion terms are computed from statistics of a large health forum (ehealthforum.com/health/health_forums.html).
An extensive experimental studies with model patients for 20 different diseases on a full-fledged prototype system for health search demonstrate the viability of our approach and its benefits.

### 39.6.8 Ranked Query Processing on Knowledge Graphs

**Investigators:** Hiba Arnaout and Gerhard Weikum in cooperation with Mohamed Yahya (Bloomberg LP), Shady Elbassuoni (American University of Beirut), Madhulika Mohanty (Indian Institute of Technology Delhi), and Maya Ramanath (Indian Institute of Technology Delhi)

RDF knowledge graphs are typically searched using the semantic query language: SPARQL. Since SPARQL follows exact match semantics, the queries may return too many, too few, or no results. In the case where too many results are returned, result ranking is needed. And when too few or no results are returned, query relaxation is needed.

**Effective Searching of RDF Knowledge Graphs**

In this work [1], we propose a general framework for effective searching of RDF knowledge graphs. Our framework allows users to search the underlying knowledge graph using both SPARQL queries, as well as SPARQL queries extended with keywords. Our framework supports result ranking by means of a novel ranking model based on statistical machine translation. To do so, we extend our case-study knowledge base: DBpedia [2], with triple weights and weighted keywords. A sample is shown in Table 39.3.

In addition, our framework supports automatic query relaxation to improve the recall of queries with no results. Our goal is to output results that at least partially match a given user query. This is done in judicious manner to ensure that the relaxed queries are close in spirit to the original query, thus improving recall without unduly sacrificing precision.

Finally, our framework provides various notions of result diversity that are achieved using the Maximal Marginal Relevance [3] re-ranking approach. The different notions cover diversifying the result triples, as well as the keywords(text) associated with the result triples.
Speculative Query Planning for Joins over Knowledge Graphs

In this work [4] we propose Spec-QP, a strategy for top-k query processing in the case where a query can have multiple relaxations. We have used a speculative approach for pruning the relaxations which are not likely to contribute answers to the top-k results. It is based on pre-computed statistics about the distribution of scores for triple pattern matches. The relaxations predicted not to contribute towards top-k answers are not processed.

We have experimented over an extended knowledge graph (XKG) [5] and Twitter hashtag data to show that Spec-QP is a cost-efficient technique for supporting relaxations 39.4. We also demonstrated the practical usability of our technique by implementing it over two popular database engines, PostgreSQL and Virtuoso.

References


39.6.9 Quantities in Text and Tables

Investigators: Yusra Ibrahim, Vinh Thinh Ho and Gerhard Weikum in cooperation with Mirek Riedewald (Northeastern University) and Demetris Zeinalipour (University of Cyprus)

Numbers are an integral part of language, though they are often overlooked by Natural Language Understanding (NLU) and Information Extraction research. Numerical quantities appear in scientific research results, financial reports, and medical records, among others. They are arranged in tabular formats or infused in natural text. Tapping into this wealth of resources by understanding quantities in web tables and text surrounding them can lay the first brick in building next-generation Information Retrieval systems that are capable of answering complex queries about quantities. In our research we target the following challenges:

- Understanding quantities in tables in the light of their surrounding text.
- Answering complex questions involving quantities.

BriQ: Bridging Quantities in Tables and Text

To understand the relation between quantity mentions in text and tables, we introduce BriQ. BriQ is a framework that computes bidirectional linking between textual mentions of quantities and the corresponding table cells. BriQ is designed to cope with the specific challenges of approximate quantities, aggregated quantities, and calculated quantities in the text that are common but cannot be directly matched in table cells.

BriQ can handle a broad range of aggregation functions, such as sum, average, difference, percentage, and change ratio. Thus, it can support advanced content summarization and faster navigation between explanations in text and details in tables. For example, in Figure 39.28a, the phrase “total of 123 patients” refers to an aggregate value, namely, the sum of the values in the sales column. In Figure 39.28b, “the least affordable option” refers to the maximum price in a column, and in Figure 39.28c, “increased by 1.5%” refers to the rate of change.

The BriQ algorithm consists of two main stages: local resolution and global resolution. The local resolution assigns a confidence score for each candidate alignment in isolation without considering other neighboring alignments. The global resolution then takes as input the candidate alignments from the previous stage and outputs the final alignment of quantities between text and tables. It uses the local resolution’s confidence score as prior weights and employs a global inference algorithm based on random walks over graphs to resolve the alignments.
A total of 123 patients who undergo the drug trials reported side effects, of which there were 69 female patients and 54 male patients. The most common side effect is depression, reported by 38 patients; and the least common side effect is eye disorder, reported by 5 patients.

<table>
<thead>
<tr>
<th>side effects</th>
<th>male</th>
<th>female</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rash</td>
<td>15</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>Depression</td>
<td>13</td>
<td>25</td>
<td>38</td>
</tr>
<tr>
<td>Hypertension</td>
<td>19</td>
<td>15</td>
<td>34</td>
</tr>
<tr>
<td>Nausea</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Eye Disorders</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

The final ratings are dominated by the PHEV from Audi (2.67) and ICE from Volkswagen (2.67). Audi A3 e-tron is the least affordable option with 37K EUR in Germany and 39K USD in the US. The Ford Focus Electric, lowest rating (1.33), is a 2K EUR (2.3K USD) cheaper alternative with 0 CO2 emission and 105 MPGe fuel consumption.

In 2013 revenue of $3.26 billion Cdn was up $70 million Cdn or 2% from the previous year. The net income of 2013 was $0.9 billion Cdn. Compared to the revenue of 2012, it increased by 1.5%.

### Figure 39.28: Examples of Web Tables with Explanatory Texts

We conducted an extensive user study for annotating web pages to train BriQ. The annotated dataset is publicly available; we presented it along with BriQ at ICDE 2019 [2].

**ExQuisiTe: Explaining Quantities in Text**

*ExQuisiTe* is a web-based system that identifies relations between quantities in text and tables. ExQuisiTe automatically detects these relations and generates an easy-to-read document where numbers in text are linked to their source tables and respective cells. It identifies simple mentions of single-cell table quantities as well as mentions of aggregate quantities. For example, in Figure 39.29 the mention “1,683” in the text refers to a simple quantity in the table; and in Figure 39.30 the mention “5.72%” refers to an aggregate quantity (percentage) in the same table.

Furthermore, ExQuisiTe can guide *Extractive Text Summarization (ETS)* systems by emphasizing sentences with aggregate quantities. Current summarization systems [4, 1] do not include table data, and ExQuisiTe opens the opportunity for them to harness table data. Once ExQuisiTe identifies references of simple and aggregate table quantities in the text, it can suggest sentences with aggregations to be included in the summary generated by the ETS algorithm.

For example, in Figure 39.30 the highlighted sentence covers more cells in the table than the other sentences. It contains more aggregate mentions, and hence it provides a better summary, with judicious consideration of the numbers in the tables.

ExQuisiTe is based on the BriQ algorithm and consists of four configurable stages: (i) Document Extraction, (ii) Local Resolution, (iii) Global Resolution, and (iv) Markup and Summary Generation. The first stage extracts text segments and their possible related tables using string similarity measures. The second stage identifies potential alignments between quantity mentions in text and tables based on local features. Then, the third stage collectively aligns quantities in the text to their relevant quantities in tables. Finally, the system generates markup for the document with the inferred alignments and selects important sentences for summarization. We demonstrated ExQuisiTe at the Web conference (WWW 2019) [3].
Answering Complex Quantity Queries

Modern search engines fail to fulfill quantity-specific queries such as “hybrid cars produced in Germany with fuel efficiency less than 70 mpg”. Hence, we aim at bringing this gap by building the next generation search engine that are capable of answering these queries. This search engine consist of the following building blocks:

- **Information source**: defines the sources of information that we can leverage to answer quantity-related queries. We consider two common sources of information on the web: natural text and web-tables. In addition, we consider a combination of these two sources.

- **Information extractor**: extracts useful facts from raw information sources. It should also standardize the collected facts, and store them in an optimized way. Such that, they are efficiently retrieved when needed to answer a user query.

- **Query parsing**: parses natural text queries to determine the intent of the user. Quantity-related queries pose a great challenge as users can express the same quantity
in multiple ways. For instance, “100,000 USD” can be written as: “$100k”, “100k US dollars”, among others. The query parser rule is to canonicalize the query into a standardized form, which can be answered.

- **Query answering**: answers the standardized user query from the facts extracted from the information sources. It retrieve the relevant information to answer the user query, filter them, and provide an aggregated ranked list of answers that match the user query.

The text is a rich source for extracting quantities-related information, but it is the most challenging to process. A single fact can be expressed in various ways with the natural text. Hence, initially we focus our research on extracting quantity-related facts from the text.

### References


### 39.7 Natural Language Understanding

**Coordinators: Paramita Mirza, Rishiraj Saha Roy, and Jannik Strötgen**

Understanding content expressed in natural language (NL) is one of the frontiers in AI today. Until a few years back, most text or voice processing methods belonged to the general family of simple feature-driven prediction algorithms, that leveraged only the tip of the actual knowledge contained in language-based information records. Examples of such information are ubiquitous: they can be text documents on Wikipedia, news sites, or the open Web, product descriptions and reviews on online marketplaces or logs of voice inputs to mobile assistants or translation websites. While rapid progress has been made by the concerted efforts of the NLP, IR, ML, and AI communities, there is still a long way to go before we can say that machines can really “understand” what is expressed in textual documents or spoken utterances and accordingly take downstream action: be it harvesting deep knowledge, or providing interactive responses with reasoning. In our group, we have tried to push the state-of-the-art for several core topics that drive advanced language understanding, with a focus on textual inputs. These research areas include understanding named entities, inferring temporal and spatial contexts, identifying counting quantifiers, analyzing emotions, modeling dialogues, and answering objective questions.

To get a qualitative feel of the vision of this line of work, let us take the following passage as an example: “The longest of Tom’s three marriages (to Mimi, Nicole, and Katie) was with
Nicole, that lasted an impressive thirteen years. He adopted Isabella and Connor when he was with Nicole, and has a biological child, Suri, from his marriage to Katie. Just yesterday in the capital, Nicole made some stunning revelations about her marriage to Tom, that got the celebrity world talking...”. Our research on named entities would try to detect names like “Tom”, “Mimi”, and “Nicole”, map them to the intended people Tom Cruise, Mimi Rogers, and Nicole Kidman (say, their respective pages in Wikipedia or Wikidata), and provide additional information like categories and salience. Going further, temporal and spatial markup would annotate the passage with time scopes of the entities and events involved, and normalize references wherever possible (“yesterday”, “the capital”). Counting quantifier extraction would give us a fact that Tom has been married “three” times, with which the number of ‘spouse’ facts for him in a Knowledge Base (KB) should be consistent. An emotion assessment will gauge if there are specific emotions associated with the text (indicated by words like “impressive”, or “stunning”), potentially indicating emotion trajectories of personal relationships between entities in the article. A dialogue model over this passage, and related documents, say, would try to sustain a conversation with a human along the following lines. User: “Wow, all of Tom’s marriages lasted more than two years!!” ; Machine: “Right! How often do we see that happening??” ; User: “What about Brad Pitt??” . Finally, a question answering system would aim to provide crisp answers to questions like “Who is Tom and Katie’s daughter?” or “How many years did the longest marriage of Tom Cruise last?”. Our research in each of these key directions is explained in the following subsections.

### 39.7.1 AmbiverseNLU

Investigators: Johannes Hoffart, Dragan Milchevski, Luciano Del Corro, Ghazaleh Torbati, in cooperation with Marco Ponza (University of Pisa)

Natural language understanding (NLU) or machine reading is enabled by the availability of both natural language processing methods (tokenization, pos tagging, dependency parsing) and large knowledge bases such as YAGO, Wikidata or DBPedia. NLU methods, such as entity linking or fact extraction, often depend on the output not only of basic NLP methods but also of other NLU tasks, e.g. facts often contain entities in their subjects or objects. The best way to research and develop, and also use NLU methods is thus as a comprehensive toolkit.

AmbiverseNLU is one such toolkit comprising numerous of the methods researched in the past years in our group: Knowledge-driven Named Entity Recognition, based on KnowNER [6]; Entity Linking, based on AIDA [4]; Concept Linking, an extension of the AIDA algorithms from named entities to conceptual nouns; Open Information Extraction, based on ClausIE [2] and MinIE (see below), and the extraction of Salient Facts (see below about SalIE).

Basic NLP tasks needed for AmbiverseNLU are covered using Apache UIMA\(^7\) and DKPro\(^8\) frameworks, which not only provide many components for immediate use, but also enable a modular development and extension of the toolkit. AmbiverseNLU has been made available as open source\(^9\) in December 2018, both for the benefit of the research community and for

---

\(^7\)[https://uima.apache.org/](https://uima.apache.org/)

\(^8\)[https://dkpro.github.io/](https://dkpro.github.io/)

\(^9\)[https://github.com/ambiverse-nlu/](https://github.com/ambiverse-nlu/)
broader development of new modules. AmbiverseNLU can be used as standalone processing toolkit, as Web Service, or directly for end users as a Web-based platform, see Figure 39.31.

**MinIE: Minimizing Facts in Open Information Extraction**

Open Information Extraction (OIE) [1] is the task of generating a structured, machine-readable representation of the information expressed in natural language text. In contrast to traditional IE, OIE systems do not require an upfront specification of the target schema (e.g., target relations) or access to background knowledge (e.g., a knowledge base). The extractions are (usually) represented in the form of surface subject-relation-object triples. OIE serves as input for deeper understanding tasks such as relation extraction, knowledge base construction, question answering, or information retrieval.

The use of OIE in downstream applications require the extractions to comply with the specifications of those applications. Most of them, have specific requirements regarding the structure of the fact and its constituents as well as the type of its arguments. For example, knowledge-base construction frameworks often require an entity-relation-entity format [e.g., they will prefer an extraction of the form ("Albert Einstein", "was born in", "Ulm") to ("Albert Einstein", "was born", "in Ulm"). Furthermore, applications will require the extractions to be as minimal as possible, meaning that they need to carry the information in the text in the most compressed possible way so that it can serve as input to automatic reasoning.

MinIE [3] is an OIE system that aims to address and trade-off both goals. It is built on top of ClausIE citeclausie, a state-of-the-art OIE system that achieves high precision...
Semantic annotations are detected via rule-based methods that are domain independent and considered to be safe. Negation, for example, is detected via a lexicon of common words (e.g., no, none, hardly, etc). If one of those words occur in the context of the relation, then the fact is considered to be negated. In addition, MinIE reformats the code so that the head of the object are entities. This is that by using principles of the English language and exploiting the structure if the dependency tree, such that all ancestors of the top-most entity in the object are pushed towards the relation so that the semantic flow of the extraction is not broken.

MinIE further minimizes the constituents of the extraction, using three different modes, trading-off accuracy, and minimality (i.e., Safe, Dictionary and Aggressive). The first one, Safe, removes words which are safe to remove as they do not carry core semantic information. For example, "great city" will become "city", and "Mr. Obama" will become "Obama". The dictionary mode goes further as only keeps frequent constituent phrases mined with the safe mode from Wikipedia and the New York Times corpus. Then, a phrase like "red meat" won’t be simplified as it is contained in the dictionary while "yellow uninteresting car" will become "car", as it is not frequent in the corpus. Finally, our aggressive mode removes all words deemed as not essential (i.e., all adverbial, adjective, possessive, and temporal modifiers), so that "man with apples" becomes "man".

We conducted extensive experiments showing that our method outperforms other state-of-the-art systems both in extraction precision (more than 10 F1 points) and extraction size.

Facts That Matter

OIE is recall-oriented in the sense that it aims to extract all facts present in a natural language text. However, not all facts in a document are important. A fact is relevant if it carries the essential information that the text conveys. A fact is not relevant per se but in a specific context. In an article about the US-Iran nuclear deal the fact ("US", "withdraws from", "Iran Nuclear Deal") is more relevant than ("Washington", "is", "US capital").

In this context, we introduced fact salience: The task of generating a machine-readable representation of the most prominent information in a text document. Additionally, we presented SalIE [5] (for salient information extraction) the first fact salience system. SalIE is unsupervised and knowledge agnostic, based on MinIE to detect facts in natural language text, PageRank to determine their relevance, and clustering to promote diversity. SalIE scores the facts produced by MinIE according to its relevance in a specific document.

We compare SalIE with several baselines (including positional, standard for saliency tasks), and in an extrinsic evaluation, with state-of-the-art automatic text summarizers. SalIE outperforms baselines and text summarizers showing that facts are an effective way to compress information.
References


39.7.2 Named Entity Recognition and Disambiguation

In this section, we cover our works on the two core NLP tasks of named entity recognition and named entity disambiguation.

Named Entity Recognition

Investigators: Luciano Del Corro, Tatiana Dembelova, Johannes Hoffart, and Gerhard Weikum in cooperation with Dominic Seyler (University of Illinois at Urbana-Champaign)

Named Entity Recognition (NER) is the task of detecting named entity mentions in text and assigning them to their corresponding type. It is a crucial component in a wide range of natural language understanding tasks, such as named entity disambiguation (NED), question answering, and information retrieval. In our work [2], we discuss the importance of external knowledge for performing Named Entity Recognition (NER). We present a novel modular framework that divides the knowledge into four categories according to the depth of knowledge they convey:

- The knowledge agnostic category contains only “local” features, i.e., features, which can be extracted directly from text without any external knowledge, e.g., about prefixes and suffixes of tokens.

562
- The name-based knowledge category contains features, which attempt to identify patterns in names and exploit the fact that the set of distinct names is limited, e.g., frequent mention tokens such as “John” or “county”.

- The knowledge-base-based knowledge category contains features extracted from knowledge bases, e.g., the Wikipedia link probability.

- The entity-based knowledge category contains features, which encode document-specific knowledge about the entities found in text to exploit the association between named entity recognition and named entity disambiguation.

Thus, each category consists of a set of features automatically generated from different information sources. Further, we show the effects on performance when incrementally adding deeper knowledge and discuss effectiveness/efficiency trade-offs. For this, we feed a linear chain CRF, a transparent, widely used method used for NER. Our results indicate that the amount of knowledge is highly correlated with NER performance. The configurations with more external knowledge systematically outperform the more agnostic ones.

**Named Entity Disambiguation**

*Investigators: Prabal Agarwal, Luciano Del Corro, Johannes Hoffart, Jannik Strötgen, and Gerhard Weikum*

Named Entity Disambiguation (NED) systems perform well on news articles and other texts covering a specific time interval. However, NED quality drops when inputs span long time periods like in archives or historic corpora. For instance, given the news headline *Schumacher convinced to win on Sunday.*, virtually all modern NED tools map the mention Schumacher onto the former Formula One champion Michael Schumacher, as the best-fitting entity from a Wikipedia-centric knowledge base. However, knowing that Sunday refers to August 14, 1949, i.e., ignoring the surface form of “Sunday” but exploiting normalized information, it becomes clear that the text actually refers to the German politician Kurt Schumacher. What is needed here is a better way of capturing temporal context, for both the mention Schumacher and each of the candidate entities. Figure 39.32 illustrates “temporal signatures” for sample entities with highly ambiguous names and shows that normalized temporal information from the input context, such as Sunday (1949-08-14), can provide additional cues for proper disambiguation.

In our work [1], we present the first time-aware method for NED that resolves ambiguities even when mention contexts give only few cues. The method is based on computing temporal signatures for entities and comparing these to the temporal contexts of input mentions. More precisely, temporal signatures are embeddings that reflect the importance of different years for entities. They are automatically constructed by extracting and normalizing temporal expressions in entity descriptions such as Wikipedia articles using the temporal tagger HeidelTime [3]. Analogously, temporal signals are captured in the contexts of textual mentions and represented by embeddings.

For our evaluation, we could not rely on standard NED benchmarks from CoNLL and TAC as they do not reflect temporal diversity at all. Thus, to demonstrate that time-aware NED methods can robustly cope with inputs from diachronic corpora, we propose a new
evaluation benchmark, based on the New York Times Archive, spanning more than 20 years, and the history collection historynet.com, spanning several centuries. To test the importance of time-awareness for NED, we use two settings. We enhance a basic NED system and the state-of-the-art system described in [4] by enriching both with temporal signatures and contexts. For both settings, the time-aware variants of the system outperform the original systems significantly.

References


39.7.3 Temporal and Spatial Expressions in Text

Temporal and geographic information is of major importance in virtually all contexts. Thus, it also occurs frequently in many types of text documents in the form of temporal and geographic expressions. An important characteristic of temporal and geographic expressions
is that they can be normalized so that their meaning is unambiguous and can be placed on a timeline or pinpointed on a map. In many research areas in which natural language processing is involved, e.g., in information retrieval, document summarization, and question answering, applications can highly benefit from having access to normalized information instead of only the words as they occur in documents. The just described work on time-aware named entity disambiguation [1] is a further example. While other tasks can benefit of the extraction and normalization of temporal and spatial information from texts, we describe in the following several works, in which we aimed at the extraction itself or the analysis of temporal and spatial information.

Temporal Tagging

Investigators: Lukas Lange and Jannik Strötgen in cooperation with Bernardo Magnini, Anne-Lyse Minard, and Manuela Speranza (Fondazione Bruno Kessler, Trento, Italy)

In recent years, temporal tagging, i.e., the extraction and normalization of temporal expressions, has become a vibrant research area [8]. Several tools have been made available, and new strategies have been developed. Due to domain-specific challenges [8], evaluations of new methods should be performed on diverse text types. Despite significant efforts towards multilinguality in the context of temporal tagging (e.g., [7]), for all languages except English, annotated corpora exist only for a single domain. In the case of German, for example, only a narrative-style corpus has been manually annotated so far, thus no evaluations of German temporal tagging performance on news articles can be made.

In collaboration with researchers of the Fondazione Bruno Kessler, we developed temporal tagging annotation guidelines for German [9]. In addition, we released KRAUTS, a new German temporally annotated corpus containing two subsets of news documents: articles from the daily newspaper Dolomiten and from the weekly newspaper Die Zeit. Overall, the corpus contains 192 documents with 1,140 annotated temporal expressions, and has been made publicly available to further boost research in temporal tagging.

Detecting Document Creation Times

Investigators: Jannik Strötgen in cooperation with Michael Gertz and Andreas Spitz (Heidelberg University)

For the temporal analysis of news articles or the extraction of temporal expressions from such documents, accurate document creation times are indispensable. While document creation times are available as time stamps or HTML metadata in many cases, depending on the document collection in question, this data can be inaccurate or incomplete in others. Especially in digitally published online news articles, publication times are often missing from the article or inaccurate due to (partial) updates of the content at a later time. In our joint work with researchers from Heidelberg University [5], we investigate the prediction of document creation times for articles in citation networks of digitally published news articles, which provide a network structure of knowledge flows between individual articles in addition to the contained temporal expressions.

For this, we explore the evolution of such networks to motivate the extraction of suitable features, which we utilize in a subsequent prediction of document creation times, framed as
a regression task. Based on our evaluation of several established machine learning regressors on a large network of English news articles, we show that the combination of temporal and local structural features allows for the estimation of document creation times from the network.

Exploiting Temporal and Spatial Information to Explore Statistical Data

Investigators: Natalia Boldyrev, Jannik Strötgen, and Gerhard Weikum in cooperation with Marc Spaniol (Université de Caen Normandi)

Authorities such as the European Commission have recognized the need to offer a unified access to the data gathered by a wide variety of providers, such as the European Statistical Organization (Eurostat) or the European Environment Agency. Its EU Open Data Portal serves as a gateway to numerical data, statistical reports, and visualization tools. While making the data available to the users from all member states and concentrating efforts on bridging the language gap, the portal still focuses on a primarily statistical perspective. That is, numerical data are explained with general terms, only. However, the related events, people, or organizations “causing” or being “affected” by the statistical observation remain concealed to the user.

In order to make statistical data better understandable, we developed the SESAME system (Statistics Explored via Semantic AlignMEnt) [3]. It relies on a novel method for identifying background information and relating it with event descriptions in Wikipedia. In particular, the Wikipedia articles are contextualized for statistical observations using three features: temporal mentions, location mentions, and outgoing links are used to perform temporal, geographic, and domain scoring, respectively. The conceptual overview is shown in Figure 39.33. Using SESAME, users can jointly browse numerical statistics, their explanation in general terms and also directly relate it to associated Wikipedia articles.

Street Names with Date References

Investigators: Rosita Andrade, Dhruv Gupta, and Jannik Strötgen

Collective memories (mémorie collective) introduced in sociology by Halbwachs [4] can be considered as collective view of the society on the past. Due to large amounts of diachronic and dynamic corpora, several text mining approaches have been suggested to study this topic on a large scale. In our works in the field of computational history [2, 6], we performed a large scale analysis of street names.

In general, street names are not only used across the world as part of addresses, but also reveal a lot about a country’s identity. Thus, they are subject to analysis in the fields of geography and social science. There, typically, a manual analysis limited to a small region is performed, e.g., focusing on the renaming of streets in a city after a political change in a country. Surprisingly, there have been hardly any automatic, large-scale studies of street names so far, although this might lead to interesting insights regarding the distribution of particular street name phenomena. In our work, we present an automated, world-wide analysis of street names with date references. Such temporal streets are frequently used to

\[ \text{http://data.europa.eu/euodp/en/data/} \]
commemorate important events and thus particularly interesting to study. In Table 39.5, we show some examples of temporal streets from various countries together with the explanation for their naming in respective regions.

We developed an approach to automatically extract street names with date references across the world from OpenStreetMaps\textsuperscript{11} data, which were tagged for all languages using HeidelTime with its automatically created resources for more than 200 languages [7]. Figure 39.34 shows the number of temporal expressions per language – which demonstrates the necessity of a highly multilingual temporal tagger. We analyze the dates’ temporal and geographic

\textsuperscript{11} \url{www.openstreetmaps.org}

<table>
<thead>
<tr>
<th>Street Name</th>
<th>Date</th>
<th>Country</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straße des 13. Januar</td>
<td>January 13, 1935</td>
<td>Germany (Saarland)</td>
<td>On January 13, 1935, the Saar status referendum took place; 90% voted for reunification with Germany.</td>
</tr>
<tr>
<td>23 Nisan Caddesi</td>
<td>April 23</td>
<td>Turkey (e.g., Ankara)</td>
<td>National sovereignty and children’s day; opening of the Grand National Assembly of Turkey at Ankara in 1920.</td>
</tr>
<tr>
<td>Rue du 8 Mai 1945</td>
<td>May 8</td>
<td>France (e.g., Paris)</td>
<td>May 8 is a French holiday; this day in 1945, de Gaulle announced the end of WWII in France.</td>
</tr>
<tr>
<td>Via XX Settembre</td>
<td>September 20, 1870</td>
<td>Italy (e.g., Rome)</td>
<td>This day in 1870, the capture of Rome ended the reign of the Papal States (754–1870).</td>
</tr>
<tr>
<td>Estádio 11 de Novembro</td>
<td>November 11</td>
<td>Angola (Luanda)</td>
<td>Stadium and surrounding street are named after the date of Angola’s independence in 1975.</td>
</tr>
</tbody>
</table>

Table 39.5: Example streets (S) with explanations (E).
distribution, and automatically gather potential explanations why specific dates occur in particular regions. To find explanations, we run multiple strategies ranging from Wikipedia look-ups for pages about street names and holidays in countries across the world, up to general approaches to find explanations in Web search results for queries of the form “region date”, with region ∈ \{city, state, country\} and date in standard English. Despite the challenges of the tasks, our evaluation demonstrates the feasibility of the street extraction and the explanation harvesting.

Figure 39.34: Coverage of dates across languages [log scale]; HeidelTime’s “auto-languages” are marked with *.

References

39.7.4 Counting Quantifiers in Text

Investigators: Paramita Mirza, Simon Razniewski, and Gerhard Weikum in cooperation with Fariz Darari and Rahmad Mahendra (University of Indonesia)

Automated construction of knowledge bases (KBs) has mainly focused on extracting fully qualified subject-predicate-object (SPO) facts such as ⟨Saarland, locatedIn, Germany⟩. However, texts often contain only counting information: the number of objects that stand in a specific relation with a certain entity, without mentioning the objects themselves, such as “Germany is divided into 16 states or Bundesländer” or “Clint Eastwood directed more than twenty movies.” This kind of knowledge can be codified into an extension of existentially quantified formulas known in AI and logics as counting quantifiers (CQs): they assert the existence of a specific number of SPO triples without fully knowing the triples themselves.

Counting information can substantially extend the scope and value of knowledge bases. First, they allow for efficiently answering queries that involve counts (e.g., number of states in Germany) or existential quantifiers (e.g., directors who made at least 20 movies). Secondly, counting information helps to identify gaps and inaccuracies in KBs, which are notoriously incomplete, contain erroneous triples, and are limited in keeping up with the pace of real-world changes. Knowing the current exact number of states in Germany or a lower bound of films directed by Eastwood are important cues to complete and enrich a KB.

Distant Supervision for Counting Quantifier Extraction

The predominant approach to extracting facts for KB population relies on distant supervision, using seeds for a relation of interest that are usually taken from an initial KB or are manually compiled (e.g., Trump and Ivanka for hasChild relation). Spotting the seeds in a text corpus allows for learning patterns for a given relation (e.g., “child of” or “(someone)’s daughter”), which in turn lead to observing new fact candidates. Distant supervision is a natural approach for extracting counting information as well, as proposed in our method for extracting counting quantifiers from text [2]: the cardinality of distinct O arguments for a given SP pair, \( n := |\{O | SPO \in KB\}| \) serves as a seed for the counting assertion, ⟨S, P, ∃n⟩. However, the task is more challenging than traditional SPO-fact extraction and we outlined several challenges:

1) Non-maximal seeds: Unlike for SPO-fact extraction, the incompleteness of KBs not only leads to a reduction in the number of seeds, but to seeds that systematically underestimate the count of facts that are valid in reality. For example, a KB that knows only a subset of Trump’s children, say three out of five, leads to a non-maximal seed that may reward spurious patterns like “owns three golf resorts” at the cost of patterns like “his five children.”
2) **Sparse and skewed observations:** For many relations, counting information is expressed in text in a sparse and highly skewed way. For example, the non-existence of children is rarely mentioned. For musicians, the first Grammy someone has won often has more mentions than later ones, hence giving undue weight to the pattern “his/her first award”.

3) **Linguistic diversity:** Counting information can be expressed in a variety of linguistic forms like
   (i) *explicit numerals* as cardinal numbers (e.g., “has five children”),
   (ii) *lower bounds* via ordinal numbers (e.g., “her third husband”),
   (iii) *number-related noun phrases* such as ‘twins’ or ‘quartet’,
   (iv) *existence-proving articles* as in “has a child”, or
   (v) *non-existence adverbs* such as ‘never’ and ‘without’.

Furthermore, counting quantifiers can be expressed in composition, as demonstrated in “Trump has three children with Ivana, a daughter with Marla, and a 10-year-old son with his current wife, Melania.” Here, the total number of children, 5, is split across three different nouns: *children, daughter* and *son*.

**The CINEX System**

We developed the first full-fledged system for *Counting Information Extraction*, called CINEX [3]. The CINEX system aims to solve the following problem: Given a text about a subject $S$, and a predicate $P$, the task of *counting quantifier (CQ) extraction* is to determine the number of objects with which $S$ stands in relation regarding $P$. For instance, given the sentence “Trump has three sons and two daughters”, the output for the predicate *numberOfChildren* is 5.

Our method is based on machine learning for sequence labeling, judiciously designed to cope with the previously outlined challenges. We counter *non-maximal seeds* (Challenge 1) by relaxing matching conditions for numbers higher than KB counts, and by reducing the training to popular, more complete entities. **Sparseness and skewed observations** (Challenge 2) are addressed by discounting uninformative numbers using entropy measures. **Linguistic variance** (Challenge 3) is handled by careful consolidation of detected mentions. Figure 39.35 gives a pictorial overview of the system architecture of CINEX. We split the overall task
into two main components: the recognition of counting information via sequence labelling and the consolidation of intermediate results into the final output of counting quantifiers.

CINEX utilizes seeds from Wikidata in a judicious way in order to train a sequence labelling model for CQ recognition, using one of two options: a conditional random field (CRF) or a bidirectional LSTM neural network. The following shows an example of the CQ recognition step:

<table>
<thead>
<tr>
<th>sentence</th>
<th>Jolie brought her twins, one daughter and three adopted children to the gala.</th>
</tr>
</thead>
<tbody>
<tr>
<td>preprocessed</td>
<td>Jolie brought her NUMTERM, CARDINAL daughter and CARDINAL adopted children to the gala.</td>
</tr>
<tr>
<td>output tags</td>
<td>O O O COUNT COMP COUNT O COMP COUNT O O O O O O</td>
</tr>
</tbody>
</table>

When applied to new text, the output of the recognition model is a set of CQ candidates, which are often fairly noisy, though. Subsequently, the second stage of CINEX – CQ consolidation – cleans and aggregates the counting information and produces the final output of CINEX. The resulting CQ triples could potentially be added to a knowledge base such as Wikidata.

Experiments with five human-evaluated relations show that CINEX can achieve 60% average precision for extracting counting information. In a large-scale experiment, we demonstrate the potential for knowledge base enrichment by applying CINEX to 2,474 frequent relations in Wikidata. CINEX can assert the existence of 2.5M facts for 110 distinct relations, which is 28% more than the existing Wikidata facts for these relations. Code and data are made available to the research community on Github (https://github.com/paramitamirza/CINEX).

### Knowledge Graph of Incidents for Counting Events and Participants

We participated at SemEval-2018 Task 5: Counting Events and Participants in the Long Tail\(^{12}\), which addresses the problem of referential quantification that requires a system to answer numerical questions about events such as (i) “How many killing incidents happened in June 2016 in San Antonio, Texas?” or (ii) “How many people were killed in June 2016 in San Antonio, Texas?”

For questions of type (i), participating systems must be able to identify the type (e.g., killing, injuring), time, location and participants of each event occurring in a given news article, and establish within- and cross-document event coreference links. In order to answer questions of type (ii), participating systems are also required to identify participant roles in each identified answer incident (e.g., victim, subject-suspect), and use such information along with victim-related counting information (“three people were killed”) mentioned in the corresponding answer documents, i.e., documents that report on the answer incident, to determine the total number of victims.

We submitted our KOI (Knowledge of Incidents) system [1], which constructs and populates a knowledge graph of incidental events mentioned in news articles, to be used to retrieve answer incidents and answer documents given numerical questions about events. The system is able to answer numerical questions about number of incidents and number of victims with 27.4% and 23.0% accuracy, respectively, outperforming other participating systems in terms of Root Mean Square Errors (5.3 and 7.7, respectively). Among the questions that were

\(^{12}\)https://competitions.codalab.org/competitions/17285
answered correctly with the exact number of victims: 52.7% of correct answers result from solely counting event participants (identified as having a role victim), 35.3% were inferred only from counting quantifiers (CQ), and the rest of 12.0% were answered by combining both victim counting and CQ mentions.

References


39.7.5 Emotion Extraction

Investigators: Paramita Mirza and Harshita Jhavar

We, as readers usually identify stories through their characters and character relationships. Relationships between story characters can be defined as factual relations similar to relations between knowledge base (KB) entities, such as hasChild or hasColleague. However, particularly in fictional narratives, the more interesting aspect of relationship between two story characters is the emotion development felt by each involved character in their interactions that occur throughout the story. We aim to analyze such emotional relationships of character pairs in terms of emotion trajectories, which in turn can be beneficial for mining typical story relationships (e.g., hasLover), and help us distinguish, for instance, hasFriend from hasEnemy relations. Such analysis is one of the important aspects of story understanding, the most challenging pursuit for an AI modelling human intelligence.

In our work, we propose a fully unsupervised approach to annotate stories with emotion mappings of different character pairs as the story evolves, taking into account the different emotional situations in which a character pair interacts with each other [2]. The direction of emotion flow from one character to another is also taken into consideration as in a particular situation, their relationships can take a different form if viewed from different characters involved, e.g., a one-sided love relationship. We have also developed a prototype that allows users to explore the annotated stories with respect to evolving emotion plots between characters identified in the story.

The EMOFIEL System

EMOFIEL (EMotion mapping of FIctional rELationships) [2] addresses the problem of understanding the emotional relationship between fictional characters, given a summary of interactions and events happening between them. The task requires extensive text analysis
to (i) identify story characters and coreference resolution to resolve character mentions in the text, (ii) extract interactions in terms of events and participating agents/patients and cluster such interactions, and (iii) mapping the emotion of such interactions to be used for analyzing emotional relationships between the interacting characters. EMOFIEL’s workflow is presented in Figure 39.36.

Given a story text as input, we identify scenes in the story by segmenting the text according to interactions between story characters: a scene is defined as a sequence of sentences describing a particular event in which a subset of story characters interact intensively. Next, we assume that all story characters occurring in a scene having roles as agents or patients, i.e. subjects or objects of sentences, interact with each other in that particular scene. Given such interaction, we infer the emotion mapping of their relationships via the mood of the scene in which they co-occur.

We followed two lines of psychological research on modelling emotions, i.e., categorical and dimensional models, to model the mood of a scene based on existing emotional words. Following categorical models, emotional states are categorized into a set of emotion categories, for instance, the basic emotions (joy, sadness, etc.) existing in the NRC Word-Emotion Association Lexicon (EmoLex) [3]. On the other hand, according to dimensional models, an emotional state is described relative to a small number of emotional dimensions, such as valence, arousal and dominance [1].

The system, which can be examined online at https://gate.d5.mpi-inf.mpg.de/emofiel/, gives an insight into emotional character relationships in a story. We took the first book summary in the Harry Potter series, and analyze the character relationships in
different scenarios, such as friends vs enemies. Figure 39.37 and 39.38 shows the relationships between Harry-Neville and Harry-Snape, which can be considered as friends and enemies respectively, according to the categorical emotion mapping. The relation of Harry and Snape is dominated by negative emotions such as sadness and anger, while Harry and Neville by more positive emotions such as surprise and joy.

References


39.7.6 Dialogue Generation

Investigators: Xiaoyu Shen in cooperation with Dietrich Klakow and Vera Demberg (Saarland University)

Recurrent neural networks (RNNs) are widely used in natural language processing tasks. However, given the history context, RNNs estimate the probability of one word at a time and does not work from a holistic sentence representation. When applied to dialogue generation, it would generate either short, safe responses or long, inconsistent sentences. These safe responses break the dialogue flow without bringing any useful information and people will easily lose interest in continuing the conversation. Variational encoder-decoders (VEDs) target this issue by learning a latent global representation for each utterance, but the model is notoriously difficult to train because of the KL-vanishing challenge. This section introduces mechanisms to improve the performance of VEDs on dialogue generation. In [2], we propose improving the training stability of VEDs through a collaborative wake-sleep procedure. The resulting model can potentially learn a much more expressive distribution while remaining easy to train. Further in [1], we propose a nexus network structure to strengthen the connection relationship between previous dialogue history and the later responses. This structure avoids the KL-annealing problem and also significantly improves the coherence of the global dialogue flow.

Improving Variational Encoder-Decoders in Dialogue Generation   KL-vanishing is a notorious problem for training the VED architecture. We [2] discussed critical issues to tackle this problem and analyze the drawbacks of current proposed methods. We argue this problem is tied with the limited expressiveness of the prior distribution used in the VED architecture. An ideal way of representing the latent variable distribution is to use a universal approximator like neural networks, then train the network with likelihood-free objects like generative adversarial networks (GANs). However, GANs itself suffer from even more unstable problems than VEDs. Combining these two structures would lead to more uncertainty and, according to our experiment, requires very careful parameter tuning to achieve acceptable performance. In our work, we try replacing the GAN phase with a VED alternative. An RNN encoder is first applied to extract the corresponding latent variable target for each dialogue turn, based on which a VED is trained to reconstruct it through context-dependent Gaussian noise. The output of the VED part are latent variables, which can represent a much broader distribution family than mean-field Gaussian. As VED is in theory less accurate than GAN because it needs to approximate the real posterior, we leverage the more powerful RNN encoder-decoders. In the autoencoding phase, they should autoencode utterences to make the real posterior easily representable by the VED part. We show the proposed model has a significantly lower loss and generates much more coherent responses than the baseline. Our human evaluation results also suggest people prefer responses generated by our model than others. Without losing generality, our model should be able to apply on any seq2seq tasks, which we leave for future work.

NEXUS Networks for Dialogue Generation  To solve the model bias towards safe responses, we propose the concept of NEXUS Networks [1]. We establish this connection by maximizing
the mutual information (MMI) of the current utterance with both the past and future contexts. In this way, generic responses can be largely discouraged as they contain no valuable information and thus have only weak correlations with the surrounding context. Specifically, we introduce an auxiliary continuous code space which is learned from the whole dialogue flow. At each time step, instead of directly optimizing discrete utterances, the current, past and future utterances are all trained to maximize the mutual information with this code space. Furthermore, a learnable prior distribution is simultaneously optimized to predict the corresponding code space, enabling efficient sampling in the testing phase without getting access to the ground-truth future conversation. Figure 39.39 depicts our model. Extensive experiments have been conducted to validate the superiority of our framework. The generated responses clearly demonstrate better performance with respect to both coherence and diversity. Moreover, the extra MMI objective can also help alleviate the KL-vanishing problem, resulting in a more stable training procedure. Our model can be considered as combining the objective of MMI and conditional VED and is compatible with current improving techniques. Future work can use a tighter bound of the mutual information or extend the code space to more than Gaussian distribution.

Figure 39.39: Framework of NEXUS Networks. Full lines indicate the generative model to generate the continuous code and corresponding responses. Dashed lines indicate the inference model where the posterior code is trained to infer the history, current and future utterances. Both parts are simultaneously trained by gradient descent.
39.7.7 Template-based Question Answering

Investigators: Abdalghani Abujabal, Rishiraj Saha Roy, Gerhard Weikum in cooperation with Mohamed Yakya (Bloomberg L.P.), Mirek Riedewald (Northeastern University) and David Ziegler (Saarland University)

With the wide availability of smart voice-controlled assistants such as Amazon Echo and Google Home, returning crisp answers to users in response to their natural language questions becomes crucial. This allows satisfying the information needs of users with minimal effort on their part. For example, the answer set for “Which actors starring in The Departed were born in Cambridge, MA?” is \{Matt Damon, Mark Wahlberg\}.

Knowledge-based question answering (KB-QA) enables this desiderata by translating natural language questions into formal query languages like SPARQL that can then be evaluated over large knowledge bases such as Yago and Freebase to retrieve crisp answers. For example, the above question is translated into the following structured SPARQL query:

```
SELECT ?x WHERE {
    ?x type movieActor .
    TheDeparted hasActor ?x .
    ?x placeOfBirth CambridgeMassachusetts
}
```

Given the numerous lexical and syntactic variations that users use in formulating natural language questions, a data-driven approach for learning the mapping between syntactic structures on the question side and semantic ones on the KB side is desired. Relying on hand-crafted rules or templates to map natural language questions to formal queries is a common characteristic of many state-of-the-art approaches to KB-QA [4, 5, 6]. However, such methods suffer from low coverage. Figure 39.40 shows an example of a hand-crafted question-query template (top) and an instantiation of the template (bottom).

As a remedy, we introduce QUINT [3], a data-driven system that automatically learns pairs of role-aligned utterance-query templates using questions paired with structured queries over knowledge bases. Templates play an important role in KB-QA, guiding the mapping of question constituents onto query components. A benefit of templates is that the mappings to the KB are traceable and can be leveraged to generate explanations for the user to understand why she receives specific answers.

Given a pair of utterance-query \((u, q)\), where \(u\) represents the dependency parse tree of the utterance, QUINT aligns the nodes in \(u\) (correspond to words in the utterance) with the

References

Who invented the Internet?

SELECT ?x WHERE {
  ?x developed WorldWideWeb
}

Who VP\text{pred} NP\text{ent}

SELECT ?x WHERE {
  ?x \text{pred} \text{ent}
}

Question Query

Question Template Query Template

<table>
<thead>
<tr>
<th>Question Template</th>
<th>Query Template</th>
</tr>
</thead>
</table>
| Who VP\text{pred} NP\text{ent} | SELECT ?x WHERE {
  ?x \text{pred} \text{ent}
  } |

Figure 39.40: An example of a hand-crafted role-aligned question-query template (top). Shared \text{pred} and \text{ent} annotations indicate an alignment between a phrase in the question template and a KB semantic item in the corresponding query template. An instantiation of the template is shown at the bottom.

semantic items in \( q \) (entities, predicates and types). We formulate the alignment problem as a constrained optimization and find the best alignment using integer linear programming.

Next, QUINT performs generalization to generate a template from a concrete pair of aligned utterance and query. It removes the concrete text in the nodes participating in the alignment and similarly for the semantic items in \( q \), keeping the annotations \text{ent}, \text{pred}, and \text{type}, thereby turning these nodes into placeholders. The result is template \( t = (u_t, q_t) \) composed of an utterance template \( u_t \), a query template \( q_t \), and an alignment between the two.

When the user issues a new question, QUINT matches its dependency parse tree against the template library created during training. For each match, the corresponding query template \( q_t \) is instantiated using the alignment and the lexicon that maps natural language phrases into KB semantic items. Finally, QUINT ranks these candidate queries using a learning-to-rank approach. The answers of the top-ranked query are returned to the user.

By showing which template was used and how it was instantiated, we can explain answers to the user. In our work in [1], we present a live system for interpretable and explainable KB-QA. When the system answers a question, it visualizes the complete derivation sequence from the natural language utterance to the final answer. The derivation provides an explanation of how the syntactic structure of the question was used to derive the structure of a SPARQL query, and how the phrases in the question were used to instantiate different parts of the query. When an answer seems unsatisfactory, the derivation provides valuable insights towards reformulating the question. The demo is available at the following URL: https://gate.d5.mpi-inf.mpg.de/quint/quint.

In TIPI [7], we use QUINT to answer complex compositional questions such as \textit{“Who won a Nobel Prize in Physics and was born in Bavaria?”} by first decomposing the question into simpler sub-questions: \textit{“Who won a Nobel Prize in Physics”} and \textit{“Who was born in Bavaria?”}, using dependency parse patterns. Next, we predict a set of expected answer types for each sub-question using an answer type predictor. The typing module is a context-aware hierarchical classifier that takes a question as input and produces a set of answer types along with confidence scores as output (like \texttt{Person: 0.8, Actor: 0.3, Scientist: 0.6}). Then, TIPI reasons over the compatibility of joint types for the sub-questions. Two types are compatible
whenever they are the same, or one is an ancestor of the other, according to some type system. For example, Scientist and Person are compatible, as the former is a descendant of the latter. Finally, we let QUINT to generate a ranked list of SPARQL queries for each sub-question, which are then used to formulate compositional SPARQL queries by stitching those respecting answer type constraints from the previous stage.

Never-Ending Learning for Open-Domain KB-QA. QA methods rely on a clear separation between an offline training phase, where a model is either learned or manually crafted, and an online phase where this model is deployed to answer users’ questions [4, 5, 6]. This necessitates the need to access reasonably large training sets with sufficient syntactic and lexical coverage representative of the kinds of questions users pose. These are expensive to construct, particularly, for new domains. Furthermore, methods are limited to the language learned at training time, therefore, they fail on questions from domains not observed previously.

We introduce NEQA [2], a continuous-learning framework for KB-QA. NEQA is initialized with a small training set and improves its performance over time. Moreover, it adapts to the language used after deployment by periodically retraining its underlying models. NEQA combines both syntax, through template-based answering, and semantics, through answering via a semantic similarity function, when templates fail to do so.

NEQA first uses its template bank, learned from a small seed using QUINT [3], to answer incoming questions. For example, it uses a template generated from “Which film awards was Brad Pitt nominated for?” to answer the syntactically isomorphic question “Which president was Lincoln succeeded by?”. Given a new question for which no matching templates were found, say, “What are the film award nominations that Brad Pitt received?”, NEQA uses a semantic similarity function to find a correctly answered and semantically-similar question from its history. By harnessing light-weight user feedback on the answers retrieved by the similarity function, NEQA learns new templates, which are then added to the collection of templates learned thus far, improving the ability of the system to directly answer questions using templates.

References

Table 39.6: Distribution of question types by source. The total is greater than 1,271 as some questions have multiple tags.

<table>
<thead>
<tr>
<th>Question Tag</th>
<th>Free917</th>
<th>WebQuestions</th>
<th>ComplexQuestions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit temporal</td>
<td>41</td>
<td>344</td>
<td>222</td>
<td>607</td>
</tr>
<tr>
<td>Implicit temporal</td>
<td>3</td>
<td>81</td>
<td>125</td>
<td>209</td>
</tr>
<tr>
<td>Temporal answer</td>
<td>88</td>
<td>254</td>
<td>51</td>
<td>393</td>
</tr>
<tr>
<td>Ordinal constraint</td>
<td>18</td>
<td>111</td>
<td>26</td>
<td>155</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>790</td>
<td>424</td>
<td>1,364</td>
</tr>
</tbody>
</table>


### 39.7.8 Temporal Question Answering

**Investigators:** Abdalghani Abujabal, Rishiraj Saha Roy, Jannik Strötgen, Gerhard Weikum in cooperation with Zhen Jia (SouthWest Jiaotong University)

Time-related information needs occur very often in web search [6], with explicit or implicit temporal conditions, for example, “Which teams did Neymar play for before joining PSG?” However, handling such complex questions poses a challenge to existing KB-QA systems, as they are geared towards simple questions without any such constraints.

Our solution, called TEQUILA [5], is built on a rule-based framework that encompasses four stages of processing. First, we detect temporal questions using a combination of rules, dictionaries and time taggers e.g., HeidelTime tagger. Second, questions are decomposed into one or more non-temporal sub-questions (returning candidate answers), and one or more temporal sub-questions (returning temporal constraints). For example, decomposing the above question into “Which teams did Neymar play for”, which returns candidate answers, and the temporal sub-question “When Neymar joining PSG?”, which returns temporal constraints to filter the candidate answers from the non-temporal sub-question. To decompose a composite question, we use a set of lexico-syntactic rules. Third, we answer sub-questions using existing KB-QA systems such as QUINT or NEQA. Finally, temporal reasoning to combine and reconcile the results of the previous stage into final answers (\{FCBarcelona, SantosFC\} for the above question).

In addition to TEQUILA, we present TempQuestions [4], a high-quality benchmark with a total of 1,271 temporal questions paired with their answers over Freebase. Questions in our
benchmark are between 4 and 15 words long, and the average question length is 8.28 words. Sample questions include “What kind of government does Iran have after 1979?” and “Who was US president when Vietnam war started?”. We collected the questions in TempQuestions by identifying temporal questions in current datasets: Free917 [3] (917 question-query pairs), WebQuestions [2] (5,810 question-answer pairs), and ComplexQuestions [1] (2,100 question-answer pairs). These datasets individually contain small fractions of temporal questions. Thus, methods that ignore temporal conditions in questions manage to get good overall performance. We provide a breakdown into the four classes of temporal questions (explicit, implicit, temporal answer and ordinals), along with the questions source, in Table 39.6. TempQuestions has a good number of questions with implicit temporal expressions (209) and ordinals (155) – both these classes require additional reasoning and ranking on part of the QA system, and thus add a level of difficulty. The total 1,364 is higher than 1,271, showing that there are several questions that belong to more than one category, and are thus quite challenging for current QA systems (like “Who was elected the first governor of Virginia in 1776?”, with both explicit and ordinal tags).

References


Factoid QA is the task of answering a question whose answer is one or a small number of entities. To advance research in QA in a manner consistent with the needs of end users, it is important to have access to benchmarks that reflect real user information needs by covering various question phenomena and the wide lexical and syntactic variety in expressing these information needs. The benchmarks should be large enough to facilitate the use of data-hungry machine learning methods. In this work, we present ComQA [1], a large dataset of 11,214 real user questions collected from the WikiAnswers community QA website. As shown in Figure 39.41, the dataset contains various question phenomena. ComQA questions are grouped into 4,834 paraphrase clusters through a major crowdsourcing effort, which capture lexical and syntactic variety. Crowdsourcing is also used to provide answers to these question clusters in the form of Wikipedia links.

Table 39.7 contrasts ComQA with other publicly available QA datasets. The foremost
issue ComQA tackles is ensuring that research is driven by information needs formulated by actual users. Most large benchmarks like SimpleQuestions [3] resort to synthetically generated questions that follow templated syntax. Other benchmarks like WebQuestions [2] utilize search logs to collect their questions, that creates a bias towards simpler questions that search engines can already answer reasonably well. In contrast, ComQA questions come from WikiAnswers, a community QA website where users pose questions to be answered by other users. This is often a reflection of the fact that such questions are beyond the capabilities of commercial search engines. Questions in our dataset exhibit a broad range of interesting aspects such as the need for addressing temporal reasoning (Figure 39.41, cluster 1), comparison (cluster 2), compositionality (cluster 3), and questions with an empty answer set (cluster 4).

Past work has demonstrated benefits of paraphrasing for QA [4]. Motivated by these works, we judiciously use crowdsourcing to obtain clean paraphrase clusters (in Figure 39.41) from WikiAnswers’ noisy question groups. Such paraphrases show both lexical and syntactic variations. The only other dataset to provide such clusters is GraphQuestions, but it is synthetically generated.

Recent research has shown that combining various resources for answering significantly improves performance. Therefore, unlike previous benchmarks, we do not pair ComQA with a specific knowledge base (KB) or text corpus for answering. We also use crowdsourcing to pair paraphrase clusters with answers. ComQA answers are Wikipedia entity URLs. This has two motivations: (i) it builds on the example of search engines that use Wikipedia as a primary way of answering entity-centric queries (e.g., through knowledge cards), and (ii) most modern KBs ground their entities in Wikipedia. Wherever the answers are temporal or measurable quantities, TIMEX3 (http://www.timeml.org) and the International System (SI) of units are used for normalization. Providing canonical answers allows for comparison of a bigger family of systems.

In summary, this benchmark makes the following salient contributions:

- We create a dataset of 11,214 real user questions collected from a community QA website. The questions exhibit a wide array of aspects that are important for users and challenging for existing QA systems.
- Using crowdsourcing, questions are grouped into 4,834 paraphrase clusters that are annotated with answers. The dataset is available at http://qa.mpi-inf.mpg.de/comqa/.

References


Graph-based Question Answering

Investigators: Rishiraj Saha Roy, Soumajit Pramanik, Abdalghani Abujabal and Gerhard Weikum in cooperation with Xiaolu Lu (RMIT University) and Yafang Wang (Ant Financial)

State-of-the-art question answering (QA) systems perform well for simple questions that involve a few relations around a single target entity. But when it comes to complex questions that contain multiple entities and relationships, brittleness in state-of-the-art is quickly exposed. As an example, consider the question: “footballers of African descent who played in the FIFA 2018 final and the Euro 2016 final?” For systems that operate over knowledge graphs (KG): (i) The natural-language question to logical-query translation tends to be highly error-prone for complex information needs; answering would work only with a perfect mapping of question terms onto KG-predicates like bornIn, playedFor, inFinal, etc. This strong assumption is rarely satisfied for such complex questions. (ii) Computing good answers depends on the completeness of the underlying KG, but KGs cannot usually keep up with the latest news and fast real-world changes.

On the other hand, even the best methods in QA over text face major obstacles: (i) To retrieve relevant passages for answer extraction, all significant question terms must be matched in the same document, ideally within short proximity. For example, if a QA system finds players in the 2018 FIFA World Cup Final and the UEFA Euro Cup 2016 Final in different news articles, and information on their birthplaces from Wikipedia, there is no single text passage for proper answering. (ii) The alternative of fragmenting complex questions into simpler sub-questions requires syntactic decomposition patterns that break with ungrammatical constructs, and also needs heuristics for stitching sub-results together.

Our proposed method, QUEST [3] (“QUEstion answering with Steiner Trees”) overcomes these limitations by providing a novel unsupervised method for combining answer evidence from multiple documents, joined together via KG-style relationships automatically gathered from text sources. QUEST first constructs an ad hoc, noisy knowledge graph by dynamically retrieving question-relevant text documents and running Open Information Extraction (Open IE) on them to produce subject-predicate-object triples. However, unlike a curated KG, these triples contain raw surface forms rather than canonicalized entities and predicates. Thus, we insert alignment edges that connect potentially synonymous names and phrases. We refer to the resulting graph as a quasi KG, which combines cues from many documents and is treated as the knowledge source for the QA algorithm (example in Fig. 39.42, showing entity, relation, and type nodes, and triple, alignment, and type edges).

Correct answers in the nodes of the quasi-KG should be well-connected with all nodes that approximately match tokens from the input question. We call these question-relevant nodes as cornerstones (nodes with thick borders in Fig. 39.42). This condition can be cast into
computing Group Steiner Trees (GST) with the cornerstones as terminals. All non-terminals in the GST are candidate answers. This computation is executed over a weighted graph, with weights based on relevance scores and extraction confidences. Finally, answers are ranked by their proximity to the terminals in the GST. In Fig. 39.42, all red and blue nodes and edges, constitute two different GSTs, respectively yielding the answers Samuel Umtiti and Blaise Matuidi (underlined).

To evaluate QUEST, we used two benchmarks of complex questions: one curated from the WikiAnswers forum (CQ-W) with 150 questions, and another from Google Trends queries (CQ-T) with 100 questions. As the corpora for building our quasi KGs, we used the top-10 documents from Google Web search, where the whole question was issued as a keyword query. As a strong neural baseline, we select DrQA [1], a very recent open-source method developed at Facebook and Stanford. As a graph-based competitor to the GST algorithm of QUEST, we adapted the breadth-first search (BFS) phase of the STAR algorithm [2] for entity-relatedness in curated KGs. To examine the importance of the optimal subgraph identified by the GST, we also compute shortest paths in the graph between every pair of terminals, where each node in a pair is from a different cornerstone group. Every non-terminal that lies on any shortest path is treated a candidate answer.

Looking at summary results in Table 39.8, we find that QUEST significantly and consistently outperforms the neural baseline DrQA, and other graph-based methods, in almost all settings. We found this performance of QUEST to be robust to variations in the underlying

<table>
<thead>
<tr>
<th>Method</th>
<th>CQ-W</th>
<th>CQ-T</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metric</strong></td>
<td>P@1</td>
<td>MRR</td>
</tr>
<tr>
<td>QUEST</td>
<td>0.267</td>
<td>0.328</td>
</tr>
<tr>
<td>DrQA [1]</td>
<td>0.193</td>
<td>0.253</td>
</tr>
<tr>
<td>BFS [2]</td>
<td>0.160</td>
<td>0.249</td>
</tr>
<tr>
<td>ShortestPaths</td>
<td>0.147</td>
<td>0.240</td>
</tr>
</tbody>
</table>
corpus. We attribute the success of the proposed method to its unique ability to stitch facts from more than one source, and the powerful GST algorithm that discovers answers in the large and very noisy quasi KGs. The task of fetching crisp text answers to complex questions directly over Web corpora is generally a very difficult one; this is reflected by relatively low values of MRR (best numbers of 0.328 for CQ-W and 0.355 for CQ-T). Anecdotal examples where QUEST retrieves the correct answer at the very first rank are: “where did sylvie vartan meet her future husband johnny hallyday?”, “which aspiring model split with chloe moretz and is dating lexi wood?”, and “which japanese baseball player was contracted for los angeles angels who also played for hokkaido nippon-ham fighters?”. To summarize: as noisy content is unavoidable with Web contents and since extensive training is infeasible for ad hoc questions, QUEST deliberately allows noise into its pipeline, and copes with it using cross-document evidence, smart answer detection, and graph-based ranking strategies.

References


39.8 Privacy, Trust and Fairness

Coordinator: Gerhard Weikum

Humans leave more and more digital traces in online systems, both laterally across a variety of platforms and longitudinally over several years. This rich online information is leveraged by platform providers to learn user profiles and enhance algorithmic services from recommenders, filters, classifiers and rankers all the way to largely automating decision-making processes such as loan requests, visa applications or invitations for job interviews. As a consequence, users are faced with increasing risks of losing privacy, receiving unwanted visibility and being treated in an unfair and potentially discriminating way in online systems.

Merely focusing on privacy protection alone (via technical contributions such as data perturbation for differential privacy and legal regulations such as GDPR) is insufficient. First, there is an inherent trade-off between the privacy gain from anonymizing data and the utility of the anonymized data for downstream applications. This is a concern for service providers, but also for users who may appreciate personalization in recommendations. Second, from the perspective of the user community, it is much more difficult to identify trustworthy peers and contents in the presence of many anonymous posts. So the interaction of privacy and trust needs to be addressed as well.
Our work in this space includes a variety of research directions, covered in the following subsections. Sections 39.8.1, 39.8.2 and 39.8.3 address various settings on assessing trustworthiness and credibility including the trade-off between anonymity and trust. Section 39.8.4 specifically investigates the trade-off between privacy (i.e., countering user profiling) and service utility. Sections 39.8.5 and 39.8.6 contribute new models for coping with potential discrimination and improving the fairness of classifiers and learning-to-rank regression models. Section 39.8.7 aims to provide user-comprehensible explanations for recommended contents in online feeds. Sections 39.8.8 and 39.8.9 study structural and dynamics properties of online discussion forums.

Much of this work is carried out in the context of the ERC Synergy Grant imPACT\(^{13}\) on Privacy, Accountability, Compliance and Trust, and the DFG CRC 1223\(^{14}\) on Methods and Tools for Understanding and Controlling Privacy. Our results contribute to trust, transparency and explainability of models and methods for the emerging paradigm of responsible data science \([1, 2]\).

References


39.8.1 Trust Assessment for Online Forums

**Investigators: Subhabrata Mukherjee, Gerhard Weikum, Stephan Günneemann (external)**

A key challenge that arises in online communities such as health discussion forums or product review platforms is to assess the trustworthiness of users and their statements. Apart from mining product reviews, prior work in this space of “truth discovery” \([1]\) had largely focused on straight factual statements, for example, on the birthplace of Barack Obama (e.g., in the context of knowledge base curation or rumor/hoax detection) and had hardly tapped thematically focused social media such as discussion forums on diseases, drugs and side effects or on good journalistic practice.

Within this research space, we have pursued a number of specific directions, leading to high-quality publications and the completion of Subhabrata Mukherjee’s dissertation \([3]\) which has won an honorable mention for the KDD dissertation award\(^{15}\).

Salient contributions are:

- a general model and inference method for jointly assessing the credibility of statements and the trustworthiness of their sources (e.g., users);

- a family of models for capturing the dynamically evolving expertise and maturity of users in online communities;

\[^{13}\]http://www.impact-erc.eu/

\[^{14}\]https://privacy-sfb.cispa.saarland/

\[^{15}\]http://videolectures.net/kdd2018_london/
Joint Model for Objectivity, Credibility and Trustworthiness

We developed a probabilistic graphical model that jointly considers the objectivity of the language in user posts, the reputation and trustworthiness of users and the credibility of statements derived from user posts. This model goes beyond established models with Conditional Random Fields (CRFs) because one statement, for example, on side effects of medical drugs, is derived by aggregating extractions from multiple posts of different users. We devised new ways of semi-supervised learning for this model, harnessing linguistic features on posts and community-level features on users, and judiciously devised joint inference techniques for applying the model. Figure 39.43 shows a pictorial overview. The work was originally published in [6] for health discussion forums and later extended to other settings. Follow-up research in our group led to a project on assessing the credibility of short claims stated in natural language such as quotes; see Section 39.8.2.

Evolution of Online Communities

On the evolution of online communities, we addressed the issue of how individual users mature over time and how their improving expertise is reflected in the language of their posts. To capture the interplay of user expertise and user posts, we devised a latent topic model for expertise and a discrete-time transition model for the evolution of expertise levels. These are components fed into a CRF model to infer how user language and expertise affect ratings of items like products, movies, restaurants, etc. We further extended this powerful model to a continuous time dimension, replacing the state transition model with a Geometric Brownian Motion model. This work was published in [4]. In terms of predicting the ratings
of users on new items, we were able to outperform strong state-of-the-art models in the prior literature (e.g., [2]) by a large margin. Moreover, our models make it easier to explain the credibility of statements or the value of ratings, by means of linguistic facets.

**Use Cases**

On the application side, we refined and experimentally tested these methods for a variety of settings. Studying health discussion forums, we obtained insights on side effects of medical drugs from the aggregate information of online communities. For a given drug, the model allows us to rank side effects by their likelihood, including adverse reactions that are labeled as rare in official drug data-bases. Moreover, for a given disease, drug or syndrome, the model also produces a ranked list of most knowledgeable users. That is, it automatically identifies users who are likely doctors but also further users who give authoritative advice.

For a second use case, we studied a news review community where “citizen journalists” discuss and rate news articles and news sources (both traditional newspapers as well as online-only sources) by different criteria. Here as well, the model was able to provide insight into credibility and trust of community members and the underlying news.

Last but not least, we developed methods for detecting non-credible posts and product review sites – a problem often referred to as “fake review detection”. Our method went beyond prior works by incorporating a notion of consistency in analyzing reviews. This in turn allowed our method to extract evidence for the credibility assessment as an element of explaining the algorithmic outcome to a user (e.g., an analyst, a forum moderator, or even an ordinary end-user).

Results from this line of research have appeared in several publications including [6, 4, 5].

**References**


The Web is a vast resource of valuable information. However, the spread of false claims in social media, other web-sources, and even in news has emerged as one of the biggest challenges of this time. The World Economic Forum has also identified “the rapid spread of misinformation online” as one of the top-10 challenges faced by the world. Increasing hoaxes and rumors on the Web have given rise to many fact-checking websites that manually assess these doubtful claims. However, the speed and scale at which misinformation spreads limit the manual verification process and to keep up with that, we need tools to automate this verification process. Hence, there is great interest in research to address this challenging problem of automatic credibility assessment.

Prior approaches for credibility assessment have focused on structured data—resolving conflict amongst multi-source data (see [2] for survey). Other credibility assessment approaches work only in restricted social media settings and rely on platform-specific features [1]. These approaches do not address the problem of assessing credibility of arbitrary textual claims, expressed freely in an open-domain setting, without making any assumptions on the structure of the claim or characteristics of the community or website where the claim is made. Most importantly, these prior approaches provide black-box techniques and lack the ability to explain why a certain statement is credible or not.

To overcome these limitations, we propose a generic framework for credibility assessment which does not make any assumptions about the structure of the claim or characteristics of the community or website where the claim is reported. In addition to the automatic assessment, our method also focuses on extracting interpretable evidence from judiciously selected web-sources making our approach for automatic credibility assessment transparent and interpretable.

**Supervised Approach for Credibility Assessment**

To assess the credibility of a given textual claim, our method (as proposed in [3]) automatically retrieves articles about the claim and models the mutual interaction between the language style of relevant articles (e.g., bias, subjectivity, etc.), their stance towards the claim (i.e., whether it supports or refutes the claim), and the trustworthiness of the underlying web-sources. Figure 39.44 gives a pictorial overview of the overall model. Extensive experiments on real-world datasets demonstrate the viability of our method and its superiority over prior works. The web interface of our credibility assessment system, CredEye [4], is available at: https://gate.d5.mpi-inf.mpg.de/credeye/

**Neural Network based Approach for Credibility Assessment**

The downside of our method proposed in [3] is that it requires substantial feature modeling and rich lexicons to detect bias and subjectivity in the language style. To overcome this limitation of our prior work, we propose an end-to-end neural network based model [5] for assessing and explaining the credibility of arbitrary claims in natural-language text form.
Given an input claim, our system searches for web articles related to the claim. It considers the context of the claim via word embeddings and the language of web articles captured via a bidirectional LSTM (biLSTM), while using an attention mechanism to focus on parts of the articles according to their relevance to the claim. Then the system aggregates all the information about claim source, web article contexts, attention weights, and trustworthiness of the underlying sources to assess the claim. It also derives informative features for interpretability, like source embeddings that capture trustworthiness and salient words captured via attention. For example, given the natural-language input claim “the gun epidemic is the leading cause of death of young African-American men, more than the next nine causes put together” by Hillary Clinton, our system draws on evidence from the Web to arrive at its verdict credible, and returns annotated snippets like the one in Table 39.9 as explanation. The overall framework of our model is given in Figure 39.45.

References

39 D5: Databases and Information Systems

Table 39.9: Example claims with credibility verdict and attention (weights) interpretation.

pp. 675–684. ACM.


39.8.3 Detecting and Classifying Fake News

Investigators: Liqiang Wang and Gerhard Weikum in cooperation with Yafang Wang (Shandong University) and Gerard de Melo (Rutgers University)

A recent large-scale study of content spreading in Twitter has shown that fake news is disseminated substantially faster, farther, deeper and more broadly than reliable content on comparable topics [2]. The ability of fake news and doubtful claims to outpace serious reporting and verified facts gives them an undue advantage in influencing public opinions. This big societal problem has motivated researchers to develop largely automated methods for assessing the truth of news and statements, leading to tools for fact checking, credibility assessment and trust analysis.
Fake news, doubtful statements and other unreliable content often come with specific intents and in different levels of misinformation. Prior work on algorithmic truth assessment has mostly pursued binary classifiers – factual vs. fake – and disregarded these finer shades of untruth. On the other hand, manual analysis of questionable content has proposed more fine-grained classification: distinguishing between hoaxes, irony and propaganda, or the six-way rating by the PolitiFact community. In our work [3], we pursue a principled approach to capture these finer shades in automatically assessing and classifying news articles and claims. Our method is based on a hierarchy of five different kinds of fakeness, and systematically explores a variety of signals from social media, capturing both the content and language of posts and the sharing and dissemination among users. This work shows results on the accuracy of our fine-grained classifier and a detailed analysis of the underlying features.

Hierarchical Taxonomy

To reflect the finer-grained classifications, we devised a taxonomic hierarchy that captures both the SHPT scheme [1] (“satire”, “hoax”, “propaganda”, “trusted”) and the PolitiFact ratings (“true”, “mostly true”, “half true”, “mostly false”, “false”, and “pants-on-fire”). This is shown in Figure 39.46. To reconcile SHPT and PolitiFact, we map their labels into our tree, and this leads to five major categories of fakeness: factual, propaganda, hoax and irony, as well as two refinements of propaganda into incomplete context and manipulative statements.

Figure 39.46: Classification hierarchy of fake content.

Features

To better understand the nature of unreliable content and misinformation, we study a number of informative features. These features include all the words, entity names, words in headlines, sentiment-bearing words (based on a lexicon), subjectivity-bearing words (based
on a lexicon), the credibility history of speakers/authors in PolitiFact, the news sharing
dynamics in Twitter, the content similarity between the news and the corresponding tweets,
tweet comments, and the stance of Twitter users with regard to a given news article.

Classifiers

Based on the taxonomic hierarchy, we devised two type of classifiers based on different
techniques: logistic regression (LR) and deep learning (DL). Our goal is to learn a model that
is interpretable so as to enable detailed analysis. We hence rely on a LR model to address
not only the task of accurately classifying the items, but also to explain the contribution of
features. To explore the effectiveness of deep learning techniques on fake news detection,
we devise two deep learning classifiers for long text and headlines, respectively. The two
classifiers are based on the concepts of convolutional neural networks (CNN) and long short
term memory (LSTM) networks. CNNs are considered more effective at capturing n-gram
features and other contextual relationships. We use the CNN-based classifier to classify
long texts such as news articles and a set of tweets. LSTM networks excel at capturing
sequential correlations, especially for short text. Accordingly, we devise an LSTM classifier
for detecting fake news with headlines and individual statements. To combine news and
comment features, we use a concatenation operation. The tweet comments for a news article
are considered as a single document in this combined model.

Experiments and Analysis

We conducted experiments on two datasets, SHPT and PolitiFact with news and tweets
data. The results indicate that the combination of content (articles or statements) and social
media information can improve the prediction quality. On most features, both LR and DL
classifiers achieve fairly good performance. The deep learning method outperforms the LR
variant on most experiment configurations, except for the SHPT dataset. The reason is that
for short text such as headlines and statements, the features are very sparse.

Based on the interpretable features of the LR method, we analyzed the influence of the
different feature groups, leading to the following observations and insights.

1. **Linguistic analysis**: Different classes of unreliable content exhibit different linguistic
   patterns, which can be used as important features for detection and classification.

2. **Tweet comments**: Certain tweet words indicate the intent of different classes of fake
   news.

3. **Topic drift from news to tweets**: From the content similarity results, we observe
   that “satire” and “trusted” news are more consistent with tweet comments. Thus,
   on these kinds of articles, people tend to really focus on the topic introduced by the
   original news.

4. **News sharing**: Often, after a few hours up to a day, there is a notable delayed burst
   for propaganda news in SHPT and hoax statements in PolitiFact, indicating that these
   kinds of content have a longer life cycle.
5. **Headline analysis:** The headline style of different kinds of news varies. For example, to gain attention, hoax news often use words such as “breaking”, “furious”, “unthinkable” etc. in the headline.

6. **Sentiment and subjectivity:** We find that trusted and satire news are more on the positive side of the sentiment spectrum, and hoax and satire news use more subjective vocabulary.

7. **Credibility history:** The credibility history feature for a given speaker or author improves the classification accuracy, but with limited gains.

8. **User stance:** As expected, propaganda and mixed-truth content lead to pronounced user stance. Especially for the PolitiFact statements, incomplete and manipulative statements gain much higher support on social media.

**References**


**39.8.4 Reconciling Privacy and Utility**

*Investigators: Asia J. Biega, Azin Ghazimatin, Rishiraj Saha Roy and Gerhard Weikum, in cooperation with Hakan Ferhatosmanoglu (University of Warwick) and Krishna P. Gummadi (Max Planck Institute for Software Systems)*

**Privacy-preserving personalization**

Search engines are often the first source people refer to when seeking information necessary for their work or hobbies, when seeking information related to health issues or personal problems, or when planning traveling. As a result, information accumulated within a single user account often draws an exact picture of a person’s life. Such massive accumulation of personal data leads to significant privacy concerns. At the same time, while caring about privacy, many users feel compelled to give such information in exchange for quality personalized results.

Our work challenges the assumption that detailed user profiles are necessary to personalize query results or recommendations, by tackling the problem of designing mechanisms for delivering personalized search results without the need for accurate user profiling. More specifically, we have proposed a framework of Mediator Accounts which supports the personalization of search results without storing full-fledged histories of user actions [2]. The Mediator splits and merges query histories of different users into synthetic user profiles, guided by the privacy-utility trade-off. Privacy is achieved by random assignments, and...
utility by keeping semantically coherent contexts intact (that is, topically related queries are kept together). Optimizing this split-and-merge grouping for a good balance of privacy and utility is based on formalized notions of profiling privacy and individual user utility. Our experimental results, with a query log synthesized from questions on the StackExchange platform, provide insights on the trade-offs from the perspective of individual users (as opposed to prior works that focused on the service provider’s view). Our results showed that it is indeed possible to reconcile big profiling privacy gains with low loss in personalization utility, particularly for users with rich profiles and diversified interests.

**Privacy of hidden profiles**

We have demonstrated the generalizability of the Mediator Accounts framework [2] by identifying and defining the problem of Privacy of Hidden Profiles [3]. Hidden profiles are the profiles of users who decided to leave an online community, but whose data was retained by the service providers for data analytics purposes. Such data still poses privacy risks for the users as it can easily be passed on beyond the original intentions, upon a governmental inquiry, a company merger, or when the infrastructure of the provider is compromised. Our results show that it is possible to protect hidden profiles by scrambling user data, while keeping the analytic utility of the data nearly unaffected.

**Sensitive Exposure in Search Systems**

When a user’s post in an online community’s search engine is returned as a top-k answer to a sensitive search query (e.g., on health problems), the user is exposed in a sensitive context. The richness and volume of the content we post online make it challenging to maintain awareness of the contexts in which our posts are returned as top-k results in search systems. Online users have very limited information about the queries that lead others to their profiles, yet – from the privacy perspective – such information is crucial if these exposing queries are of sensitive nature. Examples of such queries include those topically related to health, finance, or other personal issues, those which are unique to a user, such as phone numbers or e-mail addresses, or those which are rather uncommon for a given user profile.

To help enhance privacy awareness of users who are being searched for – search subjects, we have developed a methodology for quantifying sensitive search exposure [1]. We define search exposure as the problem of finding all queries that expose any of a user’s posts in the top-k results in a community’s search engine. With this formulation, the problem can be seen as a reverse top-k nearest-neighbor search: computing k documents closest to the search query by a given similarity or relevance metric. Identifying such queries is not enough; our empirical analysis with user profiles from the Twitter online community revealed that exposure sets for some users might be enormous and largely contain noisy and meaningless queries (e.g., queries with typos). To make the outputs useful to end users, we have designed a weakly-supervised learning-to-rank method, ordering the queries such that those at the top are most concerning. We showed that the queries can be effectively ranked using only implicit signals which are readily available to service providers.
39.8.5 Fairness in Rankings

Investigators: Asia J. Biega, Gerhard Weikum in cooperation with Krishna Gummadi (Max Planck Institute for Software Systems)

Fairness of ranked results in search systems matters especially in scenarios where rankings influence people’s lives outside of the platform. Such is the case for two-sided economy platforms, including Airbnb or Uber, or hiring support platforms such as LinkedIn. In each of these systems, ranked people seek to be displayed high in the rankings as it increases their chances of getting a real-world advantage, be it a higher income or being contacted by recruiters. With such a tangible influence over people’s lives, search engines in these scenarios should make sure their results are fair, or more specifically, that subjects get a fair share in the ranked results.

Most prior works have proposed to quantify such fair representation using different forms of diversity and exposure. In practice, exposure can be determined in eye-tracking studies by measuring the time the searchers spend investigating a result, or by estimating click probabilities for results at different ranks [2].

To be individually fair to each subject, a search system should grant subjects exposure that is proportional to their relevance. However, if many subjects have similar relevance in a given search task, it is impossible to grant everyone the attention they deserve in a single ranking. This problem arises because of a phenomenon called position bias [3] where the searchers pay disproportionately more attention to subjects ranked higher, often irrespective of their relevance. To tackle this problem, we have developed a mechanism for re-ordering rankings such that each subject in the system receives attention from the searchers that is proportional to their relevance in an amortized manner [1]. While addressing fairness concerns, re-ordered rankings will lead to accuracy loss when order is no longer determined by relevance. Trying to balance both of these dimensions, we formalized re-ordering as a constrained optimization problem, where we minimize unfairness (measured as a disparity between attention and relevance) subject to constraints on ranking accuracy loss. Choosing appropriate fairness and quality measures, the problem can be solved as an Integer Linear Program (ILP). We applied and analyzed the behavior of the proposed mechanism on synthetic and real-world data of rental apartments from the Airbnb platform.
Learning Fair Representations

Investigators: Preethi Lahoti and Gerhard Weikum in cooperation with Krishna Gummadi (MPI for Software Systems), Aristides Gionis (Aalto University), Kiran Garimella (Ecole Polytechnique Federale de Lausanne)

Algorithmic decision making involves automated and data-driven decisions both online (e.g., content recommendation, product personalization) and offline (e.g., risk prediction in banking, health, crime). As algorithmic decision making becomes pervasive in all aspects of our daily life, concerns have begun to rise about bias and fairness of the decision made for an individual based on user-attributes (which include sensitive attributes such as gender, ethnicity, and potentially sensitive attributes such as political affiliation).

The overarching question that we ask in this research is: How can we use machine learning to make decisions that affect humans, while making sure that the decisions are not exposing them to discrimination?

Learning Ideological Leaning on Twitter

People are shifting from traditional news sources to online news at an incredibly fast rate. However, the technology behind online news consumption promotes content that confirms the users’ existing point of view. Consequently, users on different ends of the ideological spectrum live in their own information filter bubbles. This phenomenon has led to polarization of opinions and intolerance towards opposing views.

In this work [2], we use a machine-learning approach to learn a liberal-conservative ideology space on Twitter, and show how we can use the learned latent space to tackle the filter bubble problem. Specifically, we propose a principled approach to infer the ideological stances (also known as ideology or polarity or leaning) of the users in a social network, and the media sources that provide news content in the network.

Our approach is based on a non-negative matrix-factorization model, which jointly decomposes the social network of users and the content they consume in a shared latent space. Finally, we demonstrate the utility of our model in real-world scenarios, by illustrating how the learned ideology latent space can be used to develop exploratory and interactive interfaces that can help users in diffusing their information filter bubble. Figure 39.47 shows popular news-media outlets and their ideology leaning scores computed by our method.

Learning Individually Fair Representations

Research on how to incorporate fairness into such tasks has prevalently pursued the paradigm of group fairness [4]: giving “fair share of beneficial outcome” to protected groups. While effective at countering group-based
Figure 39.47: Popular media outlets and their ideology leaning scores computed by our method

Figure 39.48: Overview of iFair framework and decision-making pipeline (d(.) is a pairwise distance function)

discrimination in decision outcomes, group fairness notions do not address unfairness in outcomes at the level of individual users. For instance, it is natural for individuals to compare their outcomes with those of others with similar qualifications (independently of their group membership) and perceive any differences in outcomes amongst individuals with similar standing as unfair.

In their seminal work, [1] formalized this intuition as the notion of individual fairness, where “individuals similar to each other with respect to the task at hand receive similar outcomes”. Despite its intuitive appeal and its potential to tackle unfairness beyond group-level, the alternative paradigm of individual fairness has received relatively little attention. Our recent work [3] (to appear in ICDE 2019) advances this less explored direction. Our notion of individual fairness requires that users who are similar in all task-relevant attributes such as job qualification, and disregarding all potentially discriminating attributes such as gender, should have similar outcomes. We propose iFair, an application-agnostic unsupervised framework for learning low-rank data representations that reconciles individual fairness and the utility of classifiers and rankings in downstream applications. Figure 39.48 gives a pictorial overview of our proposed framework and decision-making pipeline.

**Augmenting Representations with Human Knowledge on Fairness** One limitation of our previous work [3] is that it relies on metric distances in the input data space to capture
similarity between individuals. While, this is a reasonable approach it has obvious limitations due to the complexity of measuring similarity between users that is not easy to capture using distance metrics over feature space. In our ongoing project we aim to address this problem by incorporating various kinds of human judgments about similarities between users.

Specifically, we propose a representation learning algorithm, called \textit{PFR}, that aims to augment machine predictions by incorporating human knowledge on fairness. We investigate opportunities for obtaining such human knowledge, and formalize them into a novel notion of \textit{fairness graphs}. The problem of fair representation learning is then cast into a \textit{graph embedding problem} that aims to jointly preserve similarity in feature spaces and similarity in fairness graphs. Comprehensive experiments with synthetic and real-life data demonstrate the practical viability of our model and its advantages over state-of-the-art baselines.

References


39.8.7 Explaining Social Feed Recommendations

\textbf{Investigators: Azin Ghazimatin, Rishiraj Saha Roy, Gerhard Weikum}

By harnessing the huge volume of content in social media, algorithms can provide users with personalized stream of content. Such tailored content is referred to as social feed. Users’ feeds usually result from a complex interplay of their social contacts, their interests and their actions on the platform. Therefore, users often have no clue as to how their feeds are generated [1]. Consequently, they might wonder about the relevance of feed items to their profiles, particularly if the content of their feed is surprising or embarrassing.

Finding the relationships between users’ feed items and their profile might not be a trivial task for the users; over time, users gather thousands of actions in their profile which makes it impossible for them to remember all the details. Besides, users might not even have access to their full interaction history. These challenges necessitate a systematic approach for understanding and explaining personalized social feeds.

\textbf{Understanding Relationships between Users’ Actions and their Social Feeds}

To address the discussed challenges, we propose \textit{FAIRY}, a user-side \textit{F}ramework for \textit{A}ctivity-\textit{I}tem \textit{R}elationship discover\textit{Y} [2]. In this framework, we first build user-specific interaction
graphs exclusively using information visible to the user herself. Interaction graph is modeled as a heterogeneous information network where nodes and edges are multi-typed. In this graph, each path connecting the user to her feed item corresponds to a potential explanation for that feed item. We refer to these connections as explanation paths. Figure 39.49 illustrates the explanation paths between Quora user Alice and her feed item (a post on bomb).

![Toy interaction graph for Quora user Alice](image)

The large number of explanation paths in the interaction graph demands a subsequent ranking module. We employ a learning-to-rank model based on ordinal regression to model user’s judgments on relevance and surprisal of the explanation paths. Figure 39.49 depicts Alice’s judgments over relevance and surprisal of the paths $r_1$ and $r_2$. Longitudinal user studies on two social platforms (Quora and Last.fm) demonstrate the practical viability and user benefits of our framework.

References


39.8.8 Dense Subgraphs in Social Media

Investigators: Oana Balalau, in cooperation with Carlos Castillo (Universitat Pompeu Fabra), Maximilien Danisch (Sorbonne University) and Mauro Sozio (Télécom ParisTech)

Graphs can model a variety of types of data, spanning from road networks, social networks, and information networks, such as the connections between the terms used in a collection of tweets. Graph mining is concerned with extracting patterns that satisfy certain properties from graphs. One of the essential tasks in graph mining is finding dense subgraphs, as density can be used to measure the importance and the cohesiveness of a subgraph.

In our work, we address two main challenges:

– developing a scalable algorithm for listing $k$-cliques;

– finding mentions of disaster events in social media with high precision, via weighted clique mining.

Listing $k$-Cliqu es in Sparse Real-World Graphs

Cliqu es are dense subgraphs par excellence. The concept of a clique has been originally introduced by sociologists to measure social cohesiveness before the advent of computers. In our work, we study the problem of listing all $k$-cliques in a graph, which are subgraphs with $k$ nodes, each pair of which being connected with an edge. Recent works in the data mining and database community call for efficient algorithms for listing or counting all $k$-cliques in the input graph. In particular, in [7] the author develops an algorithm for finding subgraphs with maximum average number of $k$-cliques, with listing $k$-cliques being an important building block.

Motivated by recent research, we develop the most efficient algorithm for listing and counting all $k$-cliques in large sparse real-world graphs, with $k$ being an input parameter. Real-world graphs are often “sparse” and rarely contain very large cliques which allows us to solve such a problem efficiently. In our work, the sparsity of a graph is measured by its core value. We revisit the iconic algorithm developed by Chiba and Nishizeki [2], which is one of the most efficient algorithms for this problem. By means of several improvements on such an algorithm, we are able to provide the best asymptotic upper bound on the running time, in the case when the input graph has “low” core value. Moreover, our algorithm can be efficiently parallelized resulting in even better performances in practice.

Our extensive experimental evaluation shows that both the sequential and parallel versions of our algorithm outperform significantly state-of-the-art approaches for the same problem. In particular, our parallel algorithm is able to list all cliques in graphs containing up to tens millions edges, as well as all 10-cliques in graphs containing billions of edges, within a few minutes or a few hours, respectively, while achieving an excellent degree of parallelism.

This work has been presented at The Web Conference 2018 [3].

Detecting Disaster Mentions on Social Media

Social media have been playing increasingly a major role during crises and disasters. The American Red Cross (ARC) recognizes the effectiveness of social media and mobile apps in
Handling emergency situations such as those generated by a disaster event (e.g., earthquakes, mass shootings, etc.) [5]. Automatically acquiring valuable information about disaster events from social media would be extremely valuable, however, it presents non-trivial challenges, as social content is often noisy, inconsistent and ambiguous.

One of the challenges we address in our work, is how to provide a succinct keyword-based description of disaster events containing the most relevant information about the events, such as what happened, where, and when. According to a survey by the US Congressional Service, the administrative cost for monitoring multiple social media sources, which typically produce large amounts of noisy data, is significant [4]. Therefore, in order to alleviate the burden of analyzing social content, a succinct and informative description of the events is needed.

Our approach finds disaster event mentions on Twitter and consists of the following steps:

1. filtering of the tweets by retaining only those containing at least one term in a given lexicon. In our work, we use the lexicon constructed in [6] which contains terms related to disaster events, however, any lexicon can be used;

2. finding locations whose number of occurrences in tweets deviates significantly on a given time window from their expected frequency;

3. a graph mining approach for selecting the relevant keywords in the description, based on a novel definition of a clique and a quasi-clique in a weighted graph.

Tweets can be easily represented as weighted graphs, where words are nodes and edges represented the co-occurrence of the respective words in the collection of tweets. The intuition of this representation is that dense subgraphs in such graphs represent events occurring in the world. However, existing definitions of dense subgraphs don’t capture events in a satisfactory manner. Given this, we define the weighted (quasi) clique. A weighted clique is a completely connected subgraph where all edges have the same weight and a weighted quasi-clique is a subgraph where this constrained is relaxed.

Our extensive experimental evaluation over a period of 19 months shows that our approach outperforms state-of-the-art approaches in terms of precision and fraction of unique events retrieved, while the description provided by our algorithm is succinct and contains the most relevant information such as the location, what happened and the timeframe.

This work has been presented at the AAAI International Conference on Web and Social Media (ICWSM 2018) [1], and a journal paper is under revision.

References


39 D5: Databases and Information Systems


39.8.9 Understanding Controversies in Online Forums

Investigators: Anna Guimarães, Erisa Terolli, Oana Balalau, Gerhard Weikum

Social media networks, like Twitter and Reddit, offers a medium for users to express diverse thoughts on any topic of their interest, and an opportunity to interact with other users who may share or oppose their opinions. In such a setting, prolonged exchanges between users, via posts and replies, give way to fully fledged discussion threads. A single discussion thread may constitute a harmonic conversation where users demonstrate a mutual agreement over a certain topic or statement, or it may consist of debates containing diverging opinions. Discussions may also start out in a consensus-driven manner, but eventually become contentious or be driven off-topic by disruptive users. Over time, users may drop out due to a lack of (positive) interactions with other users, and new users may join in, reshaping the focus of the conversation.

Prior research has analyzed these online discussions piece-wise, paying special attention to specific user behavior (e.g., “trolling” [3]) or content popularity [8]. Often, the larger context of a discussion and the community in which it is inserted is overlooked. Instead of focusing on individual posts and users, in our work we study how online conversations gradually unfold. Specifically, we are interested in finding and understanding frequent patterns representing different conversational dynamics, such as long chains of disagreements or a gradual shift from one topic to another. Our objective is to discover semantic and structural indicators which differentiate between these different types of discussions, as well as their specific triggers, in an effort to better understand the key aspects that shape evolving online conversations and user engagement.

Adversarial Politics on Twitter

Politics offer a clear setting in which to observe potential controversy. Both offline and online, political discussions take the form of debates, where opposing groups advocate their individual stances on a set of issues. On social media, a recent trend is that political stakeholders themselves are often at the helm of the discussions: politicians now have an active online presence and their campaigns also run online, with the inclusion of users who voice their support for one side while strongly opposing another. This constitutes a unique and important setting in which to observe how social media discussions evolve in the midst of diverging opinions, beliefs, and ideologies.
In this context, we collected and analyzed Twitter discussions surrounding two prominent campaigns: the US presidential election in 2016, and the UK referendum on the “Brexit.” The discussions comprised more than a million tweets made by thousands of users in response to tweets by key political stakeholders themselves, throughout a period of several months. Our main objective was to find outstanding patterns in the longitudinal activity of users in these discussions, as well as insights on the impact of user activity and overarching topics on the way the discussion unfolds.

In our approach, we modeled discussions as a multi-faceted data space, where the facets of interest relate to the main topics of conversation, the inclination of users towards a particular stance, and the level of user activity. To analyze this facet space, we proposed a general framework consisting of data mining techniques. For topic discovery, we employed a Latent Dirichlet Allocation [1] model along with manual labeling of the latent topics. For stance and user activity, we trained a binary classifier with information about the age of user accounts, their total number of tweets in and outside of the political discussions at hand, the amount of replies received on their tweets, and the number of other users each account follows.

Under this framework, we were able to gain interesting insights about how political adversaries approached campaign topics, the factual or emotional nature of the discussions, and how certain user groups focus their activity on promoting or opposing a political representative. In specific, we found that winning campaigns relied more on emotional claims and sentimental messages of support or opposition. Further, these campaigns included a disproportionate amount of activity from unusually active accounts, a phenomenon that was also detected in further studies about political bots [7], and even addressed in the media.\textsuperscript{16}

Our findings were presented in the Data Mining in Politics workshop at the IEEE International Conference on Data Mining (ICDM 2017) [6].

Traits and Anomalies of Political Discussions on Reddit

Expanding on our previous work, we investigated long term discussions in two prominent political and news-centric Reddit communities. Once more, we were interested in discovering patterns in how users interacted with each other given specific and often polarizing overarching topics of conversation.

Unlike on Twitter, Reddit posts may receive feedback in the form of positive and negative votes from other users. These difference between the upvotes and downvotes a post has received appears in the form of a score, which influences the visibility of both individual posts and entire discussion threads within the community. We interpret these scores as a measure of the overall community reaction to individual posts, and are thus able to distinguish posts which received markedly negative or mixed reactions from posts which received overall positive reactions. Based on the pattern of occurrences of posts with opposing scores throughout a discussion, we then propose four interesting and frequent conversational archetypes: Harmony, Discrepancy, Disruption and Dispute.

To investigate the semantic interpretation of these score-based pattern groups, we analyzed discussions that were categorized into each group with regards to their textual features, namely sentiment and textual similarity. For the former, we identified positive, neutral

and negative sentiment in individual posts using VADER [4]. For the latter, we evaluated the topical cohesiveness of posts by measuring the similarity of their sentence embeddings [2]. We investigate how these features manifest in discussions through a suite of statistical hypotheses, which express our expectations for the pattern groups while allowing us to investigate the relationship between

Our analyses confirm that each of the proposed patterns reveals a distinct conversational dynamic, where differences between post scores were mirrored in the topic and sentiment dimensions. We found that the Harmony pattern represented discussions which lacked strong disagreements or negative sentiment, and which followed a loose format allowing for humor and off-topic discussion. In the Discrepancy pattern, posts that received negative or mixed feedback were found to be more negative in sentiment and to deviate from other posts in the same discussion. The Disruption pattern featured significant shifts in sentiment, particularly from positive to negative, as well as changes in the predominant topic of conversation. Finally, the Dispute pattern displayed a mix of sentiments and topically dissimilar posts, a likely reflection of users expressing differing opinions and using different arguments to support their individual claims.

These findings are published in the AAAI International Conference on Web and Social Media (ICWSM) [5].

References


39.9 Academic Activities

39.9.1 Journal Positions

Pauli Miettinen:
- *Data Mining and Knowledge Discovery* (member of the editorial board since 2015, action editor since 2017)
- *Frontiers in Big Data and Artificial Intelligence* (review editor since 2018)

Jilles Vreeken:
- *Data Mining and Knowledge Discovery* (member of the editorial board since 2015)
- *Journal Track of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (member of the guest editorial board since 2013)

Gerhard Weikum:
- *Member of the Editorial Board of ACM Transactions on the Web (TWEB)* (2012-2018)

39.9.2 Book Positions

Gerhard Weikum:
- *Member of the Editorial Board of the ACM Books Series* (since 2014)

39.9.3 Conference and Workshop Positions

Membership in program and organization committees

Klaus Berberich:
- 10th International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, UK, February 2017 (PC member)
- Conference on Human Information Interaction and Retrieval, CHIIR 2017, Oslo, Norway, March 2017 (PC member)
- 20th International Conference on Extending Database Technology, EDBT 2017, Venice, Italy, March 2017 (PC member)
- 26th International World Wide Web Conference, WWW 2017, Perth, Australia, April 2017 (PC member)
- 7th Temporal Web Analytics Workshop, TempWeb 2017, Perth, Australia, April 2017 (PC member)
- IEEE International Conference on Data Engineering, ICDE 2017, San Diego, USA, April 2017 (PC member)
- 21st European Conference on Advances in Databases and Information Systems, ADBIS 2017, Nicosia, Cyprus, September 2017 (PC member)
26th International Conference on Information and Knowledge Management, CIKM 2017, Singapore, Singapore, November 2017 (Senior PC member)

23rd SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2017, Halifax, Canada, August 2017 (PC member)

40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2017, Tokyo, Japan, August 2017 (PC member)

Conference on Human Information Interaction and Retrieval, CHIIR 2018, New Brunswick, USA, March 2018 (PC member)

First Workshop on Narrative Extraction from Text, Text2Story 2018, Grenoble, France, March 2018 (PC member)

The Web Conference, WWW 2018, Lyon, France, April 2018 (PC member)

8th Temporal Web Analytics Workshop, TempWeb 2018, Lyon, France, April 2018 (PC member)

41st International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2018, Ann Arbor, USA, July 2018 (PC member)

27th International Conference on Information and Knowledge Management, CIKM 2018, Turin, Italy, October 2018 (Senior PC member)

The Web Conference, WWW 2019, San Francisco, USA, May 2019 (PC member)

9th Temporal Web Analytics Workshop, TempWeb 2019, San Francisco, USA, May 2019 (PC member)

42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 2019 (PC member)

Asia J. Biega:

- ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 2019 (PC member)
- SIAM International Conference on Data Mining, SDM 2019, Calgary, Canada, May 2019 (PC member)
- ACM Conference on Fairness, Accountability, and Transparency, FAT* 2019, Atlanta, USA, January 2019 (PC member)

Mohamed Gad-Elrab:

- 32th AAAI Conference on Artificial Intelligence 2018, New Orleans, Louisiana, USA, February 2018 (PC member)

Dhruv Gupta:

- UISTDA2018: ACM IUI2018 Workshop on User Interfaces for Spatial and Temporal Data Analysis, Tokyo, Japan. March 2018 (PC member)
- First International Workshop on Narrative Extraction from Texts (Text2Story’18 @ECIR’18), Grenoble, France. March 2018 (PC member)
- ACM International Conference on Information and Knowledge Management, CIKM 2018, Turin, Italy. 22-26 October 2018 (PC member)
- UISTDA2019: ACM IUI2019 Workshop on User Interfaces for Spatial and Temporal Data Analysis, Los Angeles, California, USA. March 2019 (PC member)
- Second International Workshop on Narrative Extraction from Texts (Text2Story'19 @ECIR’19), Cologne, Germany. April 2019 (PC member)

Pauli Miettinen:

- 2019 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML-PKDD 2019, Würzburg, Germany, September 2019 (PC member)
- 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2019, Anchorage, Alaska, USA, August 2019 (PC member)
- 19th SIAM International Conference on Data Mining, SDM 2019, Calgary, Alberta, Canada, May 2019 (PC member)
- 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, January–February 2019 (senior PC member)
- 12th ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, Australia, February 2019 (PC member)
- 18th IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 2018 (PC member)
- 2018 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML-PKDD 2018, Dublin, Ireland, September 2018 (PC member)
- 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2018, London, UK, August 2018 (PC member)
- 18th SIAM International Conference on Data Mining, SDM 2018, San Diego, California, USA, May 2018 (PC member)
- 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, New Orleans, Louisiana, USA, February 2018 (PC member)
- 11th ACM International Conference on Web Search and Data Mining, WSDM 2018, Los Angeles, California, USA, February 2018 (PC member)
- 2017 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML-PKDD 2017, Skopje, Macedonia, September 2017 (PC member)
- 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2017, Halifax, Nova Scotia, Canada, August 2017 (PC member)
- 17th SIAM International Conference on Data Mining, SDM 2017, Houston, Texas, USA, April 2017 (PC member)
- 10th ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, UK, February 2017 (PC member)
Paramita Mirza:
- 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2019), Minneapolis, USA, June 2019 (PC member)
- 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018), Brussels, Belgium, November 2018 (PC member)
- 27th International Conference on Computational Linguistics (COLING 2018), Santa Fe, USA, August 2018 (PC member)
- 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018), Melbourne, Australia, July 2018 (PC member)
- International Workshop on Semantic Evaluation 2018 (SemEval 2018), New Orleans, USA, June 2018 (PC member)
- 11th edition of the Language Resources and Evaluation Conference (LREC 2018), Miyazaki, Japan, May 2018 (PC member)
- The Web Conference 2018, Lyon, France, April 2018 (PC member)
- 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, September 2017 (PC member)

Simon Razniewski:
- The Web conference (WWW), San Francisco, USA, 2019 (PC member)
- AAAI Conference on Artificial Intelligence, Honolulu, USA, 2019 (PC member)
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), Minneapolis, USA, 2019 (PC member)
- 1st Conference on Automated Knowledge Base Construction (AKBC), Amherst, USA, 2019 (PC member)
- 2nd Conference on Language, Data and Knowledge, Leipzig, Germany, 2019 (PC member)
- Data Quality in Wikidata workshop, Berlin, Germany, 2019 (PC member)
- 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, 2018 (PC member)
- 27th ACM International Conference on Information and Knowledge Management (CIKM), Turin, Italy, 2018 (PC member)
- Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 2018 (PC member)
- 23rd Conference on Database Systems for Advanced Applications (DASFAA), Gold Coast, Australia, 2018 (PC member)
- 4th Workshop on Semantic Deep Learning (SemDeep-4), Monterey, USA, 2018 (PC member)
- DL4KGs, 1st Workshop on Deep Learning for Knowledge Graphs and Semantic Technologies, Crete, Greece, 2018 (PC member)
- 1st Workshop on Quality of Open Data (QOD 2018), Berlin, Germany, 2018 (PC member)
- Reasoning on Data Workshop (RoD), Lyon, France, 2018 (PC member)
- 26th ACM International Conference on Information and Knowledge Management (CIKM), Singapore, 2017 (PC member)
- International Workshop on the Web and Databases (WebDB), Chicago, USA, 2017 (PC member)

Rishiraj Saha Roy:
- 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, July 2019 (PC member)
- 2019 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL-HLT), Minneapolis, USA, June 2019 (PC member)
- Twelfth Workshop on Dynamics on and of Complex Networks at NetSci 2019 (DOOCN-XII), Vermont, USA, May 2019 (Organizer)
- Sixth Social Networking Workshop at COMSNETS 2019 (COMSNETS SN ’19), Bangalore, India, January 2019 (Organizer)
- ACM India Joint International Conference on Data Science and Management of Data (CoDS-COMAD) 2019, Kolkata, India, January 2019 (Publications Chair)
- Conference on Empirical Methods in Natural Language Processing 2018 (EMNLP), Brussels, Belgium, November 2018 (PC member)
- Eleventh Workshop on Dynamics on and of Complex Networks at CCS 2018 (DOOCN-XI), Thessaloniki, Greece, September 2018 (Organizer)
- Fifth Social Networking Workshop at COMSNETS 2018 (COMSNETS SN ’18), Bangalore, India, January 2018 (Organizer)
- Conference on Empirical Methods in Natural Language Processing 2017 (EMNLP), Copenhagen, Denmark, September 2017 (PC member)
- Tenth Workshop on Dynamics on and of Complex Networks at NetSci 2017 (DOOCN-X), Indianapolis, USA, June 2017 (Organizer)

Daria Stepanova
- 33rd International Conference on Artificial Intelligence AAAI 2019, Honolulu, Hawaii, USA, January 2019 (PC member)
- 32nd International Conference on Artificial Intelligence AAAI 2018, New Orleans, Louisiana, USA, February 2018 (PC member)
- 15th European Semantic Web Conference, ESWC 2018, Heraklion, Crete, Greece, June 2018 (PC member)
- 31st International Conference on Artificial Intelligence AAAI 2017, San Francisco, California, USA, February 2017 (PC member)
Jannik Strötgen:
- Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 2018, (PC member),
- International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Ann Arbor, Michigan, 2018, (PC member),
- Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, 2018, (PC member),
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), New Orleans, Louisiana, 2018, (PC member),
- International Conference on Information and Knowledge Management (CIKM), Turin, Italy, 2018 (PC member),
- Digital Humanities Conference (DH), Mexico City, Mexico, 2018, (PC member),
- Temporal Web Analytics Workshop (TempWeb), Lyon, France, 2018 (PC member),
- Fifth Social Networking Workshop at COMSNETS (COMSNETS, SN), Bengaluru, India, 2018 (PC member),
- International Conference on Information and Knowledge Management (CIKM), Singapore, 2017 (PC member),
- Digital Humanities Conference (DH), Montreal, Canada, 2017, (PC member),
- Temporal Web Analytics Workshop (TempWeb), Perth, Australia, 2017 (PC member),
- Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), Vancouver, Canada, 2017 (PC member),
- Drift-a-LOD, Workshop on the Detection, Representation and Management of Concept Drift in Linked Open Data (Drift-a-LOD), Amsterdam, Netherlands, 2017 (PC member),
- EPIA Conference on Artificial Intelligence (EPIA), Porto, Portugal, 2017 (PC member),
- Workshop on Teaching NLP for Digital Humanities (Teach4DH), Berlin, Germany, 2017 (PC member).

Jilles Vreeken:
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Anchorage, Alaska, 2019 (Senior PC member),
Andrew Yates:

- 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2019), Minneapolis, USA, 2019 (PC member)
- 41st European Conference on Information Retrieval, ECIR 2019, Köln, Germany, 2019 (PC member)
- 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, 2019 (PC member)
- 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018), Melbourne, Australia, 2018 (PC member)
— 11th edition of the Language Resources and Evaluation Conference (LREC 2018), Miyazaki, Japan, 2018 (PC member)
— The Web Conference 2018, Lyon, France, 2018 (PC member)
— 32nd International Conference on Artificial Intelligence AAAI 2018, New Orleans, Louisiana, USA, February 2018 (PC member)

Gerhard Weikum:
— ACL Conference on Computational Linguistics (ACL) 2017, 2018, 2019
— ACL Conference on Empirical Methods for Natural Language Processing (EMNLP) 2017, 2018
— ACM Conference on Knowledge Discovery and Data Mining (KDD) 2018, 2019
— ACM Conference on Research and Development in Information Retrieval (SIGIR) 2017, 2018 (Senior PC), 2019 (Senior PC)
— ACM Conference on Web Search and Data Mining (WSDM) 2017 (Senior PC)
— Conference on Automatic Knowledge Base Construction (AKBC) 2019
— Conference on Innovative Data Systems Research (CIDR) 2017, 2019
— International Conference on Very Large Data Bases (VLDB) 2017 (Senior PC / Editorial Board)
— International Conference on the Semantic Web (ISWC) 2019
— The Web Conference (WWW) 2017 (Senior PC), 2018 (Senior PC)

Membership in steering and other committees

Klaus Berberich:
— Member of the steering committee: Fachgruppe Information Retrieval, German Computer Society (Gesellschaft für Informatik), since 2018

Jilles Vreeken:
— Member of the Steering Committee of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) (since 2016)

Gerhard Weikum:
— Member of the Steering Committee of ACM Conference on Web Search and Data Mining (WSDM) (since 2019)
— Member of the Steering Committee of Conference on Design of Experimental Search and Information Retrieval Systems (DESIREs) (since 2018)
39.9.4 Invited Talks and Tutorials

Asia J. Biega:

- **Enhancing Fairness and Privacy in Search Systems**, invited talk at the L3S Research Center, Hannover, Germany, January 2019
- **Enhancing Fairness and Privacy in Search Systems**, invited participation and presentation in the Microsoft Research AI Breakthroughs event, Microsoft Research Redmond, USA, September 2018
- **Enhancing Fairness and Privacy in Search Systems**, invited talk at Microsoft Research Montreal, Canada, August 2018
- **Wanted and Unwanted Exposure: Investigating Fairness and Privacy Problems in Search Systems**, invited talk Wikimedia Foundation, remote, August 2018
- **How to mentor women in IR? What works, what doesn’t, and why?**, invited panel discussant, Women in IR @ SIGIR 2018, Ann Arbor, Michigan, USA, July, 2018
- **Wanted and Unwanted Exposure: Investigating Fairness and Privacy Problems in Search Systems**, invited talk at Microsoft Research Cambridge, UK, February 2018
- **At the Intersection of Information Retrieval and User Privacy**, invited talk at Google, Zurich, Switzerland, September 2017

Panagiotis Mandros:

- **Finding approximate functional dependencies for knowledge discovery**, Invited Talk, Hasso Plattner Institute, Potsdam, Germany, October 2018

Pauli Miettinen:

- **Redescription Mining: Theory, Algorithms, and Applications**, tutorial, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), London, UK, August 2018
- **Boolean Tensor Factorisations and Beyond**, invited talk, 14th International Conference on Concept Lattices and Their Applications, CLA, Olomouc, Czech Republic, June 2018
- **Surprisingly Correlated Subgroups and Hyperbolic Communities**, invited talk, University of Eastern Finland, Kuopio, Finland, February 2018
- **Hyperbolic Communities: Modelling Communities Beyond Cliques**, invited talk, INRIA Nancy Grand–Est, France, June 2017

Paramita Mirza:

- **Mining Temporal and Causal Relations From Text**, invited lecture series at the Department of Linguistics, Ruhr-Universitaet Bochum, Germany, June 2017

Simon Razniewski:

- **Identifying and Linking Knowledge Base Counting Quantifiers**, invited talk at FU Bozen-Bolzano – Italy, 2018
What knowledge bases know (and what they don’t), invited talk at HCC lab at FU Berlin – Germany, 2018

Query-driven Data Completeness Assessment, invited talk at University Lyon 1 – France, 2017

Query-driven Data Completeness Assessment, invited talk at Telecom ParisTech University – France, 2017

Daria Stepanova:

Rule Induction and Reasoning over Knowledge Graphs, tutorial, Reasoning Web summer school (RW), Luxembourg, September 2018

Jilles Vreeken:

Causality, and How To Infer It From Observational Data, Keynote, 2nd Summer School by the European Association for Data Science, Luxemburg, September 2019

Summarizing Graphs at Multiple Scales, Tutorial, IEEE International Conference on Data Mining, Singapore, November 2018

What’s Going On In My Data? How to Discover Interpretable Correlations, Associations, and Causation, Keynote, NOMAD Summer – A hands-on course on tools for novel-materials discovery, Lausanne, Switzerland, September 2018

Simply Telling Cause from Effect, Keynote, 3rd Workshop on Formal Reasoning about Causation, Responsibility, and Explanation in Science and Technology at ETAPS 2018, Thessaloniki, Greece, April 2018

Discovering Interpretable Patterns, Correlations, and Causality, Keynote, Symposium on Managing and Exploiting the Raw Material of the 21st Century at the Spring Meeting of the German Physical Society, Berlin, Germany, March 2018

Is There Anything Out There? Efficiently Discovering Reliable Approximate Functional Dependencies, Keynote, CECAM Workshop on Big-Data Driven Materials Science, Lausanne, Switzerland, September 2017

Patterns, Sets of Patterns, and Pattern Compositions, Keynote, 14th International Conference on Formal Concept Analysis (ICFCA), Rennes, France, June 2017

Andrew Yates:

Neural Information Retrieval: Methodology and Applications, invited talk at Blekinge Institute of Technology – Sweden, September, 2018

Incorporating Positional Information and Other Domain Knowledge into a Neural IR Model, invited talk at University of Glasgow – United Kingdom, October, 2017

Gerhard Weikum:

Beyond Knowledge Graphs, Keynote at the Wiki Workshop at the Web Conference (WWW), Lyon, France, April 2018

Machine Knowledge: Encyclopedic, Scholarly, Commonsense, Invited Talk on the Occasion of the 25th Anniversary of DBLP, Trier, Germany, December 2018
39.9.5 Other Academic Activities

Asia J. Biega:
– Co-organizer: Fair Ranking Track at the Text REtrieval Conference (TREC), 2019

Pauli Miettinen:
– Peer Reviewer: Research Foundation – Flanders (FWO), 2018
– Peer Reviewer: KAUST Competitive Research Grant, 2018

Jilles Vreeken:
– Research Group Leader Representative on the Board of the Cluster of Excellence on Multimodal Computation and Interaction (since 2016)
– Peer Reviewer: Evaluation Board Member for ICT-2018 (EU Horizon 2020) 2018
– Peer Reviewer: European Research Council 2018
– Peer Reviewer: Research Foundation – Flanders (FWO) 2018

Gerhard Weikum:
– Spokesperson of the International Max Planck Research School (IMPRS) for Computer Science
– Peer reviewer for the European Research Council (ERC), German Science Foundation (DFG) and various other research funding agencies
– Member of the VLDB Awards Committee, 2015-2019
– Member of the Awards Committee for the ACM Paris Kanellakis Theory and Practice Award, 2015-2018
– Member of the Committee for the ACM CIKM Test-of-Time Award 2018
– Member of the Committee for the ACM WSDM Best Paper Award 2017 and 2019
– Member of the Working Group on “Perspectives of Computer Science” of the German Council for Science and Humanities (Wissenschaftsrat), 2018-2020

39.10 Teaching Activities

Summer Semester 2017
Courses:
Data Mining and Matrices (P. Miettinen)
Topics in Algorithmic Data Analysis (J. Vreeken)

Winter Semester 2017/2018
Courses:
Information Retrieval and Data Mining (J. Vreeken, J. Strötgen)
Tensors in Data Analysis (P. Miettinen)
Knowledge Representation for the Semantic Web (D. Stepanova)
Seminars:

Knowledge Bases (Simon Razniewski and Paramita Mirza)

Summer Semester 2018

Courses:

Topics in Algorithmic Data Analysis (J. Vreeken)

Winter Semester 2018/2019

Courses:

Elements of Statistical Learning (J. Vreeken, T. Marschall)

Seminars:

Advanced Topics in Knowledge Bases (Simon Razniewski)

Master and Bachelor Theses

Iva Farag: Efficiently Summarising Data with Patterns that Overlap, Master’s thesis, 2018, (Supervisor: Jilles Vreeken)
Thinh Vinh Ho: Rule Learning from Knowledge Graphs Guided by Embedding Models, Master’s thesis, 2018, (Supervisor: Daria Stepanova)
Vonny Pawaka: Mining Correlations of Human Activities and Feelings from Movie Scripts, Master’s thesis, 2018, (Supervisor: P. Mirza)

39.11 Dissertations, Habilitations, Awards

39.11.1 Dissertations


### 39.11.2 Offers for Faculty Positions

– Mario Boley: Lecturer (Assistant Professor) at Monash University, Melbourne, Australia
– Pauli Miettinen: Full Professor at University of Eastern Finland, Kuopio, Finland
– Simon Razniewski: Lecturer at The University of Auckland, Auckland, New Zealand
– Jilles Vreeken: Tenured Faculty at CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

### 39.11.3 Awards

– Patrick Ernst, Amy Siu and Gerhard Weikum: *Best Paper Award* at the Web Conference 2018 (WWW 2018) for *HighLife: Higher-arity Fact Harvesting*
– Dhruv Gupta and Klaus Berberich: *Poster Honorable Mention* at WWW 2018 for *Identifying Time Intervals for Knowledge Graph Facts*
– Johannes Hoffart, Fabian Suchanek, Klaus Berberich, Gerhard Weikum: *AIJ Influential Paper Award* presented at IJCAI 2017 for the paper *YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia* in the Artificial Intelligence Journal (AIJ) 2013
– Francesca A. Lisi and Daria Stepanova: *Best Poster Award* at RuleML+RR 2017 for *Combining Nonmonotonic Reasoning and Rule Learning for Link Prediction in Knowledge Graphs*
– Panagiotis Mandros, Mario Boley, and Jilles Vreeken: *Best Paper Award* at IEEE ICDM 2018 for *Discovering Reliable Dependencies from Data: Hardness and Improved Algorithms*
– Pauli Miettinen: *Best Reviewer Award* at ACM KDD 2018
– Thomas Pellissier Tanon, Daria Stepanova, Simon Razniewski, Paramita Mirza, Gerhard Weikum: *Best Student Paper Honorable Mention* at ISWC 2017 for *Completeness-aware Rule Learning from Knowledge Graphs.*
– Jannik Strötgen: *Best Reviewer Award* at EMNLP 2018
39.12 Grants and Cooperations


D5 participates in the DFG Excellence Cluster on Robust, Efficient, and Intelligent Processing of Multimodal Information (text, speech, visual data, etc.), the institute’s lead theme. Gerhard Weikum is one of the cluster’s 15 principal investigators (but receives no funding from the cluster). Until October 2018 Jilles Vreeken led an independent research group at Saarland University while also being a senior researcher of MPI-INF and associated with D5. Jilles has moved to the CISPA Helmholtz Center on campus, still keeping his affiliation with MPI-INF as well.

ERC Synergy Grant imPACT, 2015-2020

The imPACT project is funded by the European Research Council (ERC), with a total of ca. 10 Mio. Euros over 6 years (2015-2020), by awarding an ERC Synergy Grant to the four principal investigators Michael Backes (Saarland University), Peter Druschel (MPI for Software Systems), Rupak Majumdar (MPI for Software Systems), and Gerhard Weikum (MPI for Informatics). imPACT stands for “Privacy, Accountability, Compliance, and Trust for the Internet of Tomorrow”. It addresses the challenge of providing these PACT properties in the Internet. A key goal is to understand and master the different roles, interactions and relationships of users and their joint effect on the four PACT properties and the potential tensions between them. The focus is on a user-centric perspective, taking into consideration modern user behaviors in a wide spectrum of online communities and commercial services.

DFG Collaborative Research Center 1223 on Methods and Tools for Understanding and Controlling Privacy

D5 participates with two sub-projects in this DFG-funded Collaborative Research Center (CRC). Our research in this context specifically addresses models, methods and tools for analyzing and assessing privacy risks of online users, including the risks of extensive profiling and of being linked across different platforms or communities (despite using pseudonyms). This involves collaboration with Michael Backes and Jens Dittrich, both professors at Saarland University.
BigMax – MaxNet on Big-Data-Driven Material Discovery (2017-2020)

BigMax is the short-hand name for the MaxNet on Big-Data Driven Material Discovery, funded by the Max Planck Society. The network comprises partners from 9 different Max Planck Institutes, mostly in the domain of material science. Our research in this context aims to investigate theory and methods for using data mining and machine learning techniques for discovering insights from material science data.

39.13 Publications

Books and proceedings


Journal articles and book chapters


Conference and workshop articles


D. Kaltenpoth and J. Vreeken. We are not your real parents: Telling causal from confounded by MDL. In *Proceedings of the 2019 SIAM International Conference on Data Mining (SDM 2019)*, Calgary, Canada, 2019. SIAM. Accepted.


Theses


39 D5: Databases and Information Systems


Technical reports and other works


40  RG1: Automation of Logic

40.1 Personnel

Head of Group
Prof. Dr. Christoph Weidenbach

Researchers
Dr. Jasmin Christian Blanchette (Guest)
Dr. Thomas Sturm (Guest)
Dr. Sophie Tourret (October 2017–)
Dr. Uwe Waldmann

Ph.D. Students
Gábor Alagi (–August 2017)
Martin Bromberger
Albert Fiori (August 2018–)
Mathias Fleury
Fabian Kunze (–June 2017)
Ching Hoo Tang (–August 2017)
Andreas Teucke (–May 2018)
Marco Voigt
Daniel Wand (–August 2017, Guest)

Research Programmers
Michaël Mera (February 2018–April 2018)

Secretaries
Jennifer Müller

Jasmin Blanchette is an assistant professor at the university of Amsterdam and associated with the institute. He is a former member of the VeriDis project (see Sect. 40.12).

Thomas Sturm joined CNRS as a research director in January 2016 and is since then associated with the institute.

40.2 Visitors

From March 2017 to February 2019, the following researchers visited our group:

Anders Schlichtkrull 03.04.2017–04.08.2017 Technical Univ. of Denmark
40.3 Foundations of Automated Reasoning

First-order logic offers many challenges for automated reasoning: the satisfiability problem is undecidable in general and many decidable subproblems require computationally expensive reasoning in the worst case. In the reporting period we contributed to the foundations of automated reasoning, in particular in first-order logic. Our contributions range from the identification of new decidable fragments to improved reasoning technology for first-order theorem proving and first-order abductive reasoning. In Sections 40.3.1 and 40.3.7 we discuss our recent findings with regarding decidable and undecidable fragments of first-order logic with our without background theories. Sections 40.3.2, 40.3.3, 40.3.5, and 40.3.6 give an impression of the reasoning techniques we have developed or refined in the reporting period with the aim of improving the current state of the art in automated reasoning in the areas first-order theorem proving, abduction for first-order logic with equality, and SMT solving. Finally, we also outline our recent results on the representation of models in satisfiability checking for first-order clause sets in Sections 40.3.3 and 40.3.4.

40.3.1 Novel Decidable Fragments of First-Order Logic

**Investigators:** Manuel Lamotte-Schubert, Andreas Teucke, Marco Voigt, and Christoph Weidenbach

Identifying non-trivial fragments of first-order logic with a decidable satisfiability problem is a formidable task. This is the positive side of what is known as the classical decision problem [2]. Today a large number of decidable fragments is known. We list nine of the best known in Figure 40.1 (upper part), the oldest being the monadic first-order fragment, introduced in 1915 [7], the newest being the guarded-negation fragment, introduced in 2011 [1]. We succeeded in extending all nine fragments substantially into larger formula classes for which satisfiability remains decidable. To this end, we have blended each of the original definitions with a simple and elegant syntactic concept called separateness of variables. The lower part of Figure 40.1 depicts the extensions. One example is the separated fragment (SF) [11], comprising relational first-order sentences with equality in which, roughly speaking, universal and existential variables never co-occur in atoms. SF extends the Bernays–Schönfinkel–Ramsey fragment (BSR), i.e. the class of relational \( \exists^*\forall^* \)-prefix sentences with
MFO – monadic first-order fragment
BSR – Bernays–Schönfinkel–Ramsey fragment
FO² – two-variable fragment
AF – Ackermann fragment
GKS – Gödel–Kalmár–Schütte fragment
FL – fluted fragment
GF – guarded fragment
LGF – loosely guarded fragment
GNFO – guarded negation fragment

SF – separated fragment
GBSR – generalized BSR
SFO² – separated FO²
GAF – generalized AF
GGKS – generalized GKS
SFL – separated FL
SGF – separated GF
SLGF – separated LGF
SGNFO – separated GNFO

Figure 40.1: Top: Schematic overview of well-known decidable fragments of first-order logic. Only the containment of AF in GKS and of GF in LGF and the partial overlaps with MFO are depicted. All other overlaps are neglected. Bottom: Schematic overview of the extended fragments (in green) we have identified. Each of them contains MFO in addition to the respective original fragment.
equality. Evidently, separateness of first-order variables opens a new perspective on the landscape that research activity around the classical decision problem has revealed. Each sentence from one of our extended fragments can be translated into an equivalent sentence from the underlying original fragment. However, in certain cases, this translation comes at a price of a non-elementary blowup of the formula length. Such a behavior can be observed, for example, in the worst case for SF sentences that are translated into equivalent BSR sentences. This shows that certain logical properties can be expressed non-elementarily more succinct when using formulas from the extended fragments rather than from the underlying originals. Not surprisingly, the computational complexity required to decide satisfiability is affected in a similar fashion. While BSR-Sat is complete for nondeterministic exponential time (NExpTime), the satisfiability problem for SF, SF-Sat, is complete for the complexity class Tower [18] (see [8] for the definition of Tower). The class Tower contains every complexity class defined in terms of an elementary bound on running time.

In addition to the above said, we have shown closedness under Craig–Lyndon interpolation for several of the extended fragments [19, 17]. Moreover, we have shown that separateness of variables has applications beyond the classical decision problem, e.g. in second-order quantifier elimination [20] or as a basis for improved Skolemization techniques that are beneficial for first-order theorem provers in certain cases. The mentioned results on separateness and the extended fragments have not been fully published yet. However, they constitute a core part of a manuscript submitted by Voigt as a dissertation [17].

In addition to the novel decidable fragments described so far, we have also identified two new decidable clausal fragments of first-order logic. The first fragment is abbreviated BDI for bounded depth increase [5, 4, 6]. It strictly contains known fragments such as PVD (positive-variable-dominated fragment) [3]. The arity of function and predicate symbols as well as the shape of atoms is not restricted in BDI. Instead, the shape of “cycles” in resolution inferences is restricted so that the depth of terms occurring in generated clauses may increase but is still bounded. BDI is motivated by real-world problems where function terms are used to represent record structures. We have shown that the hyper-resolution calculus modulo redundancy elimination terminates on finite BDI clause sets. Employing this result to the ordered resolution calculus, we have also proven termination of ordered resolution on BDI, yielding a more efficient decision procedure.

The second novel clause fragment that we have shown to be decidable in the monadic shallow linear fragment with straight dismatching constraints, abbreviated MSL(SDC) [15, 12, 16]. It extends the monadic shallow linear Horn fragment (MSLH), which was proven to be decidable by Weidenbach in the nineties already [21]. MSLH is the class of finite sets of first-order Horn clauses without equality in which all clauses are monadic, shallow, and linear. That is, every predicate symbol has arity one, the depth of terms occurring in positive literals is at most one, and any first-order variable occurring in a positive literal occurs at most once in that literal. MSL(SDC) extends MSLH in three ways: (1) we relax the restriction to Horn clauses and allow several positive literals per clause if they are pairwise variable disjoint, (2) we allow unit clauses containing exactly one disequation $s \neq t$ over non-unifiable terms $s,t$, and (3) we allow the addition of straight dismatching constraints to clauses. MSLH and its extensions have found applications, for example, in the verification of security protocols [21, 9, 10]. Since our extension allows for non-Horn clauses, MSL(SDC) facilitates the modeling of nondeterminism in such protocols. The extension with straight dismatching
constraints, on the other hand, does not increase expressiveness, but facilitates representations of saturated clause sets that may be exponentially shorter than representations without such constraints. We exploit this property to improve our *approximation refinement* approach to automated first-order theorem proving [13, 14, 15, 12, 16]. We shall elaborate on the latter in Section 40.3.2. Moreover, MSLH and its extensions will also play a role in Section 40.3.4, where we discuss representations of finite and infinite candidate models.

References


40.3.2 The Approximation Refinement Approach to First-Order Theorem Proving

Investigators: Andreas Teucke and Christoph Weidenbach

We have developed an approximation-based approach to theorem proving using a counterexample-guided abstraction refinement loop [2, 3, 4, 1, 5]. Given any finite set of clauses $N$, we approximate it with a simpler clause set $N'$ that (a) is unsatisfiable whenever the original $N$ is unsatisfiable and (b) belongs to a decidable first-order fragment for which decision procedures are available that work well in practice. A resolution refutation generated by a decision procedure on the simplified clause set $N'$ can then either be lifted to a refutation for the original $N$, or it guides a refinement that excludes the previously found unliftable refutation. Done in this way, the approach is refutationally complete. Initially, we used the monadic shallow linear Horn fragment (MSLH) as the target class for our approximation scheme. As described in Section 40.3.1, we have identified the non-Horn extension MSL(SDC) for which satisfiability is still decidable [4, 5]. This new decidable first-order fragment, in particular the extension with straight dismatching constraints, was partially motivated by an improved refinement step of the approximation refinement framework. All needed operations on straight dismatching constraints require only linear or linear logarithmic time in the size of the constraint. Ordered resolution with straight dismatching constraints is sound
and complete and can be tuned such that it constitutes a decision procedure for MSL(SDC). We have implemented the approach based on the SPASS theorem prover called SPASS-AR [4, 5, 1]. On certain satisfiable problems, our implementation outperforms established provers such as SPASS, iProver, and Vampire.

References


40.3.3 Clause Learning from Simple Model Representations for the Bernays–Schönfinkel fragment

Investigators: Alberto Fiori and Christoph Weidenbach

The Bernays–Schönfinkel fragment (BS) is a natural generalization of propositional logic enjoying the finite model property. Viewed as a clausal fragment, a BS clause is any clause without equality in which all featured function symbols are actually constants. Several decision procedures for BS rely on explicit model representations [2, 4, 1, 3]. Since satisfiability for BS is known to be NExpTime-complete, an explicit model representation either must become exponentially large in the worst case, or satisfiability testing with respect to the model representation can, in general, not be done in polynomial time. This justifies and motivates the search for suitable model representation formalisms.

In the reporting period we developed a new decision procedure, called SCL for clause learning from simple models, partially based on earlier work of our group on NRCL [1]. SCL maintains an explicit model representation using a list of ground, i.e. variable-free, literals. The most important computations with respect to a partial model representation are the detection of a propagating literal or a false clause. Such tasks can be solved more efficiently for ground model representations than for more sophisticated representations. On the other hand, non-ground representations, such as the very expressive one used in NRCL, may be exponentially more compact. However, the mentioned computation problems become NP-complete. We were able to show that this computational overhead can be avoided for model-driven clause learning in certain cases, as any clause learned by NRCL can also be learned by SCL. This does not necessarily entail that SCL is superior to the other procedures.
Much rather, this indicates that it can be efficiently used on clause sets where the ground model representation does not become “too large”. Moreover, SCL inherits from NRCL the quality of learning only clauses that are non-redundant with respect to a so-called reasonable strategy. Practically, this implies that any clause generated by SCL based on a reasonable strategy does not need to be tested for forward redundancy. This can help avoiding lengthy computations, as experience shows that saturation-based theorem provers spend a substantial share of their runtime on testing forward redundancy. One publication treating SCL is currently under review.

References


40.3.4 Representation of Inherently Infinite Models

Investigators: Andreas Teucke, Marco Voigt, and Christoph Weidenbach

In the general setting, proving satisfiability of a first-order clause set is more difficult than proving unsatisfiability. Fully automated approaches to the former require a suitable model-representation formalism, in particular, when we attempt to prove the satisfiability of clause sets with inherently infinite models. In our approximation refinement approach to first-order theorem proving (cf. Section 40.3.2) models are represented, e.g., by finitely saturated MSLH clause sets (MSLH stands for the monadic shallow linear Horn fragment, cf. Section 40.3.1). As we have only recently shown that MSLH possess the finite model property, it became evident that such clause sets cannot represent infinite models directly. This new result has also implications for other model representation formalisms studied in the literature, e.g. [2, 1, 3].

The fact that MSLH clause sets have the finite model property does not mean that the approximation refinement (AR) calculus discussed in Section 40.3.2 cannot be used for finding infinite models of clause sets with inherently infinite models. The reason is that the MSLH model representation in [5, 6] does not directly relate to the clause set, but via an approximation. We have shown that, if the inherent infiniteness of models is due to a reflexive relation, then AR can actually terminate on clause sets with inherently infinite models. The key is an additional approximation step called reflexive relation splitting, which, roughly speaking, splits the critical relation into a reflexive and an irreflexive part and syntactically separates the two in the considered clause set. This step preserves unsatisfiability, i.e. if the approximation is satisfiable, then the original is also satisfiable. However, the existence of a
The research described in this section is not yet fully published. Reflexive relation splitting is described in [4, 7].

References

40.3.5 Prime Implicate Generation in Equational Logic

Investigators: Sophie Tourret, in cooperation with Mnacho Echenim (Univ. Grenoble Alpes, Grenoble INP, LIG) and Nicolas Peltier (Univ. Grenoble Alpes, CNRS, LIG)

Abductive reasoning is the process of inferring, given a set of axioms $\mathcal{A}$ and a formula $\phi$, a set of assertions $\mathcal{H}$ such that $\mathcal{A} \cup \mathcal{H} \models \phi$. The set $\mathcal{H}$ may be viewed as a set of hypotheses that are sufficient to ensure the validity of the entailment $\mathcal{A} \models \phi$, or as a set of explanations of $\phi$. Such hypotheses must be economical and plausible, in particular $\mathcal{H}$ must be minimal with regard to logical entailment and $\mathcal{A} \cup \mathcal{H}$ must be satisfiable. Abductive reasoning has many applications in artificial intelligence, verification and debugging, e.g. for computing missing pre-conditions, spotting and correcting errors in a logical specification, or for dealing with approximative, incomplete or spurious information. The problem has been thoroughly investigated in propositional logic [4], and very efficient algorithms have been proposed [6], but only a few approaches handle more expressive logics [3, 5]. In particular, none of them is able to deal with the equality predicate in an efficient way.

We tackled the problem of generating such a set $\mathcal{H}$, when $\mathcal{A} \cup \{\phi\}$ is a set of first-order formulas with equality and $\mathcal{H}$ is a set of ground unit clauses. In this case, by duality, $\neg \mathcal{H}$ is a clausal logical consequence of $\mathcal{A} \cup \{\neg \phi\}$, i.e., an implicate of $\mathcal{A} \cup \{\neg \phi\}$, and the problems boils down to efficiently generating sets of (entailment-minimal) implicates or
prime implicates. Thus we developed an algorithm for the generation of prime implicates in equational logic [1, 2]. This algorithm is defined by a calculus that we proved correct and complete. For the case where the considered clause set is ground, i.e., contains no variables, we devised a specialized tree data structure that efficiently detects and deletes redundant implicates. We produced termination and correctness proofs for all algorithms. An experimental evaluation was conducted in the ground case, including a comparison with state-of-the-art propositional and first-order prime implicate generation tools, that showed our approach was much better suited to handle equational problems than the other tools.

References


40.3.6 Scalable, Fine-Grained Proofs of Preprocessing for SMT Solving

Investigators: Jasmin Christian Blanchette and Mathias Fleury, in cooperation with Haniel Barbosa (U. Iowa) and Pascal Fontaine (U. Lorraine)

We have developed a framework for processing formulas in automatic theorem provers, with generation of detailed proofs. The main components are a generic contextual recursion algorithm and an extensible set of inference rules. Clausification, Skolemization, theory-specific simplifications, and expansion of ‘let’ expressions are instances of this framework. With suitable data structures, proof generation adds only a linear-time overhead, and proofs can be checked in linear time.

We implemented the approach in the SMT solver veriT. This allowed us to dramatically simplify the code base while increasing the number of problems for which detailed proofs can be produced, which is important for independent checking and reconstruction in proof assistants. This was the subject of a conference publication [2]. The work has since been extended by Mathias Fleury to include reconstruction of veriT proofs in Isabelle/HOL. The extended version of the conference paper has recently been accepted in a special issue of J. Automated Reasoning [1], with Fleury as an additional coauthor.
40.3.7 Decidable and Undecidable Fragments of First-Order Logic Modulo
Linear Arithmetic

Investigators: Matthias Horbach, Marco Voigt, and Christoph Weidenbach

In a series of papers our group has explored the decidability boundary for first-order logic modulo linear arithmetic \cite{6, 2, 8, 4, 7, 5}. The results obtained and published during the current reporting period \cite{4, 7, 5} shed new light on both sides of the boundary. On the positive side we have identified three new fragments for which satisfiability is decidable. All three are based on the Bernays–Schönfinkel–Ramsey fragment (BSR) (cf. Section 40.3.1) and are called BSR with simple linear integer constraints (BSR(SLI)), BSR with simple linear rational constraints (BSR(SLR)), and BSR with bounded difference constraints (BSR(BD)). Indeed, we have shown that checking satisfiability is $\text{NExpTime}$-complete in all three cases. Nevertheless, decision procedures that work well in practice are within reach. Indeed, we have developed a promising instantiation methodology for BSR(SLI) \cite{4}, which we expect to perform significantly better than existing approaches for similar fragments. For instance, known decidable formalisms from the field of software verification, such as the array property fragment \cite{1}, can be embedded into extensions of BSR(SLI) for which satisfiability is still decidable via our instantiation approach. In fact, certain optimizations that we have built into our instantiation methodology could be readily integrated into existing instantiation approaches for the array property fragment, such as the ones presented in \cite{1, 3}.

On the negative side of the decidability boundary, we have identified several fragments with a satisfiability problem that is undecidable. In some cases, satisfiability and unsatisfiability are not even semi-decidable \cite{5}. We have treated settings over different arithmetic domains: linear arithmetic over the natural numbers, the rationals, and the reals. In many cases it has turned out that a single uninterpreted predicate symbol of arity one suffices to encode the halting problem for two-counter machines. We have tried to keep the number of quantifier alternations and quantifiers at a minimum and restrict the allowed arithmetic atoms as much as possible. For instance, we have found that satisfiability for the $\forall^*$-fragment of Presburger arithmetic with a single uninterpreted predicate of arity one is undecidable. In fact, a single universal quantifier suffices for the encoding of the halting problem of two-counter machines. As soon as we allow a $\forall\exists$ quantifier alternation, the satisfiability problem is not semi-decidable anymore. Hence, sound reasoning procedures for this language are necessarily incomplete. In addition, we have obtained similar results for linear rational and linear real arithmetic.

References

40.4 Arithmetic

Arithmetic reasoning is a topic investigated by two independent research lines: on the one hand, we have Symbolic Computation, with a particular focus on complex problems expressible in pure first-order arithmetic; on the other hand, we have Satisfiability Checking, with a particular focus on existential decision problems that combine arithmetic with other first-order logics. In our group, we have researchers from both schools of thought so that we can best combine the experience and knowledge from both lines of research. Our work encompasses the development of new and the optimization of existing arithmetic reasoning procedures, their practical applications for industry as well as other scientific disciplines, and the realization of our theoretical research as state-of-the-art implementations. Recent results include advances in quantifier elimination procedures over the reals (Section 40.4.1), subtropical methods for real constraints (Section 40.4.2), and satisfiability modulo arithmetic theories (Sections 40.4.3 & 40.4.4). Moreover, we developed specialized methods that investigate the multistationarity of biological reaction networks (Section 40.4.5).

40.4.1 Real Quantifier Elimination by Virtual Substitution

Investigator: Thomas Sturm

Effective quantifier elimination procedures for first-order theories provide a powerful tool for generically solving a wide range of problems based on logical specifications. In contrast to general first-order provers, quantifier elimination procedures are based on a fixed set
of admissible logical symbols with an implicitly fixed semantics. This admits the use of
sub-algorithms from symbolic computation. Various invited conference talks, each with a
different focus, at SMT 2013 in Helsinki, Finland, at FroCos 2015 in Wroclaw, Poland, and
at ACA 2016 in Kassel, Germany, focused on quantifier elimination for the reals and its
applications giving examples from geometry, verification, and the life sciences. The existing
material has been finally consolidated in a journal article, which will hopefully provide an
interesting reference for the application of real algebra, symbolic computation, and related
logical methods in mathematics, engineering, and the sciences [2].

With another invited talk at ISSAC 2018 in New York City, NY, there came another
survey article specifically on the development of the virtual substitution method [4, 1] during
the past 30 years [3].

References
Universität des Saarlandes, Saarbrücken, 2016.
ed., ISSAC’18, 43rd International Symposium on Symbolic and Algebraic Computation, New
York, NY, USA, 2018, pp. 11–16. ACM.

40.4.2 Subtropical Methods for Real Constraints

Investigators: Thomas Sturm, in cooperation with P. Fontaine (University of Lorraine,
France), H. Hong (North Carolina State University, NC), M. Ogawa (JAIST, Japan),
V. K. To (University of Engineering and Technology, Hanoi, Vietnam), X. T. Vu (JAIST,
Japan)

Originally motivated by our studies of Hopf bifurcations for chemical and biological reaction
networks [1] we had developed subtropical methods for heuristically checking real constraints
in the first orthant. In its original form the method aimed at finding a witness for a single
ordering constraint, and the intermediate value theorem was applied for constructing a zero
of a single equational constraint [8]. With one of our biological models, viz. Mitogen-activated
protein kinase (MAPK), we obtained and solved polynomial equations of considerable size.
Our currently largest instance contains 863438 monomials in 10 variables with degrees up to
12.

We have newly developed a generalization of this method from a single to finitely many
ordering constraints [3, 4]. The method has been implemented in an SMT context and
performs surprisingly well on corresponding benchmark sets [2]. This has been recognized
by the SMT community, and the method has been picked up and reimplemented by other
groups, e.g., [7].

On the theoretical side we have studied the incompleteness of our method and could
prove the following characterization: The method works if and only if not only the input
constraints hold but all constraints with the same signs, but possibly other absolute values, of coefficients. To this end we developed a mathematical framework that is independent of Euclidean geometry and corresponding tedious proofs. Instead it resembles classical tropical algebra but works with real numbers and natural ordering in contrast to purely algebraic notions [5, 6].

References


40.4.3 New Techniques for Linear Arithmetic: Cubes and Equalities

*Investigators: Martin Bromberger and Christoph Weidenbach*

Polyhedra and the systems of linear arithmetic constraints $Ax \leq b$ defining them have a vast number of theoretical and real-world applications [1, 9]. It is, therefore, no surprise that the theory of linear arithmetic is one of the most popular and best investigated theories for *satisfiability modulo theories* (SMT) solving [6, 7, 8]. We have investigated several new techniques for linear arithmetic constraint solving. They are all based on the *linear cube transformation*, which allows us to efficiently determine whether a system of linear arithmetic constraints contains an axis-parallel hypercube of a given edge length.

Our first findings based on this transformation are two sound tests that find integer solutions for linear arithmetic constraints [3, 4]. The idea behind the tests is to find hypercubes that are contained inside the input polyhedron and guarantee the existence of an integer solution. The *largest cube test* finds an axis-parallel hypercube with maximum
edge length contained in the input polyhedron, determines its rational valued center, and rounds it to a potential integer solution. The unit cube test determines if a polyhedron contains an axis-parallel hypercube with edge length one, which is the minimal edge length that guarantees an integer solution. The tests are especially efficient for constraints with a large number of integer solutions, e.g., those with infinite lattice width. Inside the SMT-LIB benchmarks, we have found almost one thousand problem instances with infinite lattice width. Experimental results confirm that our tests are superior on these instances compared to several state-of-the-art SMT solvers.

We also discovered that the linear cube transformation can be used to investigate the equalities implied by a system of linear arithmetic constraints [2, 4]. This is based on two facts: (1) a system of constraints with only surface solutions, and thereby without interior solutions, has to imply an equality and vice versa; (2) a polyhedron has only a surface and no interior if the maximal edge length of any hypercube contained in the polyhedron is zero. Hence, we can efficiently verify the existence of an equality with our largest cube test. By turning all inequalities into strict inequalities, we can even further simplify the test for the detection of implied equalities. This new strict system defines exactly the interior of the original polyhedron without the surface. Therefore, if the strict system is unsatisfiable, the original system has no interior and implies an equality.

The above method is also able to return one of the implied equalities as an explanation. This allowed us to extend it into an algorithm that computes an equality basis, i.e., a finite representation of all equalities implied by the linear arithmetic constraints. For this purpose, the algorithm repeatedly applies the above method to find, collect, and eliminate equalities from our system of constraints.

The equality basis has many applications. If transformed into a substitution, it eliminates all equalities implied by our system of constraints, which results in a system of constraints with an interior and, therefore, improves the applicability of our cube tests. The equality basis also allows us to test whether a system of linear arithmetic constraints implies a given equality. We even extended this test into an efficient method that computes all pairs of equivalent variables inside a system of constraints. These pairs are necessary for the Nelson-Oppen style combination of theories [5].

A compilation of our findings on the linear cube transformation has been published in the journal Formal Methods in System Design [4]. This compilation also presents quantifier elimination as an additional application area for our cube and equality techniques.

References


40.4.4 A Reduction from Unbounded Linear Mixed Arithmetic Problems into Bounded Problems

Investigator: Martin Bromberger

We are interested in decision procedures for linear arithmetic in the context of SMT solving. In this context, we often have to handle so-called unbounded constraint system, which are particularly hard to solve linear arithmetic constraint systems. These problems occur in SMT solving because of bad encodings, necessary but complicating transformations, or the sheer complexity of the verification goal. Hence, efficient techniques for handling unbounded problems are necessary for a generally reliable arithmetic decision procedure.

For this reason, we have developed two transformations, called the Double-Bounded reduction and the Mixed-Echelon-Hermite transformation, that efficiently handle unbounded constraint systems [1]. Together the two transformations reduce any constraint system in polynomial time to an equisatisfiable constraint system that is bounded. The transformations are beneficial because many existing approaches for linear mixed and integer arithmetic, e.g., branch-and-bound, are only guaranteed to terminate on bounded constraint systems [3]. Therefore, our transformations are extensions that complete other decision procedures. However, compared to similar bounding transformations, e.g., a priori bounds [2], our transformations return constraint systems that are efficiently solvable in practice. The transformations are so efficient because they orient themselves on the structure of the input constraint system instead of computing a priori (over-)approximations out of the available constants. Experiments on the SMT-LIB benchmarks provide further evidence to the efficiency of the transformations in practice.

The combination of our transformations works as follows: First, we use the Double-Bounded reduction to eliminate all unbounded inequalities from our constraint system. Then we use the Mixed-Echelon-Hermite transformation to shift the variables of our system to ones that are either bounded or do not appear in the new inequalities and are, therefore, eliminated. Moreover, we have developed a polynomial method for converting certificates of (un)satisfiability from the transformed to the original constraint systems.
40 RG1: Automation of Logic

References


40.4.5 Multistationarity of Biological Reaction Networks

Investigators: Thomas Sturm, in cooperation with R. Bradford (University of Bath, UK), F. Boulier (University of Lille, France), F. Fages (Inria Saclay, France), S. Samal (BASF Ludwigshafen, Germany), A. Schuppert (RWTH Aachen, Germany), J. Davenport (University of Bath, UK), M. England (Coventry University, UK), H. Errami (University of Bonn, Germany), V. Gerdt (JINR, Russia), D. Grigoriev (CNRS, France), C. Hoyt (b-it, Bonn, Germany), M. Kosta (Slovak Academy of Sciences), O. Radulescu (University of Montpellier, France), W. Seiler (University of Kassel, Germany), S. Walcher (RWTH Aachen, Germany), A. Weber (University of Bonn, Germany)

The occurrence of multiple steady states (multistationarity) has important consequences on the capacity of signaling pathways to process biological signals, even in its elementary form of two stable steady states (bistability). Bistable switches can act as memory circuits storing the information needed for later stages of processing. The response of bistable signaling pathways shows hysteresis, namely dynamic and static lags between input and output. Because of hysteresis one can have, at the same time, a sharp binary response and protection against chatter noise.

Algorithmically the task is to find the positive real solutions of a parameterized system of polynomial or rational systems, since the dynamics of the network is given by polynomial systems (arising from mass action kinetics) or rational functions (arising in signaling networks when some intermediates of the reaction mechanisms are reduced). Due to the high computational complexity of this task considerable work has been done to use specific properties of networks and to investigate the potential of multistationarity of a biological network out of the network structure. This only determines whether or not there exist rate constants allowing multiple steady states. These approaches can be traced back to the origins of Feinberg’s Chemical Reaction Network Theory (CRNT) whose main result is that networks of deficiency 0 have a unique positive steady state for all rate constants [9]. However, given a bistable mechanism it is also important to compute the bistability domains in parameter space: the parameter values for which there is more than one stable steady state. The size of bistability domains gives the spread of the hysteresis and quantifies the robustness of the switches.

We have used an 11-dimensional model of a mitogen-activated protein kinases (MAPK) cascade [10] as a case study to investigate properties of the system using algorithmic methods towards the goal of semi-algebraic descriptions of parameter regions for which multiple positive steady states exist. We have shown that the determination of multistationarity of such a network can be achieved by combinations of currently available symbolic methods for
mixed equality/inequality systems if, for all but potentially one parameter, numeric values are known. As there are many very relevant systems having dimensions between 10 and 20 this points at a possible future application field for automated reasoning tools [3, 4].

In subsequent work we considered several MAPK models and introduced a graph theoretical symbolic preprocessing method to reduce the input systems. We furthermore systematically compared the quality and performance of our subsequent symbolic steps (real triangularization, cylindrical algebraic decomposition) with more traditional numerical approaches (homotopy methods). We experimentally found that our reductions offered computation savings in both worlds. In addition, the reduction avoided instability from rounding errors in the numerical approach to one of the systems. An interesting side result is that, at least for the smaller system, the symbolic approach can compete with and even outperform the numerical one, demonstrating how far such methods have progressed in recent years [7, 8].

A journal article currently in press consolidates our work in this specific area [5, 6]. Further work in this qualitative biological network analysis will take place within the French–German interdisciplinary research project SYMBIONT funded by ANR and DFG [1, 2].

References


40.5 Towards Higher-Order Automated Reasoning

In recent years, there has been considerable success in improving the degree of automation of interactive proof assistants by coupling them to automatic theorem provers that can solve subtasks without user interaction.

There is, however, a mismatch of the logics employed by these systems: Interactive proof assistants are based on expressive higher-order logics. By contrast, the two most widely employed kinds of automated provers – superposition provers and SMT (satisfiability modulo theories) solvers – work on some variant of first-order logic.

Experimental evidence shows that most subgoals of the interactive proof assistants that should be passed to the automated prover are “almost first-order” – but not quite. That means that during the translation into the input language of the automated prover, the higher-order constructs need to be encoded using combinators or using an “apply(func, arg)” encoding. This encoding, however, is necessarily incomplete and often prevents the automated prover from finding a proof.

We are therefore investigating “graceful” generalizations of high-performance first-order proof systems to higher-order logic: systems that work in higher-order logic natively without a lossy preprocessing step, but without sacrificing the efficiency of state-of-the-art first-order proof systems on the first-order part of the input.

40.5.1 Extension of the Superposition Calculus with Lambda-Free Higher-Order Terms and (Co)datatypes

Investigators: Jasmin Christian Blanchette, Sophie Tourret, and Uwe Waldmann, in cooperation with Alexander Bentkamp and Petar Vukmirović (VU Amsterdam), Simon Cruanes (Aesthetic Integration), Nicolas Peltier (U. Grenoble Alpes), and Simon Robillard (Chalmers Gothenburg)

Superposition is a highly successful calculus for reasoning about first-order logic with equality. As a stepping stone towards extending the calculus to full higher-order logic, Bentkamp et al. [1] designed a graceful generalization of the calculus to a fragment devoid of λ-abstractions, but with partial application and application of variables, two crucial higher-order features. We implemented the calculi in Simon Cruanes’s Zipperposition prover and evaluated them on TPTP benchmarks. The performance is substantially better than with the traditional, encoding-based approach.

We are currently working on extending the calculus to a logic that includes anonymous functions but excludes first-class Booleans (if desired, Booleans can be encoded using an uninterpreted type and uninterpreted “proxy” symbols corresponding to equality, the connectives, and the quantifiers).
Another extension of superposition, by Blanchette et al. [2], concerns the native support for inductive and coinductive datatypes. The ability to reason about datatypes has many applications in program verification, formalization of the metatheory of programming languages, and even formalization of mathematics.

Both lines of work aim at bridging the gap between automatic and interactive theorem provers, by increasing the expressiveness and efficiency of best-of-breed automatic first-order provers based on the superposition calculus.

References


40.5.2 Extension of Term Orders to Lambda-Free Higher-Order

Investigators: Jasmin Christian Blanchette, Mathias Fleury, Daniel Wand, and Uwe Waldmann, in cooperation with Heiko Becker (MPI-SWS) and Dmitriy Traytel (ETH Zürich)

To go hand in hand with the λ-free superposition calculus described in Section 40.5.1, we generalized the two main orders that are used in superposition-based provers today—the recursive path order (RPO) [3] and the Knuth-Bendix order (KBO) [1]. The new orders gracefully generalize their first-order counterparts and enjoy nearly all properties needed for superpositions. An exception is compatibility with contexts, which is missing for LPO and some KBO variants.

We also formalized the orders’ definitions and key properties in Isabelle/HOL. In the process, we built a formal library of results about finite nested multisets, hereditary multisets, and ordinals below ϵ₀ [2]. We applied the library formalizations of Goodstein’s theorem and the decidability of unary PCF (programming computable functions).

References

40.6 Formalizing Logic

Proof assistants for higher-order logic are increasingly used to verify hardware and software and to formalize mathematics, but they remain laborious to use. First, we try to improve them directly: Our work in this area includes improvements of the prover infrastructure with the development of (co)datatype and recursion (Sect. 40.6.1). Second, we use them in order to develop libraries and methodologies. We formalize logic, as part of a larger effort, the Isabelle Formalization of Logic (IsaFoL). This includes the formalization of an executable SAT solver and the verification of a variant of an optimizing CDCL (Sect. 40.6.2), as well as the verification of an ordered resolution prover (Sect. 40.6.3). Finally, we also work on making the automated provers more expressive, with the extension of Superposition with types and inductions (Sect. 40.6.4). This can ease the use of proof assistants by reducing the amount of work required by the user.

40.6.1 (Co)datatypes, (Co)recursion, and Generalized Nominal-Logic-Style Binders for Isabelle/HOL

Investigators: Jasmin Christian Blanchette and Mathias Fleury, in cooperation with Lorenzo Gheri (Middlesex U.), Andrei Popescu (Middlesex U.), Dmitriy Traytel (ETH Zürich), and further colleagues

The line of work that started in 2011 towards enriching Isabelle/HOL’s language with coinductive datatypes and corecursive functions has reached a natural milestone with the publication of an invited paper at FroCoS 2017 [1].

Since then, we have been working on developing a general framework for specifying and reasoning about syntax with bindings, which has been presented at POPL 2019 [2]. Abstract binder types are modeled using a universe of functors on sets, subject to a number of operations that can be used to construct complex binding patterns and binding-aware datatypes, including non-well-founded and infinitely branching types, in a modular fashion. Despite not committing to any syntactic format, the framework is “concrete” enough to provide definitions of the fundamental operators on terms (free variables, alpha-equivalence, and capture-avoiding substitution) and reasoning and definition principles.

This work is compatible with classical higher-order logic and has been formalized in the proof assistant Isabelle/HOL. It can be seen as a generalization and simplification of Nominal Logic, a popular framework for reasoning about binders following the Barendregt convention.

References


40.6.2 Verifying CDCL

Investigators: Jasmin Christian Blanchette, Mathias Fleury, Christoph Weidenbach, and Dominic Zimmer, in cooperation with Peter Lammich (TUM)

We developed [1, 2] a formal framework for conflict-driven clause learning (CDCL) using the Isabelle/HOL proof assistant. Through a chain of refinements, an abstract CDCL calculus is connected first to a more concrete calculus, then to a SAT solver expressed in a functional programming language, and finally to a SAT solver in an imperative language, with total correctness guarantees. On the most abstract level, we have formalized CDCL, including the use of restarts and forgets. The framework offers a convenient way to prove metatheorems and experiment with variants, including the Davis-Putnam-Logemann-Loveland (DPLL) calculus.

Based on this abstract calculus, we refine the CDCL calculus to add a crucial optimization used in most SAT solvers [3]: two watched literals. We formalize the data structure and the invariants. We formalize it as a refinement of the abstract calculus, such that we can inherit the termination and the correctness properties. Then, we refine the calculus further to obtain an executable SAT solver. Through a chain of refinements carried out using the Isabelle Refinement Framework, we target Imperative HOL and extract imperative Standard ML code. Although our solver, called IsaSAT, is not competitive with the state of the art, it offers acceptable performance for some applications, and heuristics can be added to improve it further. Since the publication, we have optimized it further by adding blocking literals, restarts and forget, and a better memory representation.

Based on our CDCL formalization, we want to formalize a variant of it. On paper, this usually either leads to either copy-paste or to arguments along the line of “similarly to our previous proof”. However, some subtle differences can be missed. In that spirit, we are currently formalizing an optimizing CDCL: Given a cost function on the literals, a minimal-cost model is found. It is easy to miss that only minimum total model can be generated. In our formalization, our main idea is to avoid redoing proofs, neither by copy-paste the proofs nor redoing the complete proofs. We devised and verified an encoding to reduce the search for partial models to total models.

References


40.6.3 Verifying Bachmair and Ganzinger’s Ordered Resolution Prover

Investigators: Jasmin Christian Blanchette and Uwe Waldmann, in cooperation with Anders Schlichtkrull (Technical Univ. of Denmark) and Dmitriy Traytel (ETH Zürich, Switzerland)

Bachmair and Ganzinger’s chapter on resolution theorem proving [1] in the Handbook of Automated Reasoning is a standard introduction to superposition-like calculi. It offers perhaps the most detailed treatment of the lifting of a resolution-style calculus’s static completeness to a saturation prover’s dynamic completeness. It introduces a considerable amount of general infrastructure, including different types of inference systems (sound, reductive, counterexample-reducing, etc.), theorem proving processes, and an abstract notion of redundancy. The resolution calculus, extended with a term order and literal selection, captures most of the insights underlying ordered paramodulation and superposition, but with a simpler notion of model.

We have formalized the first half of this chapter (i.e. the first four sections) in Isabelle/HOL [3, 4]. Our formal development covers the refutational completeness of two resolution calculi for ground (i.e., variable-free) clauses, general infrastructure for theorem proving processes and redundancy, and a completeness proof for a first-order prover expressed as transition rules operating on triples of clause sets. We clarify several of the fine points in the chapter’s text, emphasizing the value of formal proofs in the field of automated reasoning.

Our work in [3, 4] is the starting point of [2], where we use stepwise refinement to obtain a verified deterministic program, written in a subset of Isabelle/HOL, from which we extract purely functional Standard ML code that constitutes a semidecision procedure for first-order logic.

References


40.6.4 Superposition with Types and Datatypes

Investigator: Daniel Wand

Superposition based automated theorem provers for first-order logic are complex tools and often, e.g. in the case of SPASS [1], written in handcrafted C for performance reasons. At their core they work by applying a set of inference rules until either none is applicable anymore (saturation) or a contradiction is found. Superposition is first-order complete, thus it is guaranteed to find a contradiction if there is one. In order to be able to more
easily investigate extensions to the core calculus we created a tool called Pirate, a simple Prototype Implementation to Research Advanced Techniques and Extensions in Scala. It supports many of the standard reduction rules implemented in modern provers. Beyond this solver, Wand’s PhD thesis [3] contains two main contributions.

First we continued to explore the integration of type systems into superposition while retaining first-order completeness. Automated theorem provers usually do not have any type system support. On the other hand, the logic of verification frameworks, such as the Isabelle system [2], support a rich type system. Isabelle’s type system supports types consisting of type constructors (akin to functions for terms but for types) and type variables, so-called polymorphic types. Furthermore types can be grouped into so-called type-classes and type variables be restricted to a type-class. The type system has to be encoded into first-order logic to use automated tools. By implementing it directly into our automated prover, we avoid this significant overhead. We have implemented and formalized a first-order version of Isabelle’s type system to compare the effects to standard encodings and weaker type systems.

Secondly we further extended the type system’s expressiveness by recursive data types. Proving properties of functions over recursive data types is challenging because it almost always requires proof by induction. Therefore, most automated theorem provers do not support them. To fully support recursive data types we extended superposition with induction rules and techniques to heuristically chose where to apply induction. We also developed and implemented new Strengthening methods for inductive proof search and adapted the superposition’s existing reduction machinery to be a reliable way to purge unprovable properties.

References


40.7 Logic for Machine Learning

The comeback of Machine Learning to the foreground of the scientific and technological progress in the past ten years has been followed by an increasing demand for theoretical guarantees. In this context, we provided a formalization of a proof of the expressiveness of Deep Learning in the proof assistant Isabelle/HOL (Sect. 40.7.1), and we studied the impact of the representation of the search space on the theoretical and practical capabilities of Logic-based Machine Learning (Sect. 40.7.2).
40.7.1 Formalization of the Expressiveness of Deep Learning in Isabelle/HOL

Investigators: Jasmin Christian Blanchette, in cooperation with Alexander Bentkamp (VU Amsterdam) and Dietrich Klakow (U. Saarland)

Deep learning has had a profound impact on computer science in recent years, with applications to image recognition, language processing, bioinformatics, and more. Recently, Cohen et al. [2] provided theoretical evidence for the superiority of deep learning over shallow learning.

We formalized their mathematical proof using Isabelle/HOL [1]. The Isabelle development simplifies and generalizes the original proof, while working around the limitations of the HOL type system. To support the formalization, we developed reusable libraries of formalized mathematics, including results about the matrix rank, the Borel measure, and multivariate polynomials as well as a library for tensor analysis.

An extended version of the conference paper has been accepted in a special issue of J. Automated Reasoning.

References


40.7.2 Logic Fragments Reduction for Inductive Logic Programming

Investigators: Sophie Tourret, in cooperation with Andrew Cropper (Univ. of Oxford)

Inductive Logic Programming (ILP) is a form of machine learning which induces hypotheses from examples and background knowledge. Many forms of ILP use second-order Horn clauses as templates, also denoted as metarules, to learn logic programs, and several of them rely on SLD-resolution to produce new candidate solutions [1]. Determining which metarules to use for a given learning task is a major open problem in ILP and most approaches use clauses provided by the designers of the systems without any theoretical justifications.

We formalized the derivation reduction problem for SLD-resolution [3], the undecidable problem of finding a finite subset of a set of clauses from which the whole set can be derived using SLD-resolution. We studied the reducibility of various fragments of second-order Horn logic that are relevant in ILP and extended our results to standard resolution. We also conducted an empirical study [2] of the effects of using reduced sets of such metarules on the overall learning accuracy and time, that shows a substantial improvement over the state of the art, in addition to the theoretical guarantees offered.
40.8 Software and Applications

The logics we are interested in range from decidable fragments of first-order logic or arithmetic theories to undecidable logics such as full first-order logic or its extension with higher-order operators or arithmetic theories. Satisfiability in these logics is at least an NP-hard problem, so in order to verify advances in the theory of automated reasoning experimentation and application through implementation is indispensable. The maturity of our software efforts ranges from prototypical experimentation, Section 40.8.1, over reasonably robust libraries and small tools, Section 40.8.3, to state-of-the-art systems, Sections 40.8.2, 40.8.4, 40.8.5 that can be successfully applied to real world problems or benchmark libraries.

40.8.1 Zipperposition

General: Sophie Tourret, in cooperation with Simon Cruanes (Aesthetic Integration), Alexander Bentkamp and Petar Vukmirović (VU Amsterdam)

Zipperposition [3, 4] is an open-source superposition prover written in OCaml. Originally designed for polymorphic first-order logic (TF1 [2]), its accent is on flexibility, modularity, and simplicity rather than performance, to allow quick experimenting on automated theorem proving. This is the way we use and extend Zipperposition, in particular, for extensions towards higher-order automatic theorem proving. It was first extended with an incomplete higher-order mode based on pattern unification [5]. Bentkamp et al. [1] then added a complete λ-free higher-order mode. As a prototype, we have now implemented a Boolean-free higher-order mode based on an ongoing work that extends the results presented in Section 40.5.1. If the prototypical implementation in Zipperposition shows promising performance, we will integrate our new calculus in a state-of-the-art first-order theorem prover.

References


40 RG1: Automation of Logic


40.8.2 Reduce and Redlog

Investigators: Thomas Sturm, in cooperation with A. Dolzmann (Schloss Dagstuhl Leibniz Center for Informatics, Germany), S. Glondu (Inria Nancy, France), M. Košta (Slovak Academy of Sciences)

Reduce$^2$ is a general-purpose computer algebra system. It can be used interactively but also provides a full programming language. Historically among the oldest systems of its kind [3], Reduce, including the two most used compatible Lisp systems, went open-source in December 2008. During 2018, SourceForge counted 514 SVN commits and 11738 downloads of Reduce distributions. An incomplete list of competitors includes Sage$^3$ and Singular$^4$ on the open-source side and Maple$^5$ and Mathematica$^6$ on the commercial market. Reduce’s code base comprises around 140 KLOC C code and 300 KLOC Lisp code for two independent underlying Lisp systems plus 400 KLOC R-Lisp code for the Reduce core computer algebra system and loadable packages. Thomas Sturm is a principal developer and maintainer of Reduce participating in decisions on the overall design and actively supporting the system. He contributed several scientific packages (CGB, CLP(RL), Guardian, PGauss, Redlog), user frontends, interactive interfaces to external software (including Gurobi, Z3, Qepcad B, Mathematica, Qhull), an assertion system for the R-Lisp language, and various development tools (including a batch system, graphical cross-referencing, profiling).

The Redlog package$^7$ is part of the Reduce system distributions. It extends Reduce’s functionality from symbolic computation to efficient and robust implementations of state-of-the-art algorithms operating on formulas in interpreted first-order logic. Redlog implements procedures on first-order formulas in an abstract way based on formally specified black box algorithms for supporting various algebraic theories, aiming at high cohesion and code reuse. The Redlog system description [2] has received more than 400 citations in the scientific literature so far, mostly for applications of the system in the sciences. Redlog is quite unique with respect to its generality. The commercial computer algebra system Mathematica provides quantifier elimination (only) for the reals. There are various Maple packages related to real quantifier elimination and decision, but a rigid logical framework is not systematically supported. The Logic module of Sage has been concentrating on propositional calculus so far.

$^2$https://sourceforge.net/projects/reduce-algebra/
$^3$http://www.sagemath.org/
$^4$https://www.singular.uni-kl.de/
$^5$https://www.maplesoft.com/
$^6$https://www.wolfram.com/mathematica/
$^7$http://www.redlog.eu/
Qepcad B [1] as the reference software for real quantifier elimination using partial cylindrical decomposition is limited to that particular method. Redlog currently comprises around 65 KLOC R-Lisp code. For benchmarking the Redlog website features the well-known REMIS database with quantifier elimination benchmarks and literature citations [4]. Both Redlog and Reduce are distributed under a very liberal Free-BSD license.

Redlog won SMT-COMP 2017\(^8\) in the category non-linear real arithmetic. This is quite remarkable, as Redlog actually aims at way more general parametric problems. Furthermore a combination of Redlog with the veriT solver, which had been developed in the context of the SMArt project, performed nicely in the category for quantifier-free nonlinear arithmetic.

References


40.8.3 The SPASS Workbench

Investigators: Martin Bromberger, Mathias Fleury, Michaël Mera, Simon Schwarz, Andreas Teucke, Dominik Wagner, and Christoph Weidenbach

We have stopped the development of our first-order theorem prover SPASS. It can still be downloaded ([http://www.spass-prover.org/classic-spass-theorem-prover/](http://www.spass-prover.org/classic-spass-theorem-prover/)) and continues to be used by several groups in the world. Is is also continues to be used by us for generating first-order reasoning prototypes. The most recent prototype is the abstraction refinement extension of SPASS, called SPASS-AR, see Section 40.3.2.

Instead we have build bottom-up a new software workbench of automated reasoning libraries and tools based on our know how out of SPASS and to a small extend also based on the existing SPASS libraries. On the lower level, we have implemented a very efficient CNF transformation, currently restricted to propositional logic, but extended with non-standard logical operators, such as if-then-else. The implementation includes the small clause normal form approach [2, 3, 4] and is therefore the first CNF implementation that both supports non-standard logical operators and advanced renaming techniques. It is contained in our CDCL(LA) solver SPASS-SATT, Section 40.8.5. We have also developed our own SAT solver called SPASS-SAT that is again part of CDCL(LA) solver SPASS-SATT, Section 40.8.5. The SPASS-SAT solver is based on the CDCL paradigm [1], but with a rigorous application of certain redundancy rules based on results from [5]. In summary, it is not competitive on today’s SAT benchmarks, but behaves very well in the CDCL(LA) framework of SPASS-SATT where it is contained, Section 40.8.5.

\(^8\)http://smtcomp.sourceforge.net/2017/
The current two more high level tools in the SPASS Workbench are SPASS-IQ, Section 40.8.4, and SPASS-SATT, Section 40.8.5.

References


40.8.4 SPASS-IQ: An LA Theory Solver for SPASS

Investigators: Martin Bromberger and Christoph Weidenbach

SPASS-IQ is a linear arithmetic solver specifically constructed as a theory solver for SPASS. It is especially fine tuned to efficiently handle subsumption and constraint refutation tests. The underlying algorithm is a combination of the branch-and-bound approach and the version of the dual simplex algorithm proposed by Dutertre and de Moura for SMT solving [5]. Instead of focusing on non-chronological backtracking, as done by most branch-and-bound solvers of the SMT community, SPASS-IQ focuses on a dedicated branch selection strategy. This gives our solver the advantage of a better informed search and—together with a high rate of preprocessing, our unit cube tests [3, 4], and our bounding transformations [1]—SPASS-IQ is competitive with theory solvers found in state-of-the-art SMT solvers.

We recently extended SPASS-IQ with the preprocessing transformations called the Double-Bounded reduction and the Mixed-Echelon-Hermite transformation [1]. If SPASS-IQ encounters a constraint system $Ax \leq b$ that is not explicitly bounded, i.e., where not all variables have an explicit upper and lower bound, then it computes an equality basis for $Ax \leq 0^m$ [2, 4]. This basis is used to determine whether the system is implicitly bounded, absolutely unbounded or partially bounded, as well as which of the inequalities are bounded. Our solver only applies our two transformations if the problem is partially unbounded. The resulting equisatisfiable but bounded problem is then solved via branch-and-bound. The other two cases, absolutely unbounded and implicitly bounded, are solved respectively via the unit cube test [3, 4] and branch-and-bound on the original system. Our solver also converts any mixed solutions from the transformed system into mixed solutions for the original system. Rational conflicts are also converted between the two systems.

References


40.8.5 SPASS-SATT: A Ground Linear Arithmetic Solver

Investigators: Martin Bromberger, Mathias Fleury, Simon Schwarz, Dominik Wagner, and Christoph Weidenbach

SPASS-SATT is an automated reasoner for ground linear arithmetic (with arbitrary boolean combinations) that we have implemented as part of our SPASS Workbench. To be more precise, SPASS-SATT uses at its core a CDCL(LA) implementation that combines our CDCL implementation NSPASS (Section 40.8.3) with our theory solver SPASS-IQ (Section 40.8.4) [1, 6, 7, 10, 11]. On its own, the CDCL(LA) implementation handles only ground linear arithmetic formulas in clause normal form (conjunctions of disjunctions of linear arithmetic atoms). Therefore, SPASS-SATT also includes our small clause normal form transformation (Section 40.8.3) which extends the scope of SPASS-SATT to all ground linear arithmetic formulas.

Although SPASS-SATT is a relatively new addition to the SPASS Workbench, it is already considered a state-of-the-art solver in its areas of expertise. This is only confirmed by SPASS-SATT’s results in the 13th International Satisfiability Modulo Theories Competition (SMT-COMP 2018), where it ranked first in the category QF_LIA (quantifier free linear integer arithmetic) and ranked second in the category QF_LRA (quantifier free linear rational arithmetic).

SPASS-SATT is so efficient because it combines our own specialized decision procedures—the unit cube test (Section 40.4.3) [3, 4], as well as the Double-Bounded reduction and the Mixed-Echelon-Hermite transformation (Section 40.4.4) [2]—with various other decision procedures for linear arithmetic. This is topped off with several preprocessing techniques, e.g., for the efficient handling of if-then-else expressions [5, 9] and pseudo-boolean inequalities [8].

Binaries of the newest version of SPASS-SATT are available on http://www.spass-prover.org/spass-satt.
References


40.9 Academic Activities

40.9.1 Journal Positions

Christoph Weidenbach:
- Journal of Automated Reasoning (Editor)

Jasmin Christian Blanchette:
- Special section of International Journal on Software and Systems Modeling (SoSyM): STAF 2015 (Guest co-editor)

Thomas Sturm:
- Journal of Symbolic Computation (Editor)
- Mathematics in Computer Science (Editor)
40.9.2 Conference and Workshop Positions

Membership in program and organization committees

Christoph Weidenbach:

- The 26th International Conference on Automated Deduction, CADE-26, Gothenburg, Sweden, August 2017 (PC member),
- 9th International Joint Conference on Automated Reasoning, IJCAR 2018, Oxford, UK, July 2018 (PC member),
- 27th International Joint Conference on Artificial Intelligence and 23rd European Conference on Artificial Intelligence, IJCAI-ECAI-18, Stockholm, Sweden, July 2018 (Senior PC member),
- Summer School 2017: Verification Technology, Systems & Applications, Saarbrücken, Germany, July/August 2017 (Organizer),

Jasmin Christian Blanchette:

- Tenth NASA Formal Methods Symposium, NFM 2018, Newport News, VA, USA, April 2018 (PC member),
- 30th International Conference on Computer Aided Verification, CAV 2018, Oxford, UK, July 2018 (PC member),
- 12th International Conference on Tests & Proofs, TAP 2018, Toulouse, France, June 2018 (PC member),
- 9th International Joint Conference on Automated Reasoning, IJCAR 2018, Oxford, UK, July 2018 (PC member),
- 9th International Conference on Interactive Theorem Proving, ITP 2018, Oxford, UK, July 2018 (PC member),
- The 7th ACM SIGPLAN International Conference on Certified Programs and Proofs, CPP 2018, Los Angeles, CA, USA, January 2018 (PC member),
- Computer Aided Verification, 29th International Conference, CAV 2017, Heidelberg, Germany, July 2017 (PC member),
- The 26th International Conference on Automated Deduction, CADE-26, Gothenburg, Sweden, August 2017 (PC member),
- 11th International Conference on Tests & Proofs, TAP 2017, Marburg, Germany, July 2017 (PC member),
- 4th Conference on Artificial Intelligence and Theorem Proving, AITP 2019, Obergurgl, Austria, April 2019 (PC member),
- 13th International Workshop on the Implementation of Logics, IWIL-2018, Awassa, Ethiopia, November 2018 (PC member),
- Deduktionstreffen 2018 Esch, Luxembourg, September 2018 (PC member),
- Verification and Deduction Mentoring Workshop, VDMW 2018, Oxford, UK, July 2018 (PC member),
- 3rd Conference on Artificial Intelligence and Theorem Proving, AITP 2018, Aussois, France, March 2018 (PC member),
- 15th International Workshop on Satisfiability Modulo Theories, SMT 2017, Heidelberg, Germany, July 2017 (PC member),
- Fifth Workshop on Proof eXchange for Theorem Proving, PxTP 2017, Brasilia, Brazil, September 2017 (PC member),

Sophie Tourret:
- 27th International Joint Conference on Artificial Intelligence and 23rd European Conference on Artificial Intelligence, IJCAI-ECAI-18, Stockholm, Sweden, July 2018 (PC member),

Thomas Sturm:
- Satisfiability Checking and Symbolic Computation (SC-Square 2017), Kaiserslautern, Germany, July 2017 (PC member),
- The 19th International Workshop on Computer Algebra in Scientific Computing (CASC 2017), Beijing, China, September 2017 (PC member),
- 12th International Conference on Automated Deduction in Geometry (ADG 2018), Nanning, China, September 2018 (PC member),
- The 20th International Workshop on Computer Algebra in Scientific Computation (CASC 2018), Lille, France, September 2018 (PC member),
- 9th International Joint Conference on Automated Reasoning, IJCAR 2018, Oxford, UK, July 2018 (PC member),
- 16th International Workshop on Satisfiability Modulo Theories (SMT 2018), Oxford, UK, July 2018 (PC member),
- Satisfiability Checking and Symbolic Computation (SC-Square 2018), Oxford, UK, July 2018 (PC member),

Uwe Waldmann:
- Deduktionstreffen 2018, Esch, Luxembourg, September 2018 (PC Co-chair),
- 9th International Joint Conference on Automated Reasoning, IJCAR 2018, Oxford, UK, July 2018 (PC member),
Membership in steering and other committees

Christoph Weidenbach:
- Member of the Steering Committee for the French-German Computer Science Cooperation Agreement between INRIA, CNRS, University of Metz, University of Nancy 1, University of Nancy 2, Institut National Polytechnique de Lorraine at Nancy, University of Saarbrücken, University of Kaiserslautern, Fraunhofer Institute for Experimental Software Engineering (IESE) Kaiserslautern, Max Planck Institute for Informatics, Max Planck Institute for Software Systems, DFKI,
- Member of the Selection Committee of the Saarbrücken Graduate School in Computer Science,
- Member of Steering Committee “Bundeswettbewerb Informatik”,
- Trustee and President of CADE (Conference on Automated Deduction),
- Member of the Steering Committee of IJCAR (International Joint Conference on Automated Reasoning).

Marco Voigt:
- Member of the Admission Committee for Master or Doctoral students at MPI-INF.

Thomas Sturm:
- Member of the Steering Committee of the International Symposium on Symbolic and Algebraic Computation (ISSAC).

Uwe Waldmann:
- Member of the Admission Committee for Master students at MPI-INF.

40.9.3 Invited Talks and Tutorials

Christoph Weidenbach:
- Robust First-Order Reasoning, Mini-Symposium: Integration of Automated Deduction and Interactive Theorem Proving, November 22, 2018, Innsbruck, Austria.

Jasmin Christian Blanchette:
- Foundational (Co)datatypes and (Co)recursion for Higher-Order Logic, TABLEAUX/FroCoS/ITP 2017, September 28, 2017, Brasilia, Brazil.
- Formalizing the metatheory of logical calculi and automatic provers in Isabelle/HOL, CPP 2019, January 15, 2019, Carcais, Portugal.

Mathias Fleury:
- Programming and Reasoning with Infinite Data in Isabelle/HOL, POPL 2018 TutorialFest, January 8, 2018, Los Angeles, USA.
Sophie Tourret:


Thomas Sturm:


– *Applied Effective Quantifier Elimination*, Graduiertenkolleg Experimentelle und konstruktive Algebra, April 24, 2018, Aachen, Germany.

– *Thirty Years of Virtual Substitution*, International Symposium on Symbolic and Algebraic Computation (ISSAC 2018), July 19, 2018, New York, NY, USA.

Uwe Waldmann:


40.9.4 Other Academic Activities

40.10 Teaching Activities

Summer Semester 2017

– Lecture “Automated Reasoning II” (C. Weidenbach).

Winter Semester 2017/2018


Summer Semester 2018


Winter Semester 2018/2019

– Lecture “Automated Reasoning I” (S. Tourret and C. Weidenbach),

Master and Bachelor Theses

- Michaël Noël Divo: Formalization of Types and Programming Languages in Isabelle/HOL, Master’s Thesis, 2017 (Supervisor: C. Weidenbach and J. C. Blanchette),

40.11 Dissertations, Awards

40.11.1 Dissertations


40.11.2 Awards

- Christoph Weidenbach: Thoralf Skolem Award for On Generating Small Clause Normal Forms (with Andreas Nonnengart and Georg Rock),
- Martin Bromberger, Mathias Fleury, Simon Schwarz, and Christoph Weidenbach: First Place, Main Track in the QF_LIA division, 13th International Satisfiability Modulo Theories Competition for SPASS-SATT,
- Martin Bromberger, Mathias Fleury, Simon Schwarz, and Christoph Weidenbach: Best Newcomer Award, 13th International Satisfiability Modulo Theories Competition for SPASS-SATT.
- Haniel Barbosa and Thomas Sturm: First Place, Main Track in the NRA division, 12th International Satisfiability Modulo Theories Competition for Redlog,
- Sophie Tourret: Lise Meitner Award 2018.

40.12 Grants and Cooperations

Transregional Collaborative Research Centre 248 “Foundations of Perspicuous Software Systems”

The Transregional Collaborative Research Centre 248 “Foundations of Perspicuous Software Systems” aims at enabling comprehension in a cyber-physical world with the human in the loop.

- Starting date: January 2019.
- Duration: 4 years.
- Funding: DFG Transregional Collaborative Research Centre.
VeriDis: Modeling and Verification of Distributed Algorithms and Systems

The VeriDis project aims to exploit and further develop the advances and integration of interactive and automated theorem proving applied to the area of concurrent and distributed systems. The goal of the project is to assist algorithm and system designers to carry out formally proved developments, where proofs of relevant properties as well as bugs can be found fully automatically.

- Starting date: September 2011.
- Duration: 8 years.
- Funding: On separate request, MPI-INF and Inria Nancy, additional sponsoring by Inria Nancy.
- Max Planck Institute head: Christoph Weidenbach.
- Partners: Stephan Merz (Inria Nancy, France).

SC-SQUARE: Satisfiability Checking and Symbolic Computation

Symbolic Computation is concerned with the algorithmic determination of exact solutions to mathematical problems. Recent developments in the area of Satisfiability Checking tackle some similar problems, but with different algorithmic and technological approaches. Bridges between the communities in the form of common platforms and roadmaps are necessary to foster an exchange, and to support and to direct their interaction. Our Coordination and Support Activity initiated a wide range of corresponding activities, identify common challenges, offer global events and bilateral visits, and propose standards.

- Starting date: July 1, 2016.
- Duration: 3 years, 2 months.
- Staff at MPI-INF: Thomas Sturm.
- Partners: University of Bath (UK), RWTH Aachen (Germany), Fondazione Bruno Kessler (Italy), Università degli Studi di Genova (Italy), Maplesoft Europe Ltd (Germany), Université de Lorraine (France), Coventry University (UK), University of Oxford (UK), Universität Kassel (Germany), Universität Linz (Austria).

SMART: Satisfiability Modulo Arithmetic Theories

SMART has adapted state-of-the-art arithmetic decision procedures for use within an SMT context in order to support formal system verification. On the one hand, SMT solvers, are successfully used, e.g., for program verification, bounded and parameterized model checking, or test generation. On the other hand, computer algebra systems allow reasoning in fragments of arithmetic, which is a fundamental requirement in verification. SMART has combined both approaches in theory and software.
Matryoshka

The goal of the Matryoshka project is to make interactive verification more cost-effective by increasing the level of automation of proof assistants. We will enrich first-order calculi (superposition and SMT) with higher-order reasoning in a careful manner, to preserve their desirable properties. With higher-order superposition and higher-order SMT in place, we will develop highly automatic provers building on modern superposition provers and SMT solvers, following a novel stratified architecture. These new provers will then be integrated in proof assistants and will be available as backends to more specialized verification tools.

- Starting date: March 1, 2017.
- Duration: 5 years.
- Funding: European Research Council (ERC) Starting Grant 2016, No. 713999.
- Principal investigator: Jasmin Christian Blanchette; Senior collaborators at MPI-INF: Sophie Tourret, Uwe Waldmann.
- Partners: P. Fontaine (Nancy, France), S. Schulz (Stuttgart, Germany).

40.13 Publications

Journal articles and book chapters


**Conference articles**


**PhD theses**


Part IV

Index
Index

3D reconstruction, 261
3D scanning, 420

abductive reasoning, 647
Abujabal, A., 577, 580, 582, 584
Adhikarla, V. K., 444, 457
advanced displays, 441
adversarial
attacks on ML systems, 279
defenses, 279
game theory, 277
model stealing, 279
perturbations, 275, 277
RL agent, 276
Agarwal, P., 563
de Aguiar, E., 254
Akata, Z., 53, 267, 270, 309, 314, 343
Akhoondian Amiri, S., 219, 220
algorithm engineering, 186
algorithms and data structures, 141
all pairs shortest path, 136, 147, 202
AmbiverseNLU, 559
Andrade, R., 566
Andres, B., 55, 254, 259, 301, 305, 306
Andriluka, M., 253, 254, 259
ANewDsc, 193
annotation guidelines, 565
anomaly detection, 522
ANTICIPATE, 366
Antoniadis, A., 122, 160–162, 169, 171, 172
appearance synthesis, 437
application, 514
approximation algorithms, 140, 157–164, 166, 167, 169, 175, 188, 189, 191, 202, 222
approximation refinement calculus, 643, 644, 646
Arabadzhiiyska, E., 441, 445
arboricity, 213
arithmetic circuits, 197
Arnaout, H., 492, 553
Augustin, M., 279
Aydin, T., 454
Azimi, S. M., 348
Babaei, V., 57, 447
Balalau, O., 602, 604
Becker, R., 173, 177, 179
Beigpour, S. K., 441, 457
Bellman-Ford-like algorithms, 202
Bemana, M., 445
Benenson, R., 263, 282, 343, 345
Berberich, K., 59, 538, 543, 547
Bernard, F., 351, 416, 420, 422, 424, 427, 430, 432
Bernays–Schönfinkel fragment, 640, 641, 645, 649
Bhattacharyya, A., 289, 521
Biega, A. J., 595, 597
BigMax, 622
binary search trees, 143
biologically inspired algorithms, 179, 181, 182
Blanchette, J. C., 61, 648, 657–661, 663
blue noise, 438
Boldyrev, N., 566
Boley, M., 505, 514–516, 536
Bonifaci, V., 179
Bringmann, K., 63, 132–139, 142, 144, 147–149, 153, 175, 183, 186, 187, 189, 198

683
BriQ, 555
Bromberger, M., 652, 654, 666–668
Budhathoki, K., 505, 509, 516
Bund, J., 205, 208, 209

causal inference, 509
causation, 509
CDCL, 660
Chandrasekaran, B., 67, 386, 387
Chatterjee, A., 415, 422, 424
Chaudhury, B. R., 117, 130, 142
Chen, R., 69, 402, 404, 405
Cheung, Y. K., 118–121, 152, 168
Chiu, W.-C., 348
Choudhary, A., 190, 191
Choudhury, S., 267
Chu, C. X., 486, 488
CINEX, 569
citation networks, 565
Cityscapes, 264, 303
classical decision problem, 640, 642
commutative rank, 195
computational geometry, 183–189
Congest model, 173, 201, 214, 219
consensus, 215, 217, 218
convex optimization, 353
corpus annotation, 565
correlation, 505
CRC 1223, 621
CRC 248, 674
CredEye, 590
Croitoru, C., 176
data
  binary, 505, 509, 511, 514–516, 522
  continuous, 510–512, 514–516
  discrete, 511, 523
  event sequences, 509, 521
  graph, 524
  mixed-type, 511
  nominal, 505, 507
data contextualization, 566
data exploration, 566
Del Corro, L., 559, 562, 563
Dembelova, T., 562
Demberg, V., 575
dependency, 505
dependency elimination, 177
Dialogue Generation, 575
Didyk, P., 440, 441, 444, 445, 447, 450, 451
digital geometry processing, 402
Dirnberger, M., 180
disjoint paths, 166
distribution effects, 436
document dating, 565
DoubleFusion, 299
dynamic graphs, 438
Elbassuoni, S., 553
electrical flows, 182
Elgharib, M., 427
Elhayek, A., 254
EMOFIEL, 572
energy-efficient algorithms, 122, 169
equilibrium computation, 117–121
Ernst, P., 543, 551, 552
ExFaKT, 499
explanations, 310
grounding, 312, 317
ExQuisiTe, 555
eye tracking, 445
face performance capture, 427
Facebook Reality Labs – MPI Cooperation, 367
fairness, 128, 130
fake news, 592
Fan, Y., 265
fault tolerance, 203–205
Fieraru, M., 254
fine-grained complexity, 132–140, 142, 147, 149, 184, 187
Fiori, A., 645
fire simulations, 437
Fleszar, K., 158, 162, 166
Fleury, M., 648, 658–660, 666, 668
Fourier transform, 155
Frank-Wolfe, 353
Friedrichs, S., 202, 208
Fritz, M., 71, 251, 253, 275, 279, 280, 282, 284, 289, 316, 323, 338, 343, 348
fully homomorphic encryption, 162
Gad-Elrab, M., 497, 499
GAN, 285, 315
Garrido, P., 427
geodesic shortest path, 186
geographic expressions, 564
geometric complexity theory, 198
geometric set cover, 158
German NLP, 565
Ghazimatin, A., 595, 600
Ghosh, S., 492
Golla, B., 402
Golyanik, V., 422
Google Faculty Research Award, 368
grammar compression, 144
graph
  exploration, 524
  visualisation, 524
graph matching, 351
graph minor, 168
graph sparsification, 150, 168
graph theory, 146, 148, 151, 152, 155, 167
group theory, 200
Guimarães, A., 604
Gupta, D., 543, 547, 566
GVVPerfCapEva, 454
Habermann, M., 296, 416, 422
Habibie, I., 296, 424
hand motion capture, 430
Hasler, N., 451
He, Y., 348
HeidelTime, 563
Heydrich, S., 157
Hladky, J., 437
Ho, V. T., 497, 555
Hodges, R., 200
Hoffart, J., 485, 494, 559, 562, 563
Hoppe, S., 334
Horbach, M., 649
Hornakova, A., 301, 305
Hosang, J., 263, 343
Huang, M. X., 319, 321, 327, 329, 331, 334, 341
Hui, K., 538
human computer interaction, 178
human pose
  disentangled image generation, 251
  human action, 253
  human action anticipation, 253
  parts model, 258
  PG2, 251
  PoseTrack, 255
  reconstruction, 258
  tracking, 259
Humble, M., 518
Ibrahim, Y., 555
Ihrke, I., 454
Ikenmeyer, C., 73, 194, 198, 209
image captioning, 313, 316
image quality evaluation, 444
imPACT, 621
induction, 661
inductive logic programming, 663
Insafutdinov, E., 254, 260, 301
interactive rendering, 436
inverse rendering, intrinsic video, augmented reality, virtual reality, 415
Issac, D., 151, 152
JAMI, 306
Jhavar, H., 572
Jiang, C., 405
Jindal, G., 150, 195, 197, 198
Kalofolias, J., 516, 518
Kaltenpoth, D., 512
Karaev, S., 528
Karrenbauer, A., 75, 173, 177–179, 301
Ke, Q., 253
Kellhofer, P., 440, 441
Kesselheim, T., 124, 125
Khoreva, A., 343, 345
Kim, H., 416, 427
Kinali, A., 206, 207, 210

685
Index

Klakow, D., 575
Klauck, M., 325
Knapsack problem, 157
knowledge base, 482, 485
Kobel, A., 193
Kodric, B., 124, 127
Kolev, P., 150, 175, 179
Kolmogorov complexity, 505, 509
Krawczyk, G., 454
Kühnemann, M., 134, 135, 138, 144, 145, 147, 187

Laekhanukit, B., 140, 155, 163, 164
Lahoti, P., 598
Lamotte-Schubert, M., 640
Lange, J., 301, 305, 306
Lange, L., 565
Lapin, M., 273
large-scale scanning, 420
learning
  deep, 663
  few-shot, 267
  multilabel, 274
  weakly-supervised, 273
  zero-shot, 267, 269, 270
Leimkühler, T., 436, 438, 440
Lenzen, C., 77, 173, 181, 201–205, 208–210, 212
Levi, R., 212, 214, 215
Levinkov, E., 254, 301
Li, W., 316
Lichtblau, F., 383
light-field, 441, 444
Light-Field Archive, 457
light-field quality evaluation, 444
lightweight sensors, 420
lineage tracing, 305
linking numerical observations, 566
List, M., 305
Liu, L., 416
Local model, 212
LocVis, 456
Lorenz, T., 270
Lu, X., 584
Maghsoudlou, A., 390
Mandros, P., 505
Mantiuk, R., 454
Maple, 193
marker-less motion capture, monocular motion capture, 424
market model, 117–120, 130
Marx, A., 507
Marx, M., 510, 511
Mastermind, 146
matching problem, 128, 172
matrix approximation, 175
matrix multiplication, 145, 194, 198
Matryoshka, 676
maximum-entropy principle, 519
MDL, 505, 507, 509
mechanism design, 116, 122–124, 126
Medina, M., 204, 208, 209, 214, 215, 222
Mehlhorn, K., 117, 130, 141, 144, 146, 176, 179, 180
Mehta, D., 296, 424, 430, 432
Meka, A., 415
de Melo, G., 486, 592
Mera, M., 666
metastability, 207–209, 211
Metzler, S., 528, 534
Miettinen, P., 79, 518, 526, 528, 534
Milchevski, D., 494, 559
Mirza, P., 492, 494, 497, 548, 569, 572
Mishra, A., 543
MoCap, 294
Mohanty, M., 553
Mono2Stereo, 440
Monte Carlo rendering, 438
motion capture, 419
Mueller, F., 296, 424, 430
Mukherjee, S., 587, 590
Müller, P., 321, 334
mutual information, 505
Myszkowski, K., 81, 438–441, 444, 445, 447, 456, 457

Nag Chowdhury, S., 501
Nalbach, O., 437
named entity disambiguation, 563
named entity recognition, 562
Natale, E., 182, 215, 217, 218
natural language processing, 559, 575
natural language understanding, 559, 575
NED, 563
NER, 562
network design, 160
Neural Body Fitting (NBF), 297
Nguyen, D. B., 490
noise, 207
nonmaximum supression (NMS), 263
NSPASS, 668
Nusser, A., 187, 215
Occlusion-Robust Pose-maps (ORPM), 297
Oh, E., 184–186, 188
Oh, S. J., 275, 279, 280, 282, 316, 343
Omran, M., 263, 296, 301
online algorithms, 116, 125, 127, 169, 171–173
Oosterwijk, T. S., 126
ordered resolution, 642, 644, 661
Orekondy, T., 279, 280, 282
Pal, K., 486
Palmer, M. R., 387
Pandey, A., 195
parsing problem, 139
pattern mining, 521–523
people detection, 263
performance capture, 419
performance capture, 4D reconstruction, 422
person reidentification, 266
PFSTOOLS, 454
Pishchulin, L., 254, 258
planar graphs, 148
point location, 185
Pons-Moll, G., 85, 294, 296
Pouza, M., 559
Popat, K., 590
population protocols, 218
Pramanik, S., 584
price of anarchy, 116, 122–124
prime implicates, 648
privacy
person recognition, 282
social relation recognition, 283
user attribution, 281
visual privacy advisor, 283
visual redactions, 280
projects
ANTICIPATE, 366
BigMax, 622
CRC 248, 674
Matryoshka, 676
RVHu, 367
SC-SQUARE, 675
SMART, 675
VeriDis, 675
prophet inequalities, 125, 126
proximal bundle, 353
Quadratic Assignment Problem (QAP), 351
quantified Boolean formulas, 177
quantifier elimination, 642, 650
question answering, 582, 584
Ramanath, M., 553
random graphs, 153, 215
random walk, 150
Razniewski, S., 87, 486, 488, 492, 497, 501, 569
real-time capture, 419
realistic image synthesis, 436, 438
redescription mining, 518, 519
Redlog, 665
resource augmentation, 159
RGB-D, 420
Richter, M., 451
Ritschel, T., 436–438
Robertini, N., 422, 424
Rohrbach, A., 275, 284, 307, 309, 316, 345
Romero, J., 486
Rosenbaum, W., 205, 213
RVHu, 367
Sagraloff, M., 193
Saha Roy, R., 89, 577, 580, 582, 584, 595, 600
Index

Saidi, J. S., 390
Sakhadeo, A., 486
Sattar, H., 338
Saurabh, N., 155
SC-SQUARE, 675
scheduling, 122, 161, 169, 172
Schewior, K., 162, 169, 171, 172
Schiele, S., 251, 253, 282
Schmidt, A., 148, 167
Schulz, M., 305
Schwarz, S., 666, 668
secretary problem, 125, 127
Seidel, H.-P., 402, 436
self-stabilization, 203
semantic alignment, 566
semantic segmentation, 267, 344, 350
semi-supervised learning, 306
Sgouritsa, A., 123
Shannon Entropy, 505
shape completion, 260, 261
Sharma, G., 258, 273
Sharma, S., 254, 267
Shen, X., 575
Shetty, R., 309, 316, 348
Singh, G., 91, 435, 438, 445
single-source shortest path, 173
SIP, 295
Siren, 519
Siu, A., 551
SMART, 675
SMT, 648, 651, 652, 654
software
  AmbiverseNLU, 559
  CARL, 497
  ChC, 505
  CINEX, 569
coca, 512
  crack, 511
  CredEye, 590
cute, 509
dora, 505
  EMOFIEL, 572
  ExFaKT, 499
  Facets, 524
  fedora, 505
  NSPASS, 668
  Origo, 509
  PFSTOOLS, 454
  realKD, 514–516
  Redlog, 665
  RuLES, 497
  sci, 507
  Siren, 519
  slope, 510
  sloper, 510
  SPASS, 666, 667
  SPASS Workbench, 666
  SPASS-AR, 666
  SPASS-IQ, 667, 668
  SPASS-SATT, 668
  squish, 521
  upc, 522
  YAGO, 485
  Yago, 485
  Zipperposition, 664
software defined networks, 220
Sotnychenko, O., 296
SPASS, 645, 666, 667
SPASS Workbench, 666, 668
SPASS-AR, 666
SPASS-IQ, 667, 668
SPASS-SATT, 668
spatial expressions, 564
SPNet, 269
Sridhar, S., 296
static scene reconstruction, 419
Steil, J., 321, 331, 341
Steinberger, M., 437
Steiner tree, 163, 164
Stepanova, D., 93, 497, 499, 526
Stereo2Multiview, 440
stochastic complexity, 507
Stoll, C., 451
Streibelt, F., 383
Strong Exponential Time Hypothesis, 132–135, 137, 138, 142, 149
strongly polynomial algorithms, 147, 148
Index

Strötgen, J., 95, 547, 563, 565, 566, 580, 590
Sturm, T., 97, 650, 651, 655, 665
Stutz, D., 260, 275
subgroup discovery, 514–516
substitutable choice functions, 176
Sugano, Y., 319, 321, 323, 325, 331
Sunim, D., 444, 447, 457
Sun, Q., 251, 253, 260, 267, 280, 282, 314
superposition, 657, 658, 661
Swoboda, P., 99, 301, 351, 353
Tandon, N., 307, 486
Tang, S., 254, 259, 301, 316
Teflioudi, C., 532
temporal analysis, 565
temporal evolution, 565
temporal expressions, 564
temporal information, 563
temporal tagging, 565
Terolli, E., 552, 604
Teucke, A., 640, 644, 646, 666
Tewari, A., 280, 427
theCaptury, 451
Tigunova, A., 548
time measurement, 206
tone-mapping, 454
Tonsen, M., 331
topological data analysis, 190, 191
Torbati, G. H., 494, 559
Tourret, S., 647, 657, 663, 664
traveling salesperson problem, 162
Tretschk, E., 275, 422
Tursun, O. T., 441, 445
University of Amsterdam – MPI Cooperation, 368
Vaz, D., 128, 164
vector graphics, 437
VeriDis, 675
video editing, deep video portraits, neural rendering, 416

video processing, 414
VIP, 295
Voigt, M., 640, 646, 649
voltage droop, 210
VQA, 317
Vreeken, J., 105, 305, 505, 507, 509–512, 514–516, 521–524
Wagner, D., 666, 668
Waldmann, U., 657, 658, 661
Wand, D., 658, 661
Wang, J., 430
Wang, L., 592, 604
Wang, Y., 584, 592
Waring rank, 195
Weidenbach, C., 640, 644–646, 649, 652, 660, 666–668
weighted correlation clustering, 301
Wellnitz, P., 139, 144
Wiederhake, B., 210
Wolski, K., 444, 456
Xian, Y., 267, 270
Xu, W., 260, 280, 296, 416, 420, 422, 424, 427
YAGO, 485
Yago, 485
Yahya, M., 553, 582
Yates, A., 107, 494, 538, 547, 548, 590
Yu, H., 445
Yu, N., 279
Zayer, R., 109, 402, 407, 408, 410
zero-sum games, 121
Zhang, S., 263
Zhang, X., 319, 321, 323
Zhou, H., 159, 162
Zimmer, D., 660
Zipperposition, 664
Zollhoefer, M., 415, 420, 422, 424, 427, 432

689