Multimodal categorization

Heinrich Bülthoff, Christian Wallraven
Overview

- The question of how the human brain "makes sense" of the sensory input it receives has been at the heart of cognitive and neuroscience research for the last decades.
- One of the most fundamental perceptual processes is **categorization** - the ability to compartmentalize knowledge for efficient retrieval.
- Recent advances in computer graphics and computer vision have made it possible to both produce highly realistic stimulus material for controlled experiments in life-like environments as well as to enable highly detailed analyses of the physical properties of real-world stimuli.
Research Philosophy

- Study perception and action with stimuli as close as possible to the real world, using
  - Computer Graphics to generate natural but well controlled stimuli of objects and scenes
  - MPI Face Database (open access)
    - faces.kyb.tuebingen.mpg.de
    - vdb.kyb.tuebingen.mpg.de
  - Database of High-Dynamic-Range Images (soon to come)
- Virtual Reality to study perception and action in a closed loop
Research Philosophy

- Study perception and action with stimuli as close as possible to the real world, using
  - Computer Graphics to generate natural but well controlled stimuli of objects and scenes

- Virtual Reality
  - www.cyberneum.de
  - motion simulators
  - haptic simulators
  - walking simulators
  - immersive environments
  - panoramic projections
  - EU-projects: JAST, BACS, CyberWalk, Immersense, Wayfinding
Overview

- In this talk, we will review some of the **key challenges** in understanding categorization from a combined cognitive and computational perspective:
  - the need for spatio-temporal representations
  - perception of material properties
  - multi-modal/multi-sensory aspects of object categorization
  - coupling of perception and action
Research Paradigm

MULTISENSORY PERCEPTION

Simulate reality:
Generate complex, physically realistic stimuli, while maintaining precise control over stimulus variables

Rigorous theory:
Apply rigorous computational principles to develop theories of human visual perception

Develop heuristics:
Create perceptually inspired “short cuts” to increase efficiency, or achieve advanced effects

Biological inspiration:
Imitate design principles of biological systems to solve under-constrained vision problems

Analysis for Synthesis:
Application of segmentation, shape-from-shading, machine learning, etc. to rendering and animation

Ground Truth:
Test vision algorithms on computer generated images for which all scene parameters are known precisely

COMPUTER GRAPHICS

COMPUTER VISION
Overview

- The talk will focus on issues that so far have only started to be addressed but that are crucial for a deeper understanding of perceptual processes:
  - the need for spatio-temporal representations
  - perception of material properties
  - multi-modal/multi-sensory aspects of object categorization
  - coupling of perception and action
Representing objects: two models
Representing objects: image-based recognition
Bülthoff and Edelman [PNAS, 1992]

- Recognition of novel objects depends on the viewing conditions (→ image-based recognition)
Recognition of novel and familiar objects depends on the viewing conditions (→ image-based recognition).
The role of motion in recognition

1. Familiar motion facilitates person identification

2. Motion facilitates human target detection

3. Non-rigid motion is encoded as identity cue

Pilz, Vuong, Bülthoff, Thornton [JEP: HPP, subm]
Vuong, Hof, Bülthoff, Thornton [Journal of Vision, 2006]
Chuang, Vuong, Thornton, Bülthoff [Visual Cognition, 2006]
Quick summary (Spatio-temporal representations)

- Objects and faces are represented in an image-based fashion.
- The temporal properties of objects play an important role during learning and recognition.
- Object representations are spatio-temporal.
Overview

- In this talk, we will review some of the **key challenges** in understanding categorization from a combined cognitive and computational perspective:
  - the need for spatio-temporal representations
  - perception of material properties
  - multi-modal/multi-sensory aspects of object categorization
  - coupling of perception and action
Image-based material editing
Kahn, Reinhard, Fleming, Bülthoff [SIGGRAPH, 2006]

- Goals:
  - How do humans perceive materials?
    - Ill-posed problem
  - Can we exploit perceptual tricks to change materials in a photograph (without a 3D-model)?

- Methods:
  - Crude 3D shape reconstruction using bilateral filter (dark means deep - SFS)
    - Exploits generic viewpoint assumption as an image is consistent with many 3D models
  - Simple background-inpainting for transparency
    - Exploits masking
    - Weak model of refraction

- Results:
  - Re-texturing
  - Medium gloss to matte or glossy
  - Opaque to transparent or translucent

re-textured  transparency
Image-based material editing
Kahn, Reinhard, Fleming, Bülthoff [SIGGRAPH, 2006]
Quick summary (Material Perception)

- The brain does **not** use an inverse physics approach to perception

- Rather, the brain uses (complex) heuristics to estimate
  - Material properties
  - Shape

- By exploiting these heuristics one can create simple, but effective work-arounds to control these properties
Overview

- In this talk, we will review some of the **key challenges** in understanding categorization from a combined cognitive and computational perspective:
  - the need for spatio-temporal representations
  - perception of material properties
  - **multi-modal/multi-sensory aspects of object categorization**
  - coupling of perception and action
Sensory integration

- Humans act upon objects in order to interact with the world.
- Two studies addressed the following questions:
  - Are object representations multi-modal?
  - How can we teach artificial agents how to interact with the world?
Sensory integration

- Humans act upon objects in order to interact with the world.
- Two studies addressed the following questions:
  - Are object representations multi-modal?
  - How can we teach artificial agents how to interact with the world?
Multi-modal similarity and categorization of novel, 3D objects
Cooke, Jäkel, Wallraven, Bülthoff [Neuropsychologia, 2007]

- Goal:
  - Develop framework for understanding multi-sensory (visuo-haptic) object perception

- Methods:
  - Controlled space of visuo-haptic stimuli printed in 3D
  - Multi-Dimensional-Scaling for finding perceptual space for haptic, visual and bimodal exploration

Increasing prominence of shape
Increasing prominence of texture
Microgeometry
Macrogeometry

Photographs of printed 3D objects
The tools: Parametrically-defined stimuli & 3D printer
Cooke, Jäkel, Wallraven, Bülthoff [Neuropsychologia, 2007]
The experiment: Multi-sensory similarity
Cooke, Jäkel, Wallraven, Bülthoff [Neuropsychologia, 2007]

- 10 subjects x 3 conditions: Visual (V), Haptic (H), Visuohaptic (VH)
- Task: Similarity ratings
Results: Modality Effects
Cooke, Jäkel, Wallraven, Bülthoff [Neuropsychologia, 2007]

Common representation?
Perceptual Feature Toolbox
Wallraven, Cooke, Kannengießer [http://pft.homeunix.org/, 2007]

- **Goal:**
  - Develop toolbox for perceptual feature validation

- **Methods:**
  - **2D features from computer vision**
    - Pixel values, Edge Images, Gabor filter response, Visual Difference Predictor, Structural Similarity, Shape Context
  - **3D features from computer graphics**
    - Vertex Coordinates, Vertex Count, Perimeter, Mean Local Curvature, Shape Histograms

- **Results:**
  - Applied to Visuo-haptic similarity ratings:
    - Most 2D features model visual similarity judgments well but were not able to model human haptic perception
Sensory integration

- Humans act upon objects in order to interact with the world.
- Two studies addressed the following questions:
  - Are object representations multi-modal?
  - How can we teach artificial agents how to interact with the world?
Learning multi-modal Object Representations
Natale, Rao, Sandini & Wallraven [CogVis Project, 2004],
Wallraven & Bülthoff [Object Recognition, Attention, Action, 2007]

“How can Proprioception, Vision and Active Control make object recognition more robust?”

Self-terminating Learning .Proprioceptive View-Transition Map  Object Recognition
Learning multi-modal Object Representations
Natale, Rao, Sandini & Wallraven [CogVis Project, 2004], Wallraven & Bülthoff [Object Recognition, Attention, Action, 2007]

- A robot with stereo-cameras, an arm equipped with proprioceptive sensors (LiraLab Baby-Bot)
- A simple, view-based visual recognition framework that learns object representations from image sequences (Wallraven & Bülthoff [CVPR, 2001])
- Coupling of proprioceptive information (joint angles) with views for learning and recognition
Learning multi-modal Object Representations

Natale, Rao, Sandini & Wallraven [CogVis Project, 2004],
Wallraven & Bülthoff [Object Recognition, Attention, Action, 2007]

- Robot performs explorative motor-program for any given object to learn the multi-sensory representation

External View

Keyframes

Tracking
Learning multi-modal Object Representations
Natale, Rao, Sandini & Wallraven [CogVis Project, 2004], Wallraven & Bülthoff [Object Recognition, Attention, Action, 2007]

- **Visual matching** is sufficient to predict the best model but is not very discriminatory
- **Multi-modal matching** profile is more “sharply tuned”
- The integration of proprioceptive information adds viewer-centered 3D information
Quick summary (Sensory Integration)

- Object representations can incorporate multi-sensory information

- Common representation for vision and haptics (?)
  - Cross-modal transfer between vision and haptics
    Newell, F., M. O. Ernst, B. S. Tjan and H. H. Bülthoff *Psychological Science* [2001]

- Exploitation of common representation to develop more efficient object learning and recognition algorithms for embodied agents
Some open questions

- **Computer vision**
  - Can we go beyond image fragments ("bags of words")?
  - Do the current approaches scale to 1000s of categories?
  - How do we incorporate other modalities?

- **Computer graphics**
  - What is perceptual realism?
  - How can we make better animations?
  - Can we learn graphics?

- **Perception research**
  - Can we come up with a quantitative model for object recognition?
  - Does optimal integration hold everywhere – where does it break?
  - What is the psychophysics of higher-level cognitive functions?
Challenges

- The "Chair" challenge
The "Art" challenge: build a computer vision system that learns to interpret art images

Such a system would need to deal with abstraction

Images (c) by Robert Pepperell, see Wallraven et al. [APGV, 2007]
Challenges

- The "Pawan Sinha" challenge
  - build a computer vision system that integrates the 20 results every CV researcher should know about face recognition

Eyebrows as important features
Recognition under distortions
Caricature effect for recognition
MULTISENSORY PERCEPTION

Simulate reality:
Generate complex, physically realistic stimuli, while maintaining precise control over stimulus variables

Rigorous theory:
Apply rigorous computational principles to develop theories of human visual perception

Develop heuristics:
Create perceptually inspired “short cuts” to increase efficiency, or achieve advanced effects

Biological inspiration:
Imitate design principles of biological systems to solve under-constrained vision problems

Analysis for Synthesis:
Application of segmentation, shape-from-shading, machine learning, etc. to rendering and animation

Ground Truth:
Test vision algorithms on computer generated images for which all scene parameters are known precisely

COMPUTER GRAPHICS

COMPUTER VISION