Mathematische Bildverarbeitungsgruppe
Fakultät für Mathematik und Informatik
Universität des Saarlandes, 66041 Saarbrücken

Dissertation

zur Erlangung des Grades des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultäten
der Universität des Saarlandes

Correspondence Problems in Computer Vision

Novel Models, Numerics, and Applications

vorgelegt von
Henning Lars Zimmer

Saarbrücken, 2011

Tag des Kolloquiums:	10.02.2012
Dekan:	Prof. Dr. Holger Hermanns

Prüfungsausschuss:

Ausschussvorsitzender:	Prof. Dr. Hans-Peter Seidel
Gutachter:	Prof. Dr. Joachim Weickert (1. Gutachter)
	Prof. Dr. Daniel Cremers (2. Gutachter)
Protokollführer:	Dr. Andrés Bruhn

Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, 66041 Saarbrücken, Germany

Thesis

for obtaining the degree of a doctor of the natural sciences
of the natural-technical faculties of the Saarland University

Correspondence Problems in Computer Vision

Novel Models, Numerics, and Applications

by
Henning Lars Zimmer

Thesis Advisor: Prof. Dr. Joachim Weickert

Referees: Prof. Dr. Joachim Weickert, Prof. Dr. Daniel Cremers

Saarbrücken, Germany, 2011

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Saarbrücken, 10.02.2012

Henning Lars Zimmer

To My Grandfather *Heinrich Zimmer*

Short Abstract

Correspondence problems like optic flow belong to the fundamental problems in computer vision. Here, one aims at finding correspondences between the pixels in two (or more) images. The correspondences are described by a displacement vector field that is often found by minimising an energy (cost) function. In this thesis, we present several contributions to the energy-based solution of correspondence problems: (i) We start by developing a robust data term with a high degree of invariance under illumination changes. Then, we design an anisotropic smoothness term that works complementary to the data term, thereby avoiding undesirable interference. Additionally, we propose a simple method for determining the optimal balance between the two terms. (ii) When discretising image derivatives that occur in our continuous models, we show that adapting one-sided upwind discretisations from the field of hyperbolic differential equations can be beneficial. To ensure a fast solution of the nonlinear system of equations that arises when minimising the energy, we use the recent fast explicit diffusion (FED) solver in an explicit gradient descent scheme. (iii) Finally, we present a novel application of modern optic flow methods where we align exposure series used in high dynamic range (HDR) imaging. Furthermore, we show how the alignment information can be used in a joint super-resolution and HDR method.

Kurzzusammenfassung

Korrespondenzprobleme wie der optische Fluß, gehören zu den fundamentalen Problemen im Bereich des maschinellen Sehens (*Computer Vision*). Hierbei ist das Ziel, Korrespondenzen zwischen den Pixeln in zwei (oder mehreren) Bildern zu finden. Die Korrespondenzen werden durch ein Verschiebungsvektorfeld beschrieben, welches oft durch Minimierung einer Energiefunktion (Kostenfunktion) gefunden wird. In dieser Arbeit stellen wir mehrere Beiträge zur energiebasierten Lösung von Korrespondenzproblemen vor: (i) Wir beginnen mit der Entwicklung eines robusten Datenterms, der ein hohes Maß an Invarianz unter Beleuchtungsänderungen aufweißt. Danach entwickeln wir einen anisotropen Glattheitsterm, der komplementär zu dem Datenterm wirkt und deshalb keine unerwünschten Interferenzen erzeugt. Zusätzlich schlagen wir eine einfache Methode vor, die es erlaubt die optimale Balance zwischen den beiden Termen zu bestimmen. (ii) Im Zuge der Diskretisierung von Bildableitungen, die in unseren kontinuierlichen Modellen auftauchen, zeigen wir dass es hilfreich sein kann, einseitige *upwind* Diskretisierungen aus dem Bereich hyperbolischer Differentialgleichungen zu übernehmen. Um eine schnelle Lösung des nichtlinearen Gleichungssystems, dass bei der Minimierung der Energie auftaucht, zu gewährleisten, nutzen wir den kürzlich vorgestellten *fast explicit diffusion* (FED) Löser im Rahmen eines expliziten Gradientenabstiegsschemas. (iii) Schließlich stellen wir eine neue Anwendung von modernen optischen Flußmethoden vor, bei der Belichtungsreihen für *high dynamic range* (HDR) Bildgebung registriert werden. Außerdem zeigen wir, wie diese Registrierungsinformation in einer kombinierten *super-resolution* und HDR Methode genutzt werden kann.

Abstract

Solving correspondence problems is a fundamental task in computer vision which occurs in various applications such as optic flow estimation, 3D reconstruction or image alignment. For estimating the displacement field that describes the correspondences between the pixels in two images, energy-based approaches that minimise a certain cost function have proven most successful. In this thesis we show that despite three decades of research on correspondence problems, there is still room for significant improvements and novel ideas:

(i) Concerning the design of the energy, we develop a novel, robust data term that offers a high degree of invariance w.r.t. various kinds of illumination changes. We then propose an anisotropic smoothness term that works complementary to the data term and only smooths in directions where the data term gives no information. The proposed smoothness term further combines ideas from image- as well as flow-driven strategies, which allows to obtain sharp flow boundaries without oversegmentation problems. Finally, we present a simple, yet well-performing strategy for determining the optimal balance between the data and the smoothness term.

(ii) For discretising image derivatives that occur in our continuous models, we propose an alternative to the common central finite difference approximations. By adapting ideas from the solution of hyperbolic differential equations, we use correctly oriented, one-sided upwind differences that may lead to improved results.

The minimisation of the energy functionals comes down to solving nonlinear systems of equations. For a fast solution, we adapt the recently proposed fast explicit diffusion (FED) scheme that combines small (stable) and large (unstable) time step sizes in an explicit gradient descent scheme. Due to its simple nature, this scheme allows an efficient implementation on modern graphics hardware (GPUs) which gives highly accurate displacement fields in the fraction of a second and furthermore enables an implementation on a modern smartphone.

(iii) We present a novel application of modern optic flow approaches in the field of computational photography. By a careful modification of the data term, we are able to align exposure series that are used in the context of high dynamic range (HDR) imaging. Due to the estimated dense displacement fields, handling severe camera shake and moving objects becomes possible with our method. By exploiting the subpixel precision of the displacement fields, we can additionally enhance the spatial resolution of the result in an energy-based, joint super-resolution and HDR approach. For the latter, we propose a novel anisotropic smoothness term that gives much more appealing results compared to previous strategies.

Zusammenfassung

Die Lösung von Korrespondenzproblemen ist eine fundamentale Aufgabe im Bereich des maschinellen Sehens (*Computer Vision*), die in vielen Anwendungen wie z.B. dem optischen Fluß, bei 3D Rekonstruktionen und bei der Bildregistrierung auftaucht. Um das Verschiebungsvektorfeld, welches die Korrespondenzen zwischen den Pixeln in zwei Bildern beschreibt, zu schätzen, haben sich energiebasierte Verfahren, die eine bestimmte Kostenfunktion minimieren als am erfolgreichsten herausgestellt. In dieser Arbeit zeigen wir, dass es trotz drei Jahrzehnten Forschung im Bereich von Korrespondenzproblemen noch immer Raum für signifikante Verbesserungen und neue Ideen gibt:

(i) Bei dem Entwurf der Energie entwickeln wir einen neuen, robusten Datenterm, welcher ein hohes Maß and Invarianz unter verschiedenen Arten von Beleuchtungsänderungen aufweißt. Danach schlagen wir einen anisotropen Glattheitsterm vor, der komplementär zu dem Datenterm arbeitet und nur in Richtungen glättet, in denen der Datenterm keine Information liefert. Weiterhin kombiniert der vorgeschlagene Glattheitsterm Ideen von bild- sowie fluß-getriebenen Strategien, was scharfe Flußkanten erlaubt, ohne jedoch unter Übersegmentierung zu leiden. Schließlich stellen wir eine einfache und trotzdem gut funktionierende Strategie zur Bestimmung der optimalen Balance zwischen Daten- und Glattheitsterm vor.

(ii) Zur Diskretisierung von Bildableitungen, die in unseren kontinuierlichen Modellen auftauchen, schlagen wir eine Alternative zu den üblichen, zentralen Finite–Differenzen Approximationen vor. Wir übernehmen Ideen aus dem Bereich der Lösung von hyperbolischen Differentialgleichungen und nutzen korrekt orientierte, einseitige *upwind* Differenzen, was zur Verbesserung der Ergebnisse führen kann.

Die Minimierung der Energiefunktionale reduziert sich auf die Lösung nichtlinearer Gleichungssysteme. Um eine schnelle Lösung zu gewährleisten, nutzen wir das kürzlich vorgestellte *fast explicit diffusion* (FED) Schema, welches kleine (stabile) und große (instabile) Zeitschrittweiten in einem expliziten Gradientenabstiegsschema kombiniert. Aufgrund seiner einfachen Natur erlaubt dieses Schema eine effiziente Implementierung auf moderner Grafikhardware (GPUs), welche hoch genaue Verschiebungsfelder im Bruchteil einer Sekunde errechnet. Außerdem erlaubt dieses Schema eine Implementierung auf einem modernen Smartphone.

(iii) Wir stellen eine neue Anwendung moderner optischer Flußverfahren im Bereich der computergestützten Fotografie (*Computational Photography*) vor. Durch eine sorgsame Modifizierung des Datenterms sind wir in der Lage, Belichtungsreihen zu registrieren, die im Bereich der *high dynamic range* (HDR) Bildgebung verwendet werden. Wegen der geschätzten dichten Verschiebungsfelder kommt unsere Methode auch mit heftigem Verwackeln und sich bewegenden Objekte zurecht. Indem wir die subpixel Genauigkeit der Verschiebungsfelder ausnutzen, sind wir zusätzlich in der Lage, die räumliche Auflösung der Bilder in einem energiebasierten, gemeinsamen *super-resolution* und HDR Ansatz zu erhöhen. Bei letzterem schlagen wir einen neuen, anisotropen Glattheitsterm vor, welcher deutlich ansprechendere Ergebnisse als vorhergehende Strategien liefert.

Acknowledgements

There are many people that I have to thank because without them, this work would have never been possible.

First of all, I would like to thank Prof. Dr. Joachim Weickert for accepting me as a Ph.D. student, supervising my work, and also for sparking my interest in the fascinating field of image processing and computer vision. I am also very grateful to Prof. Dr. Daniel Cremers who agreed to be my second reviewer. I also wish to express my gratitude to all current and former members of our group for providing a nice working atmosphere and for being much more than just colleagues. Especially, I have to thank Andrés Bruhn for the countless hours that he invested in our projects, for his inspiring ideas and his fruitful comments. Similarly, I wish to thank all persons that I was granted to work with on joint projects. Those are Levi Valgaerts, Michael Breuß, Pascal Gwosdek, Oliver Demetz, Sven Grewenig, Andreas Luxenburger, and Sebastian Volz from our group, as well as Hans-Peter Seidel, Bodo Rosenhahn, Agustín Salgado, Carsten Stoll and Christian Theobalt from other groups. Furthermore, I am very grateful to the International Max Planck Research School for Computer Science (IMPRS-CS) for funding me by a three-year scholarship.

I also have to thank my friends and family for their continuous support. Especially my parents and grandparents helped me more than words can express. I am also very grateful to Nadine Moegelin for everything she did for me during our relationship. Last but not least, I want to thank our dog Daisy that eased writing up this thesis by sleeping and snoring next to me.

Contents

Chapter 1

Introduction

Per aspera ad astra.

Latin saying ("to the stars through difficulties")

Classic image processing tasks focus on enhancing or analysing a *single* image, e.g. removing noise or segmenting the image into meaningful regions. Obviously, by considering *several* images of the same scene, one should be able to do more. For example, one can combine images taken with different exposure times to create an image that shows details in very dark as well as very bright regions, which is known as high dynamic range (HDR) imaging. Another example is motion estimation which is naturally only possible if one considers a whole image sequence, i.e. a video stream. Similarly, 3D reconstructions from images also require several images to overcome the depth ambiguity that arises when projecting the 3D world onto 2D images.

If we now want to use such a multi-image methods, the most important step is to determine which pixels correspond to the same scene object, more specifically we need to establish *correspondences* between the pixels in the given images. The latter task is referred to as a *correspondence problem* which is a classic problem in computer vision research. In this thesis, we will present contributions to all aspects of research on correspondence problems: Appropriate models, efficient solvers, numerical realisation and novel applications.

In the remainder of this chapter we first discuss the basics of correspondence problems. We define the task to be solved, show examples where correspondence problems occur and briefly discuss different solution strategies. Moreover, we give an outlook on the contributions that will be presented later on.

1.1 Correspondence Problems

The basic task of correspondence problems is the following: Consider two images that roughly depict the same scene, e.g. two subsequent frames of an image sequence or two

images taken from slightly different viewpoints. For each pixel in the first image, we then aim to find its corresponding pixel in the second image. The obtained correspondences are usually described by a displacement vector field where each vector points from a pixel in the first image to its corresponding location in the second image. Consequently, the displacement vector field is the unknown to be determined when solving a correspondence problem.

Let us now present some popular examples where correspondence problems arise in the fields of image processing and computer vision. Thereby, we define the specific task that has to be solved and also give examples for application areas.

Optic Flow. Given an image sequence, we aim at computing a displacement vector field that describes the apparent motion of brightness patterns between two subsequent frames of the sequence. Usually, the displacement field is referred to as *(optic) flow field* in this context. For an example showing two frames plus the estimated flow field, we refer to Figure 1.1. As the flow vectors are projections of the actual 3D motion vectors in the 3D scene onto the 2D image plane, the flow field encodes information about motion in the scene. Thus, estimating the optic flow builds a basis for numerous applications such as driver assistance systems [KER95] (see Figure 1.2), robot navigation [dSK02], tracking [KDN93], view interpolation [LLB+10] and video compression [MKW97].

Figure 1.1: Optic flow estimation. *From left to right:* **(a)** First image. **(b)** Second image. **(c)** Computed flow field using our approach in [10, 11], visualised using the colour code shown in the inset.

Figure 1.2: Optic flow for driver assistance. *From left to right:* **(a)** First image of a sequence taken out of a car [WPZ+08]. **(b)** Second image. **(c)** Optic flow field computed using our approach in [10, 11]. Note that the pedestrian is easily distinguishable.

Stereo. In the classic binocular case we are given two images of the same scene that were captured from two different viewpoints. The magnitude of the displacements between pixels in these two images is called *disparity* and is directly related to the depth of the corresponding scene points. Consequently, disparity estimation is an important step for 3D reconstructions [FLP01, HZ00]. An example for a basic 3D reconstruction from two images is shown in Figure 1.3.

If one knows the relative pose of the two cameras, one can restrict the search for correspondences to a single line in the second image. This line is usually referred to as the epipolar line [FLP01, HZ00]. In the simplest scenario where the two cameras are orthoparallel to each other (or if the images have been rectified), all epipolar lines are horizontal. Thus, the stereo problem in the orthoparallel case can be interpreted as an optic flow problem where the vertical displacement component is equal to zero, which eases the search for correspondences. However, to obtain favourable 3D reconstructions, a large baseline distance between the two cameras is mandatory. This results in displacements that are much larger than in the optic flow case and renders stereo reconstruction a challenging task.

Figure 1.3: Stereo matching. *From left to right:* **(a)** Left image of the *Portal* scene, available at `http://cmp.felk.cvut.cz/~cechj/GCS/stereo-images/Portal/`. **(b)** Right image. **(c)** Disparity estimated with our approach in [8]. **(d)** Corresponding reconstruction (visualising the estimated disparities as heightfield).

Scene Flow. We argued that the optic flow field only describes a projection of the actual 3D motion vectors onto the 2D image plane. If one aims at recovering the full 3D displacement field, called *scene flow*, one needs to estimate both, the 3D geometry of the scene and the correspondences of 3D points between subsequent frames. Thus, scene flow computation can be seen as solving a stereo problem together with an optic flow problem which in turn requires to consider binocular image sequences, where a stereo pair is available at each time instant [PAT96].

Estimating the scene flow is for sure more complex than optic flow or stereo alone. However, the additional information can be of great help for applications like driver

assistance [WRV+08, WMR+09] as well as motion capture [5],[VBR+05], see Figure 1.4.

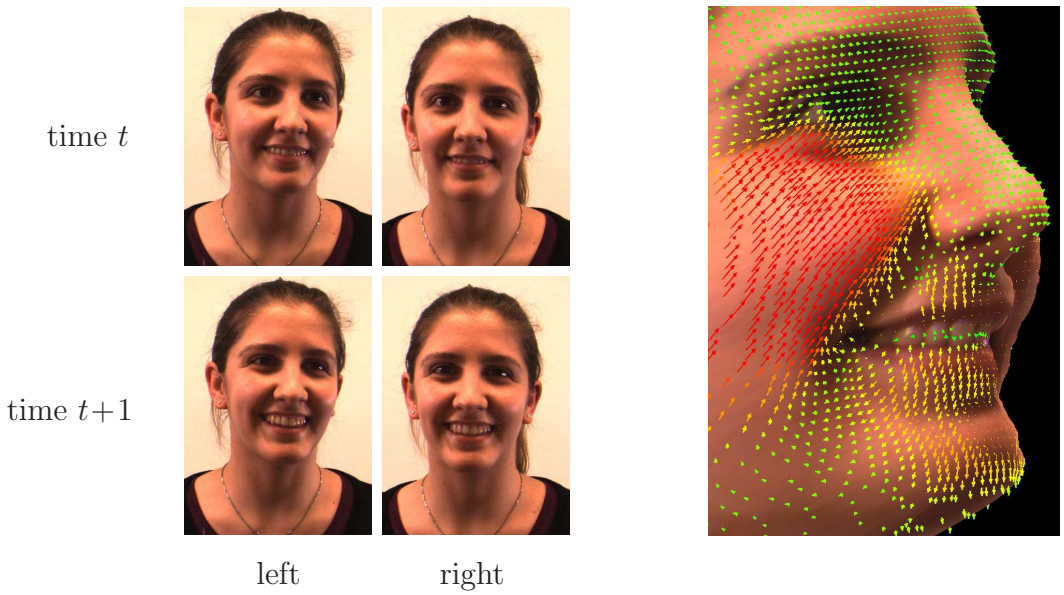


Figure 1.4: Scene flow computation. *Left:* **(a)**–**(d)** Four input images (two stereo pairs at time t and t+1). *Right:* **(e)** Reconstruction and scene flow computed with our method from [5].

Image Alignment. Photographic multi-shot techniques [GG09] fuse several images of the same scene to obtain an image with increased quality. A popular example for such a technique is *high dynamic range (HDR) imaging* [RWPD05] where one tries to tackle the problem that the dynamic range [1] of real world scenes exceeds the dynamic range of image sensors by orders of magnitude. Thus, images obtained by standard cameras often suffer from under- and oversaturated regions in dark and bright regions, respectively. By fusing the information from a series of images taken with varying exposure times, a HDR image can be computed that captures details in both, dark and bright regions. Further multi-shot techniques include super-resolution [PPK03], extended depth of field [ADA+04], flash/no-flash photography [PSA+04] and image stitching [Sze06].

A major problem is that multi-shot techniques often combine the information at the same pixel location in all the input images. Consequently, the fusion algorithms rely on perfectly aligned images without any displacements between them. However, this assumption is hardly met under real world conditions where moving objects and camera shake may cause severe displacements. In such cases, one has to compute the displacements between the input images to align them prior to processing. As the captured scene may be moving, image alignment can be seen as an optic flow problem. However, classical matching assumptions, e.g. assuming that corresponding pixels have similar colour, may not hold due to the different acquisitions of the input images. In HDR imaging,

[1] The dynamic range denotes the ratio between the darkest and the brightest scene point.

for example, all images were taken with a different exposure time. Nevertheless, we could show in [9] that a carefully adapted optic flow method is able to robustly align HDR input, see Figure 1.5 where we compare HDR images computed from a freehand exposure series with and without prior alignment.

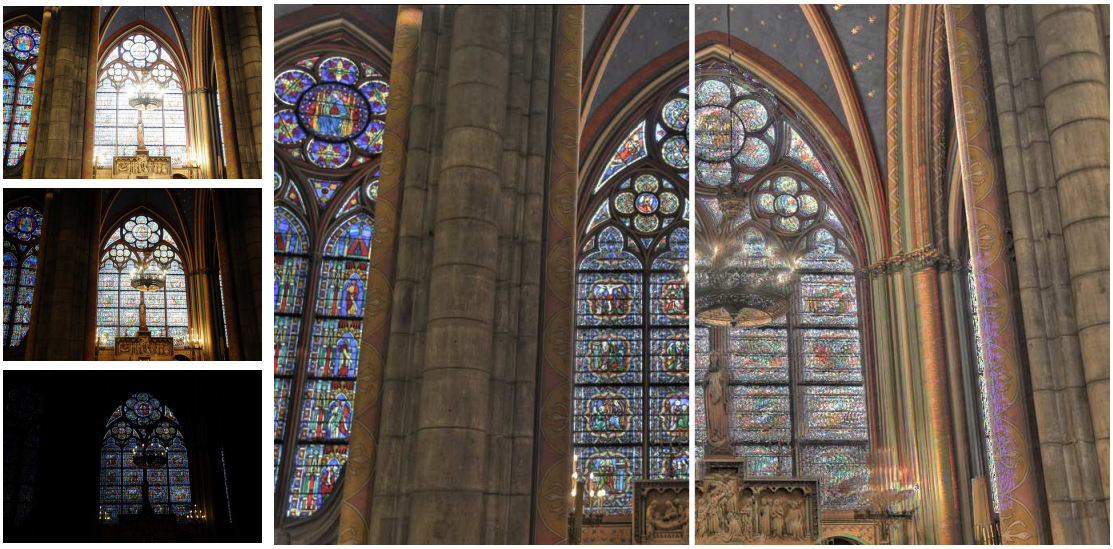


Figure 1.5: Freehand HDR imaging. *Left column:* **(a)**–**(c)** Three images of a freehand exposure series. *Right:* **(d)** Computed HDR reconstruction with and without alignment [9]. The float-valued HDR data is visualised by applying the tone mapping operator from [FLW02].

Medical Image Registration. A common problem in medical imaging is the registration of images that depict the same part of the body, but were captured with different imaging techniques. An example can be the registration of a magnetic resonance imaging (MRI) template to an ultrasound (US) reference image [FM08], see Figure 1.6. Similar to alignment problems discussed above, also here common matching assumption will fail due to the different capturing techniques. Note that we will not present specific contributions to the field of medical image registration in this thesis. However, we decided to keep this example for completeness reasons.

Particle Image Velocimetry (PIV). The basic task of PIV methods is to analyse image sequences that depict the flow of small tracer particles that were placed into a fluid [RWWK07]. The information encoded in the optic flow field of such sequences helps researchers in areas like fluid mechanics, aerodynamics, or meteorology to validate simulations or to do forecasts.

Basically, PIV can be seen as an optic flow problem. However, the estimated flow fields must obey physical laws like incompressibility, which need to be incorporated in the model; see e.g. [RSS07]. Additionally, large displacements of small particles and occurring complex motion patterns like vortices pose further challenges. An example of

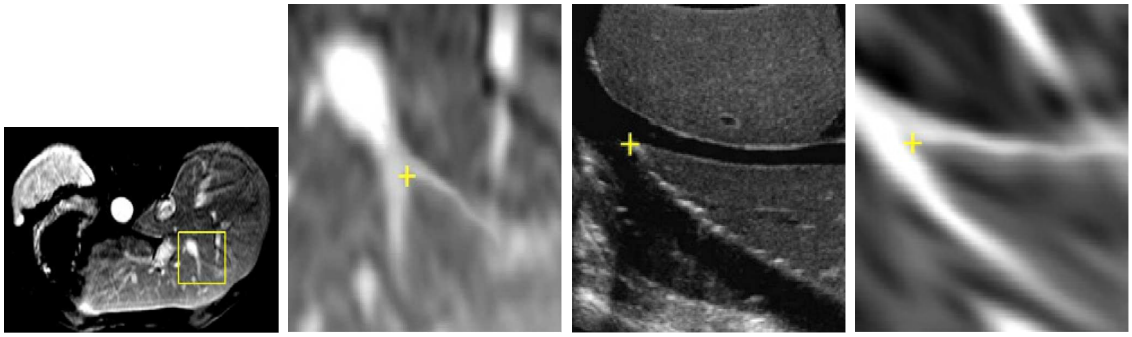


Figure 1.6: Medical image registration of MRI and US data [FM08]. *From left to right:* **(a)** MRI template of a part of a human liver. **(b)** Zoom into marked region in (a). **(c)** US reference image. **(d)** Registration of (b) onto (c).

a flow field with strong vortices is shown in Figure 1.7. As for medical image registration, we discuss PIV methods solely for completeness reasons in this chapter.

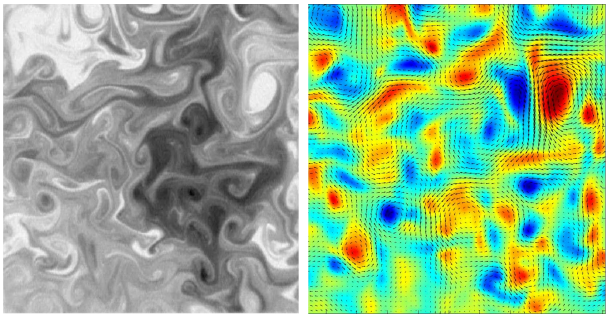


Figure 1.7: PIV example [RSS07]. *From left to right:* **(a)** One of the input images (passive scalar image). **(b)** Estimated flow field with colour-coded vorticity.

1.1.1 Our Focus

The preceding list of examples for correspondence problems shows that these problems occur in a several important and interesting research areas. One further realises that optic flow is a rather general setting for correspondence problems as it does not impose too many constraints on the displacement field. For example, optic flow subsumes stereo where the displacements have to lie on the epipolar lines. Due its generality, we decided to mainly focus on optic flow in this thesis. At some points, however, we will also show how to adapt the presented concepts to stereo in the orthoparallel case.

1.2 Solution of Correspondence Problems

Before presenting an overview of our contributions, let us discuss basic solution strategies for correspondence problems. First of all, we wish to note that solving correspondence

problems is challenging task as they belong to the class of *inverse problems*. In this sense, given an image and a displacement field, one can easily compute the second image by shifting the pixels of the first image in accordance to the displacement. However, solving the inverse problem, i.e. recovering the displacements from two given images, is hard and unfortunately this is exactly the task in correspondence problems.

The fundamental strategy for solving correspondence problems is to impose constancy assumptions on image properties. A first idea can be to assume that the image intensities do not change under their displacement. This is known as the *brightness constancy assumption* [HS81, LK81]. However, for a pixel in the first image, there are probably several pixels in the second image that have the same or a similar intensity, especially in flat image regions. Even worse, the pixel with the most similar intensity may not be necessarily the correct match, as we have to take into account image noise or illumination changes. To overcome this ambiguity, several approaches have been proposed that can be roughly classified into the two classes:

1. *Local approaches* determine the displacement for each pixel individually by finding a pixel in the second image that minimises some matching cost. For a successful resolving of the ambiguities, this cost should be as discriminative as possible. An example can be to match whole regions around pixels.

2. *Global approaches* do not only consider a matching score when assigning a displacement to a pixel, but also take into account the displacements of neighbouring pixels. A common strategy to achieve this goal is to assume that neighbouring pixels have similar displacements, i.e. assuming that the overall displacement field is smooth.

Let us now shortly discuss and compare local and global approaches.

1.2.1 Local Approaches

Block Matching. One of the simplest possibilities to overcome ambiguities is to compare whole patches around pixels instead of just a single pixel. Such techniques are referred to as *block matching* approaches; see [MPG85], for example. These methods perform an exhaustive search (maybe limited by a predefined maximal displacement limit) and match pixels that give the smallest distance between their patches. In this context, several distance metrics have been proposed, like the sum of squared distances (SSD) or the sum of absolute distances (SAD), which is more robust to outliers. Additionally, correlation based metrics such as normalised cross correlation (NCC) are popular as they are invariant under global linear illumination changes.

Regardless of the distance metric used, block matching approaches are known to produce noisy displacement fields that additionally suffer from block-like artefacts. Furthermore, the exhaustive search results in very long runtimes for larger search ranges.

Feature Matching. To improve the quality of block matching methods, one can try to only match image regions that are most discriminative. A typical choice for such *features* (also called *interest points*), are corners which are used within the well-known SIFT (Scale Invariant Feature Transform) method by Lowe [Low99, Low04]. Here, a unit feature vector is computed from the corner neighbourhood. In contrast to image patches used in basic block matching approaches, the SIFT feature vectors are invariant to image scale and rotation, and are robust under affine distortions (noise, change in viewpoint or illumination). Furthermore, the number of interest points is usually much smaller than the number of pixels, making the matching more efficient compared to block matching approaches.

Despite of this desirable properties, feature matching approaches suffer from one fundamental problem: One cannot find an interest point at each pixel and thus the resulting displacement fields are sparse. To obtain dense fields, a postprocessing by interpolation has to be performed, which is again challenging and often does not yield favourable results.

Local Energy-based Approaches. One of the first local optic flow methods was proposed by Lucas and Kanade [LK81, BM04]. The basic idea of their approach is to impose brightness constancy and additionally assuming that the displacements are constant within a small neighbourhood around each pixel. These assumptions can be expressed in a local energy that penalises deviations of the brightness constancy assumption within the considered neighbourhood in a quadratic way, leading to a least squares fit. The minimisation of the local energy then comes down to solving a 2×2 linear system of equations, which can be efficiently solved by standard methods, e.g. by applying Cramer's rule. An affine extension of the Lucas/Kanade method goes back to the work of Shi and Tomasi [ST94]. Here, it is assumed that the flow within a neighbourhood can be described by an affine function, i.e. by 6 parameters. Thus, such methods are also called parametric methods. A spatiotemporal extension of the Lucas/Kanade method with additional estimation of the temporal component of the flow field was later proposed by Bigün *et al.* [BGW91]. This method can be more robust, but requires to solve an eigenvalue problem with a 3×3 matrix.

One major issue of the discussed local energy-based methods is that the assumption of a locally constant displacement is often violated in reality, e.g. at motion discontinuities, leading to block-like artefacts in the displacement fields. Furthermore, depending on the local image structure, the equation system or eigenvalue problem may not have a unique solution. This happens for example in flat image regions and again results in sparse displacement fields.

1.2.2 Global Approaches

Approaches that yield dense displacement fields by construction are energy-based methods that find the displacement field by minimising a suitable *global* energy formulation. The latter usually consists of two terms: a data term and a smoothness term.

The data term models constancy assumptions on image features, it may for example penalise deviations from the brightness constancy assumption. The smoothness term penalises fluctuations in the displacement field, thereby regularising the result which allows to tackle problems with non-unique matches or outliers. The relative weight between the two terms is typically steered by a smoothness parameter. Although the first global energy-based method was already proposed in 1981 in the seminal work of Horn and Schunck [HS81], such methods became increasingly popular in recent years. This is mainly due to their potential for giving highly accurate results, which is witnessed by top ranking results at the Middlebury optic flow benchmark [BSL+10]. In addition to the high accuracy and the dense displacement fields, energy-based approaches are additionally attractive as they are based on transparent mathematical models and allow for an efficient solution. Concerning the latter, sequential multigrid solvers [BWKS06, KKR07], parallel implementations on modern graphics hardware (GPU) [3] and [WPZ+08, ZPB07], or even parallel multigrid schemes [GT08] rendered realtime computations of dense flow fields possible.

Within global energy-based methods, one can distinguish *discrete* methods that minimise a discrete energy function, and continuous *variational* approaches and that minimise a continuous energy functional. We now present the two approaches in more detail.

Discrete Approaches

Starting from a probabilistic model, the displacement estimation is often formulated in terms of optimising a Markov Random Field (MRF), which comes down to minimising a discrete energy function. The minimisation can be realised by graph cuts [BVZ01, Coo08], dynamic programming [LY09] or similar strategies. Examples for discrete approaches can be found in [BA96, Coo08, LY09, MP98, MB87, SRB10, SRLB08, SSB10].

Variational Approaches

Especially in the context of optic flow estimation, variational approaches are very popular; see e.g. [10, 11] and [AELS99, ADK99, BBPW04, Coh93, HS81, NE86, NBK08, Sch94, WCPB09, WPZ+08, WS01a, WPB10, WTP+09, XCJ08, XJM10, ZPB07].

The minimisation of the continuous energy functionals can be performed by different strategies, such as primal-dual approaches, e.g., [WCPB09, WPZ+08, WPB10, WTP+09, XJM10, ZPB07], or by solving the corresponding *Euler-Lagrange equations*. The latter can be considered as the most widely used technique for solving variational optic flow problems and is for example used in [10, 11] and [AELS99, ADK99, BBPW04, Coh93, HS81, NE86, NBK08, Sch94, WS01a]. As we will also follow this strategy in the remainder of this thesis, let us further detail on it.

The Euler-Lagrange Framework. The calculus of variations [Els62] states that a minimiser of the variational energy formulation necessarily has to fulfil the corresponding

Euler-Lagrange equations. Provided that the energy is strictly convex, the solution of the Euler-Lagrange equations then gives the globally optimal flow field. The Euler-Lagrange equations describe a system of coupled partial differential equations of elliptic type. After discretisation, a linear or nonlinear system of equations arises, depending on the design of the energy. In the nonlinear case, time-lagged nonlinearity methods, also known as the Kačanov–Galerkin method [FKN73], can help to reformulate the nonlinear problem as a series of linear ones. To finally solve the linear systems, one exploits the sparsity of the system matrix which allows to use iterative solvers, e.g., the Jacobi, Gauss-Seidel, or SOR method [You03].

An alternative solution strategy is to consider the elliptic problem as the steady state of a *parabolic* problem, i.e. where time tends to infinity. In this context, *explicit schemes* can be used that result in an iterative procedure that updates the solution starting from a given initialisation until a convergence criterion is met. The advantages of explicit schemes is that they are easy to deduce from the energy and very easy to implement. However, to ensure stability very small time step sizes must be chosen s.t. a large number of iterations are needed until convergence, resulting in poor computational performance. This problem can be alleviated by leveraging the simple structure of explicit schemes for parallel implementations on modern graphics hardware, which can lead to a speed up in the order of magnitudes. An alternative approach can be to resort to semi-implicit gradient descent schemes. Here, arbitrary time step sizes can be used, but one needs to solve a system of equations at each time step, as in the elliptic case discussed above.

The Warping Strategy for Handling Large Displacements. Let us give a final, but very important remark on variational approaches. To make the minimisation of variational approaches tractable, one mostly performs a linearisation of the data constraints. This proceeding is, however, only valid under the assumption of either small displacements in the sequence or for very smooth images. However, if the temporal sampling of the image sequence is too coarse, or in the stereo context where a large baseline distance is required to obtain a reasonable depth resolution, large displacements are unavoidable. Additionally, the smoothness of the given images is beyond our control.

A common remedy to this problem is to use a coarse-to-fine multiscale technique, often referred to as warping strategy; see e.g. [AWS00, Ana89, BA96, BBPW04, MP98, WTK87]. A first class of methods [Ana89, BA96, BBPW04, MP98] tackles the problems of large displacements by downsampling the image sequence in a pyramid-like structure. On a small (coarse) levels, the displacements become small enough to be estimated by a linearised approach. These displacements are then used as initialisation for the next finer level which is achieved by compensating the images for the already estimated motion, which is known as *warping*. This procedure is then continued down the pyramid until the finest level is reached. A second class of methods [AWS00, WTK87] proceeds in a similar coarse-to-fine manner, but tackles the problem by smoothing the images via a Gaussian blurring. This results in a scale-space representation (e.g. [Lin94]) instead of the multiscale pyramid used in the first class of methods.

1.3 Our Contributions

Despite the long history of research on correspondence problems that led to thousands of papers, we will show that significant progress is still possible. As stated before, our contributions will mainly focus on optic flow, where we stick to global energy-based (variational) approaches due to their accuracy and the possibility for efficient solutions. In the remainder of this thesis, we will present contributions to the complete pipeline of correspondence problems research: Modelling, solution, numerical realisation as well as applications. In the following we briefly summarise the individual contributions.

1.3.1 Modelling

We present both novel ideas for the data term and for the smoothness term.

Data Term. In [10, 11], we revisited a *normalisation* of the data constraints [LV98, SAH91, SC06] and adapted it to modern data terms. This proceeding helps to prevent an undesirable overweighting of the data term at large image gradients. This is undesirable as large gradients may be caused by unreliable structures like noise or occlusions.

Also in [10, 11], we used a *Hue-Saturation-Value (HSV) colour representation* instead of the standard RGB colour space. The benefits of the HSV representation are a high degree of photometric invariance w.r.t. illumination changes in the scene. However, each channel exhibits a distinct degree of invariance, and a high degree of invariance is attended by a loss of details. To account for this, we proposed a separate robust penalisation of each channel, which for each location downweights the influence of the less reliable channels.

Smoothness Term. Widely-used data terms such as the one resulting from the linearised brightness constancy assumption only constrain the flow in one direction, which we call data constraint direction. In the orthogonal direction, the data term gives no information. Motivated by this basic fact, we developed in [10, 11] a novel anisotropic *complementary smoothness term* that enforces smoothness mainly orthogonal to the data constraint direction. This allows to fill-in missing information orthogonal to the data constraint direction, while avoiding undesirable interference between data and smoothness term. Analysing the smoothing behaviour of our complementary smoothness term, it further turns out that it can be characterised as being *joint image- and flow-driven*: The direction of smoothing is adapted to image structures, whereas the smoothing strength depends on the flow contrast. This gives sharp flow edges and remedies oversegmentation problems of purely image-driven methods. Our smoothness term can be used in the spatial domain [11], as well as in the *spatio-temporal domain* [10], where smoothness of subsequent flow fields is additionally assumed.

Additionally, we found that slight modifications in the definition of our complementary smoothness term allow to represent a large variety of existing smoothness terms.

This motivated us to present a *taxonomy of smoothness terms* [10] in a unified notation that eases their analysis and comparison.

Finally we consider smoothness terms for stereo. For most of the optic flow smoothness term, it is straightforward to adapt them to the stereo case. However, for the anisotropic flow-driven regulariser of Weickert and Schnörr [WS01a] this is not the case: In [8], we showed that adapting the regulariser from [WS01a] to the stereo case results in an isotropic behaviour. To complete the taxonomy of stereo smoothness terms with an *anisotropic disparity-driven smoothness term*, we proposed in [8] to refrain from an energy formulation and to model a smoothness term directly on the level of the underlying Euler-Lagrange equations. This allows to use the structure tensor [FG87] of the disparity field for steering the direction of smoothing, yielding the desired anisotropic behaviour.

Automatic Selection of the Smoothness Weight. A central problem when using global energy-based methods is to find an appropriate balance between the data and the smoothness term, which comes down to find and optimal smoothness weight for the image sequence under consideration. To avoid a tedious manual adjustment of the smoothness weight, we proposed in [10] a very simple but surprisingly well-performing method for estimating the smoothness weight. Our method bases on the assumption that the flow field obtained by an optimal smoothness weight allows for the best possible prediction of the next frames in the image sequence. This novel concept we name *optimal prediction principle* (*OPP*), which allows to select the optimal smoothness weight as the one corresponding to the flow field with the best prediction quality. To judge the latter, we evaluate the data constraints between the first and the third frame of the sequence. Under mild assumptions (constant speed, linear trajectory of objects) this can be realised by simply doubling the flow vectors. Due to its simplicity, our method is easy to implement for all variational optic flow approaches, but nevertheless produces favourable results.

Using the described parameter selection in combination with our complementary smoothness term yields a variational optic flow approach with an optimal complementary behaviour plus an optimal balance between the data and the smoothness term. The resulting method thus harmonises its three components (data term, smoothness term, smoothness weight) and motivated the name *optic flow in harmony* (*OFH*) method [10].

Other Terms. Small objects that undergo large displacements are a major problem for warping strategies. As these small objects vanish at coarse levels, but their motion is too large for an estimation at finer levels, warping methods fail in estimating the displacement of small objects with large displacements. We proposed a simple possibility to overcome this problem in [5]: We first use a block matching or a SIFT matching technique to obtain a sparse displacement field. As block matching or SIFT matching approaches work on the full image resolution, they can capture the displacements of small objects with large displacements. The obtained sparse matches are then used as

an additional prior in a warping-based variational optic flow approach. This finally gives dense displacement fields that can capture large displacements of small objects.

1.3.2 Numerics

Implementing continuous variational models requires to discretise the images, the displacement fields, as well as occurring derivatives of both. In [1, 7], we showed that a sophisticated discretisation of the image derivatives can help to improve the quality of the results. As appropriately oriented one-sided (upwind) discretisations perform well at discontinuities and symmetric (central) discretisations are more appropriate in smooth regions, we came up with an adaptive scheme that blends the two discretisations based on a smoothness measure. This approach is inspired from well-established discretisations of hyperbolic partial differential equations; see [LeV92, LeV02], for example.

We use the briefly mentioned Euler-Lagrange framework (see Section 1.2.2) for minimising the energy functionals that are presented in this thesis. For solving the Euler-Lagrange equations, we use two strategies: *(i)* For an efficient solution on sequential CPU architectures, we use an elliptic strategy which requires to solve a nonlinear system of equations. To speed up the solution of the equation system, we use the nonlinear multigrid scheme proposed by Bruhn *et al.* [BWKS06]. *(ii)* Exploiting the possibility for massively parallel computing on modern graphics hardware (GPUs), we developed in [3] a method that achieves a speedup of more than one order of magnitude compared to multigrid schemes on a CPU. This allows to solve our highly accurate complementary optic flow model [10, 11] in near-realtime on sequences with 640×480 pixels. To achieve this high performance, we had to come up with a solution strategy that can efficiently be parallelised on GPUs. As this is difficult for multigrid schemes, we decided to resort to a parabolic gradient descent strategy which allows us to use the recently proposed fast explicit diffusion (FED) scheme [GWB10]. The latter is an explicit solver with varying time step sizes, where some time steps can significantly exceed the stability limit of classical explicit schemes. If a series of time step sizes is carefully chosen, the whole is still unconditionally stable. Due to the simple structure of the underlying explicit solver, FED can easily and efficiently be implemented in parallel, e.g. by using the NVidia CUDA framework [NVI10]. To obtain high performance despite the large amounts of data involved in the computation, we pay particular attention to an efficient use of on-chip memory to reduce transfers from and to global memory. To further boost the performance, we apply a coarse-to-fine strategy.

This strategy not only allows very fast solutions on GPUs, but also enabled us to implement basic optic flow methods on a modern smartphone [4].

1.3.3 Applications

Our final contributions concern novel applications of optic flow methods in the field of computational photography.

Aligning Exposure Series for HDR Imaging. In Figure 1.5, we have seen that a major obstacle for the practical applicability of HDR imaging techniques is that they assume the input images to be perfectly aligned. In [9] we showed how to adapt a modern energy-based optic flow approach to cope with the brightness changes in the given exposure series. This allows for a robust and accurate estimation of dense displacement fields between the input images and yields an alignment method that outperforms existing strategies in challenging real world scenarios. Additionally, our approach neither requires a preceding camera calibration nor knowledge of the exposure times and can be efficiently implemented on CPU and GPU architectures.

Joint Super-Resolution and HDR Reconstruction. As our HDR alignment approach yields dense displacement fields with subpixel precision, they can be used for increasing the spatial resolution of the result. To this end, one can adapt concepts from super-resolution approaches [EF97, FREM04, MPSC09, TH84] where an image with increased spatial resolution is obtained by combining the information from several images that exhibit some degree of subpixel displacement between each other. This basically allows to fuse different discrete samplings of the same continuous scene. In [9] we proposed the first energy-based *joint super-resolution and high dynamic range (SR-HDR)* approach that uses a robust data term in combination with an anisotropic smoothness term. As we could show, our model gives more appealing results than existing techniques such as [GG06, CPK09].

1.4 Organisation

The rest of this thesis is organised as follows:

- Chapter 2 reviews existing prior work that was influential for our contributions.
- The brief Chapter 3 introduces basic definitions concerning the mathematical modelling of images and flow fields.
- We then start presenting our contributions in Chapter 4 by discussing novel models for the data term (Section 4.2) and the smoothness term (Section 4.3). In the same chapter, we also show how to select the optimal smoothness weight (Section 4.4) and present a way to handle large displacements of small objects (Section 4.5).
- The first part of Chapter 5 (Section 5.1) presents the numerical approximation of image derivatives and flow derivatives that occur in the Euler-Lagrange equations. In Section 5.2 we then focus on a novel adaptive upwind scheme inspired from the numerical solution of hyperbolic partial differential equations.

  The second part (Section 5.3) first presents the warping strategy to handle large displacements and then describes different solution strategies in the elliptic case as well as in the parabolic case. For the latter, we propose an efficient strategy

based on the fast explicit diffusion (FED) solver that is implemented on parallel graphics hardware (GPU) as well as on a modern smartphone.

- In Chapter 6 we first show how to apply modern optic flow methods to the task of aligning exposure series for freehand HDR imaging, see Section 6.2. As the flow fields are of subpixel precision, they can additionally serve as input for a joint super-resolution and HDR (SR-HDR) method, which we describe in Section 6.3.
- We conclude in Chapter 7 by a summary of the presented work and an outlook to possible future research topics.
- In the Appendix A we show proofs of theorems that occur throughout the thesis.

Chapter 2

Related Work

Wer nicht von dreitausend Jahren
Sich weiß Rechenschaft zu geben,
Bleib im Dunkeln unerfahren,
Mag von Tag zu Tage leben.

Johann Wolfgang von Goethe

This chapter gives an overview of methods and approaches that were influential for our contributions. In accordance to the focus of this thesis, we mainly review developments in energy-based (variational) optic flow estimation. At some points, however, we also give remarks on stereo.

We will discuss models for the data term and the smoothness term, present solution strategies and also tackle the numerical approximation of derivatives. Finally, we also discuss applications in HDR imaging.

2.1 Modelling

2.1.1 Data Term

In their seminal work, Horn and Schunck [HS81] proposed a data term that penalises deviations from the brightness constancy assumption in a quadratic way. While this strategy facilitates the minimisation of the energy, it gives too much influence to outliers that may be caused by noise or occlusions. As a remedy, ideas from the field of robust statistics [Hub81] have been adapted which led to robust, subquadratic penaliser functions that reduce the influence of outliers [BA96, BBPW04, MP98].

Apart from noise or occlusions, also illumination changes can complicate a reliable flow estimation. To specifically tackle global additive illumination changes, it was proposed to impose higher order constancy assumptions in the data term [PBB+06]. In this context, especially the constancy of the spatial image gradient has proven useful

[BBPW04, Sch94, TP84]. As higher order constancy assumptions may be combined with the classical brightness constancy assumption, Bruhn and Weickert [BW05] proposed to apply the robust penaliser function separately to each constraint. This gives advantages in cases where a single constancy assumption produces an outlier. Recently, Xu *et al.* [XJM10] went a step further and proposed estimate a binary map that locally selects between imposing either brightness or gradient constancy. An alternative to higher-order constancy assumptions can be to preprocess the images by a structure-texture decomposition [WPZ+08].

Apart from an additive part, realistic scenarios also encompass multiplicative illumination changes [vG04]. If colour image sequences are available, this issue can be tackled by normalising the colour channels [GB97], or by using alternative colour spaces with photometric invariances [GB97, MBW07, vG04]. If one is restricted to greyscale sequences, using log-derivatives [MBW07] is possible.

Apart from the discussed efforts that aimed at enhancing the robustness of the data term, favourable effects have also been reported when normalising the data term [LV98, SC06, SAH91]. This prevents an overweighting of the data term at large image gradient locations which can be problematic as large gradients may be caused by unreliable structures like noise or occlusions.

Remarks on Stereo. Sparked by the variational optic flow approach of Horn and Schunck [HS81], researchers also investigated variational stereo approaches [ADSW00, PTK85, RD96]. It is thus not surprising that also the discussed ideas for robust data terms like subquadratic penaliser functions as well as higher-order constancy assumptions have been incorporated in recent variational stereo approaches [BAS07, SBW05].

2.1.2 Smoothness Term

Again, first ideas go back to Horn and Schunck [HS81] who proposed a smoothness term that penalises the magnitude of the flow gradients in a quadratic fashion. This results in a homogeneous regularisation that does not respect any flow discontinuities. Since different objects may move in different directions or with different velocities, it is, however, desirable to permit discontinuities in the flow field.

Image-driven Smoothness Terms. This goal can be achieved by using image-driven regularisers that take into account image discontinuities. An isotropic model was proposed by Alvarez *et al.* [AELS99] where a scalar-valued weight function reduces the regularisation at image edges. An anisotropic counterpart that also exploits the directional information of image discontinuities goes back to Nagel and Enkelmann [NE86]. Their method regularises the flow field along image edges but not across them. A theoretical analysis as well as small modifications of the Nagel and Enkelmann regulariser were later proposed by Schnörr [Sch93].

The major problem of image-driven strategies is oversegmentation: As not every image edge coincides with a flow edge, image-driven methods are prone to give oversegmentation artefacts in textured image regions.

Flow-driven Smoothness Terms. To avoid oversegmentation problems, flow-driven regularisers have been proposed that respect discontinuities of the evolving flow field and are therefore not misled by image textures. In the isotropic setting this comes down to the use of robust, nonquadratic penalisers which are closely related to line processes [BZ87]. For energy-based optic flow methods, such a strategy was used e.g. by Shulman and Hervé [SH89], and by Schnörr [Sch94]. Later, Weickert and Schnörr [WS01a] presented an anisotropic extension.

However, also flow-driven regularisers do not always give satisfactory results, as they suffer from delocalised and less sharp flow edges compared to their image-driven counterparts.

Joint Image- and Flow-driven Smoothness Terms. Comparing the properties of image- and flow-driven strategies, the idea arises to combine the advantages of both worlds. A first attempt in this direction can be found in the work of Alvarez *et al.* [AELS99] where an isotropic flow-driven regulariser is used in combination with a scalar-valued weight function that decreases the regularisation at image edges. This regulariser still reduces the smoothing at every image edge, and is also prone to oversegmentation problems. A similar strategy (with similar shortcomings) was proposed by Werlberger *et al.* [WTP+09]. They modified the anisotropic image-driven method of Nagel and Enkelmann [NE86] by reducing the amount of smoothing orthogonal to image boundaries by a scalar-valued weight function and additionally apply a nonquadratic penaliser function.

The first successful combination of image- and flow-driven strategies can be found in the discrete method of Sun *et al.* [SRLB08]. There, the authors developed an anisotropic regulariser based on directional flow derivatives that are steered by image structures. This allows to adapt the smoothing direction to the direction of image structures whereas the smoothing strength depends on the flow contrast. We call such a strategy image- and flow-driven regularisation as it combines the benefits of image- and flow-driven methods: sharp flow edges without oversegmentation problems.

Non-Local Smoothness Terms. Recently, non-local smoothing strategies [Yar85] have been introduced to the optic flow community by the works of Sun *et al.* [SRB10] as well as Werlberger *et al.* [WPB10]. In these approaches, it is assumed that the flow vector at a certain pixel is similar to the vectors in a (possibly large) spatial neighbourhood. Adapting ideas proposed by Yoon and Kweon [YK06], the similarity to the neighbours is weighted by a bilateral weight that depends on the spatial as well as on the colour value distance of the pixels. Due to the color value distance, these strategies can be classified as image-driven approaches and are thus also prone to oversegmentation

problems. However, comparing non-local strategies to the previously discussed smoothness terms is somewhat difficult: Whereas non-local methods explicitly model similarity in a certain neighbourhood, the previous smoothness terms operate on flow derivatives that only consider their direct neighbours. Nevertheless, the latter strategies model a *globally* smooth flow field as each pixel communicates with each other pixel through its neighbours.

Spatio-temporal Smoothness Terms. The so far discussed smoothness terms only assume smoothness of the flow field in the spatial domain. As image sequences may consist of more than two frames, yielding more than one flow field, it makes sense to also assume temporal smoothness of the flow fields. In a discrete setting, such spatio-temporal smoothness terms go back to Murray and Buxton [MB87]. For variational approaches, an image-driven spatio-temporal smoothness terms was proposed by Nagel [Nag90] and a flow-driven counterpart was later presented by Weickert and Schnörr [WS01b]. The flow-driven spatio-temporal smoothness term from [WS01b] was later successfully used in the method of Brox *et al.* [BBPW04] as well as in the approach of Bruhn and Weickert [BW05].

Remarks on Stereo. Equivalent regularisation strategies have also been studied for variational stereo. An isotropic image-driven regulariser was used by Kim and Sohn [KS03], whereas an anisotropic version was already earlier proposed by Mansouri *et al.* [MMK98]. Recent variational stereo approaches mostly rely on an isotropic disparity-driven regularisers [BAS07, SBW05] to remedy oversegmentation problems. An anisotropic disparity-driven regulariser has been missing so far.

2.1.3 Automatic Selection of the Smoothness Weight

It is well-known that an appropriate choice of the smoothness weight, that determines the balance between the data and the smoothness term, is mandatory for obtaining favourable results. Nevertheless, there has been remarkably little research on methods that automatically estimate the optimal smoothness weight or other model parameters.

Concerning an optimal selection of the smoothness weight for variational optic flow approaches, Ng and Solo [NS97] proposed an error measure which can be estimated from the image sequence and the flow estimate only. Using this measure, a brute-force search for the smoothness weight that gives the smallest error is performed. Computing the proposed error measure is, however, computationally expensive, especially for robust data terms. The experiments in [NS97] were hence restricted to the basic method of Horn and Schunck [HS81]. In a Bayesian framework, a parameter selection approach that can also handle robust data terms was presented by Krajsek and Mester [KM07]. This method jointly estimates the flow and the model parameters, where the latter encompass the smoothness weight and also the relative weight of different data terms. This method

does not require a brute-force search, but the minimisation of the objective function is nevertheless complicated and only feasible if certain approximations are performed.

2.1.4 Other Terms

In some situations, it can make sense to add further terms to the energy formulation, e.g. if other prior information is available.

One example for such an additional prior can be found in the work of Brox and Malik [BM11] where the authors tackled the problem that variational optic flow methods usually cannot estimate large displacements of small objects. The reason for this is that at a coarse level of the warping pyramid, these small objects vanish, whereas at a fine level their displacement is too large to be estimated. To overcome this limitation, the authors in [BM11] propose to first apply a region-based descriptor matching approach. This gives a set of sparse hypotheses for point correspondences that can, however, capture also large displacements of small objects. Then, a traditional optic flow approach [BBPW04] is guided by these hypotheses to finally obtain a dense flow field that also captures large displacements of small objects. To guide the variational approach, a prior is added that penalises deviations of the sought flow field from the given, sparse hypotheses.

In the context of incorporating feature matching in optic flow approaches, one should also mention the SIFT flow method [LYT+08]. Here, the authors construct a SIFT feature vector [Low99, Low04] for each pixel and impose constancy of the SIFT vectors in the data term of a discrete energy formulation. Although the resulting flow fields suffer from significant artefacts, the invariances of the SIFT vectors allow to match images that only roughly depict the same scene, e.g. two images from a database. This has not been possible with classical optic flow methods.

Finally, we wish to note that there exist also other strategies for handling large displacements in variational optic flow. Steinbrücker *et al.* [SPC09b] use a quadratic relaxation scheme to minimises an energy that does not perform a linearisation in the data term. To make this feasible, the minimisation w.r.t. data and smoothness term is decoupled by introducing an auxiliary variable. For the data term, this allows to perform a brute-force search for the best match on the full image resolution. This renders warping superfluous and enables the approach to estimate arbitrary large displacements at the expense of being computationally very expensive. Additionally, the simple brute-force search allows to use complicated data terms as the minimisation only requires to evaluate the data term for the different solutions. This feature was exploited in [SPC09a], where data terms that match whole patches based on the L^1-norm or normalised cross correlation (NCC) are proposed and compared. The benefits of the patch-based matching seem, however, limited and further increase the computational burden.

2.2 Energy Minimisation

There exist different possibilities for minimising the energy functionals of variational optic flow approaches.

2.2.1 The Euler-Lagrange Framework

The probably most widely used strategy is to solve the corresponding Euler-Lagrange equations, which constitute a system of coupled partial differential equations (PDE) of elliptic type. For solving the latter different techniques can be applied.

Numerics. A first step toward solving the Euler-Lagrange equations is to discretise them, which mainly comes down to sampling the images and the flow fields on some grid. However, we also need to discretise occurring derivatives of the images and the flow field. Choosing an appropriate derivative approximation offers some degree of freedom, but this issue has hardly been studied in the context of correspondence problems. If the discretisation is discussed at all, most approaches [BAS07, BBPW04] use "standard" central finite difference approximations, see [MM94]. Alternatively, some methods like [Coh93] use a finite element method [Joh87]. However, more advanced approximation schemes have been considered for a long time in variational image restoration methods [MO99, ROF92].

After discretising the PDE one ends up with a large linear or nonlinear systems of equations. Exploiting the sparsity of the system matrix, iterative solvers like the Gauss-Seidel method (used for example in [AWS00, HS81, NBK08]) or the more efficient SOR method (used for example in [BBPW04]) can be be applied. A further speed up can be achieved by using highly efficient multigrid schemes [BWKS06, GvV96, Gla84, GT08, KKR07, KR03]. Here, Bruhn *et al.* [BWKS06] came up with a bidirectional full multigrid scheme that is applicable for a variety of different models and achieves realtime performance for image sequences of size 160×120 pixels. To obtain realtime performance also for larger image sizes, El Kalmoun et al. [KKR07] proposed a parallel multigrid method on a CPU cluster. A similar approach goes back to Grossauer and Thoman [GT08] who parallelised a multigrid scheme on a GPU. However, both parallel implementations could only be realised for basic optic flow models so far.

An alternative is to consider the elliptic problem as the steady state of a parabolic PDE where the evolution time tends to infinity. In this context, explicit schemes (used for example in [AELS99]) can be used which are attractive due to their simple implementation that avoids solving large and possible nonlinear systems of equations. However, explicit schemes are only stable if the used time step size is rather small which renders them computationally burdensome. An alternative is offered by semi-implicit schemes that allow arbitrary time step sizes, but again require to solve a system of equations at every time step.

2.2.2 Primal-Dual Approaches

An alternative to the Euler-Lagrange framework that has become increasingly popular in the last years are primal-dual approaches. For optic flow computations, they were first proposed by Zach *et al.* [ZPB07] and were from then on used in a number of methods, see e.g. [WCPB09, WPZ+08, WPB10, WTP+09, XJM10].

The basic idea of primal-dual methods is to introduce an auxiliary variable to decouple the minimisation w.r.t. the data and the smoothness term. For the data term, one ends up with a simple thresholding step. For the smoothness term, a projected gradient descent algorithm similar to the algorithm of Chambolle [Cha04] can be used. As the latter was originally proposed for total variation (TV) regularised denoising problems, most of the first primal-dual approaches [WCPB09, WPZ+08, ZPB07] use the TV norm as subquadratic penaliser in the smoothness term. Applying the same penaliser in the data term (to achieve robustness under outliers) leads to an L^1 penalisation. Such methods are thus also referred to as TV-L^1 methods. As discussed above for the Euler-Lagrange framework, also here one needs to discretise occurring derivatives of the images and the flow field.

As the thresholding as well as the gradient descent step can be efficiently implemented in parallel on a GPU, primal-dual approaches are able to achieve realtime performance even for modern optic flow models and images of size 512×512 pixels [WPZ+08]. Apart from these attractive features, primal-dual approaches suffer from two major problems: *(i)* The number of data terms that allow to derive a corresponding thresholding step is rather limited, e.g. using higher-order constancy assumptions has not been realised so far. *(ii)* Adapting the gradient descent algorithm to the smoothness terms different from a TV penaliser can be difficult, especially for anisotropic regularisers, see [WTP+09].

Remarks on Stereo. As in the optic flow case, variational stereo approaches can be minimised using the Euler-Lagrange framework [BAS07, SBW05], as well as primal-dual strategies [PSG+08].

2.3 Applications

Aligning Exposure Series for HDR Imaging. A simple and fast approach for aligning exposure series is to estimate one global transformation per image pair. In its seminal work, Ward [War03] describes this transformation by a pure translation, whereas later extensions use a translation plus a rotation [Gro06, JLW08]. To cope with the brightness changes due to the varying exposures, the aforementioned approaches consider mean threshold bitmaps (MTB) obtained by a thresholding at the median of all pixel values. Using a pyramid of these images, the global displacement can then be computed by simple shift and difference operations at each pyramid level. Altough global strategies are thus very efficient, they fail in the presence of independently moving objects or for complex camera motions, such as zooming and tiling.

Similar restrictions apply to the method in [TM07] where a homography is computed from SIFT [Low04] feature matches that are invariant under the brightness changes. Such a strategy is implemented in the `align_image_stack` algorithm of the *Hugin* toolkit (`http://hugin.sourceforge.net`). However, homography-based approaches are known to fail for moving objects or camera motions different from a pure rotation.

To describe arbitrary camera motions and to handle moving objects in the scene, dense methods are needed that allow to estimate a different displacement vector for each pixel in the image. This can be achieved by multi-step methods such as the global-local alignment strategies [KUWS03, JO08] that first perform some global alignment and then refine it using the classical local optic flow approach of Lucas and Kanade [LK81]. As optic flow approaches usually assume a similar intensity at corresponding pixels, these approaches first need to transfer the pixel values to the irradiance domain. This, however, requires a preceding calibration step to estimate the camera response function. Another problem is that local optic flow approaches cannot estimate a displacement in flat image regions (aperture problem) and give blocky artefacts as they assume a constant (or parametric) displacement within a local neighbourhood.

A more advanced multi-step method was proposed in [ST04]. This approach first computes sparse correspondences by matching feature points and then computes a dense displacement field using weighted linear regression. This result is further refined by applying a local optic flow method [LK81]. A key aspect in this method is to estimate weights that allow to detect and discard mismatches. To deal with the brightness changes in exposure series, a normalisation is proposed that gives a partial invariance to the exposure changes without using the response function.

There also exist dense methods that do not need to apply several processing steps. Menzel and Guthe [MG07] propose a hierarchical matching of patches that is based on cross-correlation to ensure robustness under the brightness changes. As no smoothness assumption on the displacements is imposed, this method is prone to give noisy displacement fields, leading to artefacts in the alignment. The only approach that imposes an explicit smoothness assumption on the displacements can be found in [KP04]. Here, a stereo method based on zero-mean normalised cross-correlation is used. This, however, is only possible for static scenes without moving objects and additionally requires a pair of cameras with known epipolar geometry.

Joint Super-Resolution and HDR (SR-HDR) Reconstruction. One major challenge for SR-HDR approaches is an accurate displacement estimation with subpixel precision. Thus, some methods rely on special camera hardware to facilitate the displacement estimation: While the methods in [NN05, HTO07] use multisampled images where the pixels on the image sensor are differently exposed, Nakai *et al.* [NYUS08] influence the displacements by a controlled shift of the image sensor.

Evidently, it is more convenient to use standard cameras and to take an exposure series with varying viewpoints. This strategy is applied in [RMVS07] where the images

are aligned using a frequency domain approach that estimates a global translation and rotation, as in [Gro06, JLW08]. The SR-HDR result is then computed by simply interpolating the irradiances of the aligned images. More powerful are approaches that find the SR-HDR image by minimising an energy formulation [GG06, CPK09]. In these works it is also shown that a joint SR-HDR reconstruction is not only more elegant, but also gives better results than a sequential approach. Concerning the required displacement estimation, Choi *et al.* [CPK09] assume the displacements to be given and Gunturk and Gevrekci [GG06] estimate the displacements using a homography-based approach as in [TM07]. The main problem of existing energy-based methods is that they use a prior that enforces the result to be close to a mean image obtained by averaging the irradiances of the aligned input images. Although this prior stabilises the minimisation, it does *not* allow to fill in missing information and to smooth the resulting image. However, our experiments in Section 6.3.4 will show that an appropriate filling in of information and smoothing of the result is required to obtain favourable reconstructions.

Finally, we wish to mention the SR-HDR approach of Schubert *et al.* [SSM09]. Different from our goal (a high quality reconstruction), their focus lies on efficiency. To this end, they fuse a separately captured HDR exposure set (2 images, short and long exposure) and a SR series (several images, constant exposure, varying viewpoints).

Chapter 3

Preliminaries

> As far as the propositions of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.
>
> *Albert Einstein*

Before starting with the main part of this thesis, let us briefly present basic concepts that we will use in the remainder.

3.1 Images

We model images and consequently also image sequences and flow fields as functions. In the optic flow case, we denote the given image sequence by

$$f(\boldsymbol{x}) \,:\, \Omega \times [0,T] \to \mathbb{R} \ , \tag{3.1}$$

where $\boldsymbol{x} := (x, y, t)^\top$. In the latter, $(x, y)^\top \in \Omega$ describes the location within a rectangular *image domain* $\Omega \subset \mathbb{R}^2$ and $t \in [0, T]$ denotes time. The co-domain of f (given by $\mathbb{R}$) denotes the image intensities (greyvalues). Further note that we write vectors like $\boldsymbol{x}$ and also matrices in a bold font.

3.1.1 Colour Images

We will also consider colour images that consist of several (mostly three) colour channels. An example are colour images encoded in the popular RGB colour space. To denote the three channels of RGB image sequences in a convenient manner we use the notation $\boldsymbol{f} = (f^1, f^2, f^3)^\top$.

3.1.2 Presmoothing

We further assume that f has been presmoothed by a frame-wise spatial Gaussian convolution. Let one frame of the given image sequence be denoted as $f_0(x, y)$. Then, the presmoothed version of this frame is obtained as

$$f(x, y) = (K_\sigma * f_0)(x, y) \ , \tag{3.2}$$

where K_σ denotes a *Gaussian* of standard deviation σ:

$$K_\sigma(x, y) := \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\sqrt{x^2 + y^2}}{2\sigma^2}\right) \ , \tag{3.3}$$

and $*$ is the *convolution* operator defined as

$$(K_\sigma * f)(x, y) := \int_\Omega K_\sigma(\hat{x}, \hat{y})\, f(x - \hat{x}, y - \hat{y})\, \mathrm{d}\hat{x}\, \mathrm{d}\hat{y} \ . \tag{3.4}$$

This presmoothing step helps to reduce the influence of noise and additionally makes the image sequence infinitely many times differentiable, i.e. $f \in \mathcal{C}^\infty$. Further note that a spatio-temporal presmoothing of the whole image sequence is also possible by a straightforward extension of the above. We will use such a presmoothing strategy in the context of spatio-temporal optic flow methods in Section 4.3.4.

3.2 Flow Fields

The sought optic flow field that describes the displacement vector field between two frames at time t and $t + 1$ will be denoted by

$$\boldsymbol{u}(\boldsymbol{x}) := (u(\boldsymbol{x}), v(\boldsymbol{x}))^\top : \Omega \times [0, T] \to \mathbb{R}^2 \ . \tag{3.5}$$

To ease notation, we introduce the abbreviation $\boldsymbol{w} := (u, v, 1)^\top$. Here, we (as in the remainder of this thesis) skipped the argument $\boldsymbol{x}$ of the flow field to shorten notation.

Remarks on Stereo We will also consider stereo matching problems in the orthoparallel stereo case. Here, the disparity d is given by the magnitude of the horizontal displacement, i.e. $d = |u|$ and the vertical displacement is equal to zero, i.e. $v = 0$.

3.2.1 Error Measures

For synthetic or lab-made sequences, a ground truth flow field is often available. For the task of measuring the quality of the estimated flow $\boldsymbol{w}$ compared to the ground truth $\boldsymbol{w}_T$, different error measures have been proposed.

A popular error measure is the *average angular error* (*AAE*) [BFB94] defined as

$$\mathrm{AAE}(\boldsymbol{w}, \boldsymbol{w}_T) := \frac{1}{|\Omega|} \int_\Omega \arccos\left(\frac{\boldsymbol{w}^\top \boldsymbol{w}_T}{|\boldsymbol{w}||\boldsymbol{w}_T|}\right) \mathrm{d}x\,\mathrm{d}y \ , \tag{3.6}$$

where $|\Omega|$ denotes the area of the image domain and $|\boldsymbol{w}| := \sqrt{u^2 + v^2 + 1}$ is the Euclidean norm of the flow vector. Note that the AAE considers the average angle between two spatio-*temporal* flow vectors. Thus, if the estimated vector is a multiple of the true vector, the angle is *not* equal to zero, as often assumed incorrectly.

An alternative error measure is given by the *average endpoint error* (*AEE*) [BSL+10] defined as

$$\mathrm{AEE}(\boldsymbol{u}, \boldsymbol{u}_T) := \frac{1}{|\Omega|} \int_\Omega |\boldsymbol{u} - \boldsymbol{u}_T| \, \mathrm{d}x\,\mathrm{d}y \ , \tag{3.7}$$

Remarks on Stereo. In the stereo context, the most popular error measure is the *bad pixel error* (*BPE*) [SS02]. It gives the percentage of points that deviate more than a threshold δ from the ground truth u_T. This can be expressed as

$$\mathrm{BPE}(u, u_T) := \frac{100}{|\Omega|} \int_\Omega \mathcal{G}_\delta(|u(x,y) - u_T(x,y)|) \, \mathrm{d}x\,\mathrm{d}y \ , \tag{3.8}$$

where $\mathcal{G}_\delta(s) = 1$ if $s > \delta$, and 0 else. As proposed in [SS02], we set $\delta = 1$.

3.2.2 Visualisation

For visualising dense flow fields, a *colour code* is a probably better choice compared to arrow plots. In the colour code that we will use throughout this thesis, the hue encodes the direction and the brightness encodes the magnitude of the flow vectors, see Figure 3.1.

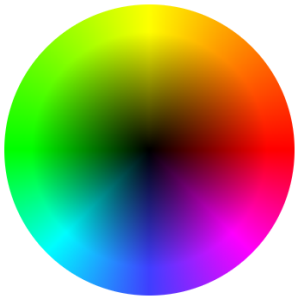


Figure 3.1: Colour code used to visualise flow fields.

3.3 Derivatives

We will often consider derivatives of the functions that describe images and flow fields. The reason for this is that the derivatives give important information about edges and the overall smoothness of the images and flow fields.

As the functions we consider usually depend on more than one variable, we will mostly deal with partial derivatives. The *partial derivative* of a function $f(x, y)$ w.r.t. the variable x will be denoted by one of the following notations

$$\frac{\partial}{\partial x} f \;=:\; \partial_x f \;=:\; f_x \;\; . \tag{3.9}$$

With this, we can define the *spatial gradient* of a function $f(x, y)$ as

$$\boldsymbol{\nabla}_2 f := \begin{pmatrix} f_x \\ f_y \end{pmatrix} \;\; , \tag{3.10}$$

and similarly we define the *spatio-temporal gradient* of $f(x, y, t)$ as

$$\boldsymbol{\nabla}_3 f := \begin{pmatrix} f_x \\ f_y \\ f_t \end{pmatrix} \;\; . \tag{3.11}$$

In the stereo context, we will use the gradient operator $\boldsymbol{\nabla}_{xt}$ defined for a function $f(x, y, t)$ as

$$\boldsymbol{\nabla}_{xt} f := \begin{pmatrix} f_x \\ f_t \end{pmatrix} \;\; . \tag{3.12}$$

Given a normalised vector $\boldsymbol{s} = (s_1, s_2)^\top$ with $|\boldsymbol{s}| = 1$, we define the *directional derivative* of $f(x, y)$ in $\boldsymbol{s}$-direction as

$$f_{\boldsymbol{s}} \;:=\; \boldsymbol{s}^\top \boldsymbol{\nabla}_2 f \;=\; s_1 \, f_x + s_2 \, f_y \;\; . \tag{3.13}$$

We will also use the *divergence* expression which for a vector-valued function $(u(x, y), v(x, y))^\top$ is defined as

$$\operatorname{div} \begin{pmatrix} u \\ v \end{pmatrix} := u_x + v_y \;\; . \tag{3.14}$$

3.4 Discretisation

Most of our models will be formulated in a continuous manner, but digital images are sampled and thus of a discrete nature. Consequently, we will have to discretise the image sequence as well as the flow field at some point in our implementation. We achieve this by sampling the functions on a spatio-temporal discrete pixel grid.

Let us illustrate this proceeding on the example of the image sequence $f(x, y, t)$. We first define spatial *grid sizes* h_1 and h_2 that represent the grid spacing in x- and y-direction, respectively. Furthermore, let h_3 denote the spacing in time direction. With this we obtain the discretisation

$$f(x_i, y_j, t_k) \approx f_{i,j}^k \ , \tag{3.15}$$

where $x_i := (i - \frac{1}{2})\, h_1$, $y_j := (j - \frac{1}{2})\, h_2$ and $t_k := k\, h_3$. In the remainder, we will mostly consider two frames $f_{i,j}^k$ and $f_{i,j}^{k+1}$ of an image sequence, entailing $h_3 = 1$. The resulting number of pixels in x- and y-direction will be denoted by n_x and n_y, respectively.

We will also have to discrete the continuous concept of derivatives. To this end, we will use *finite difference approximations*, see [MM94]. As an example, the x-derivative of the image sequence f can be approximated by the central difference expression

$$\partial_x f(x_i, y_j, t_k) \ =: \ f_x(x_i, y_j, t_k) \ \approx \ \frac{f_{i+1,j}^k - f_{i-1,j}^k}{2h_1} \ . \tag{3.16}$$

A detailed discussion on appropriate finite difference approximations will be given in Chapter 5.

Chapter 4

Modelling

> Essentially, all models are wrong, but some are useful.
>
> *George E. P. Box and Norman R. Draper*

This chapter presents our contributions to the modelling of variational optic flow approaches which have been published in [5, 8, 10, 11]. The main part of the chapter considers models for the data and the smoothness term. Here, we review classical approaches and show that our contributions can be derived as logical extensions and combinations of existing concepts. Additionally, we discuss how to automatically select an optimal smoothness weight, present an additional prior term and also give remarks on stereo in the orthoparallel case.

4.1 Variational Optic Flow

We will rely on variational approaches for computing the optic flow. These methods find the flow field $(u, v)^\top$ as minimiser of a global *energy functional* of the general form

$$E(u, v) = \int_\Omega \big(M(u, v) + \alpha\, V(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v)\big)\, \mathrm{d}x\, \mathrm{d}y \ . \tag{4.1}$$

Here, the term $M(u, v)$ denotes the data term that models how well the flow field explains the given image sequence f. Weighted by $\alpha > 0$, the term $V(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v)$ represents the smoothness term that models the assumption of a smooth flow field by penalising large flow gradients.

Note that the energy (4.1) refers to the spatial case where one aims at computing a single flow field between two frames at time t and $t+1$. The more general spatio-temporal case that uses all frames in the sequence will be discussed in Section 4.3.4.

For minimising the energy (4.1) we use the Euler-Lagrange framework. In accordance to the calculus of variations [Els62], a minimiser (u, v) of the energy (4.1) necessarily

has to fulfil the associated *Euler-Lagrange equations* given by

$$\partial_u M - \alpha \operatorname{div} \begin{pmatrix} \partial_{u_x} V \\ \partial_{u_y} V \end{pmatrix} = 0 \ , \tag{4.2}$$

$$\partial_v M - \alpha \operatorname{div} \begin{pmatrix} \partial_{v_x} V \\ \partial_{v_y} V \end{pmatrix} = 0 \ , \tag{4.3}$$

with homogeneous Neumann boundary conditions. Given that the energy functional is strictly convex, the solution of the Euler-Lagrange equations yields be the global minimum of the energy, which is identical to the "optimal" flow field w.r.t. the imposed model assumptions.

4.1.1 Remarks on Stereo

For stereo, we will consider an orthoparallel camera setup which results in purely horizontal displacements u. To be able to use a similar notation as before, we consider the left image f_ℓ of a stereo pair as a frame of an image sequence at time t, i.e. $f_\ell(x, y) \equiv f(x, y, t)$. Consequently, the right image f_r is then associated with time $t+1$, i.e. $f_r(x, y) \equiv f(x, y, t+1)$. With this, we can define an energy functional for the orthoparallel stereo case as

$$E(u) = \int_\Omega \left(M(u) + \alpha\, V(\boldsymbol{\nabla}_2 u) \right) \mathrm{d}x\, \mathrm{d}y \ , \tag{4.4}$$

where we note that the data and smoothness term term now only depend on the displacement u and its gradient $\boldsymbol{\nabla}_2 u$, respectively. For this energy functional, we only obtain a single Euler-Lagrange equation given by

$$\partial_u M - \alpha \operatorname{div} \begin{pmatrix} \partial_{u_x} V \\ \partial_{u_y} V \end{pmatrix} = 0 \ . \tag{4.5}$$

4.2 Data Term

Following our work in [10, 11], we will in the following systematically derive a sophisticated data term that is highly robust w.r.t. several degradations that often impede a reliable optic flow estimation.

A Basic Data Term. The starting point for our first data term is the classical *brightness constancy assumption* [HS81]. It states that image intensities remain constant under their displacement, which for a greyscale image sequence f can be formalised as

$$f(\boldsymbol{x}+\boldsymbol{w}) = f(\boldsymbol{x}) \ . \tag{4.6}$$

When attempting to minimise an energy functional that incorporates the constraint (4.6) in its data term, one encounters the problem that the unknown flow field $\boldsymbol{w}$ occurs implicitly in the argument of the function f. As a remedy, one can perform a Taylor linearisation: Assuming that the image sequence is smooth and that the displacements are small, a first-order Taylor expansion

$$f(\boldsymbol{x}+\boldsymbol{w}) \;\approx\; f(\boldsymbol{x}) + f_x(\boldsymbol{x})\, u + f_y(\boldsymbol{x})\, v + f_t(\boldsymbol{x}) \;\;, \tag{4.7}$$

gives a sufficiently accurate approximation. Plugging (4.7) into (4.6), rearranging the terms and dropping the argument $(\boldsymbol{x})$ one ends up with the well-known linearised *optic flow constraint* (*OFC*)

$$0 \;=\; f_x\, u + f_y\, v + f_t \;=\; (\boldsymbol{\nabla}_3 f)^\top\, \boldsymbol{w} \;\;. \tag{4.8}$$

Performing a quadratic penalisation of the constraint (4.8), we obtain our first data term

$$M_1(u,v) \;=\; \big((\boldsymbol{\nabla}_3 f)^\top\, \boldsymbol{w}\big)^2 \;=\; \boldsymbol{w}^\top \left(\boldsymbol{\nabla}_3 f (\boldsymbol{\nabla}_3 f)^\top\right) \boldsymbol{w} \;=:\; \boldsymbol{w}^\top \boldsymbol{J}_0\, \boldsymbol{w} \;\;, \tag{4.9}$$

where we define the tensor

$$\boldsymbol{J}_0 := \boldsymbol{\nabla}_3 f\, (\boldsymbol{\nabla}_3 f)^\top \;\;. \tag{4.10}$$

Note that the single equation given by the OFC (4.8) involves two unknowns u and v. It is thus not sufficient to compute a unique solution, which is known as the aperture problem [BPT88]. Nevertheless, assuming that the image gradient does not vanish, i.e. $|\boldsymbol{\nabla}_2 f| \neq 0$, the OFC allows to compute the flow component orthogonal to image edges, the so-called *normal flow*. It is defined as

$$\boldsymbol{w}_n \;:=\; \big(\boldsymbol{u}_n^\top, 1\big)^\top \;:=\; \left(-\frac{f_t}{|\boldsymbol{\nabla}_2 f|}\frac{(\boldsymbol{\nabla}_2 f)^\top}{|\boldsymbol{\nabla}_2 f|},\; 1\right)^\top \;\;. \tag{4.11}$$

Normalisation. Our experiments in Section 4.2.2 will show that a *normalisation* of the data term can be beneficial [LV98, SC06, SAH91]. To understand why this is the

case, let us we rewrite the basic data term M_1 from (4.9) as

$$\begin{aligned} M_1(u,v) &= \left((\boldsymbol{\nabla}_3 f)^\top \boldsymbol{w}\right)^2 \\ &= \left((\boldsymbol{\nabla}_2 f)^\top \boldsymbol{u} + f_t\right)^2 \\ &= \left[|\boldsymbol{\nabla}_2 f| \left(\frac{(\boldsymbol{\nabla}_2 f)^\top \boldsymbol{u}}{|\boldsymbol{\nabla}_2 f|} + \frac{f_t}{|\boldsymbol{\nabla}_2 f|}\right)\right]^2 \\ &= |\boldsymbol{\nabla}_2 f|^2 \left[\frac{(\boldsymbol{\nabla}_2 f)^\top}{|\boldsymbol{\nabla}_2 f|} \left(\boldsymbol{u} + \frac{f_t\, \boldsymbol{\nabla}_2 f}{|\boldsymbol{\nabla}_2 f|^2}\right)\right]^2 \\ &\overset{(4.11)}{=} |\boldsymbol{\nabla}_2 f|^2 \left(\frac{(\boldsymbol{\nabla}_2 f)^\top}{|\boldsymbol{\nabla}_2 f|} (\boldsymbol{u} - \boldsymbol{u}_n)\right)^2 \\ &=: |\boldsymbol{\nabla}_2 f|^2\, d^2 \ . \end{aligned} \tag{4.12}$$

The introduced term d constitutes a projection of the difference between the estimated flow $\boldsymbol{u}$ and the normal flow $\boldsymbol{u}_n$ in the direction of the image gradient $\boldsymbol{\nabla}_2 f$. In a geometric interpretation, the term d thus describes the distance from $\boldsymbol{u}$ to the line l in the uv-space that is given by

$$v = -\frac{f_x}{f_y}\, u - \frac{f_t}{f_y} \ , \tag{4.13}$$

which is a simple rewriting of the OFC (4.8). This shows that all flow vectors that lie on the line l will perfectly fulfil the OFC. Note that the normal flow $\boldsymbol{u}_n$ is actually the shortest vector that lies on this line. A sketch of this geometric interpretation of the data constraints is shown in Figure 4.1.

A logical consequence from our geometric interpretation is that one should penalise the distance d in a data term $M_2(u,v) = d^2$. The data term M_1, however, weighs this distance by the squared spatial image gradient, as $M_1(u,v) = |\boldsymbol{\nabla}_2 f|^2\, d^2$, see (4.12). This results in a stronger enforcement of the data constraint at high gradient locations. Such an overweighting may be inappropriate as large gradients can be caused by unreliable structures, such as noise or occlusions.

As a remedy, we can normalise the data term M_1 by multiplying it with a factor [LV98, SAH91]

$$\theta_0 := \frac{1}{|\boldsymbol{\nabla}_2 f|^2 + \zeta^2} \ , \tag{4.14}$$

where the regularisation parameter $\zeta > 0$ avoids division by zero. Additionally, it reduces the influence of small gradients which are significantly smaller than ζ^2, e.g. noise in flat regions. Nevertheless, the normalisation is not influenced for large gradients. Thus, it may pay off to choose a larger value of ζ in the presence of noise. A normalised

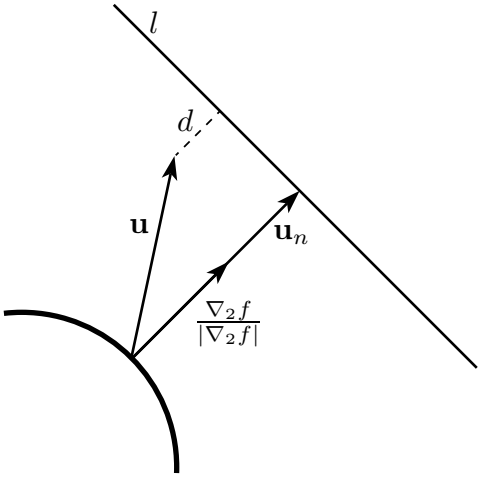


Figure 4.1: Geometric interpretation of the rewritten data term. (4.12)

version of M_1 can then be written as

$$M_2(u, v) = \boldsymbol{w}^\top \bar{\boldsymbol{J}}_0 \, \boldsymbol{w} \ , \tag{4.15}$$

using the normalised tensor

$$\bar{\boldsymbol{J}}_0 \ := \ \theta_0 \, \boldsymbol{J}_0 \ = \ \theta_0 \left(\boldsymbol{\nabla}_3 f \, (\boldsymbol{\nabla}_3 f)^\top \right) \ . \tag{4.16}$$

Gradient Constancy Assumption. The brightness constancy assumption is violated as soon as the illumination conditions in the scene change. A simple (and thus maybe not too realistic) assumption can be that the scene undergoes a global additive illumination change. To render the data term robust under such kinds of illumination changes, the *gradient constancy assumption* has been proposed [BBPW04, Sch94, TP84]. It states that image *gradients* remain constant under their displacement, i.e.

$$\boldsymbol{\nabla}_2 f(\boldsymbol{x}+\boldsymbol{w}) = \boldsymbol{\nabla}_2 f(\boldsymbol{x}) \ . \tag{4.17}$$

Performing a similar Taylor linearisation as before gives the linearised gradient constancy assumption

$$(\boldsymbol{\nabla}_3 f_x)^\top \, \boldsymbol{w} = 0 \ , \quad \text{and} \quad (\boldsymbol{\nabla}_3 f_y)^\top \, \boldsymbol{w} = 0 \ . \tag{4.18}$$

Combining both the brightness and the gradient constancy assumption in a quadratic way yields the data term

$$M_3(u, v) = \boldsymbol{w}^\top \boldsymbol{J} \, \boldsymbol{w} \ , \tag{4.19}$$

where the tensor $\boldsymbol{J}$ can be written in the *motion tensor notation* [BWKS06] that allows to combine the two constancy assumptions in a joint tensor

$$\begin{aligned} \boldsymbol{J} \ &:= \ \boldsymbol{J}_0 + \gamma \, \boldsymbol{J}_{xy} \\ &:= \ \boldsymbol{J}_0 + \gamma \, (\boldsymbol{J}_x + \boldsymbol{J}_y) \\ &:= \ \boldsymbol{\nabla}_3 f \, (\boldsymbol{\nabla}_3 f)^\top + \gamma \left(\boldsymbol{\nabla}_3 f_x \, (\boldsymbol{\nabla}_3 f_x)^\top + \boldsymbol{\nabla}_3 f_y \, (\boldsymbol{\nabla}_3 f_y)^\top \right) \ , \end{aligned} \tag{4.20}$$

where the weighting parameter $\gamma > 0$ steers the influence of the gradient constancy assumption.

To normalise M_3, we replace the motion tensor $\boldsymbol{J}$ by its normalised counterpart

$$\begin{aligned}
\bar{\boldsymbol{J}} &:= \bar{\boldsymbol{J}}_0 + \gamma\, \bar{\boldsymbol{J}}_{xy} \\
&:= \bar{\boldsymbol{J}}_0 + \gamma \left(\bar{\boldsymbol{J}}_x + \bar{\boldsymbol{J}}_y\right) \\
&:= \theta_0\, \boldsymbol{J}_0 + \gamma \left(\theta_x\, \boldsymbol{J}_x + \theta_y\, \boldsymbol{J}_y\right) \\
&:= \theta_0 \big(\boldsymbol{\nabla}_3 f\, (\boldsymbol{\nabla}_3 f)^\top\big) + \gamma \left(\theta_x \big(\boldsymbol{\nabla}_3 f_x\, (\boldsymbol{\nabla}_3 f_x)^\top\big) + \theta_y \big(\boldsymbol{\nabla}_3 f_y\, (\boldsymbol{\nabla}_3 f_y)^\top\big)\right) \quad , \qquad (4.21)
\end{aligned}$$

where the two additional normalisation factors are defined as

$$\theta_x := \frac{1}{|\boldsymbol{\nabla}_2 f_x|^2 + \zeta^2} \, , \qquad \text{and} \qquad \theta_y := \frac{1}{|\boldsymbol{\nabla}_2 f_y|^2 + \zeta^2} \, . \qquad (4.22)$$

The normalised data term M_4 is then given by

$$M_4(u, v) = \boldsymbol{w}^\top \bar{\boldsymbol{J}}\, \boldsymbol{w} \, . \qquad (4.23)$$

Before we continue, let us remark that although the assumption of a global additive illumination change maybe not very realistic, experiments show that imposing gradient constancy nevertheless improves the results in most real world scenarios.

Colour Image Sequences. In a next step we extend our data term to multi-channel sequences $\boldsymbol{f} = (f^1, f^2, f^3)$. If one uses the standard RGB colour space, the three channels represent the red, green and blue channel, respectively. We couple the three colour channels in the joint motion tensor

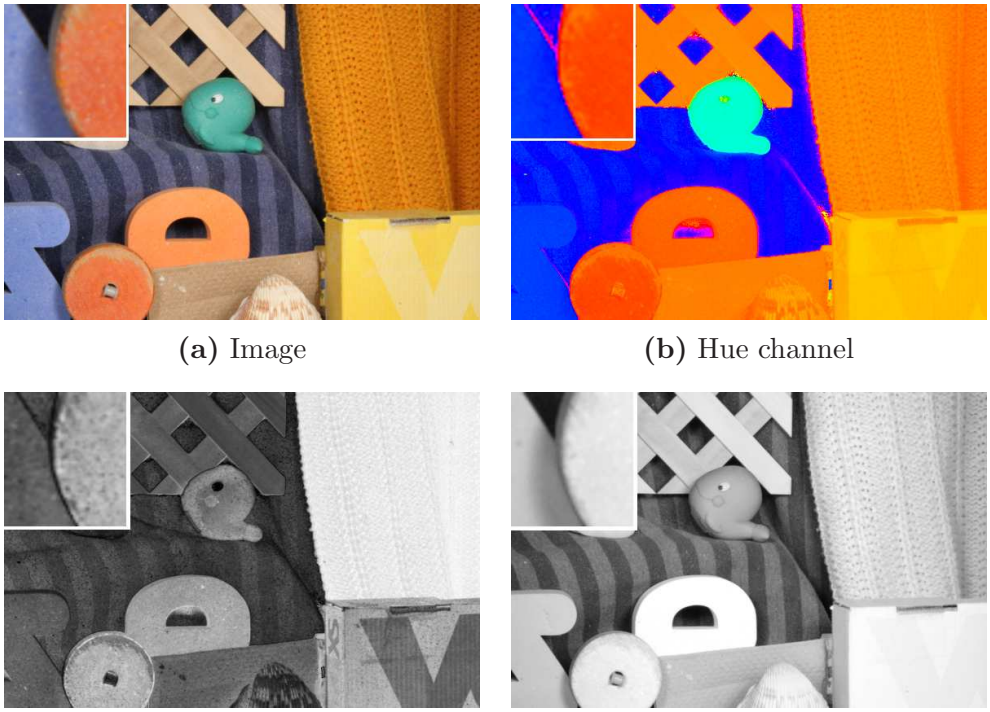
$$\begin{aligned}
\bar{\boldsymbol{J}}^{\mathcal{C}} &:= \sum_{c=1}^{3} \bar{\boldsymbol{J}}^c \\
&:= \sum_{c=1}^{3} \left[\bar{\boldsymbol{J}}_0^c + \gamma\, \bar{\boldsymbol{J}}_{xy}^c\right] \\
&:= \sum_{c=1}^{3} \left[\bar{\boldsymbol{J}}_0^c + \gamma \left(\bar{\boldsymbol{J}}_x^i + \bar{\boldsymbol{J}}_y^c\right)\right] \\
&:= \sum_{c=1}^{3} \Big[\theta_0^c \left(\boldsymbol{\nabla}_3 f^c\, (\boldsymbol{\nabla}_3 f^c)^\top\right) \qquad (4.24) \\
&\qquad + \gamma \left(\theta_x^c \left(\boldsymbol{\nabla}_3 f_x^c\, (\boldsymbol{\nabla}_3 f_x^c)^\top\right) + \theta_y^c \left(\boldsymbol{\nabla}_3 f_y^c\, (\boldsymbol{\nabla}_3 f_y^c)^\top\right)\right)\Big] \, ,
\end{aligned}$$

with normalisation factors θ^c for each colour channel f^c. The corresponding data term reads as

$$M_5(u, v) = \boldsymbol{w}^\top \bar{\boldsymbol{J}}^{\mathcal{C}}\, \boldsymbol{w} \, . \qquad (4.25)$$

Photometric Invariant Colour Spaces. It is generally agreed that realistic illumination models mainly encompass a multiplicative influence [vG04]. Consequently, solely imposing gradient constancy which is only invariant under additive illumination changes might not be sufficient. Following Golland and Bruckstein [GB97], this problem can be tackled by using the *Hue Saturation Value* (*HSV*) colour space. The hue channel is invariant under global and local multiplicative illumination changes, as well as under local additive changes. The saturation channel is invariant under global multiplicative illumination changes, and the value channel exhibits no invariances.

Taking the different levels of invariances into account, Mileva *et al.* [MBW07] proposed to only use the hue channel for optic flow computation as it is the most invariant channel. We will additionally use the saturation and value channel, because they contain information that is not present in the hue channel. As an example, consider the HSV decomposition of the *Rubberwhale* image from the Middlebury database [BSL+10] shown in Figure 4.2. As it turns out, the shadow at the left of the wheel (shown in the inset) is not visible in the hue and the saturation channel, but appears in the value channel. Nevertheless, especially the hue channel discards a lot of information, as can be observed for the striped cloth. This information is, on the other hand, available in the value channel.



(a) Image **(b)** Hue channel

(c) Saturation channel **(d)** Value channel

Figure 4.2: Example for an HSV decomposition. *First row, from left to right:* **(a)** Colour image with zoom in shadow region left of the wheel. **(b)** Same for hue channel (visualised with full saturation and value). *Second row, from left to right:* **(c)** Saturation channel. **(d)** Value channel.

Adequate Treatment of the Hue Channel. One problem when using the HSV colour space is that the hue channel f^1 describes an angle in a colour circle, i.e. $f^1 \in [0°, 360°)$. The hue channel is hence not differentiable at the interface between 0° and 360°. In practise this leads to the following problem: A large change in the angle, e.g. from 5° to 355°, leads to large derivative but corresponds to a small change in hue, only.

A possible remedy is to consider the unit vector $(\cos f^1, \sin f^1)^\top$ corresponding to the angle f^1. This results in treating the hue channel as two (coupled) channels, which are both differentiable. Practically, a change in angle is in this case expressed as the distance between the corresponding unit vectors, which does not suffer from the before mentioned problem.

If we apply this strategy to a normalised motion tensor corresponding to the brightness constancy assumption, one obtains

$$\bar{\boldsymbol{J}}_0^{\,1} := \theta_0^1 \Big(\boldsymbol{\nabla}_3 \cos f^1 \left(\boldsymbol{\nabla}_3 \cos f^1 \right)^\top + \boldsymbol{\nabla}_3 \sin f^1 \left(\boldsymbol{\nabla}_3 \sin f^1 \right)^\top \Big) \ , \tag{4.26}$$

where the normalisation factor is defined as

$$\theta_0^1 := \frac{1}{\left| \boldsymbol{\nabla}_2 \cos f^1 \right|^2 + \left| \boldsymbol{\nabla}_2 \sin f^1 \right|^2 + \zeta^2} \ . \tag{4.27}$$

Note that in the *differentiable parts* of the hue channel, the motion tensor (4.26) is equivalent to our earlier definition, as

$$\begin{aligned} & \boldsymbol{\nabla}_3 \cos f^1 \left(\boldsymbol{\nabla}_3 \cos f^1 \right)^\top + \boldsymbol{\nabla}_3 \sin f^1 \left(\boldsymbol{\nabla}_3 \sin f^1 \right)^\top \\ = \ & \sin^2 f^1 \left(\widetilde{\boldsymbol{\nabla}}_3 f^1 \left(\widetilde{\boldsymbol{\nabla}}_3 f^1 \right)^\top \right) + \cos^2 f^1 \left(\widetilde{\boldsymbol{\nabla}}_3 f^1 \left(\widetilde{\boldsymbol{\nabla}}_3 f^1 \right)^\top \right) \\ = \ & \left(\sin^2 f^1 + \cos^2 f^1 \right) \left(\widetilde{\boldsymbol{\nabla}}_3 f^1 \left(\widetilde{\boldsymbol{\nabla}}_3 f^1 \right)^\top \right) \\ = \ & \widetilde{\boldsymbol{\nabla}}_3 f^1 \left(\widetilde{\boldsymbol{\nabla}}_3 f^1 \right)^\top \ , \end{aligned} \tag{4.28}$$

where $\widetilde{\boldsymbol{\nabla}}_3$ denotes the gradient in the differentiable parts of the hue channel. The tensor $\bar{\boldsymbol{J}}_{xy}^{\,1}$ for the gradient constancy assumption is adapted accordingly:

$$\begin{aligned} \bar{\boldsymbol{J}}_{xy}^{\,1} := \quad & \theta_x^1 \left(\boldsymbol{\nabla}_3 \cos f_x^1 \ \left(\boldsymbol{\nabla}_3 \cos f_x^1 \right)^\top + \boldsymbol{\nabla}_3 \sin f_x^1 \ \left(\boldsymbol{\nabla}_3 \sin f_x^1 \right)^\top \right) \\ + \quad & \theta_y^1 \left(\boldsymbol{\nabla}_3 \cos f_y^1 \ \left(\boldsymbol{\nabla}_3 \cos f_y^1 \right)^\top + \boldsymbol{\nabla}_3 \sin f_y^1 \ \left(\boldsymbol{\nabla}_3 \sin f_y^1 \right)^\top \right) \ , \end{aligned} \tag{4.29}$$

with the normalisation factors

$$\theta_x^1 := \frac{1}{\left| \boldsymbol{\nabla}_2 \cos f_x^1 \right|^2 + \left| \boldsymbol{\nabla}_2 \sin f_x^1 \right|^2 + \zeta^2} \ , \tag{4.30}$$

and

$$\theta_y^1 := \frac{1}{\left| \boldsymbol{\nabla}_2 \cos f_y^1 \right|^2 + \left| \boldsymbol{\nabla}_2 \sin f_y^1 \right|^2 + \zeta^2} \ . \tag{4.31}$$

Robust Penalisers. To provide robustness of the data term against outliers (possibly caused by noise or occlusions), Black and Anandan [BA96] proposed to refrain from a quadratic penalisation as it gives too much influence to outliers. Instead they use a subquadratic penaliser function $\Psi_M(s^2)$, where s^2 denotes the quadratic data term. Using such a robust penaliser within our data term yields

$$M_6(u,v) = \Psi_M\Big(\boldsymbol{w}^\top \bar{\boldsymbol{J}}^{\mathcal{C}}\, \boldsymbol{w}\Big) \quad . \tag{4.32}$$

Good results are reported by Brox *et al.* [BBPW04] for the subquadratic penaliser $\Psi_M(s^2) := \sqrt{s^2+\varepsilon^2}$, that uses a small regularisation parameter $\varepsilon > 0$.

Bruhn and Weickert [BW05] later proposed a *separate* penalisation of the brightness and the gradient constancy assumption, which is advantageous in cases where only one assumption produces an outlier, but not the other. Incorporating this strategy into our approach gives the data term

$$M_7(u,v) = \Psi_M\Big(\boldsymbol{w}^\top \bar{\boldsymbol{J}}_0^{\mathcal{C}}\, \boldsymbol{w}\Big) + \gamma\, \Psi_M\big(\boldsymbol{w}^\top \bar{\boldsymbol{J}}_{xy}^{\mathcal{C}}\, \boldsymbol{w}\big) \quad , \tag{4.33}$$

where the separate motion tensors are defined as

$$\bar{\boldsymbol{J}}_0^{\mathcal{C}} := \sum_{c=1}^{3} \bar{\boldsymbol{J}}_0^{c} \, , \quad \text{and} \quad \bar{\boldsymbol{J}}_{xy}^{\mathcal{C}} := \sum_{c=1}^{3} \bar{\boldsymbol{J}}_{xy}^{c} \quad . \tag{4.34}$$

We will go a step further by proposing a separate robust penalisation of each channel in the HSV colour space. This is justified by the distinct information content of each of the three channels, see Figure 4.2. The separate robust penalisation then allows to only consider the most reliable channel at each location, and to reduce the influence of the less reliable channels.

Final Data Term. Incorporating the separate robust penalisation into M_7 brings us to our final data term

$$M(u,v) = \sum_{c=1}^{3} \Psi_M\big(\boldsymbol{w}^\top \bar{\boldsymbol{J}}_0^{c}\, \boldsymbol{w}\big) + \gamma \left(\sum_{c=1}^{3} \Psi_M\big(\boldsymbol{w}^\top \bar{\boldsymbol{J}}_{xy}^{c}\, \boldsymbol{w}\big) \right) \quad , \tag{4.35}$$

where the motion tensors $\bar{J}$ [1] are adapted to the HSV colour representation as described before on page 40. Note that our final data term is *(i)* normalised, *(ii)* combines the brightness and gradient constancy assumption, and *(iii)* uses the HSV colour space with *(iv)* a separate robust penalisation of all channels.

Contribution to the Euler-Lagrange Equations. The contributions of our final data term (4.35) to the Euler-Lagrange equations (4.2) and (4.3) are given by

$$\begin{aligned}\partial_u M = \quad & \sum_{c=1}^{3}\left(\Psi'_M\left(\boldsymbol{w}^\top \bar{\boldsymbol{J}}^c_0\, \boldsymbol{w}\right)\cdot\left(\left[\bar{\boldsymbol{J}}^c_0\right]_{1,1} u + \left[\bar{\boldsymbol{J}}^c_0\right]_{1,2} v + \left[\bar{\boldsymbol{J}}^c_0\right]_{1,3}\right)\right) \\ + \gamma\Bigg(\quad & \sum_{c=1}^{3}\left(\Psi'_M\left(\boldsymbol{w}^\top \bar{\boldsymbol{J}}^c_{xy}\, \boldsymbol{w}\right)\cdot\left(\left[\bar{\boldsymbol{J}}^c_{xy}\right]_{1,1} u + \left[\bar{\boldsymbol{J}}^c_{xy}\right]_{1,2} v + \left[\bar{\boldsymbol{J}}^c_{xy}\right]_{1,3}\right)\right)\Bigg)\ ,\end{aligned} \tag{4.36}$$

$$\begin{aligned}\partial_v M = \quad & \sum_{c=1}^{3}\left(\Psi'_M\left(\boldsymbol{w}^\top \bar{\boldsymbol{J}}^c_0\, \boldsymbol{w}\right)\cdot\left(\left[\bar{\boldsymbol{J}}^c_0\right]_{1,2} u + \left[\bar{\boldsymbol{J}}^c_0\right]_{2,2} v + \left[\bar{\boldsymbol{J}}^c_0\right]_{2,3}\right)\right) \\ + \gamma\Bigg(\quad & \sum_{c=1}^{3}\left(\Psi'_M\left(\boldsymbol{w}^\top \bar{\boldsymbol{J}}^c_{xy}\, \boldsymbol{w}\right)\cdot\left(\left[\bar{\boldsymbol{J}}^c_{xy}\right]_{1,2} u + \left[\bar{\boldsymbol{J}}^c_{xy}\right]_{2,2} v + \left[\bar{\boldsymbol{J}}^c_{xy}\right]_{2,3}\right)\right)\Bigg)\ ,\end{aligned} \tag{4.37}$$

where $[\boldsymbol{J}]_{m,n}$ denotes the entry in row m and column n of the matrix $\boldsymbol{J}$, and $\Psi'_M(s^2)$ is the derivative of $\Psi_M(s^2)$ w.r.t. its argument. For our proposed penaliser $\Psi_M(s^2) = \sqrt{s^2+\varepsilon^2}$ one for example obtains $\Psi'_M(s^2) = 1\,/\,(2\sqrt{s^2+\varepsilon^2})$.

Analysing the terms (4.36) and (4.37), we can analyse the effect of the separate robust penalisation of the HSV channels: If a specific channel violates the imposed constancy assumption at a certain location, the corresponding argument of the decreasing function Ψ'_M will be large, which reduces the influence of this channel. Other channels that satisfy the constancy assumption then have a dominating influence on the solution.

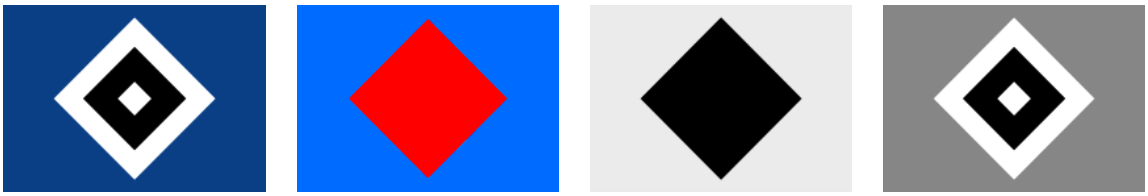
4.2.1 Remarks

Let us finish this section by some remarks on the presented concepts.

A Note on Photometric Invariances. Using a photometric invariant colour representation (like our proposed HSV colour space) renders the estimation of the optic flow invariant under lighting artefacts like shadows. Is this, however, desirable? If one is interested in recovering the optic flow in its original definition where it describes the apparent motion of brightness patterns between two images [Hor86], such an invariance is undesirable as it hinders the estimation of apparent motion that is due to lighting changes. Estimating these motions can, on the other hand, also be of great importance, e.g. for frame interpolation applications.

Nevertheless, photometric invariances are often desirable. This is the case if one aims at recovering the actual motion of objects in the scene, i.e. the projection of the 3D motion field (scene flow) onto the 2D image plane. An example for such a scenario are driver assistance systems, that should not be misled by moving shadows.

Issues with an HSV colour representation. Apart from its (mostly) desirable photometric invariances, the HSV colour representation may also pose problems. One such problem is that greyscales are not distinguishable in the hue and the saturation channel. An example for this problem is given in Figure 4.3 that shows the HSV decomposition of the logo of the Hamburger Sportverein (a German football club, interestingly also called HSV). Apparently, the black and white areas in the centre are indistinguishable in the hue as well as the saturation channel. In the hue channel, the angle corresponding to a greyscale is even not defined. For visualisation purposes, we assign an angle of 0°, corresponding to a red hue.



(a) Image **(b)** Hue **(c)** Saturation **(d)** Value

Figure 4.3: Problems of an HSV decomposition. *From left to right:* **(a)** Colour image depicting the logo of the Hamburger Sportverein (`http://www.hsv.de`). **(b)** Hue channel, visualised with full saturation and value. **(c)** Saturation channel. **(d)** Value channel.

A RGB Variant of our Data Term. We have seen that in some situations, an HSV colour representation may not be suitable. As a remedy, we propose to resort to a standard RGB colour representation. In this case, however, a separate robust penalisation of each colour channel makes no sense as none of the RGB channels exhibits photometric invariances. Consequently, we only separately apply the robust penaliser function to the brightness and the gradient constancy assumption, as in the data term M_7 from (4.33).

Linearisation versus Large Displacements. In order to make the minimisation of the resulting energy functional feasible, we performed a linearisation of the data constraints. This proceeding is, however, only valid under the assumption of small displacements in the sequence or for very smooth image data. However, those assumptions are often violated under real world conditions, e.g. when the frame rate of the camera used to record an image sequence is too low. A popular strategy for handling large displacements is to use a coarse-to-fine multiscale warping approach, as for example in [AWS00, Ana89, BA96, BBPW04, MP98, WTK87]. At a coarse (downsampled or smoothed) version of the image sequence, the displacements become small enough to be estimated by a linearised approach. The obtained coarse solution then serves as initialisation for the next finer level, resulting in an incremental estimation of the final flow estimate.

To simplify the presentation of our ideas, we stick to linearised data terms. In Section 5.3.1 we will detail on how to incorporate our models into a warping scheme to handle large displacements correctly.

Data Terms for Stereo. Transferring the presented concepts to stereo in the orthoparallel case is straightforward. Recall that in the orthoparallel case, one only experiences horizontal displacements u. Thus, to transfer our optic flow data terms to orthoparallel stereo, we only have to set the vertical flow component to zero, i.e. $v = 0$. For the brightness constancy assumption this yields

$$f(x+u, y, t+1) = f(x, y, t) \ . \tag{4.38}$$

Performing a linearisation and a quadratic penalisation, one ends up with the data term

$$(f_x\, u + f_t)^2 = \left((\boldsymbol{\nabla}_{xt} f)^\top \begin{pmatrix} u \\ 1 \end{pmatrix} \right)^2 = (u, 1) \left(\boldsymbol{\nabla}_{xt} f \, (\boldsymbol{\nabla}_{xt} f)^\top \right) \begin{pmatrix} u \\ 1 \end{pmatrix} \ . \tag{4.39}$$

As analogue to the the basic motion tensor J_0 from (4.10), we can now define a basic *stereo tensor* as

$$\boldsymbol{J}_0^{\text{Stereo}} := \boldsymbol{\nabla}_{xt} f (\boldsymbol{\nabla}_{xt} f)^\top \ . \tag{4.40}$$

Comparing the stereo data term to its to its optic flow counterpart, we realise that the gradient operator $\boldsymbol{\nabla}_3$ was replaced by $\boldsymbol{\nabla}_{xt}$ and the flow field $\boldsymbol{w}$ was replaced by the vector $(u, 1)^\top$. Performing these two modifications, all of the above concepts can be transferred from optic flow to orthoparallel stereo.

Normalisation. The normalisation factors for stereo are obtained from their optic flow counterparts by dropping the contributions of f_y. The normalisation factor for the brightness constancy part consequently reads as

$$\frac{1}{f_x^2 + \zeta^2} \ . \tag{4.41}$$

Note that this factor makes sense because of the following rewriting

$$(f_x\, u + f_t)^2 \ = \ \left(f_x \left(u + \frac{f_t}{f_x} \right) \right)^2 \ = \ f_x^2 \left(u + \frac{f_t}{f_x} \right)^2 =: \ f_x^2 \, (u - u_n)^2 \ , \tag{4.42}$$

where $u_n := -(f_t / f_x)$ is the stereo counterpart of the normal flow, as it solves the linearised constraint (4.39). We see that also in the stereo context, a quadratic penalisation of the brightness constancy assumption implicitly results in an undesirable weighting with f_x^2. As a remedy, one can normalise the data term by multiplication with the factor from (4.41).

4.2.2 Experiments

We now present some experiments that prove the merits of our data term design. For the smoothness term, we use our novel complementary regulariser that will be presented in Section 4.3.2.

Benefits of Normalisation and the HSV Colour Space. We proposed two main innovations in the data term: constraint normalisation and using an HSV colour representation. In our first experiment, we thus compare our method against variants of it where we *(i)* do not perform data term normalisation, and *(ii)* where we use the RGB instead of the HSV colour space. Recall that in the latter case, we only separately robustify the brightness and the gradient constancy assumption, as a separate robustification of the RGB channels makes no sense. In Figure 4.4 we show the results for the *Snail* sequence that we have created. Note that it is a rather challenging sequence due to severe shadows and large displacements up to 25 pixels.

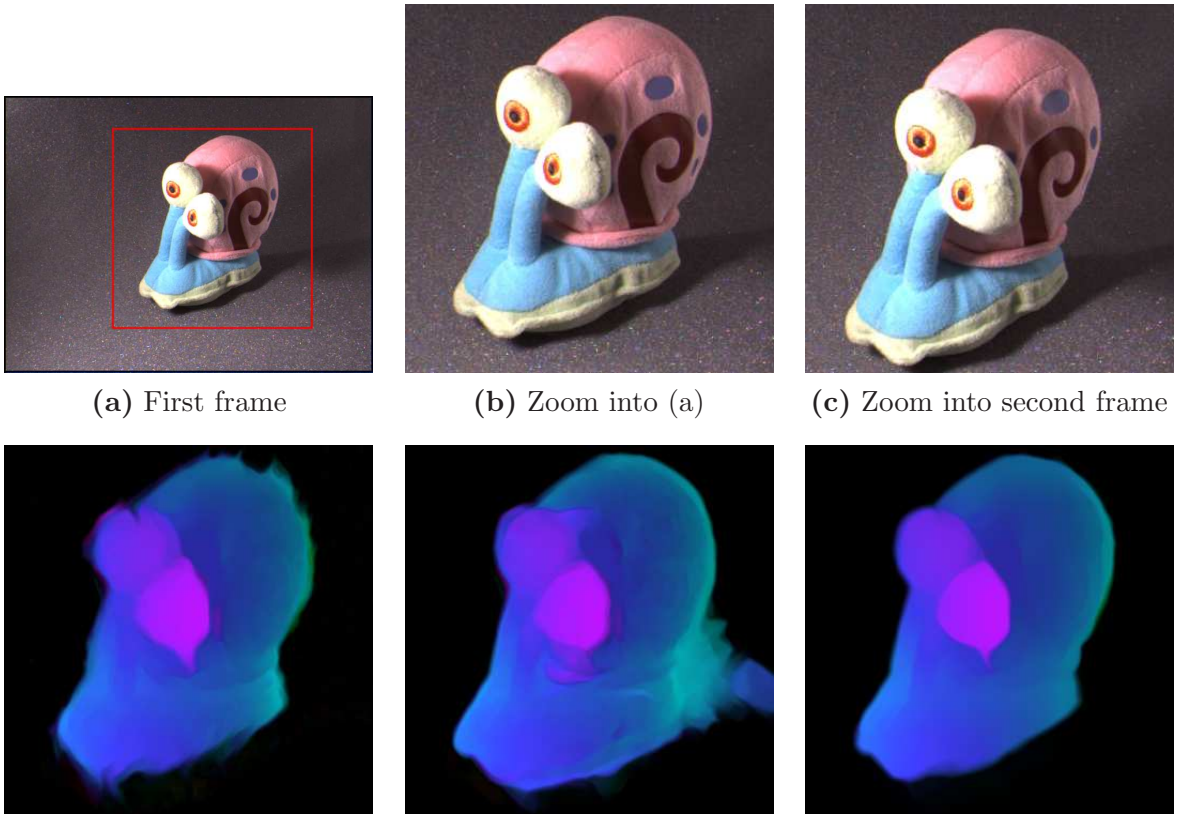
When comparing the results to our final result in Figure 4.4 (f), the following drawbacks of the modified versions become obvious: Without data term normalisation (Figure 4.4 (d)), unpleasant artefacts at image edges arise, even when using a large smoothness weight α. This is explained by the fact that without normalisation, the data term is overweighted at these locations and thus noise and occlusions lead to artefacts. When relying on the RGB colour space (Figure 4.4 (e)), a phantom motion in the shadow region at the right border is estimated.

Effect of the Separate Robust Penalisation. This experiment illustrates the desirable effect of our separate robust penalisation of the HSV channels. Using the *Rubberwhale* sequence from the Middlebury database [BSL+10], we show in Figure 4.5 the data term weights $\Psi'_M(\boldsymbol{w}^\top \bar{\boldsymbol{J}}_0^c \boldsymbol{w})$ for the brightness constancy assumption on the hue, the saturation and the value channel ($c=1,..,3$). Here, brighter pixels correspond to a larger weight and we only show a zoom for better visibility.

As we can see, the weight of the value channel is reduced in the shadow regions (left of the wheel, of the orange toy and of the clam). This is desirable as the value channel is not invariant under shadows, see Figure 4.2.

A CLG Variant of Our Method. Our next experiment is concerned with a CLG variant [BWS05] of our data term. In this case we perform (as for the regularisation tensor) a Gaussian convolution of the motion tensor entries. Formally, we thus replace the motion tensor $\bar{\boldsymbol{J}}_0^c$ by $K_\rho * \bar{\boldsymbol{J}}_0^c$ and $\bar{\boldsymbol{J}}_{xy}^c$ by $K_\rho * \bar{\boldsymbol{J}}_{xy}^c$. From this it also becomes clear that for $\rho = 0$ we fall back to our original model. Consequently, a CLG variant of our method refers to a setting with $\rho > 0$ that led to the best results.

First, we compare our method against a CLG variant for some Middlebury sequences, see Table 4.1. We find that the CLG variant always leads to worse results and conclude that for the considered test sequences, this modification seems not to be useful, i.e. setting $\rho = 0$ seems most appropriate. However, a CLG variant of can actually be useful in the presence of severe noise in the image sequence. To prove this, we compare in Table 4.2 the performance of our method to its CLG counterpart on noisy versions of the *Yosemite* sequence. As it turns out, the CLG variant improves the results at large noise scales, but deteriorates the quality for low noise scenarios. This also explains the experienced behaviour on the Middlebury data sets, which hardly suffer from noise.



(a) First frame **(b)** Zoom into (a) **(c)** Zoom into second frame

(d) No normalisation **(e)** RGB colour space **(f)** Proposed method

Figure 4.4: Results on the *Snail sequence* with different variants of our data term. *First row, from left to right:* **(a)** First frame. **(b)** Zoom into marked region of first frame. **(c)** Same for second frame. *Second row, from left to right:* **(d)** Flow field in marked region, without normalisation ($\alpha = 5000.0$). **(e)** Same for RGB colour space ($\alpha = 300.0$). **(f)** Same for proposed method that uses HSV colour space and normalisation ($\alpha = 2000.0$). All results used the fixed parameters $\sigma = 0.5, \gamma = 20.0, \rho = 4.0, \eta = 0.95$.

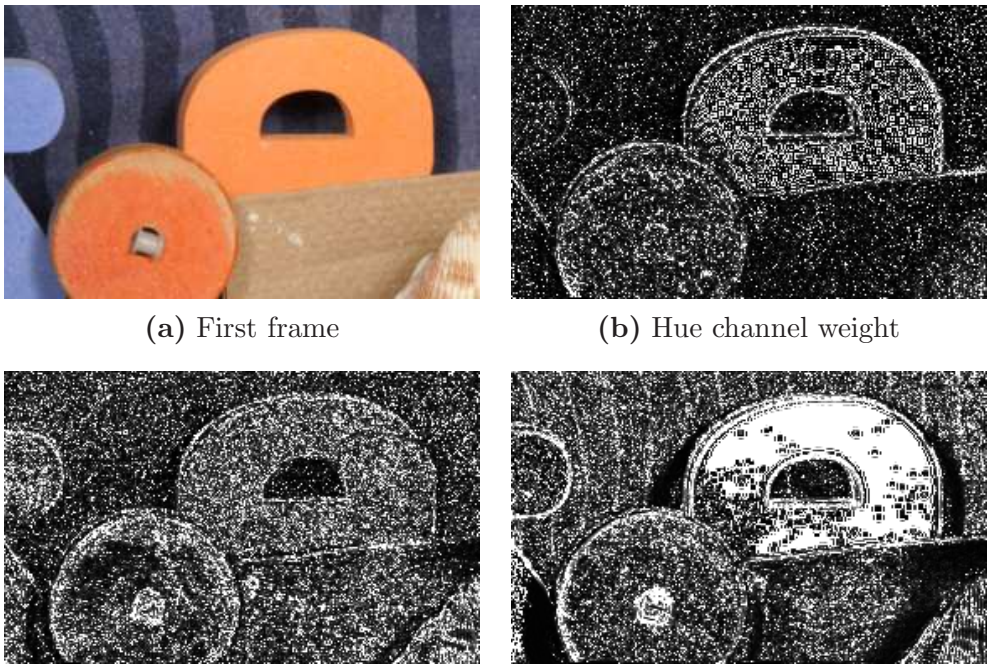
4.3 Smoothness Term

This section first reviews existing smoothness terms that were influential for the design of our complementary regulariser [10, 11]. After presenting the latter and its extension to the spatio-temporal domain, we finally also discuss comparable smoothness terms for stereo.

4.3.1 A General Framework for Smoothness Terms

To ease notation and comparison of all discussed regularisers, we rewrite them in a novel framework [10] that unifies their notation and eases their comparison.

Preliminaries for our General Framework. Let us first some basic concepts that will be used in our general framework.



(a) First frame **(b)** Hue channel weight

(c) Saturation channel weight **(d)** Value channel weight

Figure 4.5: Effect of our separate robust penalisation of the HSV channels. *First row, from left to right:* **(a)** Zoom in first frame of the *Rubberwhale* sequence. **(b)** Visualisation of the corresponding hue channel weight. Brighter pixels correspond to a larger weight. *Second row, from left to right:* **(c)** Same for the saturation channel. **(d)** Same for the value channel.

Table 4.1: Comparison of our data term to a CLG variant (using the AAE)

Sequence	*Rubberwhale*	*Dimetrodon*	*Grove2*	*Urban2*
CLG ($\rho > 0$)	3.00°	1.59°	2.20°	2.67°
Original ($\rho = 0$)	**2.77°**	**1.54°**	**2.16°**	**2.49°**

Table 4.2: Comparison of our data term to a CLG variant on noisy versions of the *Yosemite* sequence (using the AAE). We have added Gaussian noise with zero mean and standard deviation σ_n

σ_n	0	10	20	40
CLG ($\rho > 0$)	1.75°	4.01°	**5.82°**	**8.14°**
Original ($\rho = 0$)	**1.64°**	**3.82°**	6.08°	8.55°

Image-driven regularisers take into account image structures. To analyse the latter, one can consider the *structure tensor* [FG87]

$$\boldsymbol{S}_\rho \;:=\; K_\rho * \left[\boldsymbol{\nabla}_2 f \; (\boldsymbol{\nabla}_2 f)^\top\right] \;=:\; \sum_{i=1}^{2} \mu_i \;\boldsymbol{s}_i \, \boldsymbol{s}_i^\top \;\;, \tag{4.43}$$

with an integration scale $\rho > 0$. The structure tensor is a symmetric, positive semi-definite 2×2 matrix that possesses two orthonormal eigenvectors $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ with corresponding eigenvalues μ_1 and μ_2, where w.l.o.g. $\mu_1 \geq \mu_2 \geq 0$. Directional information of image structures is encoded in the eigenvectors: The vector $\boldsymbol{s}_1$ points across image structures, whereas the vector $\boldsymbol{s}_2$ points along them. The corresponding eigenvalues μ_1 and μ_2 describe the contrast along these directions.

In the case $\rho = 0$, i.e. without considering neighbourhood information, one obtains

$$\boldsymbol{s}_1^0 = \frac{\boldsymbol{\nabla}_2 f}{|\boldsymbol{\nabla}_2 f|} \;, \quad \text{and} \quad \boldsymbol{s}_2^0 = \frac{(\boldsymbol{\nabla}_2 f)^\perp}{|\boldsymbol{\nabla}_2 f|} \;\;, \tag{4.44}$$

where $(\boldsymbol{\nabla}_2 f)^\perp := (-f_y, f_x)^\top$ denotes the vector orthogonal to $\boldsymbol{\nabla}_2 f$. The corresponding eigenvalues in this case are $\mu_1^0 = |\boldsymbol{\nabla}_2 f|^2$ and $\mu_2^0 = 0$. For $\rho = 0$ we also have that $|\boldsymbol{\nabla}_2 f|^2 = \text{tr } \boldsymbol{S}_0$, with tr denoting the trace operator. Consequently, a large value of tr $\boldsymbol{S}_0$ indicates the presence of an image edge.

Most regularisers impose smoothness of the flow field by penalising the magnitude of the flow gradients. As $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ constitute an orthonormal basis, we can rewrite the (squared) magnitude of the flow gradients as

$$|\boldsymbol{\nabla}_2 u|^2 \;=\; u_x^2 + u_y^2 \;=\; u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 \;\;, \tag{4.45}$$

$$|\boldsymbol{\nabla}_2 v|^2 \;=\; v_x^2 + v_y^2 \;=\; v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2 \;\;, \tag{4.46}$$

using directional derivatives in $\boldsymbol{s}_1$- and $\boldsymbol{s}_2$-direction.

To analyse the smoothing behaviour of the presented regularisers we will consider the corresponding Euler-Lagrange equations (4.2) and (4.3). For all regularisers discussed here, the Euler-Lagrange equations can be written in the form

$$\partial_u M - \alpha \, \text{div} \left(\boldsymbol{D} \, \boldsymbol{\nabla}_2 u\right) \;=\; 0 \;\;, \tag{4.47}$$

$$\partial_v M - \alpha \, \text{div} \left(\boldsymbol{D} \, \boldsymbol{\nabla}_2 v\right) \;=\; 0 \;\;, \tag{4.48}$$

with a diffusion tensor $\boldsymbol{D}$ that steers the smoothing of the flow components u and v. Specifically, the eigenvectors of $\boldsymbol{D}$ give the smoothing direction, whereas the corresponding eigenvalues determine the magnitude of the smoothing. Thus, the eigenstructure of the diffusion tensor determines the smoothing behaviour of the corresponding regulariser. For analysing the eigenstructure, we consider the eigendecomposition

$$\boldsymbol{D} = \sum_{i=1}^{2} m_i \, \boldsymbol{d}_i \boldsymbol{d}_i^\top \;\;, \tag{4.49}$$

where $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ are the eigenvectors of $\boldsymbol{D}$ with corresponding eigenvalues $m_1 \geq m_2 \geq 0$.

Homogeneous Regularisation. The first and most basic idea for a regulariser in variational optic flow estimation goes back to Horn and Schunck [HS81]. They proposed a homogeneous regulariser which we rewrite in our framework as

$$\begin{aligned} V_{\mathrm{H}}(\nabla_2 u, \nabla_2 v) &:= |\nabla_2 u|^2 + |\nabla_2 v|^2 \\ &= u_{s_1}^2 + u_{s_2}^2 + v_{s_1}^2 + v_{s_2}^2 \ . \end{aligned} \tag{4.50}$$

The corresponding diffusion tensor and its eigendecomposition are given by

$$\begin{aligned} \boldsymbol{D}_{\mathrm{H}} &= \boldsymbol{I} := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= 1 \cdot \boldsymbol{e}_x \boldsymbol{e}_x^\top + 1 \cdot \boldsymbol{e}_y \boldsymbol{e}_y^\top \ , \end{aligned} \tag{4.51}$$

where $\boldsymbol{e}_x := (1,0)^\top$ and $\boldsymbol{e}_y := (0,1)^\top$ denote the unit vectors in x- and y-direction, respectively. It becomes obvious that this regulariser performs a homogeneous smoothing with the full strength of 1 at each location. Thus, it will blur important flow edges.

Image-Driven Regularisation. To obtain sharp flow edges, image-driven methods like [AELS99, NE86] reduce the smoothing at image edges, which manifest themselves by large values of $|\nabla_2 f|^2$.

An isotropic image-driven regulariser was proposed by Alvarez et al. [AELS99] and can be written as

$$\begin{aligned} V_{\mathrm{II}}(\nabla_2 u, \nabla_2 v) &:= g\big(|\nabla_2 f|^2\big) \, \big(|\nabla_2 u|^2 + |\nabla_2 v|^2\big) \\ &= g(\operatorname{tr} \boldsymbol{S}_0) \, \big(u_{s_1}^2 + u_{s_2}^2 + v_{s_1}^2 + v_{s_2}^2\big) \ , \end{aligned} \tag{4.52}$$

where g is a decreasing, strictly positive weight function. Considering the eigendecomposition of the corresponding diffusion tensor

$$\begin{aligned} \boldsymbol{D}_{\mathrm{II}} &= g(\operatorname{tr} \boldsymbol{S}_0) \, \boldsymbol{I} \\ &= g(\operatorname{tr} \boldsymbol{S}_0) \cdot \boldsymbol{e}_x \boldsymbol{e}_x^\top + g(\operatorname{tr} \boldsymbol{S}_0) \cdot \boldsymbol{e}_y \boldsymbol{e}_y^\top \ , \end{aligned} \tag{4.53}$$

it turns out that the weight function allows to decrease the smoothing in accordance to the strength of image edges. However, the smoothing in x- as well as in y-direction is equally reduced.

To additionally steer the direction of smoothing and more importantly, to allow for a different strength of smoothing in different directions, anisotropic strategies have been proposed. The anisotropic image-driven regulariser of Nagel and Enkelmann [NE86] prevents smoothing of the flow field across image boundaries but encourages smoothing along them. This is achieved by the regulariser

$$V_{\mathrm{AI}}(\nabla_2 u, \nabla_2 v) := (\nabla_2 u)^\top \boldsymbol{P}(\nabla_2 f) \, \nabla_2 u + (\nabla_2 v)^\top \boldsymbol{P}(\nabla_2 f) \, \nabla_2 v \ , \tag{4.54}$$

where $\boldsymbol{P}(\boldsymbol{\nabla}_2 f)$ denotes a regularised projection matrix perpendicular to the image gradient $\boldsymbol{\nabla}_2 f$, i.e. along image edges. The projection matrix is defined as

$$\boldsymbol{P}(\boldsymbol{\nabla}_2 f) := \frac{1}{|\boldsymbol{\nabla}_2 f|^2 + 2\kappa^2} \left((\boldsymbol{\nabla}_2 f)^{\perp} \, (\boldsymbol{\nabla}_2 f)^{\perp^\top} + \kappa^2 \boldsymbol{I} \right) \quad , \tag{4.55}$$

with a regularisation parameter $\kappa > 0$. In our framework, the regulariser of Nagel and Enkelmann can be written as

$$V_{\mathrm{AI}}(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) = \frac{\kappa^2}{\operatorname{tr} \boldsymbol{S}_0 + 2\kappa^2} \left(u^2_{\boldsymbol{s}^0_1} + v^2_{\boldsymbol{s}^0_1} \right) + \frac{\operatorname{tr} \boldsymbol{S}_0 + \kappa^2}{\operatorname{tr} \boldsymbol{S}_0 + 2\kappa^2} \left(u^2_{\boldsymbol{s}^0_2} + v^2_{\boldsymbol{s}^0_2} \right) \quad . \tag{4.56}$$

The correctness of the rewriting (4.56) is shown in the Appendix A and is based on the observations that $\boldsymbol{s}^0_1$ and $\boldsymbol{s}^0_2$ are the eigenvectors of $\boldsymbol{P}$, and that the factors in front of $(u^2_{\boldsymbol{s}^0_1} + v^2_{\boldsymbol{s}^0_1})$ and $(u^2_{\boldsymbol{s}^0_2} + v^2_{\boldsymbol{s}^0_2})$ are the corresponding eigenvalues. The diffusion tensor for this regulariser is identical to the projection matrix, which leads to the eigendecomposition

$$\begin{aligned} \boldsymbol{D}_{\mathrm{AI}} &= \boldsymbol{P}(\boldsymbol{\nabla}_2 f) \\ &= \frac{\kappa^2}{\operatorname{tr} \boldsymbol{S}_0 + 2\kappa^2} \, \boldsymbol{s}^0_1 \left(\boldsymbol{s}^0_1 \right)^\top + \frac{\operatorname{tr} \boldsymbol{S}_0 + \kappa^2}{\operatorname{tr} \boldsymbol{S_0} + 2\kappa^2} \, \boldsymbol{s}^0_2 \left(\boldsymbol{s}^0_2 \right)^\top \quad . \end{aligned} \tag{4.57}$$

Let us consider the limiting case where the regularisation parameter tends to zero, i.e. $\kappa \to 0$. The regulariser then becomes

$$V_{\mathrm{AI}}(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) = u^2_{\boldsymbol{s}^0_2} + v^2_{\boldsymbol{s}^0_2} \quad , \tag{4.58}$$

and the eigendecomposition of the diffusion tensor simplifies to

$$\boldsymbol{D}_{\mathrm{AI}} = \boldsymbol{s}^0_2 \left(\boldsymbol{s}^0_2 \right)^\top \quad . \tag{4.59}$$

In this limit, we thus obtain a smoothing of the flow field solely in $\boldsymbol{s}^0_2$-direction, i.e. along image edges. In the definition of the normal flow (4.11) we have seen that a data term that models the brightness constancy assumption constraints the flow only orthogonal to image edges. In the limiting case, the regulariser of Nagel and Enkelmann can hence be interpreted as a first complementary smoothness term that fills in information orthogonal to the data constraint direction. This complementarity has been emphasised by Schnörr [Sch93], who contributed a theoretical analysis as well as modifications of the original Nagel and Enkelmann functional.

As not every image edge necessarily corresponds to a flow edge, image-driven strategies are prone to give oversegmentation artefacts, especially in textured image regions.

Flow-Driven Regularisation. To remedy the oversegmentation problem, it makes sense to adapt the smoothing process to the *flow* edges instead of the image edges.

In the isotropic setting, Shulman and Hervé [SH89] as well as Schnörr [Sch94] proposed to use rotationally invariant, subquadratic penaliser functions Ψ_V, resulting in

$$\begin{aligned} V_{\mathrm{IF}}(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) &:= \Psi_V\big(|\boldsymbol{\nabla}_2 u|^2 + |\boldsymbol{\nabla}_2 v|^2\big) \\ &= \Psi_V\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\big) \quad . \end{aligned} \tag{4.60}$$

The associated diffusion tensor is given by

$$\begin{aligned} \boldsymbol{D}_{\mathrm{IF}} &= \Psi_V'\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\big)\, \boldsymbol{I} \\ &= \Psi_V'\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\big) \cdot \boldsymbol{e}_x \boldsymbol{e}_x^\top \\ &\quad + \Psi_V'\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\big) \cdot \boldsymbol{e}_y \boldsymbol{e}_y^\top \quad . \end{aligned} \tag{4.61}$$

The eigenstructure of this diffusion tensor is very similar to the one in the isotropic image-driven setting, see (4.53). However, in the flow-driven case, the derivative Ψ' of the penaliser function serves as decreasing weight function that reduces the smoothing of the flow field. The weighting thus depends on the contrast of the evolving flow field, and not anymore on image edges.

The underlying diffusion processes on the flow field are given by $u_t = \operatorname{div}(\boldsymbol{D}_{\mathrm{IF}} \boldsymbol{\nabla}_2 u)$ and $v_t = \operatorname{div}(\boldsymbol{D}_{\mathrm{IF}} \boldsymbol{\nabla}_2 v)$, respectively, which can be characterised as nonlinear isotropic diffusion processes. In order to introduce *total variation* (TV) regularisation [ROF92] to the optic flow case, Cohen [Coh93] used the convex penaliser

$$\Psi_V(s^2) := \sqrt{s^2 + \varepsilon^2} \quad , \tag{4.62}$$

with the diffusivity

$$\Psi_V'(s^2) = \frac{1}{2\sqrt{s^2 + \varepsilon^2}} \approx \frac{1}{2\,|s|} \quad . \tag{4.63}$$

using a small regularisation parameter $\varepsilon > 0$. Another possible choice is the non-convex *Perona-Malik regulariser* (*Lorentzian*) [BA96, PM90] given by

$$\Psi_V(s^2) := \lambda^2 \, \log\left(1 + \frac{s^2}{\lambda^2}\right) \quad , \tag{4.64}$$

that results in *Perona-Malik diffusion* [PM90] with the diffusivity

$$\Psi_V'(s^2) = \frac{1}{1 + \frac{s^2}{\lambda^2}} \quad , \tag{4.65}$$

with a contrast parameter $\lambda > 0$.

An extension of flow-driven strategies to the anisotropic setting was presented by Weickert and Schnörr [WS01a]. They proposed the regulariser

$$V_{\mathrm{AF}}(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) := \operatorname{tr} \Psi_V\big(\boldsymbol{\nabla}_2 u\, (\boldsymbol{\nabla}_2 u)^\top + \boldsymbol{\nabla}_2 v\, (\boldsymbol{\nabla}_2 v)^\top\big) \quad , \tag{4.66}$$

where the application of the function Ψ_V to the tensor $\boldsymbol{F} := \boldsymbol{\nabla}_2 u \, (\boldsymbol{\nabla}_2 u)^\top + \boldsymbol{\nabla}_2 v \, (\boldsymbol{\nabla}_2 v)^\top$ is realised by applying Ψ_V to the eigenvalues of $\boldsymbol{F}$, i.e.

$$\boldsymbol{F} := \sum_{i=1}^{2} \nu_i \, \boldsymbol{v}_i \boldsymbol{v}_i^\top \qquad \Rightarrow \qquad \Psi_V(\boldsymbol{F}) := \sum_{i=1}^{2} \Psi_V(\nu_i) \, \boldsymbol{v}_i \boldsymbol{v}_i^\top \ . \tag{4.67}$$

The eigenvectors $\boldsymbol{v}_i$ describe the direction of local flow structures, and the corresponding eigenvalues ν_i encode the contrast of the flow field in the eigenvector direction. If one analytically computes the eigenvalues and eigenvectors of $\boldsymbol{F}$, one finds that they cannot be expressed using our rewriting of the flow gradients from (4.45) and (4.46). Hence, this regulariser cannot be written in our framework. As shown in [WS01a], the diffusion tensor corresponding to V_{AF} is given by

$$\begin{aligned} \boldsymbol{D}_{\mathrm{AF}} &= \Psi_V'(\boldsymbol{F}) \\ &= \sum_{i=1}^{2} \Psi_V'(\nu_i) \, \boldsymbol{v}_i \boldsymbol{v}_i^\top \ . \end{aligned} \tag{4.68}$$

The smoothing direction is thus adapted in accordance to flow structures encoded in the eigenvectors $\boldsymbol{v}_i$. A real anisotropic smoothing behaviour is achieved as the smoothing strength depends on the in general not identical eigenvalues ν_1 and ν_2.

Despite the fact that flow-driven methods solve the oversegmentation problem, they suffer from another drawback: The flow edges are not as sharp and well localised as with image-driven strategies.

Joint Image- and Flow-Driven Regularisation. Our review of image- and flow-driven strategies has revealed that the two classes exhibit complementary strengths and weaknesses: Image-driven methods give sharp flow edges, but suffer from oversegmentation artefacts. Flow-driven strategies remedy the oversegmentation problem, but give less pleasant flow edges. It would thus be desirable to combine the advantages of both strategies to obtain sharp flow edges without oversegmentation problems.

A first attempt to combine ideas from image- and flow-driven methods in the isotropic setting can be found in the work of Alvarez *et al.* [AELS99]. They proposed the regulariser

$$\begin{aligned} V_{\mathrm{IIF}}(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) &:= g\big(|\boldsymbol{\nabla}_2 f|^2\big) \ \Psi_V\big(|\boldsymbol{\nabla}_2 u|^2 + |\boldsymbol{\nabla}_2 v|^2\big) \\ &= g(\mathrm{tr}\ \boldsymbol{S}_0) \ \Psi_V\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\big) \ , \end{aligned} \tag{4.69}$$

with the corresponding diffusion tensor

$$\begin{aligned} \boldsymbol{D}_{\mathrm{IIF}} &= g(\mathrm{tr}\ \boldsymbol{S}_0) \, \Psi_V'(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2) \, \boldsymbol{I} \\ &= \quad g(\mathrm{tr}\ \boldsymbol{S}_0) \cdot \Psi_V'\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\big) \cdot \boldsymbol{e}_x \boldsymbol{e}_x^\top \\ &\quad + g(\mathrm{tr}\ \boldsymbol{S}_0) \cdot \Psi_V'\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\big) \cdot \boldsymbol{e}_y \boldsymbol{e}_y^\top \ . \end{aligned} \tag{4.70}$$

As one can see, the smoothing is still reduced at every image edge via the decreasing function $g(\operatorname{tr} \boldsymbol{S}_0)$. Thus, one obtains similar oversegmentation artefacts as purely image-driven methods, and the proposed strategy fails to combine the advantages of image- and flow-driven strategies.

In the work of Werlberger at al. [WTP+09], a subquadratic penaliser function is used within an anisotropic image-driven regulariser, thereby combining concepts from anisotropic image-driven and isotropic flow-driven methods. The corresponding regulariser is given by

$$V_{\mathrm{AIF}_1}(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) := \Psi_V\Big((\boldsymbol{\nabla}_2 u)^\top \; \boldsymbol{Q}(\boldsymbol{\nabla}_2 f) \; \boldsymbol{\nabla}_2 u\Big) + \Psi_V\Big((\boldsymbol{\nabla}_2 v)^\top \; \boldsymbol{Q}(\boldsymbol{\nabla}_2 f) \; \boldsymbol{\nabla}_2 v\Big) \quad , \tag{4.71}$$

with the tensor

$$\boldsymbol{Q}(\boldsymbol{\nabla}_2 f) := g\left(|\boldsymbol{\nabla}_2 f|^\beta\right) \; \frac{\boldsymbol{\nabla}_2 f}{|\boldsymbol{\nabla}_2 f|} \frac{(\boldsymbol{\nabla}_2 f)^\top}{|\boldsymbol{\nabla}_2 f|} + \frac{(\boldsymbol{\nabla}_2 f)^\perp}{|\boldsymbol{\nabla}_2 f|} \frac{(\boldsymbol{\nabla}_2 f)^{\perp^\top}}{|\boldsymbol{\nabla}_2 f|} \quad . \tag{4.72}$$

Similar to the rewriting of the Nagel and Enkelmann regulariser, we can express the regulariser from (4.71) in our general framework as

$$V_{\mathrm{AIF}_1}(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) := \Psi_V\Big(g(\operatorname{tr} \boldsymbol{S}_0)\, u_{\boldsymbol{s}_1^0}^2 + u_{\boldsymbol{s}_2^0}^2\Big) + \Psi_V\Big(g(\operatorname{tr} \boldsymbol{S}_0)\, v_{\boldsymbol{s}_1^0}^2 + v_{\boldsymbol{s}_2^0}^2\Big) \quad , \tag{4.73}$$

where we set $\beta = 2$. For this regulariser, we obtain two different diffusion tensors, one for each of the Euler-Lagrange equations. For $p \in \{u, v\}$, they read as

$$\begin{aligned} \boldsymbol{D}_{\mathrm{AIF}_1}^p &= \Psi_V'\Big(g(\operatorname{tr} \boldsymbol{S}_0)\, p_{\boldsymbol{s}_1^0}^2 + p_{\boldsymbol{s}_2^0}^2\Big) \; \boldsymbol{Q}(\boldsymbol{\nabla}_2 f) && (4.74)\\ &= \Psi_V'\Big(g(\operatorname{tr} \boldsymbol{S}_0)\, p_{\boldsymbol{s}_1^0}^2 + p_{\boldsymbol{s}_2^0}^2\Big)\, g(\operatorname{tr} \boldsymbol{S}_0)\, \boldsymbol{s}_1^0 \; \left(\boldsymbol{s}_1^0\right)^\top && (4.75)\\ &\quad + \Psi_V'\Big(g(\operatorname{tr} \boldsymbol{S}_0)\, p_{\boldsymbol{s}_1^0}^2 + p_{\boldsymbol{s}_2^0}^2\Big) \; \boldsymbol{s}_2^0 \; \left(\boldsymbol{s}_2^0\right)^\top \quad . \end{aligned}$$

As it turns out, the anisotropy of this regulariser is solely determined by the function $g(\operatorname{tr} \boldsymbol{S}_0)$, which only depends on the image gradient. As seen for flow-driven methods, it is, however, necessary to reduce the smoothing strength w.r.t. flow gradients to overcome oversegmentation problems. Thus, also this regulariser does not allow to combine the advantages of image- and flow driven methods.

The first method that actually allows to combine the advantages of image- and flow-driven methods goes back to the discrete approach of Sun et al. [SRLB08]. A continuous version of their regulariser can be written as

$$V_{\mathrm{AIF}_2}(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) := \Psi_V\big(u_{\boldsymbol{s}_1}^2\big) + \Psi_V\big(v_{\boldsymbol{s}_1}^2\big) + \Psi_V\big(u_{\boldsymbol{s}_2}^2\big) + \Psi_V\big(v_{\boldsymbol{s}_2}^2\big) \quad . \tag{4.76}$$

Note that in contrast to earlier approaches [NE86, WTP+09] who considered $(\boldsymbol{\nabla}_2 f)^\perp / |\boldsymbol{\nabla}_2 f|$ to obtain directional information of image structures, the eigenvectors $\boldsymbol{s}_i$ of the structure

tensor $\boldsymbol{S}_\rho$ are here considered to obtain a more robust direction estimation. We again obtain two diffusion tensors, that for $p \in \{u, v\}$ read as

$$\boldsymbol{D}^p_{\mathrm{AIF}_2} = \Psi'_V\big(p^2_{\boldsymbol{s}_1}\big)\ \boldsymbol{s}_1\,\boldsymbol{s}_1^\top + \Psi'_V\big(p^2_{\boldsymbol{s}_2}\big)\ \boldsymbol{s}_2\,\boldsymbol{s}_2^\top\ . \tag{4.77}$$

We observe that these tensors allow to obtain the desired behaviour: The regularisation direction is adapted to the image structure directions $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, whereas the magnitude of the regularisation depends on the flow contrast encoded in $p^2_{\boldsymbol{s}_i} = (\boldsymbol{s}_i^\top \boldsymbol{\nabla}_2 p)^2$. As a result, one obtains the same sharp flow edges as image-driven methods but does not suffer from oversegmentation problems.

4.3.2 Our Novel Complementary Regulariser

In spite of its sophistication, the anisotropic image- and flow-driven model of Sun *et al.* [SRLB08] given in (4.76) still suffers from a few shortcomings. We introduced in [10, 11] three amendments that we will discuss now.

Regularisation Tensor. A first remark w.r.t. the model from (4.76) is that the directional information from the structure tensor $\boldsymbol{S}_\rho$ is not consistent with the more advanced data constraints, like the ones of our data term (4.35). It is more natural to take into account directional information provided by the motion tensor (4.24) and to steer the anisotropic regularisation process w.r.t. *"constraint edges"* instead of image edges. To this end we propose to analyse the eigenvectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ of the *regularisation tensor*

$$\boldsymbol{R}_\rho := \sum_{c=1}^{3} K_\rho * \left[\theta_0^c \left(\boldsymbol{\nabla}_2 f^c \left(\boldsymbol{\nabla}_2 f^c\right)^\top\right) + \gamma \left(\theta_x^c \left(\boldsymbol{\nabla}_2 f_x^c \left(\boldsymbol{\nabla}_2 f_x^c\right)^\top\right) + \theta_y^c \left(\boldsymbol{\nabla}_2 f_y^c \left(\boldsymbol{\nabla}_2 f_y^c\right)^\top\right)\right)\right]\ , \tag{4.78}$$

which can be regarded as a generalisation of the structure tensor. Note that the regularisation tensor differs from the motion tensor $\bar{J}^{\mathcal{C}}$ from (4.24) by two aspects: *(i)* It integrates neighbourhood information via the Gaussian convolution. Note that if we also perform a Gaussian convolution of the motion tensor, we would obtain a combined local-global (CLG) data term in the spirit of [BWS05]. Our experiments in Section 4.2.2 have shows that such a modification of our data term is mainly useful to cope with severe noise in the image sequence. *(ii)* The regularisation tensor uses the spatial gradient operator $\boldsymbol{\nabla}_2$ instead of the spatio-temporal operator $\boldsymbol{\nabla}_3$. The latter is due to the spatial regularisation. In Section 4.3.4 we will extend our regulariser to the spatio-temporal domain, yielding a regularisation tensor that also uses the spatio-temporal gradient $\boldsymbol{\nabla}_3$.

Taking the mentioned two modifications into account, we can devise a general strategy for computing the regularisation tensor from the motion tensor. To this end, we denote a *general motion tensor* encoding n constancy assumptions on features p_i $(i = 1, .., n)$ by

$$\boldsymbol{J} = \sum_{i=1}^{n} \gamma_i\ \Theta_i \left(\boldsymbol{\nabla}_3 p_i \left(\boldsymbol{\nabla}_3 p_i\right)^\top\right)\ , \tag{4.79}$$

where $\gamma_i > 0$ are weighting factors and Θ_i denotes normalisation factors that are equal to 1 if we do not perform normalisation. The *general regularisation tensor* corresponding to the general motion tensor (4.79) then reads as

$$\boldsymbol{R}_\rho = K_\rho * \left[\sum_{i=1}^{n} \gamma_i \, \Theta_i \left(\nabla_2 p_i \, (\nabla_2 p_i)^\top \right) \right] \quad . \tag{4.80}$$

Rotational Invariance. The smoothness term V_{AIF_2} from (4.76) lacks the desirable property of rotational invariance because the directional derivatives of u and v in the eigenvector directions are penalised separately. We propose to jointly penalise the directional derivatives, yielding

$$V_{\mathrm{AIF}-\boldsymbol{R}_\rho-\mathrm{RI}}(\nabla_2 u, \nabla_2 v) := \Psi_V\Big(u_{\boldsymbol{r}_1}^2 + v_{\boldsymbol{r}_1}^2\Big) + \Psi_V\Big(u_{\boldsymbol{r}_2}^2 + v_{\boldsymbol{r}_2}^2\Big) \quad , \tag{4.81}$$

where we also used the eigenvectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ of the regularisation tensor.

Single Robust Penalisation. The above regulariser $V_{\mathrm{AIF}-\boldsymbol{R}_\rho-\mathrm{RI}}$ performs a *twofold robust penalisation* in both eigenvector directions. However, we have seen that the data term mainly constraints the flow in the direction of the largest eigenvalue of the regularisation tensor, i.e. in $\boldsymbol{r}_1$-direction. We hence propose a *single robust penalisation* in $\boldsymbol{r}_1$-direction. In the orthogonal $\boldsymbol{r}_2$-direction, we opt for a quadratic penalisation yielding a strong filling-in effect of missing information. Incorporating the single robust penalisation finally yields our *complementary regulariser* [10, 11]

$$V_{\mathrm{CR}}(\nabla_2 u, \nabla_2 v) := \Psi_V\Big(u_{\boldsymbol{r}_1}^2 + v_{\boldsymbol{r}_1}^2\Big) + u_{\boldsymbol{r}_2}^2 + v_{\boldsymbol{r}_2}^2 \quad . \tag{4.82}$$

For the penaliser function Ψ_V, we propose the to use the Lorentzian as defined in (4.64). For our complementary regulariser, we obtain a single, joint diffusion tensor given by

$$\boldsymbol{D}_{\mathrm{CR}} = \Psi_V'\big(u_{\boldsymbol{r}_1}^2 + v_{\boldsymbol{r}_1}^2\big) \cdot \; \boldsymbol{r}_1 \boldsymbol{r}_1^\top + 1 \cdot \boldsymbol{r}_2 \boldsymbol{r}_2^\top \quad , \tag{4.83}$$

with Ψ_V' as given in (4.65). The derivation of this diffusion tensor is presented in Appendix A.

Discussion. To understand the advantages of our complementary regulariser compared to the closest existing regulariser (4.76), we compare our joint diffusion tensor from (4.83) to its counterparts in (4.77). In this way, the following innovations become apparent: *(i)* The smoothing direction is adapted to constraint edges instead of image edges, as the eigenvectors of the regularisation tensor $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are used instead of the eigenvectors of the structure tensor. *(ii)* We achieve rotational invariance by coupling the two flow components in the argument of Ψ_V'. *(iii)* We only reduce the smoothing *across* constraint edges, i.e. in $\boldsymbol{r}_1$-direction. Along them, always a strong diffusion with strength 1 is performed, resembling edge-enhancing anisotropic diffusion [Wei96].

Let us also detail on the benefits of the anisotropic image- and flow-driven smoothing behaviour of our complementary regulariser. Considering the eigenstructure of our joint diffusion tensor, it becomes clear that the smoothing strength across constraint edges is determined by the expression $\Psi_V'(u_{\boldsymbol{r}_1}^2 + v_{\boldsymbol{r}_1}^2)$. Here we can distinguish two scenarios: At a flow edge that corresponds to a constraint edge, the flow gradients will be large and almost parallel to $\boldsymbol{r}_1$. Thus, the argument of the decreasing function Ψ_V' will be large, yielding a reduced diffusion which preserves this important edge. At "deceiving" texture edges in flat flow regions, however, the flow gradients are small which gives a small argument for Ψ_V'. This leads to almost homogeneous diffusion with a pronounced smoothing in both directions and helps to avoid oversegmentation artefacts.

4.3.3 Summary

To conclude this section, Table 4.3 summarises the discussed regularisers rewritten in our general framework. It also compares the way directional information is obtained for anisotropic strategies, and indicates if the regulariser is rotational invariant. Although these regularisers were developed within almost three decades, our rewriting shows their structural similarities. We also wish to note that this comparison extends the work of Weickert and Schnörr [WS01a] who already compared regularisation strategies in a unified framework in 2001.

4.3.4 Extension to a Spatio-Temporal Smoothness Term

All smoothness terms we have discussed so far model the assumption of a *spatially* smooth flow field. As image sequences in general encompass more than two frames, yielding several flow fields, it can make sense to also assume a temporal smoothness of the flow fields. This leads to the so-called spatio-temporal regularisers. A spatio-temporal (ST) version of the general energy functional (4.1) reads as

$$E^{\mathrm{ST}}(u,v) = \int_{\Omega\times[0,T]} \left[M(u,v) + \alpha\, V(\boldsymbol{\nabla}_3 u, \boldsymbol{\nabla}_3 v)\right] \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \ . \tag{4.84}$$

Compared to the spatial energy (4.1) we note an additional integration over the time interval $[0,T]$ and that the smoothness term now depends on the spatio-temporal flow gradient $\boldsymbol{\nabla}_3$. The data term remains unchanged. Extending spatial regularisers to the spatio-temporal domain is thus very simple, we just have to replace the spatial gradient operator $\boldsymbol{\nabla}_2$ by its spatio-temporal counterpart $\boldsymbol{\nabla}_3$. For an homogeneous regulariser

$$V_{\mathrm{H}}(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) = |\boldsymbol{\nabla}_2 u|^2 + |\boldsymbol{\nabla}_2 v|^2 \ , \tag{4.85}$$

one for example obtains the spatio-temporal counterpart as

$$V_{\mathrm{H}}(\boldsymbol{\nabla}_3 u, \boldsymbol{\nabla}_3 v) = |\boldsymbol{\nabla}_3 u|^2 + |\boldsymbol{\nabla}_3 v|^2 \ . \tag{4.86}$$

Table 4.3: Comparison of regularisation strategies. The vectors $\boldsymbol{s}_i$ and $\boldsymbol{r}_i$ denote the eigenvectors of the structure tensor $\boldsymbol{S}_\rho$ from (4.43) and the vectors regularisation tensor $\boldsymbol{R}_\rho$ from (4.78), respectively. The next to last column names the tensor that is analysed to obtain directional information for anisotropic strategies, and the last column indicates if the corresponding regulariser is rotational invariant

Strategy	Regulariser V	Directional Adaptation	Rotational Invariance
Homogeneous [HS81]	$u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2$	—	✓
Isotropic image-driven [AELS99]	$g(\operatorname{tr} \boldsymbol{S}_0)\ \left(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\right)$	—	✓
Anisotropic image-driven [NE86]	$u_{\boldsymbol{s}_2^0}^2 + v_{\boldsymbol{s}_2^0}^2$, for $\kappa \to 0$	$\boldsymbol{S}_0$	✓
Isotropic flow-driven [SH89]	$\Psi_V\left(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\right)$	—	✓
Isotropic image- and flow-driven [AELS99]	$g(\operatorname{tr} \boldsymbol{S}_0)\ \Psi_V\left(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2 + v_{\boldsymbol{s}_1}^2 + v_{\boldsymbol{s}_2}^2\right)$	—	✓
Anisotropic image and flow-driven [WTP+09]	$\Psi_V\left(g(\operatorname{tr} \boldsymbol{S}_0)\, u_{\boldsymbol{s}_1^0}^2 + u_{\boldsymbol{s}_2^0}^2\right) + \Psi_V\left(g(\operatorname{tr} \boldsymbol{S}_0)\, v_{\boldsymbol{s}_1^0}^2 + v_{\boldsymbol{s}_2^0}^2\right)$	$\boldsymbol{S}_0$	—
Anisotropic image and flow-driven [SRLB08]	$\Psi_V\left(u_{\boldsymbol{s}_1}^2\right) + \Psi_V\left(v_{\boldsymbol{s}_1}^2\right) + \Psi_V\left(u_{\boldsymbol{s}_2}^2\right) + \Psi_V\left(v_{\boldsymbol{s}_2}^2\right)$	$\boldsymbol{S}_\rho$	—
Anisotropic complementary image- and flow-driven [10, 11]	$\Psi_V\left(u_{\boldsymbol{r}_1}^2 + v_{\boldsymbol{r}_1}^2\right) + u_{\boldsymbol{r}_2}^2 + v_{\boldsymbol{r}_2}^2$	$\boldsymbol{R}_\rho$	✓

For our complementary regulariser from (4.82), we first have to define the spatio-temporal regularisation tensor

$$\begin{aligned} \boldsymbol{R}_\rho^{\mathrm{ST}} &:= \sum_{c=1}^{3} \Big[\theta_0^c \left(\boldsymbol{\nabla}_3 f^c \left(\boldsymbol{\nabla}_3 f^c \right)^\top \right) \\ &\quad + \gamma \left(\theta_x^c \left(\boldsymbol{\nabla}_3 f_x^c \left(\boldsymbol{\nabla}_3 f_x^c \right)^\top \right) + \theta_y^c \left(\boldsymbol{\nabla}_3 f_y^c \left(\boldsymbol{\nabla}_3 f_y^c \right)^\top \right) \right) \Big] \end{aligned} \tag{4.87}$$

$$= K_\rho * \bar{\boldsymbol{J}}^{\mathcal{C}} \ . \tag{4.88}$$

For $\rho = 0$ it is identical to the motion tensor $\bar{\boldsymbol{J}}^{\mathcal{C}}$ from (4.24). Note in this context that the Gaussian convolution with K_ρ is now performed in the spatio-temporal domain, which also holds for the presmoothing of the image sequence. The spatio-temporal regularisation tensor is a 3×3 matrix that possesses three orthonormal eigenvectors $\boldsymbol{r}_1, \boldsymbol{r}_2$ and $\boldsymbol{r}_3$. With their help, we define the spatio-temporal complementary regulariser

$$V_{\mathrm{CR}}(\boldsymbol{\nabla}_3 u, \boldsymbol{\nabla}_3 v) := \Psi_V\Big(u_{\boldsymbol{r}_1}^2 + v_{\boldsymbol{r}_1}^2 \Big) + u_{\boldsymbol{r}_2}^2 + v_{\boldsymbol{r}_2}^2 + u_{\boldsymbol{r}_3}^2 + v_{\boldsymbol{r}_3}^2 \ , \tag{4.89}$$

with its spatio-temporal diffusion tensor

$$\boldsymbol{D}_{\mathrm{CR}}^{\mathrm{ST}} = \Psi_V'\big(u_{\boldsymbol{r}_1}^2 + v_{\boldsymbol{r}_1}^2 \big) \ \boldsymbol{r}_1 \boldsymbol{r}_1^\top + \boldsymbol{r}_2 \boldsymbol{r}_2^\top + \boldsymbol{r}_3 \boldsymbol{r}_3^\top \ . \tag{4.90}$$

4.3.5 Experiments

In the following experiments, we compare the discussed regularisation strategies.

Comparison with Other Regularisers. In Figure 4.6, we again use our *Snail* sequence to compare our complementary regulariser from (4.82) with other regularisers: *(i)* A basic homogeneous regulariser [HS81]. *(ii)* The anisotropic image-driven regulariser of Nagel and Enkelmann [NE86]. *(iii)* The popular TV regulariser; see (4.60) and (4.62). *(iv)* The anisotropic image and flow-driven regulariser from [SRLB08], which built the basis of our complementary regulariser. Here, we use a rotationally invariant formulation that can be obtained from (4.81) by replacing the eigenvectors $\boldsymbol{r}_i$ of the regularisation tensor by the eigenvectors $\boldsymbol{s}_i$ of the structure tensor. As data term, we use our final data term from (4.35).

Let us compare the obtained results: A homogeneous regularisation (Figure 4.6 (d)) clearly gives a oversmoothed result. Using an image-driven strategy (Figure 4.6 (e)) gives relatively sharp edges, but suffers from oversegmentation. Note that the latter leads to the cloud-like artefacts because of the strongly textured background. With a flow-driven TV regulariser (Figure 4.6 (f)), we obtain blurred and badly localised flow edges. Using the image-and flow-driven regulariser from [SRLB08] (Figure 4.6 (g)), the flow edges are already more appealing, but unpleasant staircasing artefacts deteriorate the result. This can be explained by the twofold robust penalisation that is performed in this regulariser. Our result is depicted in Figure 4.6 (h) and obviously gives an appealing flow field without severe degradations.

Table 4.4: Smoothness weight α and AAE measures for our spatio-temporal method on the *Marble* sequence, see Figure 4.7. All results used the fixed parameters $\sigma = 0.5, \gamma = 0.5, \rho = 1.0, \lambda = 0.1, h_3 = 1.5, \eta = 0.5$. When using more than two frames, the convolutions with K_σ and K_ρ are performed in the spatio-temporal domain

Number of frames (from – to)	2 (16–17)	4 (15–18)	6 (14–19)	8 (13–20)
Smoothn. weight α	75.0	50.0	50.0	50.0
AAE	4.85°	2.63°	**1.86°**	2.04°

Table 4.5: Smoothness weight α and AAE measures for our spatio-temporal method on the *Yosemite* sequence from Middlebury. All results used the fixed parameters $\sigma = 1.0, \gamma = 20.0, \rho = 1.5, \lambda = 0.1, h_3 = 1.0, \eta = 0.5$. Here, better results could be obtained when disabling the temporal presmoothing

Number of frames (from – to)	2 (10–11)	4 (9–12)	6 (8–13)	8 (7–14)
Smoothn. weight α	2000.0	1000.0	1000.0	1000.0
AAE	1.65°	1.16°	1.05°	**1.01°**

Optic Flow in the Spatio-Temporal Domain. Let us now turn to the spatio-temporal extension of our complementary smoothness term. Concerning the test sequences to use, one could consider the data sets from the Middlebury benchmark [BSL+10]. The latter mostly consist of 8 frames and thus a spatio-temporal method would in general be applicable. However, the displacements between two subsequent frames are often rather large and thus flow vectors at the same pixel location in different flow fields do not necessarily correspond to the same object. This leads to a violation of the assumption of a temporally smooth flow field and consequently spatio-temporal methods do not improve the results, which has as already been noted in [WTP+09]. In our experiments, we use the *Marble* sequence (available at `http://i21www.ira.uka.de/image_sequences/`) and the *Yosemite* sequence from the Middlebury datasets. These sequences exhibit relatively small displacements and our spatio-temporal method allows to obtain notably better results, see Figure 4.7 and Tables 4.4 and 4.5. Note that when using more than two frames, a smaller smoothness weight α has to be chosen and that a too large temporal window may also deteriorate the results again.

Temporal Smoothing for Large Displacements. We wish to note that in recent work [6] we successfully incorporated temporal information from several images despite the presence of large displacements. To achieve this, we parametrise all flow fields w.r.t. one reference frame, e.g. the middle frame. In this case, assuming smoothness of flow vectors in "time" direction (at the same pixel location in all flow fields) results in

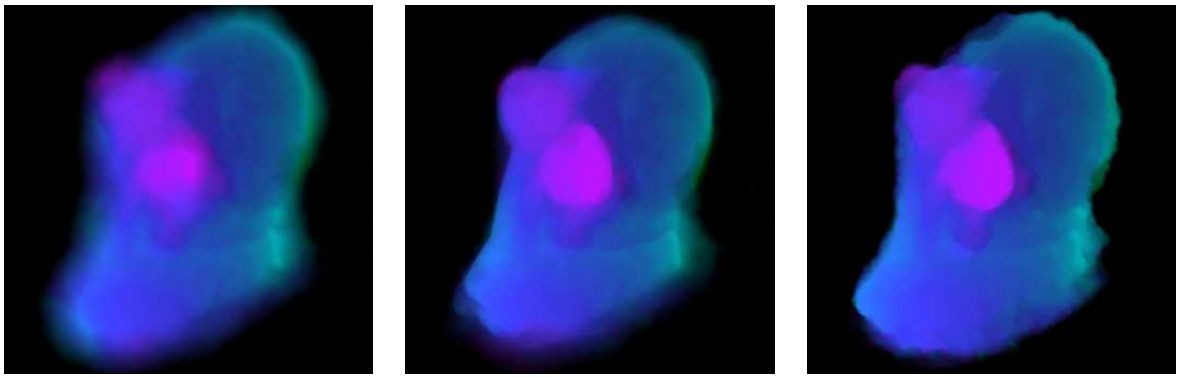
Table 4.6: Error measures (AAE and AEE) and parameter settings for the Middlebury training datasets. We used the fixed parameters $\lambda = 0.1, \eta = 0.95$. Corresponding results are shown in Figures 4.9 and 4.10, respectively

	AAE	AEE	α	σ	γ	ρ
Rubberwhale	2.77°	0.083	850.0	0.3	20.0	2.0
Hydrangea	1.78°	0.146	150.0	0.3	3.0	1.0
Dimetrodon	1.54°	0.079	2600.0	0.7	26.0	2.0
Venus	5.33°	0.344	500.0	0.8	10.0	2.0
Grove2	2.16°	0.151	35.0	0.5	0.2	1.0
Grove3	4.87°	0.487	20.0	0.5	0.1	1.0
Urban2	2.49°	0.245	126.0	0.5	1.0	1.5
Urban3	2.77°	0.298	75.0	0.7	1.0	1.5

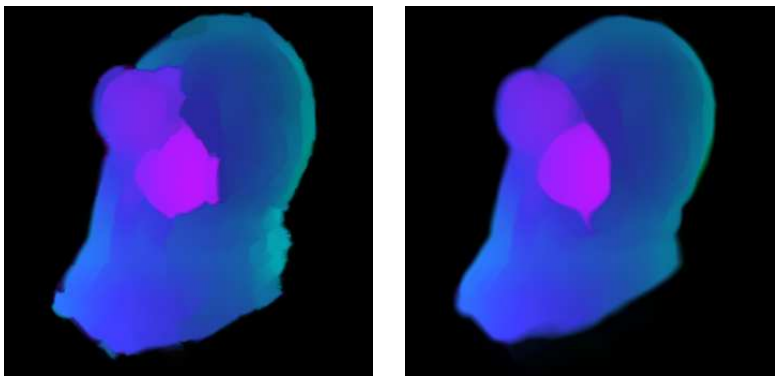
a smoothing along motion *trajectories* which makes sense for arbitrarily large displacements. Furthermore, our parametrisation allows to add a temporal smoothing behaviour to the *spatial* regularisers by assuming that all flow fields have joint flow edges at the same locations. Our experiments on the Middlebury sequences showed that the described *trajectorial model* yields significant improvements compared to a purely spatial method, also in the presence of large displacements, see Figure 4.8.

Results for the Middlebury Training Sequences. We conclude this experimental section by presenting results for our complementary optic flow method [10, 11] on the Middlebury training sequences with ground truth. Recall that the complementary optic flow method uses the robust data term from (4.35) in combination with the complementary smoothness term from (4.82). The manually tuned parameter settings and the AAE and AAE measures can be found in Table 4.6, whereas the flow fields are compared to the ground truth in Figure 4.9 and 4.10, respectively. As one can see, our obtained flow fields closely resemble the ground truth. However, minor problems arising from incorporating image information in the smoothness term are visible at a few locations, e.g. at the bottom of the *Grove 2* result (Figure 4.10 (c)) as well as at the lower part of the leftmost building in the *Urban 3* result (Figure 4.10 (l)). Furthermore, some tiny flow details are smoothed out, e.g. at the leafs in the *Grove 2* and the *Grove 3* result (Figure 4.10 (c) and (f)). As recently shown by Xu *et al.* [XJM10], the latter problem could be resolved by using a more sophisticated warping strategy.

(a) First frame **(b)** Zoom into (a) **(c)** Zoom into second frame

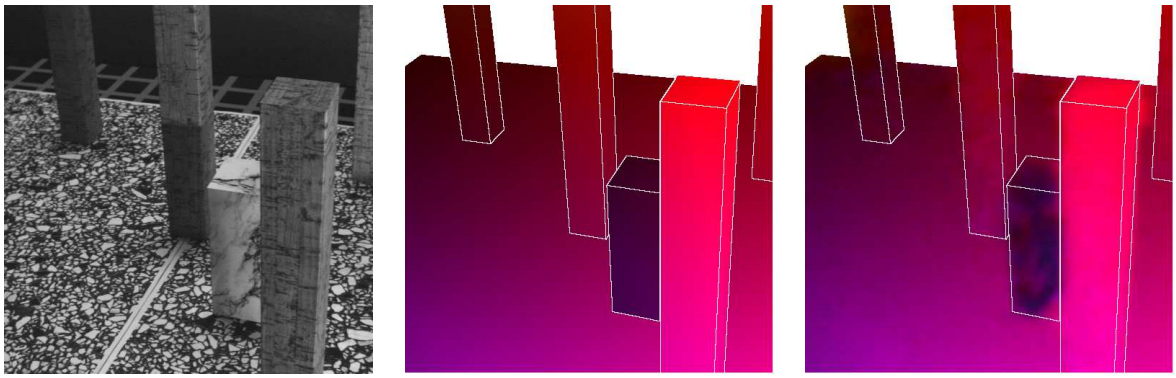


(d) Homogeneous **(e)** Anisotropic image-driven **(f)** Isotropic flow-driven

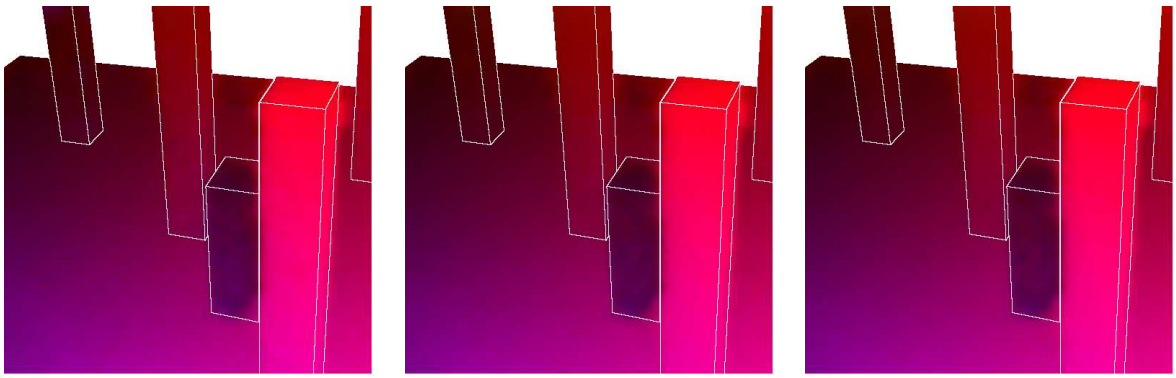


(g) Anisotropic image- and flow-driven **(h)** Anisotropic complementary

Figure 4.6: Results for our *Snail sequence* with different smoothness terms. *First row, from left to right:* **(a)** First frame. **(b)** Zoom in marked region of first frame. **(c)** Same for second frame. *Second row, from left to right:* **(d)** Flow field in marked region, with homogeneous regularisation [HS81] ($\alpha = 200.0$). **(e)** Same for anisotropic image-driven regularisation [NE86] ($\alpha = 500.0, \kappa = 0.1$). **(f)** Same for isotropic flow-driven TV regularisation [Sch94, SH89] ($\alpha = 50.0, \varepsilon = 0.001$). *Third row, from left to right:* **(g)** Same for anisotropic image- and flow-driven regularisation [SRLB08] ($\alpha = 2000.0, \rho = 4.0, \lambda = 0.1$). **(h)** Same for our anisotropic complementary regulariser [10, 11] ($\alpha = 2000.0, \rho = 4.0, \lambda = 0.1$). All results used the fixed parameters $\sigma = 0.5, \gamma = 20.0, \eta = 0.95$.



(a) Reference frame **(b)** Ground truth **(c)** Flow (2 frames)



(d) Flow (4 frames) **(e)** Flow (6 frames) **(f)** Flow (8 frames)

Figure 4.7: Results for the *Marble* sequence with our spatio-temporal method. *First row, from left to right:* **(a)** Reference frame (frame 16). **(b)** Ground truth (white pixels mark locations where no ground truth is available). **(c)** Result using 2 frames (16–17). *Second row, from left to right:* **(d)** Same for 4 frames (15–18). **(e)** Same for 6 frames (14–19). **(f)** Same for 8 frames (13–20).

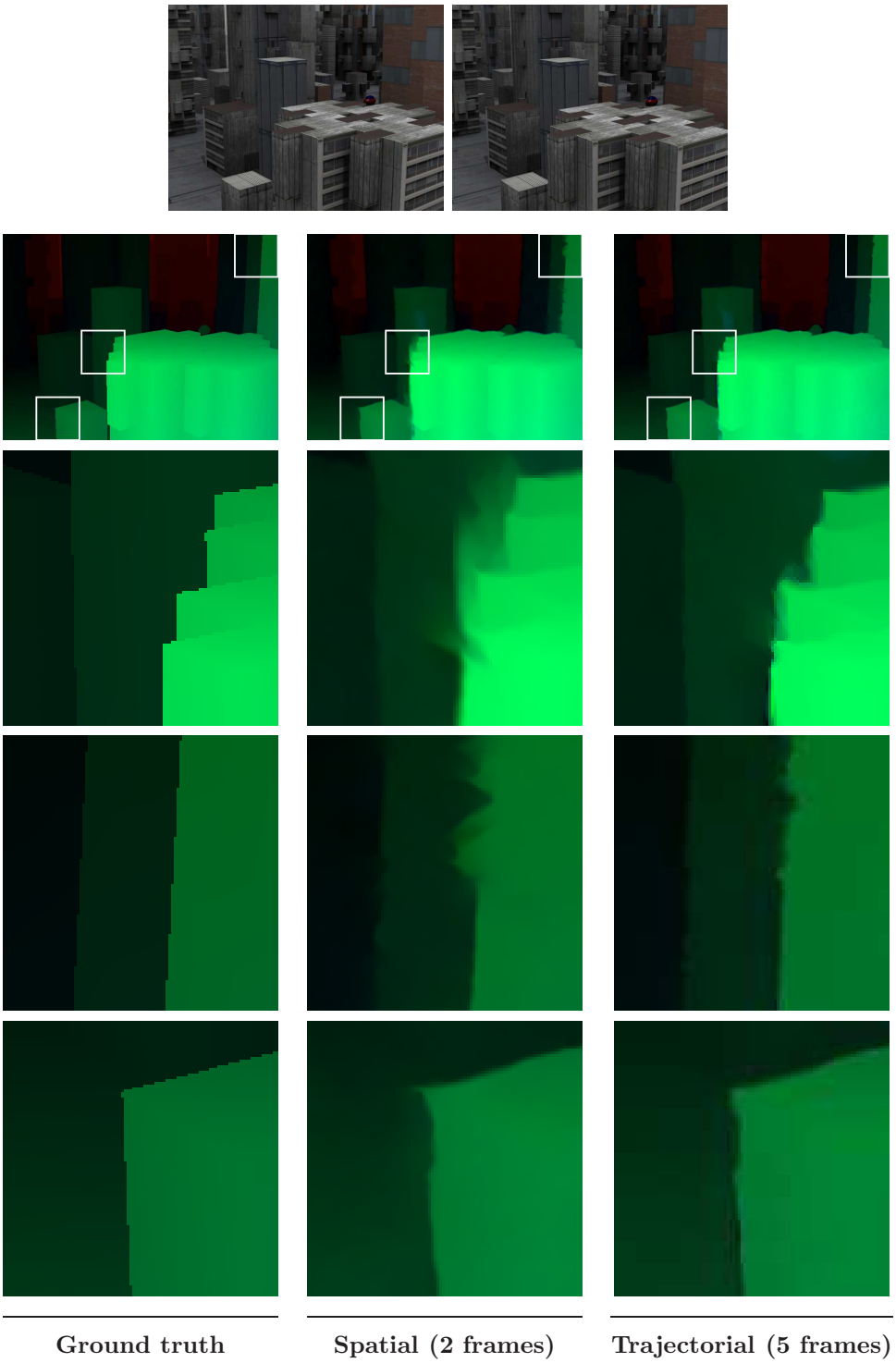


Figure 4.8: Comparison of a spatial method (2 frames) [10, 11] to a trajectorial method (5 frames) [6] on the *Urban2* sequence. *First row, from left to right:* First and second image. *Second to fifth row, from left to right:* Ground truth, spatial method (AAE = 2.75°), trajectorial method (AAE = 2.00°). *Second to fifth row, from top to bottom:* Flow field and three magnifications.

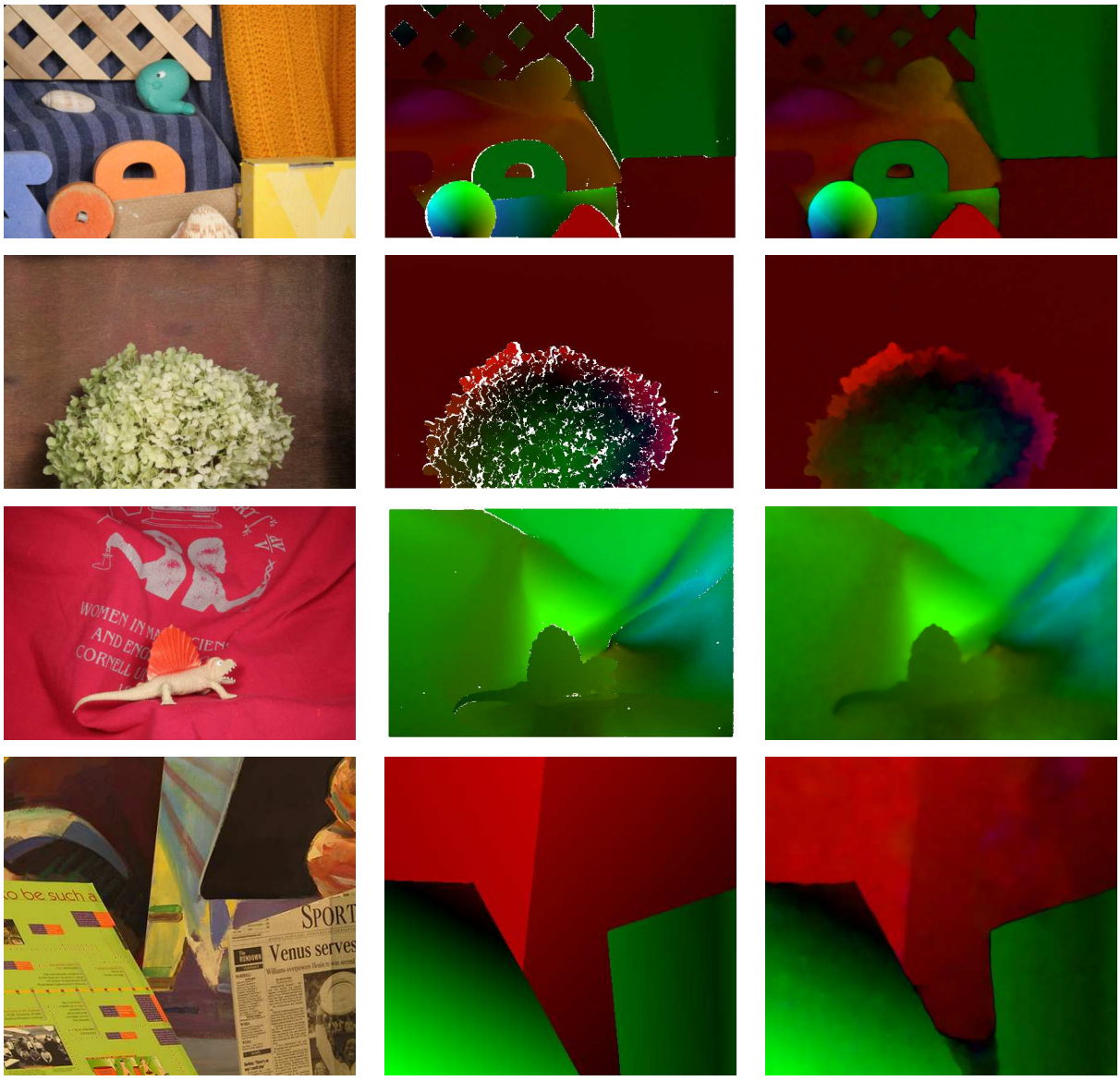


Figure 4.9: Results for our complementary optic flow method [10, 11] on the Middlebury training sequences with ground truth. *First column:* Reference frame. *Second column:* Ground truth (white pixels mark locations where no ground truth is given). *Third column:* Result with our method. *From top to bottom: Rubberwhale*, *Dimetrodon*, *Hydrangea* and *Venus* sequence.

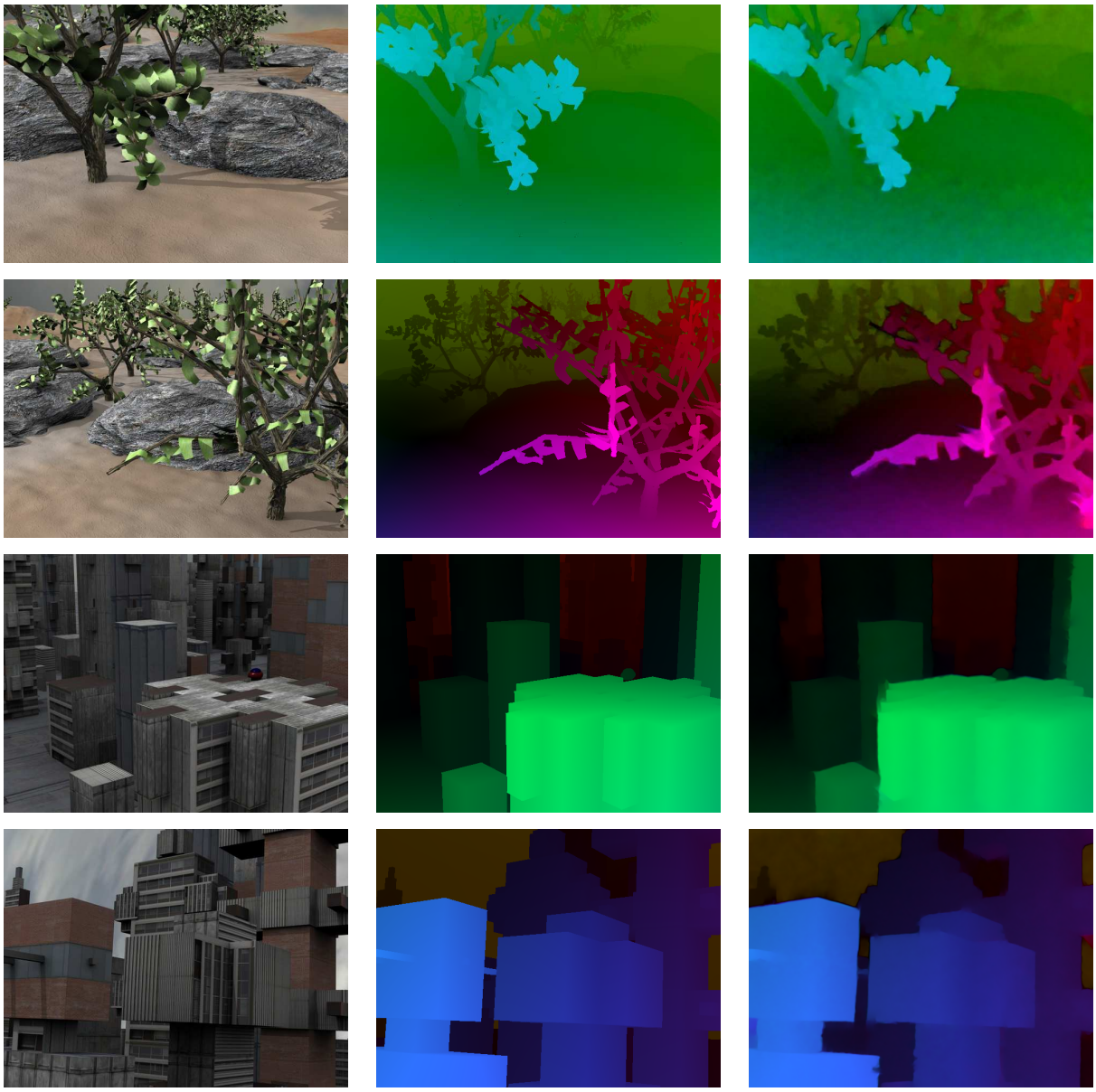


Figure 4.10: Results for our complementary optic flow method [10, 11] on the Middlebury training sequences with ground truth. *First column:* Reference frame. *Second column:* Ground truth (white pixels mark locations where no ground truth is given). *Third column:* Result with our method. *From top to bottom: Grove2*, *Grove3*, *Urban2* and *Urban3* sequence.

4.3.6 Smoothness Terms for Stereo

In the variational stereo context, similar regularisation strategies as for optic flow have been proposed, which we will review in this subsection. Recall that in the considered orthoparallel case, we will only experience horizontal displacements u.

Image-driven Regularisation. In the isotropic setting, an image-driven regulariser was proposed by Kim and Sohn [KS03]. It can be written as

$$\begin{aligned} V_{\mathrm{II}}(\boldsymbol{\nabla}_2 u) &:= g\big(|\boldsymbol{\nabla}_2 f|^2\big)\ |\boldsymbol{\nabla}_2 u|^2 \\ &= g(\operatorname{tr} \boldsymbol{S}_0)\ \big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2\big) \quad , \end{aligned} \tag{4.91}$$

and leads to the same diffusion tensor as in (4.53).

Anisotropic image-driven smoothness terms for stereo became popular through the works of Mansouri *et al.* [MMK98] and Alvarez *et al.* [ADSW02]. They proposed regularisers of the form

$$V_{\mathrm{AI}}(\boldsymbol{\nabla}_2 u) = \Psi_V\left(\boldsymbol{\nabla}_2 u^\top \boldsymbol{P}(\boldsymbol{\nabla}_2 f)\ \boldsymbol{\nabla}_2 u\right) \quad , \tag{4.92}$$

with a projection matrix $\boldsymbol{P}$ similar as in the Nagel and Enkelmann regulariser (4.55).

Disparity-driven Regularisation. As seen for flow-driven optic flow regularisers, disparity-driven strategies allow to remedy oversegmentation problems. Isotropic image-driven strategies were introduced to the stereo community by Slesareva *et al.* [SBW05] and Ben-Ari and Sochen [BAS07]. They used the regulariser

$$\begin{aligned} V_{\mathrm{ID}}(\boldsymbol{\nabla}_2 u) &:= \Psi_V\left(|\boldsymbol{\nabla}_2 u|^2\right) \\ &= \Psi_V\left(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2\right) \quad , \end{aligned} \tag{4.93}$$

leading to the diffusion tensor

$$\begin{aligned} \boldsymbol{D}_{\mathrm{ID}} &= \Psi_V'\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2\big)\ \boldsymbol{I} \\ &= \Psi_V'\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2\big) \cdot \boldsymbol{e}_x \boldsymbol{e}_x^\top + \Psi_V'\big(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2\big) \cdot \boldsymbol{e}_y \boldsymbol{e}_y^\top \quad . \end{aligned} \tag{4.94}$$

Completing the Taxonomy of Stereo Regularisers

So far, all presented stereo regularisers could be obtained in a straightforward manner from their optic flow counterparts. Let us also try this for an anisotropic disparity-driven regulariser by adapting the anisotropic flow-driven regulariser of Weickert and Schnörr [WS01a] from (4.66). A stereo counterpart can be written as

$$V_{\mathrm{AD}}(\boldsymbol{\nabla}_2 u) := \operatorname{tr} \Psi_V\big(\boldsymbol{\nabla}_2 u\,(\boldsymbol{\nabla}_2 u)^\top\big) \quad , \tag{4.95}$$

with the diffusion tensor

$$\begin{aligned} \boldsymbol{D}_{\mathrm{AD}} &= \Psi_V' \left(\boldsymbol{\nabla}_2 u \, (\boldsymbol{\nabla}_2 u)^\top \right) \\ &=: \sum_{i=1}^{2} \Psi_V'(\nu_i) \, \boldsymbol{v}_i \boldsymbol{v}_i^\top \ . \end{aligned} \tag{4.96}$$

In the optic flow case, the tensor in the argument of Ψ_V was given by $\boldsymbol{\nabla}_2 u \, (\boldsymbol{\nabla}_2 u)^\top + \boldsymbol{\nabla}_2 v \, (\boldsymbol{\nabla}_2 v)^\top$, resulting in a coupling of the two flow components. As a consequence, the desired anisotropic behaviour can be ensured as the eigenvectors of the tensor are in general not parallel to the gradients of both flow components.

In the stereo case, however, the resulting tensor $\boldsymbol{\nabla}_2 u \, (\boldsymbol{\nabla}_2 u)^\top$ is equal to the structure tensor of the disparity with a zero integration scale. As we have seen in (4.44) and the following, the eigenstructure of this tensor is trivial. In above notation, it can be written as $\nu_1 = |\boldsymbol{\nabla}_2 u|^2, \nu_2 = 0$ and

$$\boldsymbol{v}_1 = \frac{\boldsymbol{\nabla}_2 u}{|\boldsymbol{\nabla}_2 u|}, \quad \text{and} \quad \boldsymbol{v}_2 = \frac{\boldsymbol{\nabla}_2 u^\perp}{|\boldsymbol{\nabla}_2 u|} \ . \tag{4.97}$$

With this, let us rewrite the diffusion part of the corresponding Euler-Lagrange equation:

$$\mathrm{div} \left(\boldsymbol{D}_{\mathrm{AD}} \, \boldsymbol{\nabla}_2 u \right) = \mathrm{div} \left(\Psi_V' \left(\boldsymbol{\nabla}_2 u \, (\boldsymbol{\nabla}_2 u)^\top \right) \boldsymbol{\nabla}_2 u \right) \tag{4.98}$$

$$= \mathrm{div} \left(\left(\sum_{i=1}^{2} \Psi_V'(\nu_i) \, \boldsymbol{v}_i \boldsymbol{v}_i^\top \right) \boldsymbol{\nabla}_2 u \right) \tag{4.99}$$

$$= \mathrm{div} \Bigg(\Bigg(\frac{\Psi_V'(|\boldsymbol{\nabla}_2 u|^2)}{|\boldsymbol{\nabla}_2 u|^2} \, \boldsymbol{\nabla}_2 u \boldsymbol{\nabla}_2 u^\top + \frac{\Psi_V'(0)}{|\boldsymbol{\nabla}_2 u|^2} \, \boldsymbol{\nabla}_2 u^\perp (\boldsymbol{\nabla}_2 u^\perp)^\top \Bigg) \boldsymbol{\nabla}_2 u \Bigg) \tag{4.100}$$

$$\overset{(*)}{=} \mathrm{div} \left(\frac{\Psi_V'(|\boldsymbol{\nabla}_2 u|^2)}{|\boldsymbol{\nabla}_2 u|^2} \, |\boldsymbol{\nabla}_2 u|^2 \, \boldsymbol{\nabla}_2 u + \mathbf{0} \right) \tag{4.101}$$

$$= \mathrm{div} \left(\Psi_V'(|\boldsymbol{\nabla}_2 u|^2) \, \boldsymbol{\nabla}_2 u \right) \tag{4.102}$$

$$= \mathrm{div} \left(\Psi_V'(u_{\boldsymbol{s}_1}^2 + u_{\boldsymbol{s}_2}^2) \, \boldsymbol{\nabla}_2 u \right) = \mathrm{div} \left(\boldsymbol{D}_{\mathrm{ID}} \, \boldsymbol{\nabla}_2 u \right) \ , \tag{4.103}$$

where $(*)$ makes use of the facts that $\boldsymbol{\nabla}_2 u^\top \boldsymbol{\nabla}_2 u = |\boldsymbol{\nabla}_2 u|^2$ and $(\boldsymbol{\nabla}_2 u^\perp)^\top \boldsymbol{\nabla}_2 u = 0$. As it turns out, adapting the anisotropic flow-driven regulariser (4.66) to the stereo context results in an *isotropic* disparity-driven behaviour as for the regulariser (4.93) with the diffusion tensor (4.94).

Towards Anisotropic Disparity-driven Regularisation. To model an anisotropic disparity-driven smoothing process for stereo we follow our work in [8]: We refrain from designing a regulariser V_{AD} and directly model the diffusion tensor $\boldsymbol{D}_{\mathrm{AD}}$.

In order to obtain a truly anisotropic behaviour we adapted ideas from edge enhancing anisotropic diffusion (EED) filtering [Wei94]. Specifically, we consider a stereo variant of the structure tensor (4.43), given by

$$\boldsymbol{S}_\rho\left(\boldsymbol{\nabla}_2 u_\sigma\right) := K_\rho * \left(\boldsymbol{\nabla}_2 u_\sigma \boldsymbol{\nabla}_2 u_\sigma^\top\right) \quad , \tag{4.104}$$

where $u_\sigma := K_\sigma * u$ denotes a Gaussian smoothed version of the disparity u. Compared to the previously considered tensor $\boldsymbol{\nabla}_2 u \boldsymbol{\nabla}_2 u^\top$, we observe two innovations: *(i)* The disparity u is regularised by the Gaussian convolution with standard deviation σ, and *(ii)* neighbourhood information is integrated by convolving the tensor entries with a Gaussian kernel of standard deviation ρ. Note that regularisation of the unknown u by Gaussian convolution with noise scale σ was first proposed in the context of nonlinear diffusion to reduce staircaising artefacts and problems with noise [CLMC92]. Despite the fact that $\boldsymbol{\nabla}_2 u_\sigma$ is a useful edge detector, the problem still remains that it is sensitive under noise for small σ, while an increased σ can lead to undesired cancellation effects. This can be overcome by an additional convolution of the tensor entries with the integration scale ρ.

Our modified structure tensor $\boldsymbol{S}_\rho(\boldsymbol{\nabla}_2 u_\sigma)$ is a symmetric, positive semi-definite matrix with two orthonormal eigenvectors $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ which give the directions of the local disparity structure. The corresponding eigenvalues $\mu_1 \geq \mu_2 \geq 0$, give the average contrast along these directions. Consequently, we propose to use the following diffusion tensor which takes into account the structural information contained in $\boldsymbol{S}_\rho(\boldsymbol{\nabla}_2 u_\sigma)$:

$$\boldsymbol{D}_{\mathrm{AD}} := \Psi_V'\left(\boldsymbol{S}_\rho\left(\boldsymbol{\nabla}_2 u_\sigma\right)\right) = \sum_{i=1}^{2} \Psi_V'(\mu_i)\, \boldsymbol{s}_i\, \boldsymbol{s}_i^\top \quad . \tag{4.105}$$

For the diffusivity Ψ_V' we propose to use the Perona-Malik diffusivity from (4.65) as it is known to make backward diffusion possible and thereby enhance edges even more.

Let us now show that our method indeed exhibits the described anisotropic behaviour:

- In *flat disparity regions:*
  $\mu_1 \approx \mu_2 \approx 0 \quad \Rightarrow \quad \Psi_V'(\mu_1) \approx 1,\ \Psi_V'(\mu_2) \approx 1.$
  This leads to a homogeneous smoothing in both directions.

- At a *straight disparity edge in $\boldsymbol{s}_2$-direction:*
  $\mu_1 \gg \mu_2 \approx 0 \quad \Rightarrow \quad \Psi_V'(\mu_1) \approx 0,\ \Psi_V'(\mu_2) \approx 1.$
  This leads to an anisotropic smoothing in edge direction, but not across.

- At *disparity corners:*
  $\mu_1 \geq \mu_2 \gg 0 \quad \Rightarrow \quad \Psi_V'(\mu_1) \approx 0,\ \Psi_V'(\mu_2) \approx 0.$
  This prevents smoothing.

Complementary Regularisation for Stereo. Adapting our complementary regulariser from (4.82) to the stereo case can be done in a straightforward manner.

Although the stereo tensor (4.39) uses the gradient operator $\boldsymbol{\nabla}_{xt}$, the regularisation tensor for stereo is identical to the optic flow regularisation tensor $\boldsymbol{R}_\rho$ from (4.78). This choice can be justified as also in the stereo case we want to use directional information from the spatial domain, i.e., from the x- and y-direction, and not from the x- and t-direction. The complementary regulariser for stereo is defined as

$$V_{\mathrm{CR}}(\boldsymbol{\nabla}_2 u) = \Psi_V\Big(u_{\boldsymbol{r}_1}^2\Big) + u_{\boldsymbol{r}_2}^2 \ , \tag{4.106}$$

with the diffusion tensor

$$\boldsymbol{D}_{\mathrm{CR}} = \Psi_V'\big(u_{\boldsymbol{r}_1}^2\big) \cdot \ \boldsymbol{r}_1\boldsymbol{r}_1^\top + 1 \cdot \boldsymbol{r}_2\boldsymbol{r}_2^\top \ . \tag{4.107}$$

4.3.7 Experiments

Similar to our experiment for the optic flow case in Section 4.3.5, Figure 4.6, we now want to compare the performance of the presented stereo regularisers. As test image pair, we chose the popular *Tsukuba* pair (available at `http://bj.middlebury.edu/~schar/stereo/data/tsukuba/`). In Figure 4.11 we show the disparity maps obtained when using a stereo version of our robust data term (4.35) in combination with different regularisation strategies. Table 4.7 additionally summarises the parameter settings and the BPE error measures.

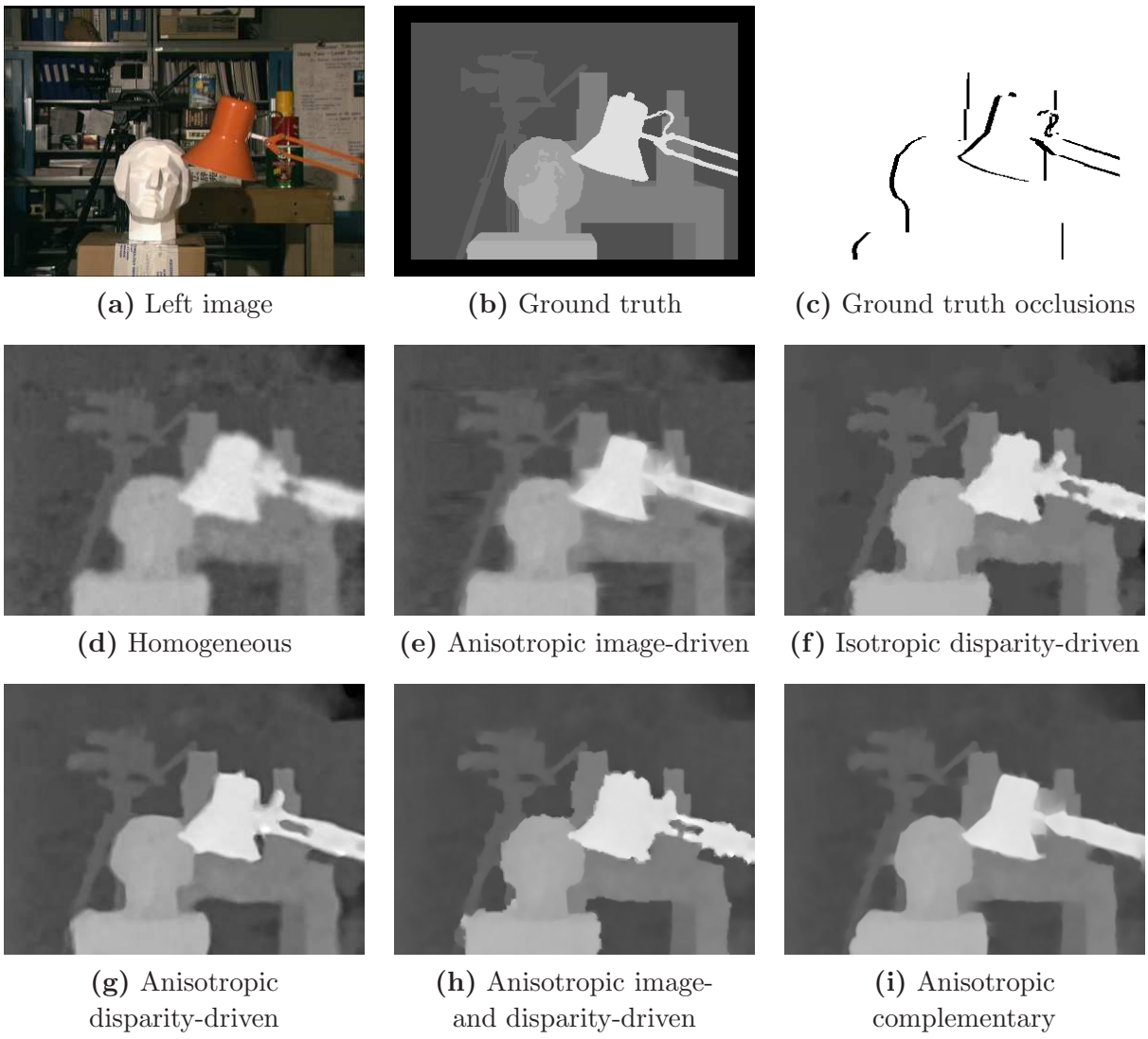
Comparing the results, similar observations as in the optic flow case can be made. A homogeneous regularisation as in [HS81] (Figure 4.11 (d)) gives oversmoothed result. With an anisotropic image-driven regulariser [ADSW02, MMK98] (Figure 4.11 (e)), the disparity edges become sharper, but oversegmentation artefacts deteriorate the result. As an example, consider the artefact arising at the left of the lamp. A disparity-driven TV regulariser [BAS07, SBW05] (Figure 4.11 (f)) remedies oversegmentation problems, but clearly gives delocalised disparity edges. Using our anisotropic disparity-driven regulariser [8] (Figure 4.11 (g)) gives more appealing disparity edges due to the smoothing along edges. To obtain even sharper edges, one can use the image- and disparity-driven regulariser by adapting the smoothness term from [SRLB08] (Figure 4.11 (h)). However, this regulariser gives unpleasantly jagged disparity edges. When using the stereo variant (4.106) of our complementary regulariser [10, 11] (Figure 4.11 (i)), we obtain sharp and well-localised disparity edges.

The above observations are further supported by the bad pixel error (BPE) measures resulting from the different regularisation strategies, see Table 4.7. Finally, we wish to note that all image-driven strategies (Figure 4.11 (e),(h),(i)) perform especially well in the occluded areas, which are shown in Figure 4.11 (c). This observation can be explained as follows: Occlusions are outliers in the data term, as no matching based on data constancy assumptions is possible in occluded regions. Consequently, a robust data term will disable its influence and the smoothness has to fill-in the disparity there.

Table 4.7: Error measures (BPE) and parameter settings for the *Tsukuba* pair, see Figure 4.11. All results used the fixed parameters $\sigma = 0.5, \gamma = 5.0, \eta = 0.95$

Regulariser	BPE	Parameters
Homogeneous [HS81]	11.73	$\alpha = 50.0$
Anisotropic image-driven [ADSW02, MMK98]	10.14	$\alpha = 200.0, \kappa = 0.1$
Isotropic disparity-driven [BAS07, SBW05]	8.57	$\alpha = 10.0, \varepsilon = 0.001$
Anisotropic disparity-driven [8]	8.23	$\alpha = 150.0, \sigma = 1.0, \rho = 2.0, \lambda = 0.1$
Anisotropic image- and disparity-driven [SRLB08]	8.12	$\alpha = 275.0, \rho = 1.5, \lambda = 0.1$
Anisotropic complementary image- and disparity-driven [10, 11]	7.12	$\alpha = 260.0, \rho = 1.5, \lambda = 0.1$

For the filling-in, it is crucial to consider image instead of disparity information, because the disparity estimate is less reliable at occlusions. Thus, image-driven strategies have advantages.



(a) Left image **(b)** Ground truth **(c)** Ground truth occlusions

(d) Homogeneous **(e)** Anisotropic image-driven **(f)** Isotropic disparity-driven

(g) Anisotropic disparity-driven **(h)** Anisotropic image- and disparity-driven **(i)** Anisotropic complementary

Figure 4.11: Results for the *Tsukuba* pair with different smoothness terms. *First row, from left to right:* **(a)** Left image. **(b)** Ground truth disparity. **(c)** Ground truth occlusions. *Second row, from left to right:* **(d)** Disparity map with homogeneous regularisation [HS81]. **(e)** Same for anisotropic image-driven regularisation [MMK98, ADSW02] **(f)** Same for isotropic disparity-driven TV regularisation [SBW05, BAS07]. *Third row, from left to right:* **(g)** Same for our anisotropic disparity-driven regularisation [8]. **(h)** Same for anisotropic image- and disparity-driven regularisation [SRLB08]. **(i)** Same for our anisotropic complementary regulariser [10, 11].

4.4 Automatic Selection of the Smoothness Weight

To achieve favourable results, an appropriate selection of the model parameters is mandatory. Especially for the smoothness weight α that determines the balance between the data and the smoothness term, an appropriate choice is crucial: Too small values lead to noisy flow fields whereas too large values lead to an over-smoothing. An automatic selection of the smoothness weight is thus not only convenient, but especially important in real world applications of optic flow such as robot navigation or driver assistance. Also in less time-critical applications like video re-timing, an automatic selection of model parameters for sure eases the use and boost the acceptance of optic flow-based approaches.

Note that in the mentioned real world scenarios, no ground truth flow is known in general. If this would be the case, a parameter selection would be very simple: One could just select the parameters that give the flow field with the smallest deviation from the ground truth.

4.4.1 A Novel Concept

Our parameter selection approach uses an error measure that allows to judge the quality of a flow field *without* knowing the ground truth. This error measure is based on a novel concept, the *optimal prediction principle* (*OPP*) that we introduced in [10]. The OPP states that the flow field obtained with an optimal smoothness weight allows for the best prediction of the next frames in the image sequence. This makes sense as a too small smoothness weight would lead to an overfit to the first two frames and consequently result in a bad prediction of further frames. For too large smoothness weights, the flow fields would be too smooth and thus also lead to a bad prediction.

Following the OPP, our error measure needs to judge the quality of the prediction achieved with a given flow field. To this end, we evaluate the imposed data constraints between the first and the *third* frame of the image sequence, resulting in an *average data constancy error* (*ADCE*) measure. To compute this measure, we assume that the motion of the scene objects is of more or less constant speed and that it describes linear trajectories within the considered three frames. Under these assumptions, we simply double the flow vectors to evaluate the data constraints between the first and the third frame. Due to the favourable performance of our parameter selection method that is demonstrated in Section 4.4.3, we conjecture that the made assumptions hold for common test sequences. In accordance to this strategy, we can define the ADCE

between frame 1 and 3 as

$$\begin{aligned}\mathrm{ADCE}_{1,3}(\boldsymbol{w}_\alpha) := \frac{1}{|\Omega|} \int_\Omega \Bigg[\sum_{c=1}^{3} \Psi_M\big(\theta_0^c \left(f^c(\boldsymbol{x}+2\boldsymbol{w}_\alpha) - f^c(\boldsymbol{x})\right)^2\big) \\ + \gamma \left(\sum_{c=1}^{3} \Psi_M\Big(\theta_x^c \left(f_x^c(\boldsymbol{x}+2\boldsymbol{w}_\alpha) - f_x^c(\boldsymbol{x})\right)^2 \right. \\ \left. + \theta_y^c \left(f_y^c(\boldsymbol{x}+2\boldsymbol{w}_\alpha) - f_y^c(\boldsymbol{x})\right)^2 \Big) \right) \Bigg] \,\mathrm{d}x\,\mathrm{d}y \ ,\end{aligned} \tag{4.108}$$

where $\boldsymbol{w}_\alpha$ denotes the flow field obtained with a smoothness weight α. The integrand of above expression is (apart from the doubled flow field) a variant of our final data term (4.35) where no linearisation of the constancy assumptions has been performed. To evaluate the images at subpixel locations $f^i(\boldsymbol{x}+2\boldsymbol{w}_\alpha)$ we use Coons patches based on bicubic interpolation [Coo67]. Note that when using the HSV colour space, the normalisation factors Θ^1 of the hue channel have to be computed as in (4.27). Also the difference expression for the hue channel have to be computed accordingly. For example, the expression $|f^1(\boldsymbol{x}+2\boldsymbol{w}) - f^1(\boldsymbol{x})|^2$ has to be replaced by $|(\cos f^1)(\boldsymbol{x}+2\boldsymbol{w}) - (\cos f^1)(\boldsymbol{x})|^2 + |(\sin f^1)(\boldsymbol{x}+2\boldsymbol{w}) - (\sin f^1)(\boldsymbol{x})|^2$.

Determining the Best Parameter. In general, the relation between α and the ADCE is not convex, which excludes the use of gradient descent-like approaches for finding the optimal value of α w.r.t. our error measure.

Similar to the approach in [NS97] we propose a brute-force method that first computes the error measures for a "sufficiently large" set of flow fields obtained with different values of α. Then, we select the very α that gave the smallest error. To reduce the number values to test for α, we propose to start from a given, standard value α_0. This value is incremented/decremented n_α times by multiplying/dividing it with a stepping factor $a > 1$, yielding in total $2\,n_\alpha + 1$ tests. This strategy results in testing more values close to α_0 and, more importantly, less very small or very large values that hardly give reasonable results.

4.4.2 Scale Selection in the Optic Flow Scale Space

In a recent work [2] we investigated the scale space behaviour of variational optic flow methods, the *optic flow scale space*. Intuitively, it is clear that our smoothness parameter α serves a similar role as the scale parameter in classic image scale spaces (see e.g. [Lin94]) as larger values of α lead to smoother, i.e. simplified flow fields. In our work, we reformulated the classic Horn and Schunck energy functional in a signal regularisation-like manner by rewriting the data term as a similarity term to the normal flow. To make this possible, we need to penalise deviations from the normal flow in a spatially varying,

matrix weighted norm. If we apply an explicit scheme to solve the corresponding Euler-Lagrange equations (see Section 5.3.4 for more details), we end up with a diffusion-like evolution that starts from the normal flow and builds up the optic flow scale space.

It is clear that in the optic flow case, there exists one distinct scale α that corresponds to the flow field of highest accuracy. To select this optimal scale, which is an important issue for any scale space, see e.g. [MN03], we utilised our optimal prediction principle. The big advantage in the context of the optic flow scale space is that the explicit evolution gives a flow field after each iteration, i.e. after k iterations a flow field corresponding to the smoothness weight $\alpha = k \cdot \tau$ is available, where τ denotes the time step size of the explicit scheme. Thus, one can easily compute the average data constancy error (ADCE) after each iteration and finally select the scale with the smallest ADCE as the optimal scale.

4.4.3 Experiments

Performance of our Proposed Error Measure. We first show that our proposed data constancy error between frame 1 and 3 ($\text{ADCE}_{1,3}$, see (4.108)) is a good approximation of the popular average angular error (AAE) measure. To this end, we compare the two error measures for the *Grove2* sequence in Figure 4.12 (a) and (b). Additionally, we also show results obtained with a simpler error measure that evaluates the data constancy assumptions between the first and second frame ($\text{ADCE}_{1,2}$), see Figure 4.12 (c).

It becomes obvious that our proposed error measure (Figure 4.12 (b)) indeed exhibits a shape very close to the angular error shown in Figure 4.12 (a). As our error measure reflects the quality of the prediction with the given flow field, our result further substantiates the validity of the proposed OPP. Furthermore, we see that a simple error measure that uses only frame one and two (Figure 4.12 (c)) gives unusable results as it favours way too small values of the smoothness weight. This behaviour could be expected as the constancy assumptions are imposed between frame one and two in the data term and thus they are best fulfilled when the influence of the smoothness term, which is steered by the smoothness weight, becomes small.

Benefits of an Automatic Parameter Selection. Next, we want show that our automatic parameter selection works well for a large variety of different test sequences. To this end we apply our method to the Middlebury data sets [BSL+10]. Note that we cannot give results for the *Dimetrodon* and *Venus* sequence here, as those only encompasses two frames and thus our $\text{ADCE}_{1,3}$ measure cannot be computed.

In Table 4.8, we summarise the AAE when *(i)* setting the smoothness weight α to a fixed value ($\alpha = \alpha_0 = 400.0$), *(ii)* using our automatic parameter selection method, and *(iii)* selecting the (w.r.t. the AAE) optimal value of α under the tested proposals. As we can see, estimating the smoothness weight with our proposed method allows to improve the results compared to a fixed value of α in almost all cases. The only exception is the *Grove 3* sequence where the fixed value of α accidentally coincides with the optimal

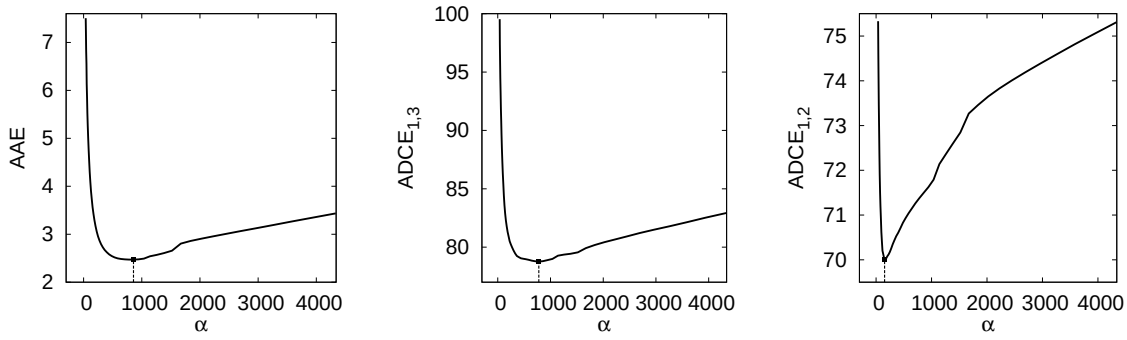


Figure 4.12: Automatic selection of the smoothness weight α for the *Grove2* sequence. *From left to right:* **(a)** Angular error (AAE) for 51 values of α, computed from $\alpha_0 = 400.0$ and using a stepping factor $a = 1.1$. The remaining parameters were set fixed to $\sigma = 0.5, \gamma = 20.0, \rho = 4.0, \eta = 0.95$. The dotted line marks the minimum of the graph. **(b)** Same for the proposed data constancy error ($\mathrm{ADCE}_{1,3}$). **(c)** Same for a simple data constancy error ($\mathrm{ADCE}_{1,2}$).

value. Compared to the results achieved with an optimal value of α, our results are on average 3% and at most 10% worse than the optimal result. When actually comparing the value of α chosen by our method to the optimal value, it turns out that we often choose a value that seems way too small. Nevertheless, the AAE of the corresponding flow field is still close to the optimum and our method thus fulfils its purpose.

We wish to note that a favourable performance of the proposed parameter selection of course depends on the validity of our assumptions (constant speed and linear trajectories of objects). For all considered test sequences these assumptions were obviously valid. In order to show the limitations of our method, we applied our approach again to the same sequences, but replaced the third frame by a copy of the second one. This violates the assumption of a constant speed and leads to a severe impairment of results that reach up to 30%.

Comparison to State-of-the-Art Methods. In order to evaluate the performance of an optic flow method compared to the state-of-the-art, one can submit the results to the popular Middlebury benchmark [BSL$^+$10]. The results that we submitted were obtained with our optic flow in harmony (OFH) method [10] that uses our robust data term from (4.35) in combination with the complementary smoothness term from (4.82). The optimal value of the smoothness weight α is determined by our proposed method based on the OPP.

We found that for the provided benchmark sequences, using an HSV colour representation is not as beneficial as seen in our experiment from Figure 4.4 on page 46. As the Middlebury sequences hardly suffer from difficult illumination conditions, we cannot profit from the photometric invariances of the HSV colour space. On the other hand, some sequences even pose problems in their HSV representation. As an example, consider the results for the the *Teddy* sequence in the first row of Figure 4.13. Here we see that the small white triangle beneath the chimney causes unpleasant artefacts in the flow field. This results from the problem that greyscales do not have a unique represen-

Table 4.8: Results (AAE) for some Middlebury sequences when *(i)* fixing the smoothness weight ($\alpha = 400.0$), *(ii)* estimating α with our proposed method, and *(iii)* with the optimal value of α

Sequence	Fixed α	Estimated α	Optimal α
Rubberwhale	3.43°	3.00°	3.00°
$\alpha =$	400.0	644.2	943.2
Hydrangea	1.96°	1.94°	1.86°
$\alpha =$	400.0	440.0	943.2
Grove2	2.59°	2.43°	2.43°
$\alpha =$	400.0	532.4	779.5
Grove3	5.50°	5.62°	5.50°
$\alpha =$	400.0	532.4	400.0
Urban2	3.22°	2.84°	2.66°
$\alpha =$	400.0	532.4	1255.4
Urban3	3.44°	3.37°	3.35°
$\alpha =$	400.0	484.0	644.2
Yosemite	2.56°	1.89°	1.71°
$\alpha =$	400.0	1037.5	2224.0
Marble	5.73°	5.05°	4.94°
$\alpha =$	400.0	1838.0	3581.7

Table 4.9: Estimated values of the smoothness weight α, using our automatic parameter selection method

Sequence	*Army*	*Grove*	*Mequon*	*Schefflera*	*Urban*	*Wooden*	*Yosemite*
Smoothness weight	277.8	277.8	277.8	691.2	480.0	995.3	1433.3

tation in the hue as well as the saturation channel, see our experiment in Figure 4.3 on page 43. Nevertheless, there are also sequences where an HSV colour representation is beneficial. For the *Mequon* sequence (second row of Figure 4.13) an HSV colour representation removes artefacts in the shadows left of the toys that are shown in the insets. The bottom line is, however, that for the whole set of benchmark sequences, we obtain slightly better results when using the RGB colour space. Thus, we use this variant of our method for evaluation at the Middlebury benchmark.

For our submission we used in accordance to the guidelines a fixed set of parameters: $\sigma = 0.5, \gamma = 20.0, \rho = 4.0, \eta = 0.95$. For determining the smoothness weight α, we used the settings $n_\alpha = 8, \alpha_0 = 400.0, a = 1.2$ (see page 73). For the *Teddy* sequence with only two frames, we set $\alpha = \alpha_0$, as our parameter estimation method is not applicable in this case. The resulting running time for the *Urban* sequence (640×480 pixels) was 620 s on a standard PC (3.2 GHz Intel Pentium 4). For the parameter selection we computed $2 \cdot 8 + 1 = 17$ flow fields, corresponding to approximately 36 s per flow field. Following the trend of parallel implementations on modern GPUs [GT08, SBK10, WPB10, ZPB07] we recently came up with a GPU version of our method that can compute flow fields of size 640×480 pixels in less than one second [3] .

At the time of writing (January 2011), our method achieved the 7th place w.r.t. the AAE and the 6th place w.r.t. the AEE measure among 46 listed methods. Note that our previous Complementary Optic Flow method [11] only ranked 9th for the AAE and 13th for the AEE, which demonstrates the benefits of the proposed novelties, like the automatic parameter selection. In Table 4.9 we additionally summarise the estimated values of α resulting from our automatic parameter selection method. As desired, for sequences with small details in the flow field (*Army*, *Grove*, *Mequon*) a small smoothness weight is chosen. On the other hand, sequences like *Wooden* and *Yosemite* with a rather smooth flow yield significantly larger values for the smoothness weight. Note that we did not submit our stereo results to the Middlebury stereo benchmark. Despite our proposed improvements, variational stereo approaches are still not competitive to recent discrete methods. The main reason for this probably lies in the coarse-to-fine strategy that is needed to handle large displacements in variational approaches. This strategy hinders the correct estimation of comparatively small details in the disparity map. The minimisation strategies of discrete methods on the other hand mostly proceeds on the finest grid only.

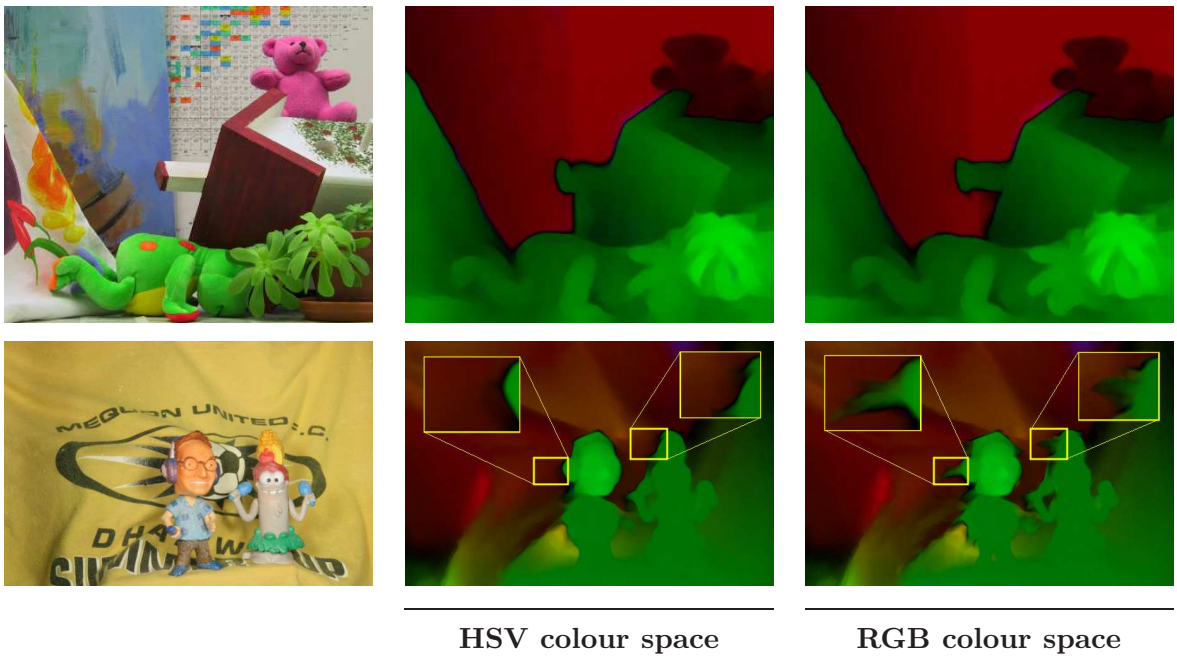


Figure 4.13: Comparison of results obtained with HSV or RGB colour representation. *From left to right:* Reference frame. Result with HSV colour space. Result with RGB colour space. *From top to bottom: Teddy* sequence (HSV $\Rightarrow$ AAE = 3.94°, RGB $\Rightarrow$ AAE = 2.64°). *Mequon* sequence (HSV $\Rightarrow$ AAE = 2.28°, RGB $\Rightarrow$ AAE = 2.84°).

4.5 Prediction Prior

By construction, optic flow methods using a coarse-to-fine warping scheme cannot estimate large displacements of *small objects*. More precisely, the displacement of an object must not be larger than the object itself. However, such large displacements may occur due to fast motion, low frame rate of the camera, or large baseline distances in the stereo context. For our scene flow method in [5] we had to cope with the latter problem. As a remedy, we followed the ideas in [BM11] and added a *prediction prior* to our usual energy functional. This prior models the knowledge of an initial guess (a "prediction") of the flow field we aim to compute, e.g. the prior could be the result from some other matching strategy that does not suffer from the afore mentioned problems. To incorporate this prior knowledge into our model we add a third term to the general energy functional (4.1), which results in the modified energy functional [BM11]

$$E(u,v) = \int_\Omega \big(M(u,v) + \alpha\, V(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) + \beta\, P(u,v)\big)\, \mathrm{d}x\, \mathrm{d}y \quad , \tag{4.109}$$

where $P(u,v)$ denotes the prediction prior, and the weight $\beta > 0$ steers its influence on the overall flow estimation. An obvious idea to define the P is

$$P(u,v) = \Psi_P\big((u-\tilde{u})^2 + (v-\tilde{v})^2\big) \quad , \tag{4.110}$$

where $(\tilde{u}, \tilde{v})^\top$ denotes the given predictor flow field. For the robust penaliser function we propose to use $\Psi_P(s^2) = \sqrt{s^2 + \varepsilon^2}$, which helps to cope with outliers (wrong estimates) in the predictor.

In some cases, e.g. when using SIFT matches [Low99, Low04] as a predictor, we are not given a prediction at any point of the image domain. To account for this, we multiply the prediction prior P with a mask $c_P(\boldsymbol{x})$ that is equal to one if we are given a prediction at the point $\boldsymbol{x} \in \Omega$ and is else equal to zero. Incorporating the mask in (4.109) gives the energy

$$E(u,v) = \int_\Omega \big(M(u,v) + \alpha\, V(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) + \beta\, c_P\, P(u,v)\big)\, \mathrm{d}x\, \mathrm{d}y \ . \tag{4.111}$$

The corresponding Euler-Lagrange equations for energy (4.111) when using the prior from (4.110) are given by

$$\partial_u M - \alpha\, \mathrm{div} \begin{pmatrix} \partial_{u_x} V \\ \partial_{u_y} V \end{pmatrix} + \beta\, c_P\, \Psi'_P\big((u-\tilde{u})^2 + (v-\tilde{v})^2\big)\ (u-\tilde{u}) \ = \ 0 \ , \tag{4.112}$$

$$\partial_v M - \alpha\, \mathrm{div} \begin{pmatrix} \partial_{v_x} V \\ \partial_{v_y} V \end{pmatrix} + \beta\, c_P\, \Psi'_P\big((u-\tilde{u})^2 + (v-\tilde{v})^2\big)\ (v-\tilde{v}) \ = \ 0 \ . \tag{4.113}$$

4.5.1 Experiments

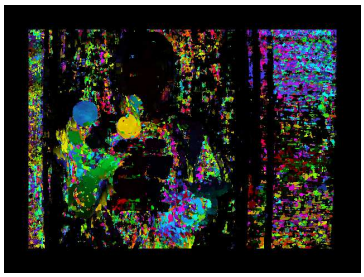
We now show experiments that illustrate the benefits of a prediction prior. For our first experiment we consider two frames of the *Beanbag* sequence shown in the top row of Figure 4.14. Note the rightmost ball undergoes a large displacement to the top, which is indeed larger than its own size. Consequently, optic flow approaches relying on warping cannot capture this displacement, as can be observed in Figure 4.14 (c). On the other hand, even a simple block matching approach, see Figure 4.14 (d), can estimate the displacement as this technique does not require to work on coarse representations of the images. Using the block matching result as a predictor within a warping-based optic flow approach then finally allows to obtain a smooth flow field that can capture the displacement of the right ball, as witnessed in Figure 4.14 (e). A small remaining problem is that some of the artefacts in the upper right corner are still present in the final result.

In our second experiment, we consider a synthetic sequence that we created in the context of our work on scene flow from uncalibrated cameras [5]. The scene depicts a textured sphere which is situated in front of a likewise textured background. In Figure 4.15 (a) and (b) we show the left and right image of the corresponding stereo pair at time t. Note that we do not have an orthoparallel camera setup in this case, which means that the displacements are not purely horizontal. This can easily be seen at the ground truth displacement in Figure 4.15 (c). For estimating the displacements, we first try to apply our Complementary optic flow approach [10, 11]. As the displacements of the small-scale textures are rather large (about 30 pixels), the resulting displacement

(a) First image

(b) Second image

(c) Complementary OF

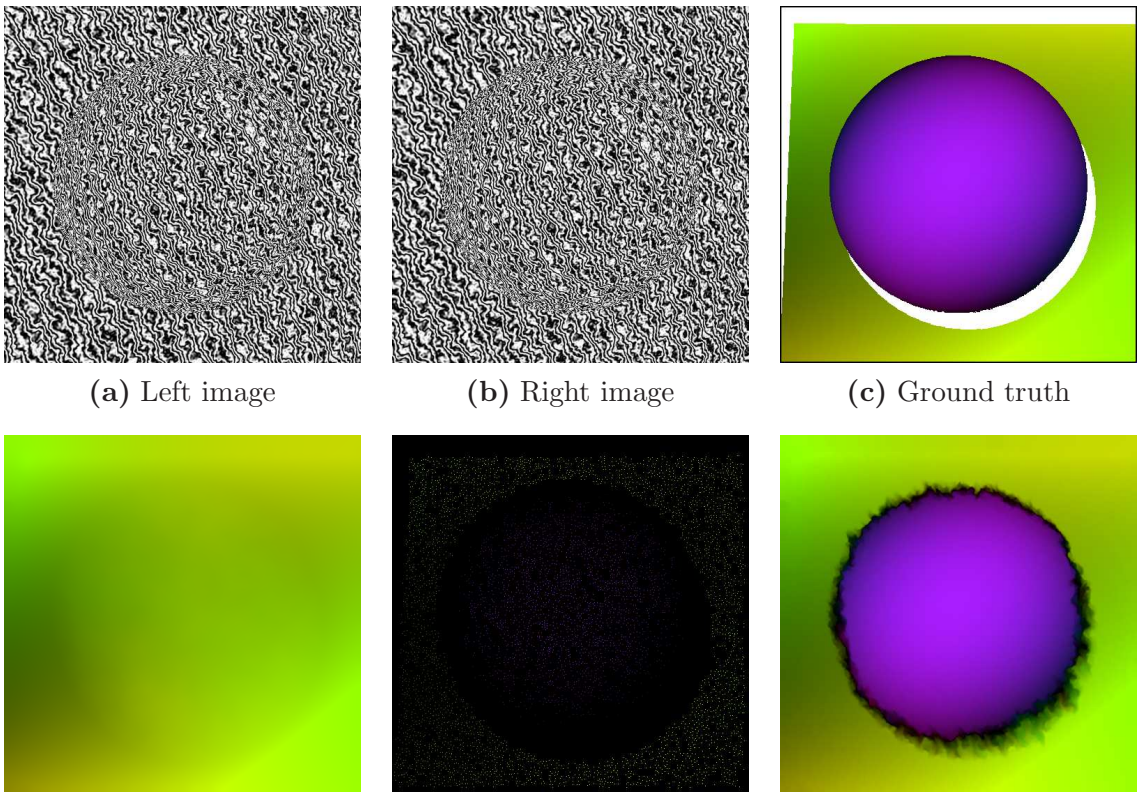


(d) Block matching (NCC)

(e) Complementary OF + prediction (NCC)

Figure 4.14: Results for the *Beanbag* sequence. *First row, from left to right:* **(a)** First image. **(b)** Second image. *Second row, from left to right:* **(c)** Flow field obtained with our Complementary Optic Flow approach [10, 11] ($\alpha = 200.0, \sigma = 0.3, \gamma = 10.0, \rho = 1.0, \eta = 0.95$). **(d)** Same for a simple block matching approach (normalised cross correlation (NCC), 9×9 blocks). **(e)** Same for our Complementary Optic Flow approach with additional prediction prior using previous block matching result ($\alpha = 800.0, \beta = 20.0$).

estimation fails, see Figure 4.15 (d). Due to the dominant texture, the present sequence is on the other hand well-suited for feature matching approaches like SIFT [Low99, Low04], which find a large number of matches that are shown in Figure 4.15 (e). These correspondences were computed using the binary available under `http://www.cs.ubc.ca/~lowe/keypoints/`. Incorporating the sparse correspondences as a prediction prior, a favourable estimation of the displacements can be achieved, see Figure 4.15 (f).



(a) Left image **(b)** Right image **(c)** Ground truth

(d) Complementary OF **(e)** SIFT matches **(f)** Complementary OF + prediction (SIFT)

Figure 4.15: Results for the *general sphere* sequence [5]. *First row, from left to right:* **(a)** Left image. **(b)** Right image. **(c)** Ground truth (white pixels mark locations where no ground truth is available). *Second row, from left to right:* **(d)** Flow field obtained with our Complementary Optic Flow approach [10, 11] ($\alpha = 100.0, \sigma = 0.9, \gamma = 0.3, \rho = 1.0, \eta = 0.95 \Rightarrow$ AAE $= 77.78°$). **(e)** Sparse SIFT matches [Low99, Low04] (matching threshold $= 0.6$). **(f)** Flow field obtained with our Complementary Optic Flow approach with additional prediction prior using previous SIFT matches ($\alpha = 10.0, \beta = 4.0 \Rightarrow$ AAE $= 5.13°$).

Chapter 5

Numerics

> All the mathematical sciences are founded on relations between physical laws and laws of numbers, so that the aim of exact science is to reduce the problems of nature to the determination of quantities by operations with numbers.
>
> *James C. Maxwell*

This chapter encompasses two major parts. We first discuss the numerical approximation of derivatives that occur in our continuous models from the previous section. To this end, we make use of the popular concept of finite differences, see e.g. [MM94]. For the discretisation of *image* derivatives, we additionally present a novel finding from our work in [1, 7]. Here, we showed that an appropriate discretisation strategy borrowed from the theory of hyperbolic differential equations can have a tangible influence on the quality of results.

The second part is mainly concerned with reviewing strategies for the solution of the Euler-Lagrange equations, including the embedding in a coarse-to-fine warping framework for handling large displacements. As an illustrative example, we consider the Euler-Lagrange equations arising in our complementary optic flow approach [10, 11]. Additionally, we present two novel contributions from our side: (i) A fast solver that is based on a parallel implementation of a fast explicit diffusion (FED) [GWB10] gradient descent scheme on modern graphics hardware [3]. (ii) An implementation of the previous solver on a modern smart phone [4].

5.1 Discretisation

Implementing the continuous variational models that we have discussed in the previous chapter requires to *discretise* the images and the flow field as well as their derivatives.

The discretisation of the images was already presented in Section 3.4 and the discretisation of the flow fields follows in complete accordance. The discretisation of derivatives is more involved and will be described in the following.

5.1.1 Discretising Derivatives by Finite Differences

In this thesis, we use the concept of *finite difference approximations* for discretising derivatives. As an example, consider the approximation of the x-derivative of an image $f(x, y)$. To this end, finite differences offer three basic choices: *forward*, *backward* and *central* differences, which are defined as

$$\mathcal{D}_x^+ f_{i,j}^k := \frac{1}{h_1}\left(f_{i+1,j}^k - f_{i,j}^k\right) , \qquad \mathcal{D}_x^+ f_{i,j}^{k+1} := \frac{1}{h_1}\left(f_{i+1,j}^{k+1} - f_{i,j}^{k+1}\right) , \tag{5.1}$$

$$\mathcal{D}_x^- f_{i,j}^k := \frac{1}{h_1}\left(f_{i,j}^k - f_{i-1,j}^k\right) , \qquad \mathcal{D}_x^- f_{i,j}^{k+1} := \frac{1}{h_1}\left(f_{i,j}^{k+1} - f_{i-1,j}^{k+1}\right) , \tag{5.2}$$

$$\mathcal{D}_x^0 f_{i,j}^k := \frac{1}{2h_1}\left(f_{i+1,j}^k - f_{i-1,j}^k\right) , \qquad \mathcal{D}_x^0 f_{i,j}^{k+1} := \frac{1}{2h_1}\left(f_{i+1,j}^{k+1} - f_{i-1,j}^{k+1}\right) , \tag{5.3}$$

where h_1 denotes the grid size in x-direction and the finite difference operators $\mathcal{D}^+$, $\mathcal{D}^-$ and $\mathcal{D}^0$ denote forward, backward and central differences, respectively. Further note that these expressions can be computed at the time level k and $k+1$. The corresponding differences for the y-derivative are obtained by simply switching the roles of i and j in above equations. The approximation error of the one-sided differences (forward and backward) is in $\mathcal{O}(h)$, where $h \in \{h_1, h_2\}$. Their central counterparts only involve an error of $\mathcal{O}(h^2)$, which, together with the unbiased stencil orientation, explains why central differences are a popular "standard" choice in image processing applications.

In order to increase the accuracy, one may use *averaged differences* that take into account differences from both time levels k and $k+1$. One possibility is to use averaged central differences that for the x-derivative are defined as

$$\begin{aligned} \mathcal{D}_x^0 f_{i,j}^{k+\frac{1}{2}} &:= \frac{1}{2}\left(\mathcal{D}_x^0 f_{i,j}^k + \mathcal{D}_x^0 f_{i,j}^{k+1}\right) \\ &= \frac{1}{4h_1}\left(f_{i+1,j}^k - f_{i-1,j}^k + f_{i+1,j}^{k+1} - f_{i-1,j}^{k+1}\right) . \end{aligned} \tag{5.4}$$

In the remainder of this chapter, such a central difference approximation will be referred to as a "standard" derivative approximation.

Note that averaging the two signals at time k and $k+1$ only makes sense if the displacements between the signals are small, in fact smaller than one pixel. Only if this precondition is satisfied, our averaging takes into account information originating from the same points in space. In our case, the multiscale warping approach (see Section 5.3.1) results in displacements in the desired range.

A means to further reduce the approximation error is to use approximations of fourth-order that involve more pixels than just the two direct neighbours of the pixel

under consideration. An averaged fourth-order approximation of the x-derivative is for example defined as

$$\begin{aligned} \mathcal{D}_x^{00} f_{i,j}^{k+\frac{1}{2}} &:= \frac{1}{2}\left(\mathcal{D}_x^{00} f_{i,j}^{k} + \mathcal{D}_x^{00} f_{i,j}^{k+1}\right) \\ &:= \frac{1}{24h_1}\Big(-f_{i+2,j}^{k} + 8f_{i+1,j}^{k} - 8f_{i-1,j}^{k} + f_{i-2,j}^{k} \\ &\qquad -f_{i+2,j}^{k+1} + 8f_{i+1,j}^{k+1} - 8f_{i-1,j}^{k+1} + f_{i-2,j}^{k+1}\Big) \ . \end{aligned} \tag{5.5}$$

Discretisation of Flow Derivatives

Having presented the basics of finite difference approximations on the examples of image derivatives, we can turn to the discretisation of the flow derivatives.

Discretising Spatial and Temporal Flow Derivatives. When using flow-driven or image- and flow-driven regularisers, the spatial flow derivatives u_x, u_y and v_x, v_y will occur in the diffusion tensor $\boldsymbol{D}$. For discretising the flow derivatives, we use a central difference approximation of second order. For example, discretising the derivative u_x is done via

$$u_x \ \approx \ \mathcal{D}_x^0 u_{i,j} \ = \ \frac{1}{2h_1}\left(u_{i+1,j} - u_{i-1,j}\right) \ . \tag{5.6}$$

For spatio-temporal regularisers discussed in Section 4.3.4, we also need to discretise temporal flow derivatives u_t and v_t. To this end, we use the approximation

$$u_t \ \approx \ \frac{1}{h_3}\left(u_{i,j}^{k+1} - u_{i,j}^{k}\right) \ , \quad \text{and} \quad v_t \ \approx \ \frac{1}{h_3}\left(v_{i,j}^{k+1} - v_{i,j}^{k}\right) \ , \tag{5.7}$$

respectively. Here, it makes sense to adapt the value of h_3 to the given image sequence to allow for an appropriate scaling of the temporal direction compared to the spatial directions [WS01b].

Discretising the Divergence Expression. Considering the Euler-Lagrange equations (5.56) and (5.57) shows that we need to discretise divergence expression of the form $\operatorname{div}(\boldsymbol{D}\,\boldsymbol{\nabla}_2 u)$. Denoting the tensor entries by

$$\boldsymbol{D} := \begin{pmatrix} a & b \\ b & c \end{pmatrix} \ , \tag{5.8}$$

we can write the divergence expression as

$$\operatorname{div}\left(\boldsymbol{D}\,\boldsymbol{\nabla}_2 u\right) \ = \ \operatorname{div}\begin{pmatrix} au_x + bu_y \\ bu_x + cu_y \end{pmatrix} \tag{5.9}$$

$$\ = \ \partial_x(au_x) + \partial_x(bu_y) + \partial_y(bu_x) + \partial_y(cu_y) \ . \tag{5.10}$$

The pure x- and y-terms ($\partial_x(au_x)$ and $\partial_y(cu_y)$) are discretised by using forward differences inside the bracket and backward differences for the outer derivative, i.e.

$$\partial_x(au_x) \approx \mathcal{D}_x^- \left(\frac{a_{i+1,j} + a_{i,j}}{2} \, \frac{u_{i+1,j} - u_{i,j}}{h_1} \right) \tag{5.11}$$

$$\approx \frac{a_{i+1,j} + a_{i,j}}{2h_1^2} (u_{i+1,j} - u_{i,j}) - \frac{a_{i,j} + a_{i-1,j}}{2h_1^2} (u_{i,j} - u_{i-1,j}) \quad , \tag{5.12}$$

and correspondingly

$$\partial_y(cu_y) \approx \frac{c_{i,j+1} + c_{i,j}}{2h_2^2} (u_{i,j+1} - u_{i,j}) - \frac{c_{i,j} + c_{i,j-1}}{2h_2^2} (u_{i,j} - u_{i,j-1}) \quad . \tag{5.13}$$

For the mixed terms $\partial_x(bu_y)$ and $\partial_y(bu_x)$, we can use a standard approximation by central differences, resulting in

$$\partial_x(bu_y) \approx \mathcal{D}_x^0 \left(b_{i,j} \, \frac{u_{i,j+1} - u_{i,j-1}}{2h_2} \right) \tag{5.14}$$

$$\approx \frac{b_{i+1,j}}{4h_1h_2} (u_{i+1,j+1} - u_{i+1,j-1}) - \frac{b_{i-1,j}}{4h_1h_2} (u_{i-1,j+1} - u_{i-1,j-1}) \quad , \tag{5.15}$$

and correspondingly

$$\partial_y(bu_x) \approx \frac{b_{i,j+1}}{4h_1h_2} (u_{i+1,j+1} - u_{i-1,j+1}) - \frac{b_{i,j-1}}{4h_1h_2} (u_{i+1,j-1} - u_{i-1,j-1}) \quad . \tag{5.16}$$

This shows that the divergence expression can be discretised as a *weighted sum of differences*

$$\operatorname{div}(\boldsymbol{D} \, \boldsymbol{\nabla}_2 u) \approx \sum_{\tilde{i},\tilde{j} \in \mathcal{N}_{i,j}} [w]_{\tilde{i},\tilde{j}} \left([u]_{\tilde{i},\tilde{j}} - [u]_{i,j} \right) \quad , \tag{5.17}$$

where $\mathcal{N}_{i,j}$ denotes the set of pixels in the 8-neighbourhood of the pixel (i,j) and the weights $[w]_{\tilde{i},\tilde{j}}$ are defined as

$$[w]_{i-1,j} = \frac{a_{i-1,j} + a_{i,j}}{2\,h_1^2} \quad , \qquad [w]_{i+1,j} = \frac{a_{i+1,j} + a_{i,j}}{2\,h_1^2} \quad , \tag{5.18}$$

$$[w]_{i,j-1} = \frac{c_{i,j+1} + c_{i,j}}{2\,h_2^2} \quad , \qquad [w]_{i,j+1} = \frac{c_{i,j-1} + c_{i,j}}{2\,h_2^2} \quad , \tag{5.19}$$

$$[w]_{i-1,j-1} = \frac{-b_{i-1,j} - b_{i,j+1}}{4\,h_1h_2} \quad , \qquad [w]_{i+1,j-1} = \frac{b_{i+1,j} + b_{i,j+1}}{4\,h_1h_2} \quad , \tag{5.20}$$

$$[w]_{i-1,j+1} = \frac{b_{i-1,j} + b_{i,j-1}}{4\,h_1h_2} \quad , \qquad [w]_{i+1,j+1} = \frac{-b_{i+1,j} + b_{i,j-1}}{4\,h_1h_2} \quad . \tag{5.21}$$

In order to obtain sharper results, we follow [Wei98] and modify above weights in accordance to the non-negative stencil proposed in [Wei98], chapter 3.4. This gives the weights

$$[w]_{i-1,j} = \frac{a_{i-1,j} + a_{i,j}}{2\,h_1^2} - \frac{|b_{i-1,j}| + |b_{i,j}|}{2\,h_1 h_2} \quad , \tag{5.22}$$

$$[w]_{i+1,j} = \frac{a_{i+1,j} + a_{i,j}}{2\,h_1^2} - \frac{|b_{i+1,j}| + |b_{i,j}|}{2\,h_1 h_2} \quad , \tag{5.23}$$

$$[w]_{i,j-1} = \frac{c_{i,j+1} + c_{i,j}}{2\,h_2^2} - \frac{|b_{i,j+1}| + |b_{i,j}|}{2\,h_1 h_2} \quad , \tag{5.24}$$

$$[w]_{i,j+1} = \frac{c_{i,j-1} + c_{i,j}}{2\,h_2^2} - \frac{|b_{i,j-1}| + |b_{i,j}|}{2\,h_1 h_2} \quad , \tag{5.25}$$

$$[w]_{i-1,j-1} = \frac{|b_{i-1,j+1}| - b_{i-1,j+1} + |b_{i,j}| - b_{i,j}}{4\,h_1 h_2} \quad , \tag{5.26}$$

$$[w]_{i+1,j-1} = \frac{|b_{i+1,j+1}| + b_{i+1,j+1} + |b_{i,j}| + b_{i,j}}{4\,h_1 h_2} \quad , \tag{5.27}$$

$$[w]_{i-1,j+1} = \frac{|b_{i-1,j-1}| + b_{i-1,j-1} + |b_{i,j}| + b_{i,j}}{4\,h_1 h_2} \quad , \tag{5.28}$$

$$[w]_{i+1,j+1} = \frac{|b_{i+1,j-1}| - b_{i+1,j-1} + |b_{i,j}| - b_{i,j}}{4\,h_1 h_2} \quad . \tag{5.29}$$

5.2 Upwind Discretisation of Image Derivatives

All experiments presented so far used the *central* difference approximation (5.5) for discretising occurring image derivatives. The only exception is the computation of the regularisation tensor (4.78). Here, we do not average the differences, but only use the image at time level k to consider directional information from the reference image, only.

However, our investigations in [1, 7] showed that using *correctly oriented one-sided* differences can tangibly improve the results compared to central difference approximations. In the context of *hyperbolic partial differential equations* (*HDEs*) this fact is well-known and used in the popular *upwind schemes*. The major advantage of upwind discretisations is that they prevent oscillations at edges as they belong to the class of so-called monotone schemes that satisfy a discrete maximum-minimum principle. In this class, upwind schemes are known to be the most accurate schemes.

5.2.1 A 1D Example

To illustrate the influence of the applied finite difference operator, we use a simple 1D example. Here, we consider a 1D signal sequence $f(x,t)$ where $x \in \Omega_1$ denotes the

position in the interval $\Omega_1 \subset \mathbb{R}$. To compute the unknown displacement function $u(x)$ we minimise the energy functional

$$E(u) = \int_{\Omega_1} \left((f_x\, u + f_t)^2 + \alpha\, u_x^2 \right) \mathrm{d}x \ , \tag{5.30}$$

which is identical to a 1D version of the Horn and Schunck model [HS81]. Equivalently to the brightness constancy assumption, the data term imposes a constancy on the signal values. In the smoothness term, we penalise the 1D pendant of the gradient magnitude, given by the x-derivative of the displacement function. Both term use a basic quadratic penalisation. The Euler-Lagrange equation corresponding to the energy (5.30) reads

$$f_x^2\, u + f_x f_t - \alpha\, u_{xx} = 0 \ . \tag{5.31}$$

Let us now present a 1D experiment that shows the importance of an appropriate choice for the finite difference approximation of f_x. Consider the two frames of a signal sequence in Figure 5.1 (a). Here, the signal is displaced by exactly one position to the right in its middle part and stays unchanged otherwise. This is illustrated in the ground truth displacement shown in Figure 5.1 (b). Note that this example comprises smooth as well as discontinuous signal and displacement regions which make it rather indicative.

We further wish to note that our example also exhibits the so-called *occlusion problem*. This arises if a foreground object is shifted and occludes parts of the background. Thus, one cannot find any correspondence for the regions that are only visible in the first frame and then become occluded in the second frame. In our example this happens at Point 9. Therefore, any computed displacements u will be corrupted at this occlusion. However, while we comment in this way on the expected computational results, the occlusion problem is not a topic in the focus of our work. In practical computations, the occlusion problem is dealt with separately; see e.g. [BBH03] and the references therein.

In Figure 5.1 (c)–(e) we depict the computed displacements using various finite difference discretisations for the signal derivative f_x. The displacements were obtained as the solution of a linear system of equations that arises from the discrete Euler-Lagrange equation (5.31). As the corresponding system matrix is tridiagonal, the Thomas algorithm [Tho49] can be used as a simple and efficient solver. In our experiment, we set the smoothness weight to the small value of $\alpha = 10^{-4}$ in order to clearly show the influence of the discretisation of the signal derivative that occurs in the data term.

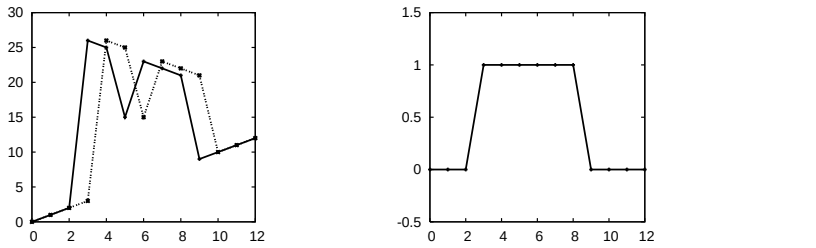
Comparing the resulting displacements in Figure 5.1 (c)–(e), the effect of the different derivative approximations become obvious: Central differences only perform well in the smooth signal regions, i.e. at the left and right boundaries. At discontinuities they produce severe oscillations. One-sided differences perform either favourably or fail totally. Obviously, the correct orientation matters here. When using correctly oriented one-sided differences, the displacement almost coincides with the ground truth, except at one point. As indicated above, this is not a fault of the method, but is caused by the occlusion at the jump in the displacement. Note that as the displacement of an

occluded point is in general undefined, we set the ground truth at occlusions to the value of its right neighbour. The observed behaviour in our experiment is in accordance with the theory of numerical methods for HDEs [LeV92, LeV02]. There, so called *upwind schemes* are a widely used concept for the discretisation of transport equations, where the term *upwind* refers to correctly oriented one-sided differences. The correct orientation of an upwind stencil in our case is opposite to the displacement direction; see our experiment.

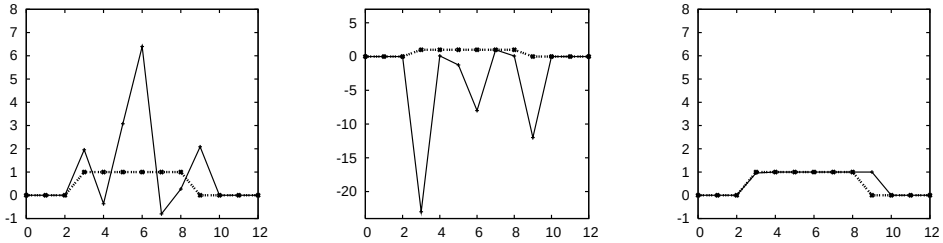
In the hyperbolic theory, central difference approximations as in (5.5) are known to lead to oscillations. They can even be unconditionally unstable, as for example shown in [Tor99]. The reason why they are still a common choice in correspondence problems is that here only one time step is performed, and thus the instability is "only" observable in terms of oscillations near strong gradients. Note that there exist stable central discretisations like the monotone Lax-Friedrichs and the Lax-Wendroff scheme, see [LeV02]. However, the Lax-Friedrichs scheme is the most dissipative one, it actually involves the most possible blurring of all monotone schemes. The Lax-Wendroff scheme is not monotone (it is second-order accurate) and is known to generate oscillations as by the Gibbs phenomenon. Finally, let us remark that the stability of explicit schemes for HDEs involves a condition on the time step size called CFL condition, see [LeV02, Tor99]. This CFL condition can be translated into the assumption, that in one time step, information can only be translated up to one pixel. Therefore, the CFL condition for HDEs is related to the assumption of small displacements in correspondence problems. As discussed earlier, the CFL condition is always satisfied because we use a multiscale warping approach that results in displacements in the range of one pixel at all pyramid levels.

Relation to HDEs. Interpreting the data constancy assumption $f_x\,u + f_t = 0$ as a PDE for the temporal evolution of f leads to a transport process in the form of a *hyperbolic colour equation*, see [LeV02]. The name of this equation is derived from a specific application in fluid dynamics where coloured tracer particles are given into a flowing fluid in order to visualise the flow lines. The PDE governing the transport of the particle density of the tracer is the colour equation. It is often assumed that the fluid moves always in the same way, i.e. the velocity of the fluid is in a steady state. Therefore, the velocity function of the colour is varying in space (as the displacement u in correspondence problems does), but not in time.

As we can observe, the colour equation determines the evolution of f at hand of a given displacement field u. In hyperbolic problems, the roles of known and unknown are thus switched compared to correspondence problems. We will show, however, that the same discretisation principles apply in both cases. A last remark is that the colour equation is a PDE given in the framework of an initial value problem or an initial-boundary value problem. In our setting, the initial state $f(x,t)$ is evolved in time. The role of the other given state $f(x,t+1)$ will be *(i)* to determine the displacement direction and *(ii)* to provide data for accurate discretisations.



(a) Signal at k and $k+1$ **(b)** Ground truth displacement



(c) Central differences **(d)** Forward differences **(e)** Backward differences

Figure 5.1: Comparison of different derivative approximations on a simple 1D example. *Top row, from left to right:* **(a)** Signal at time k (solid) and $k+1$ (dotted). **(b)** Ground truth displacement. *Bottom row, from left to right:* **(c)** Displacement computed using standard *central* differences averaged between time level k and $k+1$ (solid), compared to the ground truth (dotted). **(d)** Same for one-sided *forward* differences. **(e)** Same for one-sided *backward* differences.

5.2.2 A Novel Adaptive Upwind Scheme

As we have seen, the one-sided, low-order upwind differences perform well at signal discontinuities. However, in smooth regions the higher order of accuracy of central differences will give a better resolution of the displacement function. Hence, a natural idea is to combine the two types of schemes by using high-order central approximation in smooth signal parts and upwinding at discontinuities.

This idea has been successfully used for the construction of so-called *high-resolution methods* [LeV02, Tor99] for HDEs. They use a nonlinear blending of low- and high-order approximations, steered by a smoothness measure. In the classic framework for HDEs, the blending is performed by a limiter function that can be constructed to result in the total variation stability of the high-resolution (HR) scheme. This construction results in a non-uniform order but effectively in a discretisation of high quality, which explains the naming of HR methods. The adaptation of this methodology to our variational framework results in an *adaptive high-resolution-type* (*HRT*) discretisation scheme for correspondence problems. This scheme was subject of our papers [1, 7] and will be presented in the following.

Before proceeding with the scheme description, let us give some comments on similarities and differences of our method to the classic HR schemes, as we will not apply an off-the-shelf-approach. While it is very useful to consider the hyperbolic colour equation

as a distinct important part to be re-interpreted and discretised, the final aim is to compute the displacement u. Especially, while a non-oscillatory resolution of edges in u is obviously important as seen by the experiment in Figure 5.1, we do not have to spend too much attention to structural properties of a discretisation, e.g. its total variation stability. This would only be important for a long-time integration of f, while in our case f is already given and only two time levels k and $k+1$ need to be considered.

Measuring Smoothness. First we discuss how to determine the smooth and discontinuous regions of a signal which will be needed for steering the blending of the two considered schemes. To this end, we introduce a *smoothness measure*

$$\Theta_i := \left|\mathcal{D}_x^- f_i^k - \mathcal{D}_x^+ f_i^k\right| + \left|\mathcal{D}_x^- f_i^{k+1} - \mathcal{D}_x^+ f_i^{k+1}\right| \quad , \tag{5.32}$$

that is close to 0 in smooth regions where backward and forward differences are almost identical, and large at discontinuities of the signal. Here, one of the differences to an usual setup of high-resolution methods for HDEs becomes obvious: In correspondence problems one already has the final state of the evolving signal f at hand, and so we can base our smoothness measure on both time levels k and $k+1$. By using the multiscale warping approach, such a combination of data from different frames will not result in taking data from entirely different locations as the displacements at each level of the warping pyramid are in the range of one pixel.

Determining the Upwind Directions. Next we need to determine the appropriate upwind directions for discretising the derivative f_x. This is not straightforward, since the upwind direction depends on the direction of the displacement field, and this is exactly the unknown we aim to compute.

As a remedy, we propose to compute a *predictor solution* $\tilde{u}$ whose sign determines the upwind direction. This predictor is computed using the high-order standard approximation f_x^H of the derivative f_x. It is given by the averaged central difference approximation

$$\left(f_x^H\right)_i := \mathcal{D}_x^0 f_i^{k+\frac{1}{2}} \quad . \tag{5.33}$$

In order to avoid oscillations as seen in our previous experiment, we use a comparatively large smoothness weight for this computation ($\alpha = 1$). If we later consider the extension of our scheme to optic flow, we can use the same value of α for computing the predictor as well as the final solution, as common choices of α are large enough in general.

With the help of the predictor solution $\tilde{u}$, the *low-order upwind approximation* f_x^L of f_x is defined as

$$\left(f_x^L\right)_i := \begin{cases} \mathcal{D}_x^- f_i^k \ , & \text{if } \tilde{u}_i > 0 \ , \\ \mathcal{D}_x^+ f_i^k \ , & \text{if } \tilde{u}_i < 0 \ , \\ (f_x^H)_i \ , & \text{if } \tilde{u}_i = 0 \ . \end{cases} \tag{5.34}$$

Revisiting our experiment in Figure 5.1, one confirms that this definition agrees with the results obtained there. Note that we compute in (5.34) the upwind stencils employing

solely the frame k. This is theoretically justified since the linearised brightness constancy assumption $f_x\, u + f_t = 0$ involves a linearisation at that time level.

The Blending Function. Now we can define the *blending function* $\Phi(\Theta_i)$ that realises the switch between high-order and low-order approximations in accordance to the value of Θ_i. The idea is that Θ_i shall be close to 1 in smooth signal regions, leading to a high-order approximation there. At discontinuities, the blending function shall be close to 0 which results in a low-order upwind approximation that is better suited at such locations. For the actual choice of $\Phi(\Theta_i)$ we propose

$$\Phi(\Theta_i) := \begin{cases} 1 - \Theta_i\,, & \text{if } 0 \le \Theta_i < 1\,, \\ 0\,, & \text{else} \quad . \end{cases} \tag{5.35}$$

The blending is performed in a different way than in the usual setup of high-resolution schemes for HDEs [LeV02, Tor99]. Other blending functions – especially those standard in the field of HDEs – do not lead to better results. We tested this but do not comment on it in more detail here.

The High-Resolution-Type (HRT) Discretisation Scheme. Now everything is prepared to define the adaptive HRT discretisation scheme as

$$(f_x)_i^k := \left(f_x^L\right)_i + \Phi\left(\Theta_i\right)\left[\left(f_x^H\right)_i - \left(f_x^L\right)_i\right]\,, \tag{5.36}$$

using the function $\Phi(\Theta_i)$ to blend between the high-order derivative approximation f_x^H and its low-order counterpart f_x^L.

Note that applying the HRT scheme to the signal sequence from Figure 5.1 gives the same result as with the appropriate upwind scheme, hence we omit an additional figure. However, our experiments in Section 5.2.4 will show that for more challenging 2D problems such as optic flow, the HRT scheme will give results that are superior to a pure upwind scheme.

5.2.3 Extending the Scheme to 2D Optic Flow Problems

Now we adapt the presented HRT discretisation scheme to the 2D optic flow case. Note that the orthoparallel stereo case is implicitly encompassed in these considerations as just less derivatives occur there.

Smoothness Measures. The first extension for the 2D case are distinct smoothness measures Θ_x, Θ_y for the x- and the y-direction, respectively. For Θ_x we use the according expression (5.32) from the 1D case, where we only need to replace the pixel index i by i, j. The measure Θ_y is obtained by using y- instead of x-differences. As our models use the gradient constancy assumption, also the second order and mixed derivatives $f_{xx}, f_{yy}, f_{xy}, f_{xz}$ and f_{yz} occur in the motion tensor, see (4.21), and need to be treated.

Thus, we also need to define a smoothness measure for the mixed xy-direction. Given the smoothness measures Θ_x and Θ_y, we define the mixed expression as $\Theta_{xy} := \Theta_x + \Theta_y$.

Second Order Derivative Approximations. More involved are the high-order and the (one-sided) low-order approximations of the second order derivatives.

High-Order: The high-order approximations of f_{xx} and f_{yy} are defined as

$$f_{xx} \approx \mathcal{D}_x^- \left(\mathcal{D}_x^+ f_{i,j}\right) = \frac{1}{h_1^2}\left(f_{i+1,j} - 2f_{i,j} + f_{i-1,j}\right) \quad , \tag{5.37}$$

and

$$f_{yy} \approx \mathcal{D}_y^- \left(\mathcal{D}_y^+ f_{i,j}\right) = \frac{1}{h_2^2}\left(f_{i,j+1} - 2f_{i,j} + f_{1,j-1}\right) \quad , \tag{5.38}$$

respectively. The mixed derivative $f_{xy} = \partial_y f_x$ is approximated in the finite difference case as

$$\begin{aligned} f_{xy} &\approx \mathcal{D}_y^0 \left(\mathcal{D}_x^0 f_{i,j}^k\right) && (5.39)\\ &= \mathcal{D}_y^0 \left(\frac{1}{2h_1}\left(f_{i+1,j}^k - f_{i-1,j}^k\right)\right) && (5.40)\\ &= \frac{1}{4h_1h_2}\left(f_{i+1,j+1}^k - f_{i-1,j+1}^k - \left(f_{i+1,j-1}^k - f_{i-1,j-1}^k\right)\right) \quad . && (5.41) \end{aligned}$$

An averaged version taking into account both time levels is then obtained via

$$\begin{aligned} f_{xy} &\approx \frac{1}{2}\left(\mathcal{D}_y^0 \left(\mathcal{D}_x^0 f_{i,j}^k\right) + \mathcal{D}_y^0 \left(\mathcal{D}_x^0 f_{i,j}^{k+1}\right)\right) && (5.42)\\ &= \frac{1}{8h_1h_2}\Big(f_{i+1,j+1}^k - f_{i-1,j+1}^k - \left(f_{i+1,j-1}^k - f_{i-1,j-1}^k\right) && (5.43)\\ &\qquad + f_{i+1,j+1}^{k+1} - f_{i-1,j+1}^{k+1} - \left(f_{i+1,j-1}^{k+1} - f_{i-1,j-1}^{k+1}\right)\Big) \quad . \end{aligned}$$

Similarly, we approximate f_{xt} by

$$\begin{aligned} f_{xt} &\approx \mathcal{D}_t^+ \left(\mathcal{D}_x^0 f_{i,j}^k\right)\\ &= \frac{1}{2h_1}\left(f_{i+1,j}^{k+1} - f_{i-1,j}^{k+1} - (f_{i+1,j}^k - f_{i-1,j}^k)\right) \quad , && (5.44) \end{aligned}$$

where $\mathcal{D}_t^+ f_{i,j}^k := f_{i,j}^{k+1} - f_{i,j}^k$ realises a temporal difference operator. Analogously we proceed for f_{yt}, yielding

$$f_{yt} \approx \frac{1}{2h_2}\left(f_{i,j+1}^{k+1} - f_{i,j-1}^{k+1} - (f_{i,j+1}^k - f_{i,j-1}^k)\right) \quad , \tag{5.45}$$

Low-order: In the low-order upwind case, the sign of the predictor $(\tilde{u}, \tilde{v})^\top$ indicates which one-sided difference should be used: We consider $\tilde{u}$ for x-derivatives and $\tilde{v}$ for y-derivatives, respectively. For approximating f_{xx} we make use of the corresponding upwind difference for f_x:

$$\tilde{u} > 0 \quad : \quad f_{xx} \approx \mathcal{D}_x^- \left(\mathcal{D}_x^- f_{i,j}^k\right) = \frac{1}{h_1^2}\left(f_{i,j}^k - 2f_{i-1,j}^k + f_{i-2,j}^k\right) \ , \tag{5.46}$$

$$\tilde{u} < 0 \quad : \quad f_{xx} \approx \mathcal{D}_x^+ \left(\mathcal{D}_x^+ f_{i,j}^k\right) = \frac{1}{h_1^2}\left(f_{i+2,j}^k - 2f_{i+1,j}^k - f_{i,j}^k\right) \ . \tag{5.47}$$

In case $\tilde{u} = 0$, we use the corresponding high-order approximation described before. For f_{yy}, we proceed accordingly, taking into account the predictor $\tilde{v}$:

$$\tilde{v} > 0 \quad : \quad f_{yy} \approx \mathcal{D}_y^- \left(\mathcal{D}_y^- f_{i,j}^k\right) = \frac{1}{h_2^2}\left(f_{i,j}^k - 2f_{i,j-1}^k + f_{i,j-2}^k\right) \ , \tag{5.48}$$

$$\tilde{v} < 0 \quad : \quad f_{yy} \approx \mathcal{D}_y^+ \left(\mathcal{D}_y^+ f_{i,j}^k\right) = \frac{1}{h_2^2}\left(f_{i,j+2}^k - 2f_{i,j+1}^k - f_{i,j}^k\right) \ . \tag{5.49}$$

Concerning the mixed derivative f_{xy} we have to use the two predictors $\tilde{u}$ and $\tilde{v}$. If a predictor is equal to zero, we use the corresponding high-order approximation, and if it is non-zero, its sign determines which one-sided upwind approximation to use. This leads to the case distinction summarised in Table 5.1. If $\tilde{u} = 0$ and $\tilde{v} = 0$ holds, we again use the high-order approximation of f_{xy}.

For f_{xt} we use the same approach as presented above in the high-order case but just use one-sided upwind differences for approximating f_x. This gives

$$\tilde{u} > 0 \quad : \quad f_{xt} \approx \mathcal{D}_t^+ \left(\mathcal{D}_x^- f_{i,j}^k\right) = \frac{1}{h_1}\left(f_{i,j}^{k+1} - f_{i-1,j}^{k+1} - (f_{i,j}^k - f_{i-1,j}^k)\right) \ , \tag{5.50}$$

$$\tilde{u} < 0 \quad : \quad f_{xt} \approx \mathcal{D}_t^+ \left(\mathcal{D}_x^+ f_{i,j}^k\right) = \frac{1}{h_1}\left(f_{i+1,j}^{k+1} - f_{i,j}^{k+1} - (f_{i+1,j}^k - f_{i,j}^k)\right) \ . \tag{5.51}$$

Accordingly we proceed for f_{yt}, resulting in

$$\tilde{v} > 0 \quad : \quad f_{yt} \approx \mathcal{D}_t^+ \left(\mathcal{D}_y^- f_{i,j}^k\right) = \frac{1}{h_2}\left(f_{i,j}^{k+1} - f_{i,j-1}^{k+1} - (f_{i,j}^k - f_{i,j-1}^k)\right) \ , \tag{5.52}$$

$$\tilde{v} < 0 \quad : \quad f_{yt} \approx \mathcal{D}_t^+ \left(\mathcal{D}_y^+ f_{i,j}^k\right) = \frac{1}{h_2}\left(f_{i,j+1}^{k+1} - f_{i,j}^{k+1} - (f_{i,j+1}^k - f_{i,j}^k)\right) \ . \tag{5.53}$$

5.2.4 Experiments

We now present experiments that illustrate the benefits of the proposed HRT scheme. In order to clearly see the influence of the numerics, we start by a considering simple, basic optic flow model and then turn to more advanced, state-of-the-art models.

Table 5.1: Upwind-type (one-sided) discretisations of the mixed derivative f_{xy}

Case	Discretisation of f_{xy}
$\tilde{u} = 0\,,\, \tilde{v} > 0$	$\frac{1}{2h_1h_2}\left(f^k_{i+1,j} - f^k_{i-1,j} - (f^k_{i+1,j-1} - f^k_{i-1,j-1})\right)$
$\tilde{u} = 0\,,\, \tilde{v} < 0$	$\frac{1}{2h_1h_2}\left(f^k_{i+1,j+1} - f^k_{i-1,j+1} - (f^k_{i+1,j} - f^k_{i-1,j})\right)$
$\tilde{u} > 0\,,\, \tilde{v} = 0$	$\frac{1}{2h_1h_2}\left(f^k_{i,j+1} - f^k_{i-1,j+1} - (f^k_{i,j-1} - f^k_{i-1,j-1})\right)$
$\tilde{u} < 0\,,\, \tilde{v} = 0$	$\frac{1}{2h_1h_2}\left(f^k_{i+1,j+1} - f^k_{i,j+1} - (f^k_{i+1,j-1} - f^k_{i,j-1})\right)$
$\tilde{u} > 0\,,\, \tilde{v} > 0$	$\frac{1}{h_1h_2}\left(f^k_{i,j} - f^k_{i-1,j} - (f^k_{i,j-1} - f^k_{i-1,j-1})\right)$
$\tilde{u} > 0\,,\, \tilde{v} < 0$	$\frac{1}{h_1h_2}\left(f^k_{i,j+1} - f^k_{i-1,j+1} - (f^k_{i,j} - f^k_{i-1,j})\right)$
$\tilde{u} < 0\,,\, \tilde{v} > 0$	$\frac{1}{h_1h_2}\left(f^k_{i+1,j} - f^k_{i,j} - (f^k_{i+1,j-1} - f^k_{i,j-1})\right)$
$\tilde{u} < 0\,,\, \tilde{v} < 0$	$\frac{1}{h_1h_2}\left(f^k_{i+1,j+1} - f^k_{i,j+1} - (f^k_{i+1,j} - f^k_{i,j})\right)$

Horn and Schunck Model

Our first experiments are conducted with one of the most simple variational optic flow models, the one proposed in the seminal work of Horn and Schunck [HS81]. The corresponding energy functional models the brightness constancy assumption in the data term and uses a homogeneous regulariser in the smoothness term. In the original formulation only greyscale images are considered, leading to the energy

$$E(u,v) = \int_\Omega \left(\boldsymbol{w}^\top \boldsymbol{J}_0\, \boldsymbol{w} + \alpha \, \left(|\boldsymbol{\nabla}_2 u|^2 + |\boldsymbol{\nabla}_2 v|^2 \right) \right) \mathrm{d}x\, \mathrm{d}y \;\;, \tag{5.54}$$

using the motion tensor notation introduced in Section 4.2.

At first we consider a simple synthetic sequence we have created, see Figure 5.2. The sequence depicts a rectangle that is displaced by one pixel to the right and one pixel to the bottom, as shown in the ground truth flow field in Figure 5.2 (c). In the same figure we also compare the results obtained with two different derivative approximations: *(i)* A standard scheme using central derivative approximations of second order, and *(ii)* our proposed adaptive HRT scheme. Additionally, we show two error maps that visualise the AAE and where brighter pixels correspond to larger errors. To ease comparison of these maps, we visualise in Figure 5.2 (d) the difference of the error maps. Here, green encodes improvements of the HRT scheme over the standard scheme and red

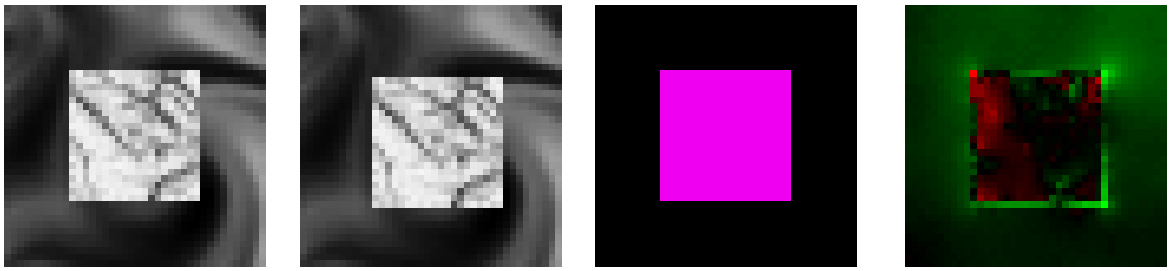
Table 5.2: Error measures (AAE) for several sequences using the method of Horn and Schunck. We compare a standard derivative approximation to our proposed adaptive HRT scheme.

	Rectangle	Marble	Yosemite	Street
Standard	31.93°	9.11°	10.72°	9.38°
HRT	28.40°	8.50°	9.53°	9.00°

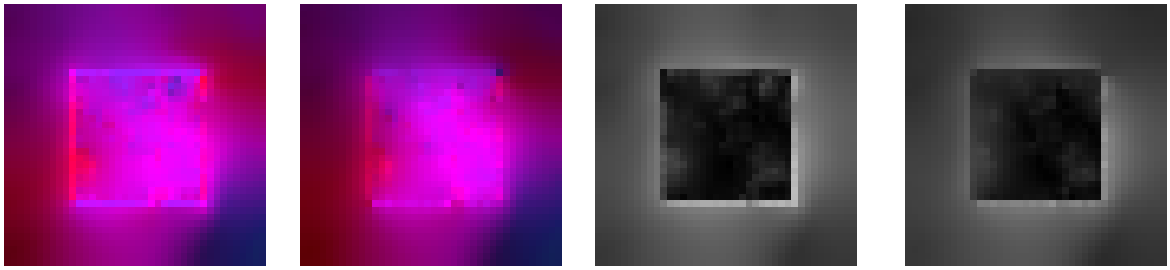
encodes impairments. The brightness of the pixels corresponds to the magnitude of the difference. Inspecting the error maps and their difference, the expected benefits of the HRT scheme become obvious: Especially at the lower and right boundary of the rectangle, i.e. at regions with large image discontinuities, the HRT scheme reduces the error. This observation is validated by the AAE measures given in Table 5.2.

This table also lists the AAE for more complex sequences, like the *Marble*, the *Yosemite without clouds*, and the *Street* sequence. The latter is available at `http://of-eval.sourceforge.net`. For these sequences, a comparison of error maps resulting from a standard derivative approximation and our adaptive scheme is shown in Figure 5.3. It turns out that the HRT scheme allows to decrease the errors in regions with strong discontinuities. For *Marble* this is the case at the ground floor, for *Yosemite* we see an improvement at the lower left boundary, and for *Street* the error decreases at the leaves of the tree.

Finally, we wish to evaluate if the upwind directions were correctly estimated using the predictor solution. To this end, Figure 5.4 visualises the errors by greyscale maps where black pixels mark locations with correct upwind direction, i.e. the sign of the predictor solution equals the sign of the ground truth flow in this locations. The grey pixels mark locations with an error in the direction estimation. The brightness of the pixels corresponds to the magnitude of the ground truth flow at this location. The latter is motivated by the observation that errors in the upwind direction estimation are on the one hand hardly avoidable, but also less harmful, if the flow is very small. Inspecting the error maps in Figure 5.4, we verify that the computed predictor allows for a very reliable estimation of the correct upwind directions. Note that we do not show an error map for our previous experiment in Figure 5.2 as the error is zero there. For the upcoming experiments, we also refrain from showing error maps for the upwind direction as this is difficult when using a multiscale warping framework (see Section 5.3.1) where flow increments are computed at each level.



(a) First frame **(b)** Second frame **(c)** Ground truth **(d)** Difference between (g) and (h)



(e) Standard scheme **(f)** HRT scheme **(g)** Error map of (e) **(h)** Error map of (f)

Figure 5.2: Results for the method of Horn and Schunck on our *Rectangle* sequence. We compare a standard derivative approximation to our adaptive HRT scheme. *First row, from left to right:* **(a)** First frame. **(b)** Second frame. **(c)** Ground truth. **(d)** Difference of the error maps in (g) and (h). Green marks improvements, red impairments of our HRT scheme compared to the standard scheme. *Second row, from left to right:* **(e)** Flow field with a standard derivative approximation. **(f)** Same with our adaptive HRT scheme. **(g)** Error map for (e). **(h)** Error map for (f).

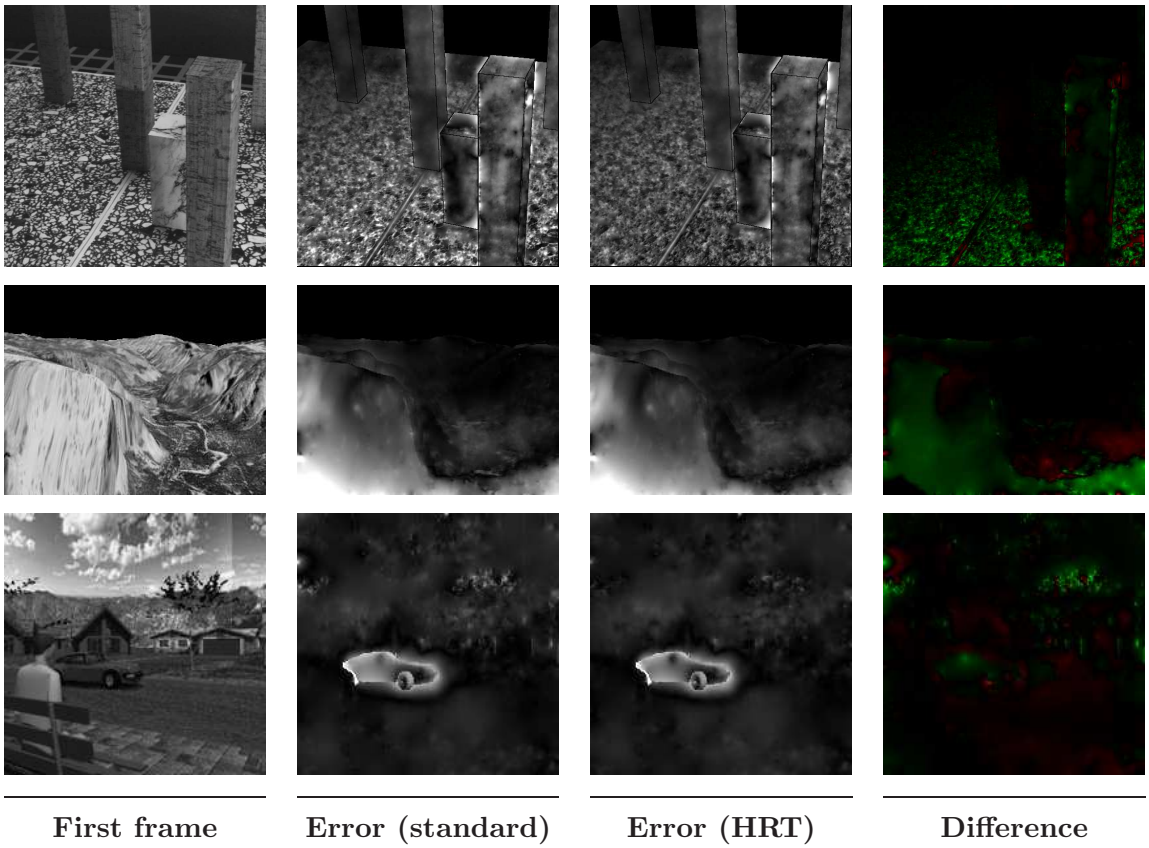


Figure 5.3: Results for the method of Horn and Schunck. We compare a standard derivative approximation to our proposed adaptive HRT scheme. *From top to bottom: Marble*, *Yosemite without clouds* and *Street* sequence. *From left to right:* First frame, error map with a standard derivative approximation, same for our adaptive HRT scheme, difference of the two error maps. Green marks improvements, red impairments of our HRT scheme compared to the standard scheme.

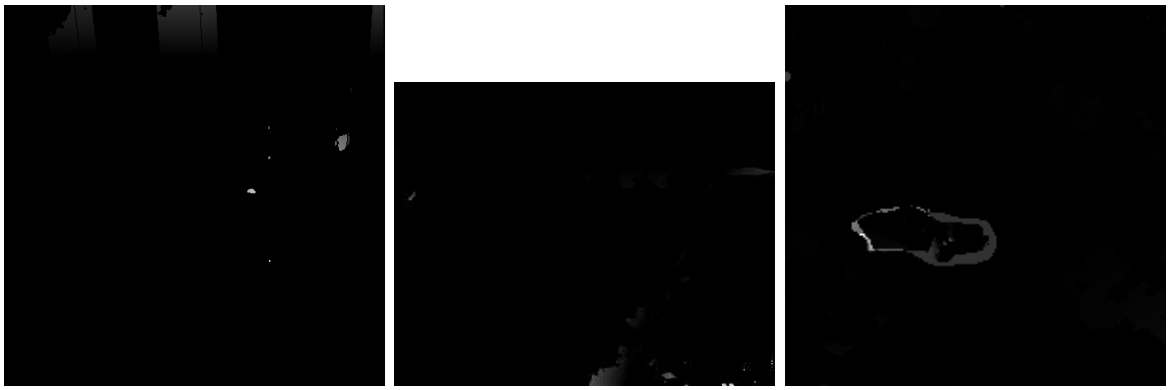


Figure 5.4: Error in the estimated upwind directions for the sequences shown in Figure 5.3. Black pixels mark correct direction. Grey pixels mark errors, where the brightness encodes to the magnitude of the ground truth flow at the corresponding location. *From left to right:* **(a)** *Marble*, **(b)** *Yosemite without clouds*, **(c)** *Street* sequence.

Table 5.3: Error measures (AAE) for several Middlebury sequences using the method of Brox *et al.*. We compare a standard derivative approximation scheme to a pure upwind scheme ($\Phi(\Theta_i) = 0$), and our proposed adaptive HRT scheme.

	Rubber-whale	Hydrangea	Dime-trodon	Grove2	Grove3	Urban2	Urban3	Venus
Standard	3.69°	2.18°	1.94°	2.49°	6.17°	2.84°	5.71°	4.79°
Upwind	4.87°	2.39°	3.06°	3.13°	6.75°	3.77°	4.58°	4.58°
HRT	3.55°	2.16°	1.88°	2.46°	6.21°	2.74°	4.11°	4.45°

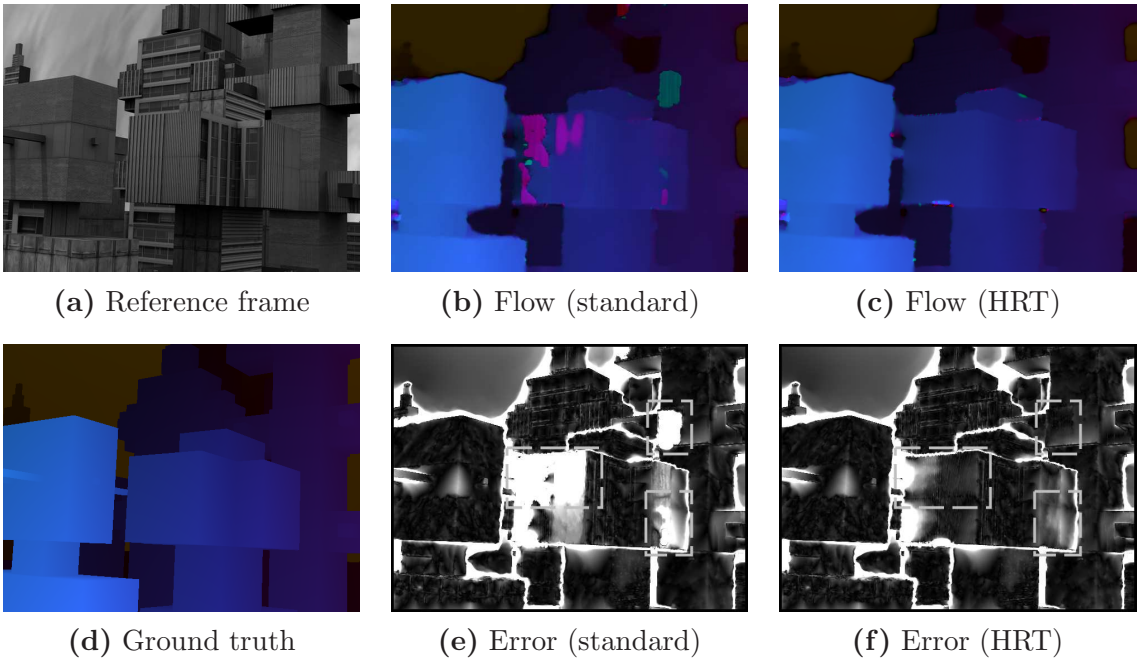
Brox *et al.* Model

In order to show results with a more advanced mode,l we conducted some experiments using our HRT scheme within the method of Brox *et al.* [BBPW04]. Here, compared to the model of Horn and Schunck [HS81], the data term additionally imposes the gradient constancy assumption and a subquadratic L_1 penalisation is performed. As smoothness term, an isotropic flow-driven TV regulariser is used, leading to the energy

$$E(u,v) = \int_\Omega \left(\Psi_M\Big(\boldsymbol{w}^\top (\boldsymbol{J}_0 + \gamma\, \boldsymbol{J}_{xy})\; \boldsymbol{w}\Big) + \alpha\; \Psi_V\big(|\boldsymbol{\nabla}_2 u|^2 + |\boldsymbol{\nabla}_2 v|^2\big) \right) \mathrm{d}x\, \mathrm{d}y \;\;, \tag{5.55}$$

with $\Psi_M(s^2) = \Psi_V(s^2) = \sqrt{s^2 + \varepsilon^2}$ and again using the motion tensor notation from Section 4.2. The minimisation of this energy is embedded in a coarse-to-fine multiscale warping approach to handle large displacements, see Section 5.3.1.

Due to the more advanced model we can now consider more challenging test data, like the Middlebury sequences that we have also used in the previous chapter. In Figure 5.5, we show results on the *Urban3* sequence. As we can see, also for the method of Brox *et al.*, the HRT scheme allows to improve the results at locations with strong discontinuities (marked in the images). In this context, we also refer to Figure 5.6 where we show a plot of the gradient magnitude to support the latter observation. The qualitative improvement is confirmed by the corresponding AAE measures in Table 5.3. This table also lists other Middlebury sequences and gives errors for a pure upwind scheme using first-order, one-sided differences. Such a scheme can easily obtained from our HRT scheme (5.36) by setting $\Phi(\Theta_i) = 0$. Analysing the results in Table 5.3 shows: *(i)* For 7 out of the 8 test sequences, the HRT scheme improves the quality over the standard scheme. Only for *Grove 3*, the standard scheme is slightly better. *(ii)* For the considered challenging sequences, the blending between high-order and low-order approximations of the HRT scheme gives significantly better results than a pure upwind scheme. Note that for some sequences, e.g. *Dimetrodon*, the upwind scheme produces significantly worse results. This occurs in the presence of large smooth areas in the images. There, central derivative approximations are best suited and thus an upwind scheme that uses the less appropriate first-order approximation leads to bad results.



(a) Reference frame **(b)** Flow (standard) **(c)** Flow (HRT)

(d) Ground truth **(e)** Error (standard) **(f)** Error (HRT)

Figure 5.5: Results for the method of Brox *et al.* on the *Urban3* sequence. We compare a standard derivative approximation to our proposed adaptive HRT scheme. *First row, from left to right:* **(a)** Reference frame. **(b)** Flow field with a standard derivative approximation. **(c)** Same with our adaptive HRT scheme. *Second row, from left to right:* **(d)** Ground truth flow. **(e)** Error map with a standard derivative approximation. **(f)** Same with our adaptive HRT scheme. Boxes indicate regions of significantly better results with the HRT scheme.

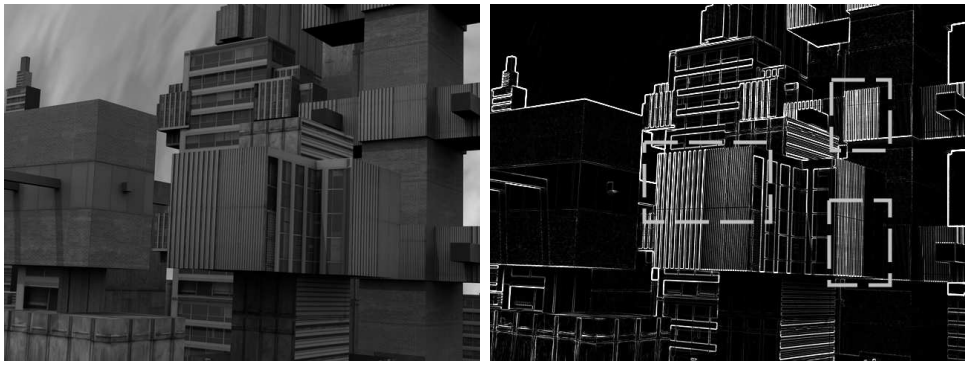


Figure 5.6: Plot of the gradient magnitude for the *Urban3* sequence. *From left to right:* **(a)** Reference frame. **(b)** Corresponding gradient magnitude, scaled to the range from 0 to 255. The boxes indicate regions where the HRT scheme performs significantly better than the standard approach, see Figure 5.5. We observe that these are regions with strong gradients.

Table 5.4: Error measures (AAE) for several Middlebury sequences using our Complementary Optic Flow method. We compare a standard derivative approximation scheme to a pure upwind scheme ($\Phi(\Theta_i) = 0$), and our proposed adaptive HRT scheme.

	Rubber-whale	Hydrangea	Dime-trodon	Grove2	Grove3	Urban2	Urban3	Venus
Standard	3.47°	2.30°	1.84°	2.54°	5.34°	2.74°	3.09°	5.26°
Upwind	4.40°	2.45°	2.87°	3.02°	5.72°	4.23°	3.25°	5.24°
HRT	3.28°	2.27°	1.87°	2.51°	5.34°	2.75°	2.96°	5.41°

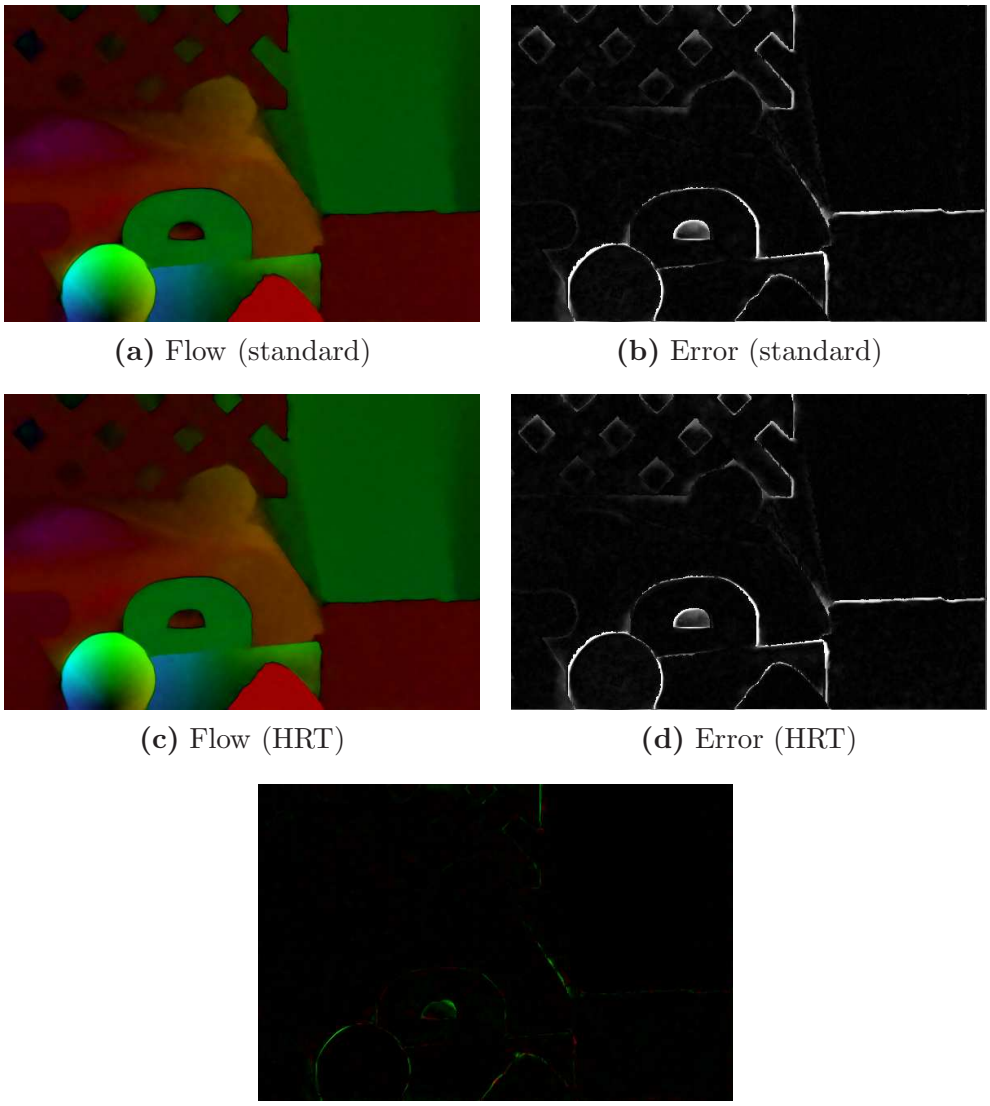
Complementary Optic Flow Model

We conclude our experiments on derivative approximations by applying the HRT scheme within our Complementary Optic Flow model [10, 11] that has been presented in the previous chapter.

In Figure 5.7, we compare results for the *Rubberwhale* sequence obtained with the HRT scheme to results with a standard central derivative approximation of second order. Note that we usually used central differences of fourth order for this model, but as the high-order approximations in our HRT scheme are of second order, comparing to fourth-order approximations would not be fair. To illustrate the improvements of our HRT scheme, we again give error maps visualising the AAE and show a colour-coded map visualising the differences in error. Considering the difference map, we see that our HRT scheme improves the result especially at the edges in the flow field. This improvement becomes also visible in the corresponding AAE measures that are given in Table 5.4, which also lists results for other Middlebury sequences and further encompasses results obtained with a pure upwind scheme. Comparing the error measures, it turns out that the HRT scheme allows to improve the results for 4 out of the 8 sequences, in 2 cases the results are almost identical, and for the remaining 2 sequences, the results of our HRT scheme are slightly worse. Compared to the upwind scheme, the HRT scheme again gives considerably better results.

Discussion

We wish to stress one important conclusion that we can draw from comparing the results of above experiments. Obviously, the advanced numerics of our HRT scheme gives a much stronger improvement in terms of quality when applied within *basic* models, e.g. in the Horn and Schunck model [HS81]. It seems that the more sophisticated the continuous models become, the less influential is an appropriate numerical approximation of derivatives, as shown in the results with our Complementary Optic Flow model [10, 11].



(a) Flow (standard)

(b) Error (standard)

(c) Flow (HRT)

(d) Error (HRT)

(e) Difference in error

Figure 5.7: Comparison of a "standard", central derivative approximation of second order to our proposed HRT scheme (on the *Rubberwhale* sequence). *Top row, from left to right:* **(a)** Flow field obtained with standard derivative approximation. **(b)** Corresponding error map visualising the AAE (brighter pixels mark higher errors). *Middle row, from left to right:* **(c)** Flow field obtained with our proposed HRT scheme. **(d)** Corresponding error map. *Bottom row:* **(e)** Difference of the error maps in (b) and (d). Green marks improvements, red impairments of our HRT scheme compared to the standard scheme.

5.3 Solving the Euler-Lagrange Equations

After our detailed discussion on discretisation aspects, we now consider the actual solution of the Euler-Lagrange equations. We start by presenting the coarse-to-fine warping strategy that is used for handling large displacements and which leads to modified Euler-Lagrange equations. Then we discuss various solvers for the modified equations.

5.3.1 Handling Large Displacements – Warping

The derivation of the optic flow constraint (4.8) by means of a Taylor linearisation is only valid under the assumption of either small displacements in the image sequence or for very smooth image data. However, if the temporal sampling of the sequence is too coarse, the precondition of small displacements is violated and a linearised approach will fail. To overcome this problem, a popular strategy [Ana89, BA96, BBPW04, MP98] is to use a coarse-to-fine multiscale warping strategy that uses downsampled images: At each warping level ℓ, the flow field is split into $(u^\ell, v^\ell, 1)^\top + (\delta u^\ell, \delta v^\ell, 0)^\top$, where $(u^\ell, v^\ell, 1)^\top =: \boldsymbol{w}^\ell$ denotes the already computed solution from coarser levels and $(\delta u^\ell, \delta v^\ell, 0)^\top$ is the unknown flow increment to be computed at the current level. As the increments are small, they can be computed by our discussed linearised approach. As soon as the increments are computed, the flow at the next finer level $\ell+1$ is then obtained by adding the two quantities, i.e. $(u^{\ell+1}, v^{\ell+1}, 1)^\top := (u^\ell, v^\ell, 1)^\top + (\delta u^\ell, \delta v^\ell, 0)^\top$. This proceeding is continued until the finest level (original resolution) is reached.

To illustrate how above strategy works in practise we apply it to our complementary optic flow model [10, 11] that is obtained by using our final data term (4.35) in combination with our complementary regulariser (4.82). The resulting Euler-Lagrange equations that need to be solved to obtain the increment $(\delta u^\ell, \delta v^\ell, 0)^\top$ then read as

$$
\begin{aligned}
0 \;=\; & \sum_{c=1}^{3} \Psi'_M\Big(\Big(\delta\boldsymbol{w}^\ell\Big)^\top \big(\bar{\boldsymbol{J}}^c_0\big)^\ell\, \delta\boldsymbol{w}^\ell\Big) \cdot \Big(\Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{1,1} \delta u^\ell + \Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{1,2} \delta v^\ell + \Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{1,3}\Big) \qquad (5.56)\\
+\,\gamma \Bigg(& \sum_{c=1}^{3} \Psi'_M\Big(\Big(\delta\boldsymbol{w}^\ell\Big)^\top \big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\, \delta\boldsymbol{w}^\ell\Big) \cdot \Big(\Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{1,1} \delta u^\ell + \Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{1,2} \delta v^\ell + \Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{1,3}\Big)\Bigg)\\
& -\alpha\, \mathrm{div}\Big(\boldsymbol{D}^\ell_{\mathrm{CR}}\, \boldsymbol{\nabla}_2\Big(u^\ell + \delta u^\ell\Big)\Big)\;,
\end{aligned}
$$

$$
\begin{aligned}
0 \;=\; & \sum_{c=1}^{3} \Psi'_M\Big(\Big(\delta\boldsymbol{w}^\ell\Big)^\top \big(\bar{\boldsymbol{J}}^c_0\big)^\ell\, \delta\boldsymbol{w}^\ell\Big) \cdot \Big(\Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{1,2} \delta u^\ell + \Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{2,2} \delta v^\ell + \Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{2,3}\Big) \qquad (5.57)\\
+\,\gamma \Bigg(& \sum_{c=1}^{3} \Psi'_M\Big(\Big(\delta\boldsymbol{w}^\ell\Big)^\top \big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\, \delta\boldsymbol{w}^\ell\Big) \cdot \Big(\Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{1,2} \delta u^\ell + \Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{2,2} \delta v^\ell + \Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{2,3}\Big)\Bigg)\\
& -\alpha\, \mathrm{div}\Big(\boldsymbol{D}^\ell_{\mathrm{CR}}\, \boldsymbol{\nabla}_2\Big(v^\ell + \delta v^\ell\Big)\Big)\;,
\end{aligned}
$$

where we defined $\delta \boldsymbol{w}^\ell := (\delta u^\ell, \delta v^\ell, 1)^\top$ to ease notation. Further note that the warped image sequence $f(\boldsymbol{x} + \boldsymbol{w}^\ell)$ is used in the computation of the motion tensors J^ℓ and the diffusion tensor

$$\boldsymbol{D}^\ell_{\mathrm{CR}} := \left(\boldsymbol{r}^\ell_1 | \boldsymbol{r}^\ell_2\right) \begin{pmatrix} \Psi'_V\left((u^\ell + \delta u^\ell)^2_{\boldsymbol{r}^\ell_1} + (v^\ell + \delta v^\ell)^2_{\boldsymbol{r}^\ell_1}\right) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \left(\boldsymbol{r}^\ell_1\right)^\top \\ \left(\boldsymbol{r}^\ell_2\right)^\top \end{pmatrix} . \tag{5.58}$$

The vectors $\boldsymbol{r}^\ell_1$ and $\boldsymbol{r}^\ell_2$ are the eigenvectors of the regularisation tensor R^ℓ_ρ which also needs to be computed from the warped image sequence $f(\boldsymbol{x} + \boldsymbol{w}^\ell)$.

We want to stress that Brox *et al.* [BBPW04] showed that the presented warping scheme is a mathematically consistent strategy for minimising an energy functional with a data term where no linearisation is performed. Such a version of our final data term (4.35) is given by

$$\begin{aligned} M(u,v) = & \sum_{c=1}^{3} \Psi_M\left(\theta^c_0 \left(f^c(\boldsymbol{x}+\boldsymbol{w}) - f^c(\boldsymbol{x})\right)^2\right) \\ & + \gamma \left(\sum_{c=1}^{3} \Psi_M\left(\theta^c_x \left(f^c_x(\boldsymbol{x}+\boldsymbol{w}) - f^c_x(\boldsymbol{x})\right)^2 + \theta^c_y \left(f^c_y(\boldsymbol{x}+\boldsymbol{w}) - f^c_y(\boldsymbol{x})\right)^2 \right) \right) . \end{aligned} \tag{5.59}$$

As large displacements occur in most real world image sequences, all results shown in this thesis have been computed using the presented warping scheme.

Implementation Details

To evaluate the images at the warped subpixel locations $f(\boldsymbol{x}+\boldsymbol{w}^\ell)$ we again use Coons patches based on bicubic interpolation [Coo67].

To obtain a coarse representation of our problem, we downsample the image sequence by a factor $\eta \in [0.5, 1)$. In order to prevent aliasing problems, we apply a low-pass filter to the images prior to downsampling [SW49, Uns00]. The latter is realised by a Gaussian convolution with standard deviation $\sqrt{2}/(4\eta)$. The actual downsampling is then realised by an area-based resampling strategy [BWF+05] that operates on the smoothed images. In most cases, the aliasing artefacts will not significantly deteriorate the flow estimation, which can be attributed to the robust data term. However, for the *Urban* sequence from the official Middlebury benchmark[BSL+10], anti-aliasing is crucial for obtaining reasonable results, see Figure 5.8. As it turns out, the large displacement of the building in the lower left corner can only be estimated when using anti-aliasing. We explain this by the high frequent stripe pattern on the façade of the building.

In order to warp the image sequence at the next finer grid, we need to upsample the already computed flow field. This is again done by area-based resampling.

Adapting the Smoothness Weight to the Warping Level. The influence of the data term usually becomes smaller at coarser levels of our multiscale framework. This

(a) Frame 10 **(b)** Frame 11 **(c)** No anti-aliasing **(d)** Anti-aliasing

Figure 5.8: Importance of anti-aliasing on the example of the *Urban* sequence. *From left to right:* **(a)** Frame 10. **(b)** Frame 11. **(c)** Our result without anti-aliasing. **(d)** Same with the proposed anti-aliasing strategy. All results used the fixed parameters $\alpha = 500.0, \sigma = 0.5, \gamma = 20.0, \rho = 4.0, \eta = 0.95$. For this sequence, no ground truth is publicly available.

is due to the smoothing properties of the downsampling that leads to smaller values of the image gradients at coarse levels. Such a behaviour is in fact desirable as the data term might not be reliable at coarse levels. Our proposed data term normalisation leads, however, to image gradients that are approximately in the same range at each level. To recover the previous reduction of the data term at coarse levels, we propose to adapt the smoothness weight α to the warping level ℓ. This is achieved by setting $\alpha(k) = \alpha/\eta^\ell$ which results in larger values of α and an emphasis of the smoothness term at coarse levels.

5.3.2 Different Solution Strategies

Let us now turn to solution strategies for the Euler-Lagrange equations (5.56) and (5.57) that have to be solved at each level of the warping scheme.

5.3.3 Elliptic Approach

A common solution strategy is to discretise the *elliptic* Euler-Lagrange equations and then solve the arising system of equations.

Discretising the Euler-Lagrange equations on the one hand requires to discretise the motion tensor entries $\boldsymbol{J}$ which comes down to approximating the occurring image derivatives. We realise this by using finite difference approximations, see Section 5.1.1. If not stated otherwise, we used central difference approximations of fourth order in our experiments. The discretisation of the flow vectors (u^ℓ, v^ℓ) as well as $(\delta u^\ell, \delta v^\ell)$ is straightforward and was shortly sketched in our preliminaries chapter, see Section 3.4. To denote the discrete versions of the flow fields evaluated at pixel (i, j), we use the notation $[\cdot]_{i,j}$, e.g. $u^\ell(x, y)$ will be discretised as $[u^\ell]_{i,j}$. Similarly, the discrete version of the motion tensor entry $[J]_{m,n}(x, y)$ in row m and column n is denoted by $[J]_{m,n;\,i,j}$. Finally, we need to discretise the divergence expressions $\operatorname{div}(\boldsymbol{D}^\ell_{\mathrm{CR}} \boldsymbol{\nabla}_2(u^\ell + \delta u^\ell))$ and $\operatorname{div}(\boldsymbol{D}^\ell_{\mathrm{CR}} \boldsymbol{\nabla}_2(v^\ell + \delta v^\ell))$. As shown in Section 5.1.1 we can write the discrete divergence expressions as a weighted sum of differences. In the present case, i.e. within the warping

framework, we get

$$\operatorname{div}\Big(\boldsymbol{D}^{\ell}_{\mathrm{CR}}\,\boldsymbol{\nabla}_2\big(u^{\ell}+\delta u^{\ell}\big)\Big) \approx \sum_{\tilde{i},\tilde{j}\in\mathcal{N}_{i,j}} [w]^{\ell}_{\tilde{i},\tilde{j}}\,\Big(\big[\delta u^{\ell}\big]_{\tilde{i},\tilde{j}}-\big[\delta u^{\ell}\big]_{i,j}+\big[u^{\ell}\big]_{\tilde{i},\tilde{j}}-\big[u^{\ell}\big]_{i,j}\Big) \quad ,(5.60)$$

$$\operatorname{div}\Big(\boldsymbol{D}^{\ell}_{\mathrm{CR}}\,\boldsymbol{\nabla}_2\big(v^{\ell}+\delta v^{\ell}\big)\Big) \approx \sum_{\tilde{i},\tilde{j}\in\mathcal{N}_{i,j}} [w]^{\ell}_{\tilde{i},\tilde{j}}\,\Big(\big[\delta v^{\ell}\big]_{\tilde{i},\tilde{j}}-\big[\delta v^{\ell}\big]_{i,j}+\big[v^{\ell}\big]_{\tilde{i},\tilde{j}}-\big[v^{\ell}\big]_{i,j}\Big) \quad , (5.61)$$

where the weights $[w]^{\ell}_{\tilde{i},\tilde{j}}$ now need to be computed with the current flow estimate $(u^{\ell}+\delta u^{\ell}, v^{\ell}+\delta v^{\ell})$ as well as with the warped image sequence $f(\boldsymbol{x}+\boldsymbol{w}^{\ell})$.

With this notation, a discrete version of the Euler-Lagrange equations (5.56) and (5.57) reads as

$$\begin{aligned}
0 = \quad & \sum_{c=1}^{3} \Psi'_M\Big(\big[\delta\boldsymbol{w}^{\ell}\big]^{\top}_{i,j}\Big[\big(\bar{\boldsymbol{J}}^{c}_{0}\big)^{\ell}\Big]_{i,j}\big[\delta\boldsymbol{w}^{\ell}\big]_{i,j}\Big) \\
& \qquad \cdot\Big(\Big[\big(\bar{\boldsymbol{J}}^{c}_{0}\big)^{\ell}\Big]_{1,1;\,i,j}\big[\delta u^{\ell}\big]_{i,j}+\Big[\big(\bar{\boldsymbol{J}}^{c}_{0}\big)^{\ell}\Big]_{1,2;\,i,j}\big[\delta v^{\ell}\big]_{i,j}+\Big[\big(\bar{\boldsymbol{J}}^{c}_{0}\big)^{\ell}\Big]_{1,3;\,i,j}\Big) \\
+\,\gamma\,\Bigg(& \sum_{c=1}^{3} \Psi'_M\Big(\big[\delta\boldsymbol{w}^{\ell}\big]^{\top}_{i,j}\Big[\big(\bar{\boldsymbol{J}}^{c}_{xy}\big)^{\ell}\Big]_{i,j}\big[\delta\boldsymbol{w}^{\ell}\big]_{i,j}\Big) \\
& \qquad \cdot\Big(\Big[\big(\bar{\boldsymbol{J}}^{c}_{xy}\big)^{\ell}\Big]_{1,1;\,i,j}\big[\delta u^{\ell}\big]_{i,j}+\Big[\big(\bar{\boldsymbol{J}}^{c}_{xy}\big)^{\ell}\Big]_{1,2\,,i,j}\big[\delta v^{\ell}\big]_{i,j}+\Big[\big(\bar{\boldsymbol{J}}^{c}_{xy}\big)^{\ell}\Big]_{1,3;\,i,j}\Big)\Bigg) \\
-\alpha & \sum_{\tilde{i},\tilde{j}\in\mathcal{N}_{i,j}} [w]^{\ell}_{\tilde{i},\tilde{j}}\,\Big(\big[\delta u^{\ell}\big]_{\tilde{i},\tilde{j}}-\big[\delta u^{\ell}\big]_{i,j}+\big[u^{\ell}\big]_{\tilde{i},\tilde{j}}-\big[u^{\ell}\big]_{i,j}\Big) \quad ,
\end{aligned} \tag{5.62}$$

$$\begin{aligned}
0 = \quad & \sum_{c=1}^{3} \Psi'_M\Big(\big[\delta\boldsymbol{w}^{\ell}\big]^{\top}_{i,j}\Big[\big(\bar{\boldsymbol{J}}^{c}_{0}\big)^{\ell}\Big]_{i,j}\big[\delta\boldsymbol{w}^{\ell}\big]_{i,j}\Big) \\
& \qquad \cdot\Big(\Big[\big(\bar{\boldsymbol{J}}^{c}_{0}\big)^{\ell}\Big]_{1,2;\,i,j}\big[\delta u^{\ell}\big]_{i,j}+\Big[\big(\bar{\boldsymbol{J}}^{c}_{0}\big)^{\ell}\Big]_{2,2;\,i,j}\big[\delta v^{\ell}\big]_{i,j}+\Big[\big(\bar{\boldsymbol{J}}^{c}_{0}\big)^{\ell}\Big]_{2,3;\,i,j}\Big) \\
+\,\gamma\,\Bigg(& \sum_{c=1}^{3} \Psi'_M\Big(\big[\delta\boldsymbol{w}^{\ell}\big]^{\top}_{i,j}\Big[\big(\bar{\boldsymbol{J}}^{c}_{xy}\big)^{\ell}\Big]_{i,j}\big[\delta\boldsymbol{w}^{\ell}\big]_{i,j}\Big) \\
& \qquad \cdot\Big(\Big[\big(\bar{\boldsymbol{J}}^{c}_{xy}\big)^{\ell}\Big]_{1,2;\,i,j}\big[\delta u^{\ell}\big]_{i,j}+\Big[\big(\bar{\boldsymbol{J}}^{c}_{xy}\big)^{\ell}\Big]_{2,2\,,i,j}\big[\delta v^{\ell}\big]_{i,j}+\Big[\big(\bar{\boldsymbol{J}}^{c}_{xy}\big)^{\ell}\Big]_{2,3;\,i,j}\Big)\Bigg) \\
-\alpha & \sum_{\tilde{i},\tilde{j}\in\mathcal{N}_{i,j}} [w]^{\ell}_{\tilde{i},\tilde{j}}\,\Big(\big[\delta v^{\ell}\big]_{\tilde{i},\tilde{j}}-\big[\delta v^{\ell}\big]_{i,j}+\big[v^{\ell}\big]_{\tilde{i},\tilde{j}}-\big[v^{\ell}\big]_{i,j}\Big) \quad .
\end{aligned} \tag{5.63}$$

Solving the discretised Euler-Lagrange equations (5.62) and (5.63) for the unknowns δu^{ℓ} and δv^{ℓ} comes down to solving a large nonlinear system of equations. Note that

large here means that for an image sequences with a resolution of 1 Megapixel, storing the system matrix would require several Terabyte of memory. Thus, a direct solution is intractable and one has to resort to an iterative solver, such as the Jacobi, Gauss-Seidel or SOR solver [You03]. Let us briefly explain how to apply such iterative solvers to our problem.

The first observation is that the mentioned iterative solvers can only handle *linear* systems of equations, but our discretised Euler-Lagrange equations (5.62) and (5.63) lead to nonlinear systems as the unknowns δu^ℓ and δv^ℓ occur in the argument of the Ψ'_M function and implicitly also in the weights $[w]^\ell_{\tilde{i},\tilde{j}}$ that are computed from the diffusion tensor $\boldsymbol{D}^\ell_{\mathrm{CR}}$. A common remedy to this problem is to embed the solution of the nonlinear system into a series of linear systems that can be solved by the discussed methods. This strategy is known as time-lagged nonlinearity, or Kačanov–Galerkin method [FKN73]. To obtain the linear systems, we evaluate the nonlinear terms with a current solution of the flow field (initially we use a zero solution) and after some iterations of the linear solver, we recompute the nonlinearities with the updated solution. Then the iterative solver is again applied to the updated linear system. These steps are iterated in two nested loops (the outer loop updates the nonlinearities, the inner loop updates the solution by some solver iterations) until some stopping criterion, usually a maximal number of inner and outer iterations, is met.

Let us briefly sketch how the discussed iterative solvers for the linear systems work. We consider a general linear equation system given as

$$\boldsymbol{A}\,\boldsymbol{u} = \boldsymbol{b} \ , \tag{5.64}$$

where $\boldsymbol{A} \in \mathbb{R}^{2N\times 2N}$ denotes the block-diagonal system matrix (with $N := n_x \cdot n_y$ denoting the number of pixels), $\boldsymbol{u} \in \mathbb{R}^{2N}$ is the unknown flow field, and the right-hand side is denoted by $\boldsymbol{b} \in \mathbb{R}^{2N}$. As the matrix $\boldsymbol{A}$ is positive definite for non-constant image sequences, the solution of the system is given by

$$\boldsymbol{u} = \boldsymbol{A}^{-1}\,\boldsymbol{b} \ , \tag{5.65}$$

which requires to compute the inverse $\boldsymbol{A}^{-1}$ of the system matrix. This, however, is in general a difficult and time-consuming task.

The basic idea of iterative solvers is to find an efficient way to compute an approximation of this inverse. A first step towards this goal is the decomposition of the system matrix into $\boldsymbol{A} = \boldsymbol{A}_1 + \boldsymbol{A}_2$. With this, we can rewrite (5.64) as

$$\boldsymbol{A}_1\,\boldsymbol{u} + \boldsymbol{A}_2\,\boldsymbol{u} = \boldsymbol{b} \ . \tag{5.66}$$

This in turn enables us to come up with the fixed point iteration

$$\boldsymbol{A}_1\,\boldsymbol{u}^{k+1} = \boldsymbol{b} - \boldsymbol{A}_2\,\boldsymbol{u}^k \ , \tag{5.67}$$

where k denotes the iteration index, i.e. $\boldsymbol{u}^k$ is the already computed solution and $\boldsymbol{u}^{k+1}$ is the unknown result of the next iteration. Next, we rewrite equation (5.67) as

$$\boldsymbol{u}^{k+1} = \boldsymbol{A}_1^{-1}(\boldsymbol{b} - \boldsymbol{A}_2\,\boldsymbol{u}^k) \ , \tag{5.68}$$

where $\boldsymbol{A}_1^{-1}$ should be a reasonable and efficient to compute approximation of $\boldsymbol{A}^{-1}$.

A frequent approach for an efficient approximation of $\boldsymbol{A}^{-1}$ bases on the decomposition $\boldsymbol{A} = \boldsymbol{A}_d - \boldsymbol{A}_l - \boldsymbol{A}_u$, where $\boldsymbol{A}_d$ denotes the diagonal part, $\boldsymbol{A}_l$ the strictly lower triangular part and $\boldsymbol{A}_u$ the strictly upper triangular part of $\boldsymbol{A}$. A popular solver relying on this decomposition is the Gauss-Seidel method, which is similar to the Jacobi method, but more efficient. The Gauss-Seidel method defines $\boldsymbol{A}_1 = \boldsymbol{A}_d - \boldsymbol{A}_l$, which is a triangular matrix and is hence easy to invert. Consequently, $\boldsymbol{A}_2 = -\boldsymbol{A}_u$, as we imposed that $\boldsymbol{A} = \boldsymbol{A}_1 + \boldsymbol{A}_2 = \boldsymbol{A}_d - \boldsymbol{A}_l - \boldsymbol{A}_u$. Using these decomposition in the fixed point scheme (5.68) yields

$$\boldsymbol{u}^{k+1} = (\boldsymbol{A}_d - \boldsymbol{A}_l)^{-1}(\boldsymbol{b} + \boldsymbol{A}_u\, \boldsymbol{u}^k) \ . \tag{5.69}$$

For our Euler-Lagrange equations (5.62) and (5.63), this results in the following Gauss-Seidel iteration step

$$\begin{aligned}
\left[\delta u^\ell\right]_{i,j}^{k+1} = \Bigg[& \sum_{c,a} \left(\gamma(a) \left[\Psi'_M \left(\bar{\boldsymbol{J}}_a^c\right)\right]_{i,j}^{\ell,k} \left(\left[\left(\bar{\boldsymbol{J}}_a^c\right)^\ell \right]_{1,3;\,i,j} + \left[\left(\bar{\boldsymbol{J}}_a^c\right)^\ell \right]_{1,2;\,i,j} \left[\delta v^\ell\right]_{i,j}^k \right) \right) \\
& - \alpha \left(\sum_{\tilde{i},\tilde{j} \in \mathcal{N}_{i,j}^-} [w]_{\tilde{i},\tilde{j}}^{\ell,k} \left(\left[\delta u^\ell\right]_{\tilde{i},\tilde{j}}^{k+1} + \left[u^\ell\right]_{\tilde{i},\tilde{j}} \right) + \sum_{\tilde{i},\tilde{j} \in \mathcal{N}_{i,j}^+} [w]_{\tilde{i},\tilde{j}}^{\ell,k} \left(\left[\delta u^\ell\right]_{\tilde{i},\tilde{j}}^{k} + \left[u^\ell\right]_{\tilde{i},\tilde{j}} \right) \right) \Bigg] \\
& \cdot \left[\alpha\, [w]_{i,j}^{\ell,k} - \left(\sum_{c,a} \left(\gamma(a) \left[\Psi'_M \left(\bar{\boldsymbol{J}}_a^c\right)\right]_{i,j}^{\ell,k} \left[\left(\bar{\boldsymbol{J}}_a^c\right)^\ell \right]_{1,1;\,i,j} \right) \right) \right]^{-1} ,
\end{aligned} \tag{5.70}$$

$$\begin{aligned}
\left[\delta v^\ell\right]_{i,j}^{k+1} = \Bigg[& \sum_{c,a} \left(\gamma(a) \left[\Psi'_M \left(\bar{\boldsymbol{J}}_a^c\right)\right]_{i,j}^{\ell,k} \left(\left[\left(\bar{\boldsymbol{J}}_a^c\right)^\ell \right]_{2,3;\,i,j} + \left[\left(\bar{\boldsymbol{J}}_a^c\right)^\ell \right]_{1,2;\,i,j} \left[\delta u^\ell\right]_{i,j}^k \right) \right) \\
& - \alpha \left(\sum_{\tilde{i},\tilde{j} \in \mathcal{N}_{i,j}^-} [w]_{\tilde{i},\tilde{j}}^{\ell,k} \left(\left[\delta v^\ell\right]_{\tilde{i},\tilde{j}}^{k+1} + \left[v^\ell\right]_{\tilde{i},\tilde{j}} \right) + \sum_{\tilde{i},\tilde{j} \in \mathcal{N}_{i,j}^+} [w]_{\tilde{i},\tilde{j}}^{\ell,k} \left(\left[\delta v^\ell\right]_{\tilde{i},\tilde{j}}^{k} + \left[v^\ell\right]_{\tilde{i},\tilde{j}} \right) \right) \Bigg] \\
& \cdot \left[\alpha\, [w]_{i,j}^{\ell,k} - \left(\sum_{c,a} \left(\gamma(a) \left[\Psi'_M \left(\bar{\boldsymbol{J}}_a^c\right)\right]_{i,j}^{\ell,k} \left[\left(\bar{\boldsymbol{J}}_a^c\right)^\ell \right]_{2,2;\,i,j} \right) \right) \right]^{-1} ,
\end{aligned} \tag{5.71}$$

where we initialise the iterations with zero, i.e. $\delta u^{\ell,0} = 0$ and $\delta v^{\ell,0} = 0$ and the weights $[w]^{\ell,k}$ are now computed from a diffusion tensor $\boldsymbol{D}_{\mathrm{CR}}^{\ell,k}$ that is obtained with the current flow estimate $\boldsymbol{w}^\ell + \delta\boldsymbol{w}^{\ell,k}$. Note that we furthermore introduced the following abbreviations to ease notation: The sum

$$\sum_{c,a} := \sum_{c=1}^{3} \sum_{a \in \{0, xy\}} \ , \tag{5.72}$$

sums over the image channels c and the two constancy assumptions $a \in \{0, xy\}$, where 0 denotes the brightness constancy assumption and xy the gradient constancy assumption. The parameter $\gamma(a)$ is defined as $\gamma(0) := 1$ and $\gamma(xy) := \gamma$. To denote the derivative of the penaliser in the data term, we define

$$\left[\Psi'_M \left(\bar{\boldsymbol{J}}_a^c\right)\right]_{i,j}^{\ell,k} := \Psi'_M \left(\left[\delta\boldsymbol{w}^\ell\right]_{i,j}^{k\ \top} \left[\left(\bar{\boldsymbol{J}}_a^c\right)^\ell \right]_{i,j} \left[\delta\boldsymbol{w}^\ell\right]_{i,j}^k \right) \ . \tag{5.73}$$

Finally, $\mathcal{N}_{i,j}^-$ and $\mathcal{N}_{i,j}^+$ denote the already processed and the still to be processed neighbours of (i,j), respectively.

A popular strategy for speeding up the presented Gauss-Seidel scheme is to perform an overrelaxation as in the *SOR* (*Successive Overrelaxation*) method. The basic idea of the latter method is to extrapolate the result of one Gauss-Seidel step. If we denote by $\boldsymbol{u}_{GS}^{k+1}$ the result of a Gauss-Seidel iteration, then the SOR scheme reads as

$$\boldsymbol{u}^{k+1} \;=\; (1-\omega)\,\boldsymbol{u}^k + \omega\,\boldsymbol{u}_{GS}^{k+1} \;=\; \boldsymbol{u}^k + \omega\left(\boldsymbol{u}_{GS}^{k+1} - \boldsymbol{u}^k\right) \;, \tag{5.74}$$

with an overrelaxation parameter $\omega \in [0,2)$. For $\omega = 1$ one falls back to the Gauss-Seidel scheme. We propose to set $\omega = 1.95$ which gives good results and allows for a speedup of two orders of magnitude compared to the original Gauss-Seidel scheme.

A further speed up can be achieved when updating the nonlinearities $[\Psi'_M(\bar{\boldsymbol{J}}_a^c)]_{i,j}^{\ell,k}$ and $[w]_{\tilde{i},\tilde{j}}^{\ell,k}$ not after every iteration, but keeping them fixed for some iterations. This is implemented using two nested loops, as described in the context of the time-lagged nonlinearity method (see page 107).

Multigrid Solver

A further speed up compared to the SOR solver can be achieved by using highly efficient multigrid solvers, see e.g. [Bra77, BWKS06, Hac85, KKR07, Ter86]. For the experiments in this thesis, we used the full multigrid solver of Bruhn *et al.* [BWKS06], unless stated otherwise.

Full multigrid schemes combine ideas from unidirectional and bidirectional multigrid methods. The basic strategy of unidirectional methods is to start the iterative solution with a good initialisation to reduce the required number of iterations. To obtain this initialisation, the problem is transferred to a smaller (coarser) pixel grid and solved there, which should be relatively fast. The coarse solution is then interpolated and serves as initialisation on the next finer grid. Consequently, unidirectional methods belong to the class of coarse-to-fine strategies. In contrast to unidirectional schemes, bidirectional strategies do not only go from coarse to fine levels, but also revisit coarser levels to obtain correction steps. This is motivated by the fact that iterative solvers eliminate high frequency errors relatively fast, but require a very large number of iterations until low-frequent errors vanish. However, the low frequencies occur as high frequencies on coarser grids. There, one then computes the *error* (not the solution) that was still present in the fine grid solution and then corrects the fine grid solution for the estimated error.

For our experiments, we use a Gauß-Seidel type solver with alternating line relaxation [BWKS06] to solve the problems arising at each grid level. Please also note that an in-detail explanation of sophisticated multigrid schemes would by far exceed the scope of this thesis. For a comprehensive overview on multigrid methods for optic flow computation, we refer the interested reader to [Bru06].

5.3.4 Parabolic Approach

An alternative to the previously discussed elliptic approach is to apply a gradient descent scheme on the Euler-Lagrange equations. This comes down to solving a system of two coupled diffusion-reaction equations of *parabolic* type. The latter can be derived from the Euler-Lagrange equations (5.56) and (5.57) and read as

$$\begin{aligned}\partial_\theta\,\delta u^\ell \;=\; & \operatorname{div}\Big(\boldsymbol{D}^\ell_{\mathrm{CR}}\,\boldsymbol{\nabla}_2\Big(u^\ell+\delta u^\ell\Big)\Big) \\ & -\frac{1}{\alpha}\Bigg[\sum_{c=1}^{3}\Psi'_M\Big(\Big(\delta\boldsymbol{w}^\ell\Big)^\top\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\,\delta\boldsymbol{w}^\ell\Big)\cdot\Big(\Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{1,1}\delta u^\ell+\Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{1,2}\delta v^\ell+\Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{1,3}\Big) \\ & +\gamma\left(\sum_{c=1}^{3}\Psi'_M\Big(\Big(\delta\boldsymbol{w}^\ell\Big)^\top\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\,\delta\boldsymbol{w}^\ell\Big)\cdot\Big(\Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{1,1}\delta u^\ell+\Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{1,2}\delta v^\ell+\Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{1,3}\Big)\right)\Bigg]\;,\end{aligned} \tag{5.75}$$

$$\begin{aligned}\partial_\theta\,\delta v^\ell \;=\; & \operatorname{div}\Big(\boldsymbol{D}^\ell_{\mathrm{CR}}\,\boldsymbol{\nabla}_2\Big(v^\ell+\delta v^\ell\Big)\Big) \\ & -\frac{1}{\alpha}\Bigg[\sum_{c=1}^{3}\Psi'_M\Big(\Big(\delta\boldsymbol{w}^\ell\Big)^\top\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\,\delta\boldsymbol{w}^\ell\Big)\cdot\Big(\Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{1,2}\delta u^\ell+\Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{2,2}\delta v^\ell+\Big[\big(\bar{\boldsymbol{J}}^c_0\big)^\ell\Big]_{2,3}\Big) \\ & +\gamma\left(\sum_{c=1}^{3}\Psi'_M\Big(\Big(\delta\boldsymbol{w}^\ell\Big)^\top\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\,\delta\boldsymbol{w}^\ell\Big)\cdot\Big(\Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{1,2}\delta u^\ell+\Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{2,2}\delta v^\ell+\Big[\big(\bar{\boldsymbol{J}}^c_{xy}\big)^\ell\Big]_{2,3}\Big)\right)\Bigg]\;,\end{aligned} \tag{5.76}$$

where $\partial_\theta\,\delta u^\ell$ and $\partial_\theta\,\delta v^\ell$ denote derivatives in a *numerical time* direction. To derive above diffusion reaction equations (5.75) and (5.76) from the Euler-Lagrange equations (5.56) and (5.57), we proceeded in two steps: First we replace the zeros on the left-hand side by the numerical time derivatives, which makes sense as in the steady state ($\theta\to\infty$), the numerical time derivatives tends to zero. Then, we switch the signs on the right-hand side of the Euler-Lagrange equations to obtain a gradient *descent* behaviour.

To solve the diffusion-reaction equations (5.75) and (5.76), it is common to use *iterative schemes*. The latter compute the sought flow increment $\delta\boldsymbol{w}^\ell$ starting from an initialisation $\delta\boldsymbol{w}^{\ell,0}:=(\delta u^{\ell,0},\delta v^{\ell,0},1)^\top$. As before, we initialise the iterations with zero. The result of the next iteration $\delta\boldsymbol{w}^{\ell,k+1}:=(\delta u^{\ell,k+1},\delta v^{\ell,k+1},1)^\top$ is then computed from the previous result $\delta\boldsymbol{w}^{\ell,k}:=(\delta u^{\ell,k},\delta v^{\ell,k},1)^\top$. The iterations are stopped as soon as a stopping criterion is met, e.g. when the difference between the results of two subsequent iterations drops below a given threshold. If this is the case we can consider the scheme to have reached the desired steady state where $\partial_\theta\,\delta u^\ell\approx 0$ and $\partial_\theta\,\delta v^\ell\approx 0$.

Stabilised Explicit Scheme. In order to obtain an iterative scheme, we first discretise the numerical time derivatives by the forward differences

$$\partial_\theta\,\delta u^\ell\approx\frac{\delta u^{\ell,k+1}-\delta u^{\ell,k}}{\tau}\,,\qquad\text{and}\qquad\partial_\theta\,\delta v^\ell\approx\frac{\delta v^{\ell,k+1}-\delta v^{\ell,k}}{\tau}\;, \tag{5.77}$$

where τ denotes the numerical time step size.

The simplest form of an iterative scheme is then given by an explicit scheme that uses the old value at iteration k for all occurrences of the unknowns at the right-hand side. To stabilise such a scheme, one can use the new value at iteration $k+1$ in the reaction part. This gives the following stabilised explicit scheme (where we mark the stabilisation in red):

$$\begin{aligned}\frac{\delta u^{\ell,k+1}-\delta u^{\ell,k}}{\tau} &= \sum_{\tilde{i},\tilde{j}\in\mathcal{N}_{i,j}} [w]_{\tilde{i},\tilde{j}}^{\ell,k}\left(\left[\delta u^{\ell,k}\right]_{\tilde{i},\tilde{j}}-\left[\delta u^{\ell,k}\right]_{i,j}+\left[u^{\ell}\right]_{\tilde{i},\tilde{j}}-\left[u^{\ell}\right]_{i,j}\right) \\ &-\frac{1}{\alpha}\left[\sum_{c=1}^{3}\Psi_M'\left(\left(\delta\boldsymbol{w}^{\ell,k}\right)^{\top}\left(\bar{\boldsymbol{J}}_0^c\right)^{\ell}\delta\boldsymbol{w}^{\ell,k}\right)\cdot\left(\left[\left(\bar{\boldsymbol{J}}_0^c\right)^{\ell}\right]_{1,1}\delta u^{\ell,k+1}+\left[\left(\bar{\boldsymbol{J}}_0^c\right)^{\ell}\right]_{1,2}\delta v^{\ell,k}+\left[\left(\bar{\boldsymbol{J}}_0^c\right)^{\ell}\right]_{1,3}\right)\right. \\ &\left.+\gamma\left(\sum_{c=1}^{3}\Psi_M'\left(\left(\delta\boldsymbol{w}^{\ell,k}\right)^{\top}\left(\bar{\boldsymbol{J}}_{xy}^c\right)^{\ell}\delta\boldsymbol{w}^{\ell,k}\right)\cdot\left(\left[\left(\bar{\boldsymbol{J}}_{xy}^c\right)^{\ell}\right]_{1,1}\delta u^{\ell,k+1}+\left[\left(\bar{\boldsymbol{J}}_{xy}^c\right)^{\ell}\right]_{1,2}\delta v^{\ell,k}+\left[\left(\bar{\boldsymbol{J}}_{xy}^c\right)^{\ell}\right]_{1,3}\right)\right)\right], \end{aligned} \tag{5.78}$$

$$\begin{aligned}\frac{\delta v^{\ell,k+1}-\delta v^{\ell,k}}{\tau} &= \sum_{\tilde{i},\tilde{j}\in\mathcal{N}_{i,j}} [w]_{\tilde{i},\tilde{j}}^{\ell,k}\left(\left[\delta v^{\ell,k}\right]_{\tilde{i},\tilde{j}}-\left[\delta v^{\ell,k}\right]_{i,j}+\left[v^{\ell}\right]_{\tilde{i},\tilde{j}}-\left[v^{\ell}\right]_{i,j}\right) \\ &-\frac{1}{\alpha}\left[\sum_{c=1}^{3}\Psi_M'\left(\left(\delta\boldsymbol{w}^{\ell,k}\right)^{\top}\left(\bar{\boldsymbol{J}}_0^c\right)^{\ell}\delta\boldsymbol{w}^{\ell,k}\right)\cdot\left(\left[\left(\bar{\boldsymbol{J}}_0^c\right)^{\ell}\right]_{1,2}\delta u^{\ell,k}+\left[\left(\bar{\boldsymbol{J}}_0^c\right)^{\ell}\right]_{2,2}\delta v^{\ell,k+1}+\left[\left(\bar{\boldsymbol{J}}_0^c\right)^{\ell}\right]_{2,3}\right)\right. \\ &\left.+\gamma\left(\sum_{c=1}^{3}\Psi_M'\left(\left(\delta\boldsymbol{w}^{\ell,k}\right)^{\top}\left(\bar{\boldsymbol{J}}_{xy}^c\right)^{\ell}\delta\boldsymbol{w}^{\ell,k}\right)\cdot\left(\left[\left(\bar{\boldsymbol{J}}_{xy}^c\right)^{\ell}\right]_{1,2}\delta u^{\ell,k}+\left[\left(\bar{\boldsymbol{J}}_{xy}^c\right)^{\ell}\right]_{2,2}\delta v^{\ell,k+1}+\left[\left(\bar{\boldsymbol{J}}_{xy}^c\right)^{\ell}\right]_{2,3}\right)\right)\right], \end{aligned} \tag{5.79}$$

Solving these equations for the unknowns $\delta u^{\ell,k+1}$ and $\delta v^{\ell,k+1}$ is straightforward and gives the desired iterative scheme to be implemented.

Obviously, explicit schemes are appealing because they are very easy to deduce and to implement: In contrast to the previously described elliptic approach, it is not necessary to solve a large nonlinear system of equations, but an iterative solution scheme can directly be deduced. However, there is also a price to pay: The stability of explicit schemes can only be guaranteed when using rather small time step sizes τ. This results in a poor performance as many iterations are required to reach the desired steady state. In the considered 2D case, the time step limit τ_{lim} is given by

$$\tau_{\text{lim}} = \frac{1}{m_{\max}\left(\frac{2}{h_1^2}+\frac{2}{h_2^2}\right)}\ , \tag{5.80}$$

where we recall that h_1 and h_2 denote the grid size in x- and y-direction, respectively. By $m_{\max}$ we denote the maximum of the larger eigenvalue over all diffusion tensors $\boldsymbol{D}_{\text{CR}}^{\ell,k}$. Considering (5.80) we realise that large eigenvalues of the diffusion tensors will decrease the time step size limit even further. When using our complementary smoothness term (4.82), the eigenvalues of the resulting diffusion tensor $\boldsymbol{D}_{\text{CR}}^{\ell,k}$ are given by 1 and $\Psi_V'(s^2)$, respectively. As we use the Perona-Malik penaliser, we have $\Psi_V'(s^2)\leq 1$, see (4.65), and consequently $m_{\max}\leq 1$. In this sense, our complementary regulariser is well-suited for an explicit scheme.

Scheme	Step Sizes for 1000 Iterations		Stopping Time
Explicit	0.5 ➤ 0.5 ➤ ... ➤ 0.5 ➤ 0.5	=	500.00
FED	0.250000 ➤ 0.250001 ➤ ... ➤ 25 355.72 ➤ 101 422.61	=	166 833.33

Figure 5.9: Comparing the stopping time of a standard explicit scheme after 1000 iterations to a corresponding FED cycle. The resulting speed-up factor is 334. Note that this example considers the 1D case where the step size limit is 0.5 for an explicit scheme. *Illustration*: P. Gwosdek.

5.3.5 A Fast Explicit Solver

Our discussion on explicit schemes show that it is desirable to modify an explicit scheme s.t. larger time steps can be used and thus fewer iterations are needed until a steady state is reached. This would combine the easy implementation of explicit schemes with the efficiency of the sophisticated solvers for elliptic approaches. This goal was recently achieved by Grewenig *et al.* [GWB10]. They presented a *Fast Explicit Diffusion (FED)* scheme, an explicit scheme that uses some extremely large time steps to ensure a fast convergence. As could be shown, the combination of large (unstable) and small (stable) time steps within one *cycle* allows to guarantee unconditional stability of the complete approach. An illustrated example for an FED cycle compared to a standard explicit scheme is shown in Figure 5.9. Hence, FED schemes outperform elliptic schemes in terms of efficiency and are additionally much easier to implement, especially on massively parallel architectures like GPUs.

In our paper [3], we showed that although FED schemes were originally developed for solving diffusion problems, they can easily be adapted for solving the Euler-Lagrange equations of variational optic flow approaches. An intuitive reason for this is that the Euler-Lagrange equations actually are of diffusion-reaction type. We apply the FED scheme within our modified explicit scheme from (5.78) and (5.79) by replacing the previously fixed time step size τ by a varying FED step size τ_k which is computed as [GWB10]

$$\tau_k = \tau_{\text{lim}} \cdot \frac{1}{2\,\cos^2\!\left(\pi\,\frac{2k+1}{4n+2}\right)} \quad , \tag{5.81}$$

where n denotes the number of iterations in the current cycle.

Remarks. The result of FED iterations only makes sense at the end of a cycle, where the unstable time steps have been compensated by stable ones. We thus can only update the nonlinearities in the weights and the penaliser function Ψ'_M only if a cycle is completed and not in between a cycle.

The number of individual time steps n in a cycle is given by $\min\{n \in \mathbb{N}^+ \mid (n^2 + n)/12 \geq T\}$, where T denotes the desired stopping time of the cycle. For $n \geq 3$, one can show that an FED cycle reaches this stopping time T faster than any other explicit scheme with n stable time step sizes.

Moreover, the ordering of steps within one FED cycle is irrelevant from a theoretical point of view, but can in practise affect the influence of rounding errors on the result. However, it is possible to find permutations of the set $\{\tau_k \mid 0 \leq k < n\}$ that are more robust w.r.t. floating-point inaccuracies than others. Given the next larger prime number p to k and let $\kappa < p$, a series $\{\tau_{\tilde{k}} \mid \tilde{k} = ((k+1) \cdot \kappa) \bmod p,\ \tilde{k} < n\}$ is known to give good results [GS78, GWB10]. In order to find a suitable value for the parameter κ, we analysed a simple 1-D problem and choose the one κ that minimises the error between the FED output and the analytic reference solution. These values were once computed for all practical choices of n to set up a lookup table which is used throughout our implementation.

Further Improvements. The already high performance of our FED optic flow approach can further be boosted by using a simple unidirectional coarse-to-fine multigrid scheme as described above in Section 5.3.3. We call the resulting solver *cascadic FED* (*CFED*).

Finally, our whole approach is parallelised on a GPU using the NVidia CUDA architecture [NVI10] which is well-suited for FED and CFED solvers due to their algorithmic simplicity. To obtain a high performance despite the large amounts of data involved in our computations, we pay particular attention to an efficient use of on-chip memory to reduce transfers from and to global memory. For more details on the GPU implementation, we refer to our paper [3].

A Note on Semi-Implicit Schemes. An alternative to overcome the strict limitations on the time step size can be to use semi-implicit schemes which are in fact unconstrained in their time step sizes. There, one uses the new value from the current iteration $k + 1$ also in the divergence expression on the right-hand side. However, this makes it impossible to explicitly solve the arising equations for the unknowns $\delta u^{\ell,k+1}$ and $\delta v^{\ell,k+1}$. Instead, if one rearranges the equations s.t. all occurrences of the unknowns are shifted to the left-hand side, one again ends up with a linear system of equations that has to be solved at each iteration by some iterative solvers, e.g. the discussed SOR solver. As parabolic semi-implicit schemes play no role in the present work, we do not go into detail here.

Remarks on Stereo

The solution schemes for orthoparallel stereo can be derived from above optic flow schemes in a straightforward manner. The only differences are that: *(i)* We obtain only a single equation for the unknown δu^ℓ. *(ii)* All occurrences of the vertical flow component are removed by setting $\delta v^\ell = v^\ell = 0$. *(iii)* A stereo tensor in the form (4.40)

is used instead of the motion tensors. Here, one should keep in mind that the entry $1,1$ of the stereo and motion tensor are identical, but the entry $1,2$ of the stereo tensor is equivalent to the entry $1,3$ of the motion tensor, and the entry $2,2$ of the stereo tensor is equivalent to the entry $3,3$ of the motion tensor.

5.4 Experiments

The following experiments compare efficient CPU and GPU implementations and also show some results that could be obtained on a modern smartphone.

5.4.1 Multigrid on CPU versus CFED on GPU

We start by comparing CPU and GPU implementations. On the CPU side, we use the full multigrid solver of Bruhn et al. [BWKS06], whereas a parallel CUDA implementation of our cascadic fast explicit diffusion (CFED) solver [3] is used on a GPU architecture. We compare the two solvers on the Middlebury training data sets and use our complementary optic flow method [10, 11]. In contrast to the original formulation, we refrain from an HSV colour representation and resort to the standard RGB colour space out of simplicity and efficiency reasons. Additionally, we restrict the number of warping levels to 40 for the *Urban 2*, *Urban 3* and the *Venus* sequence, and to 10 for the remaining sequences.

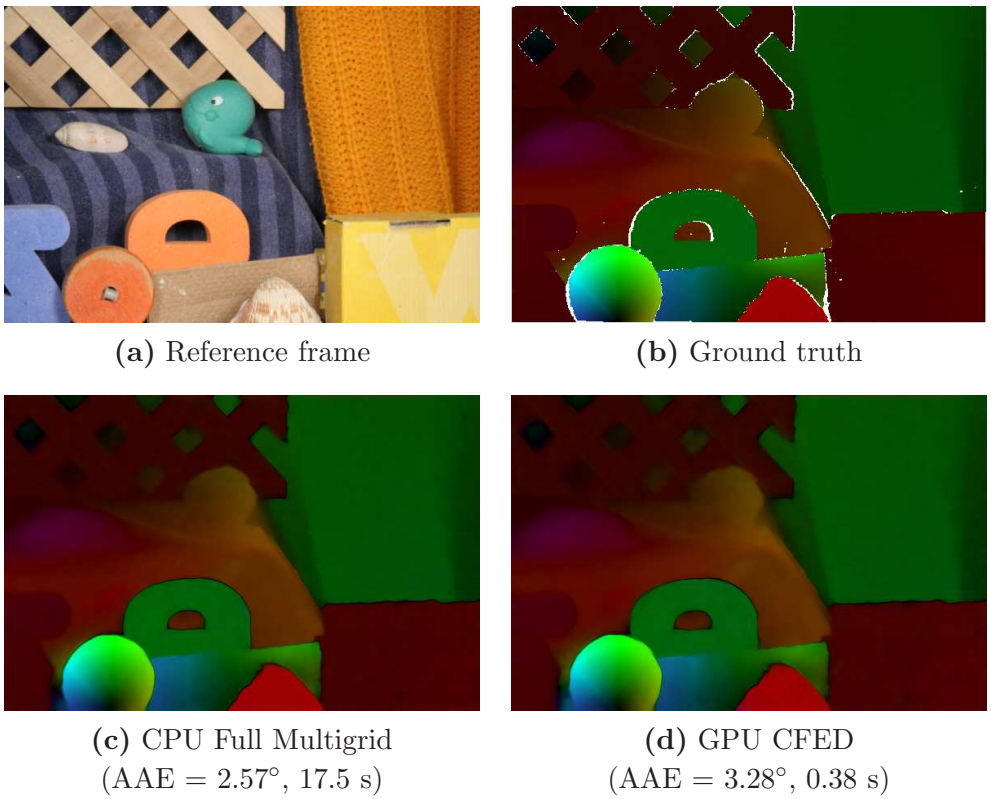
Considering the results shown in Table 5.5 and Figure 5.10, it turns out that our GPU algorithm allows to obtain speedup factors of up to 55. The quality of results can be considered to be comparable to the CPU implementation, where some degree of impairment had to be accepted for preserving the possibility of an efficient parallel implementation. We furthermore want to note that due to a better GPU occupancy, the speed up factors are even higher for larger image sizes. This is illustrated in Figure 5.11 where we plotted the runtimes for images of size ratio 4:3 for different image sizes. As we can see, our method scales linearly for image sizes larger than 0.5 Megapixel. Below, the efficiency is reduced.

5.4.2 Results on a Mobile Phone

The simple structure of an explicit scheme in combination with the fast CFED solver does not only enable implementation on powerful GPUs, but also on mobile platforms, like modern smartphones. As a proof of concept, we showed in [4] that it is possible to implement basic variational optic flow methods like the method of Horn and Schunck [HS81] and the CLG method [BWS05] on an Android smartphone. For small test images of size 316×252 we achieved interactive runtimes of about 10 seconds, see Figure 5.12.

Table 5.5: Comparison of a sequential full multigrid (FMG) solver [BWKS06] on the CPU to our parallel cascadic fast explicit diffusion (CFED) solver [3] on the GPU. We use the Middlebury training sequences and give the AAE measure together with the corresponding runtimes and the speed up factor achieved by the CFED solver. Our CPU code was executed on a Intel Pentium 4 (3.2 GHz), the GPU code on a GeForce GTX 480

Sequence	Size	FMG (CPU)	CFED (GPU)	Speedup
Rubberwhale	584 × 388	2.57° (17.5 s)	3.28° (0.38 s)	46
Hydrangea	584 × 388	1.81° (17.5 s)	2.05° (0.38 s)	46
Dimetrodon	584 × 388	1.57° (17.5 s)	1.56° (0.38 s)	46
Grove 2	640 × 480	2.43° (23.5 s)	2.43° (0.43 s)	55
Grove 3	640 × 480	5.41° (23.5 s)	5.30° (0.43 s)	55
Urban 2	640 × 480	2.59° (38.0 s)	2.74° (1.04 s)	37
Urban 3	640 × 480	2.94° (38.0 s)	3.35° (1.04 s)	37
Venus	420 × 380	3.89° (20.0 s)	4.24° (0.82 s)	24



(a) Reference frame **(b)** Ground truth

(c) CPU Full Multigrid (AAE = 2.57°, 17.5 s) **(d)** GPU CFED (AAE = 3.28°, 0.38 s)

Figure 5.10: Comparison of results on the *Rubberwhale* sequence (584×388) obtained with a sequential full multigrid (FMG) solver [BWKS06] on the CPU to our parallel cascadic fast explicit diffusion (CFED) solver [3] on the GPU. *First row, from left to right:* **(a)** Reference frame. **(b)** Ground truth (white pixels mark locations where no ground truth is available). *Second row, from left to right:* **(c)** Flow field obtained with the multigrid scheme on the CPU. **(d)** Same for our CFED scheme on the GPU.

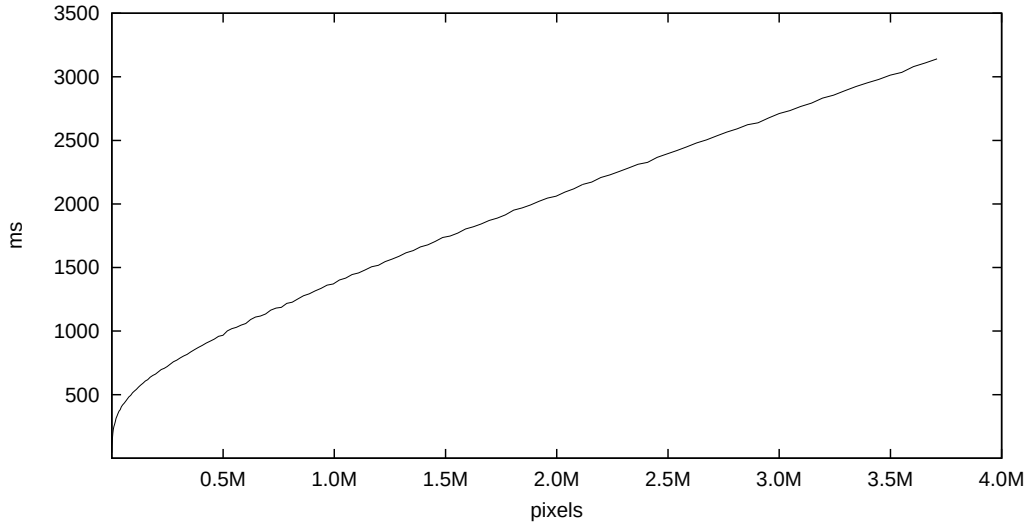
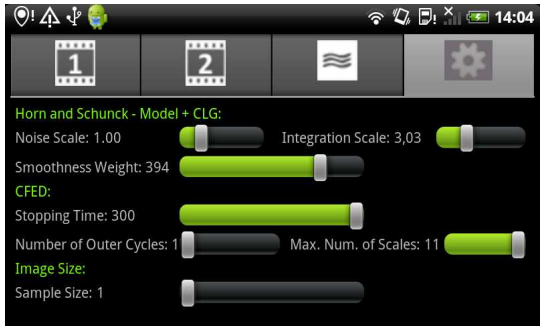
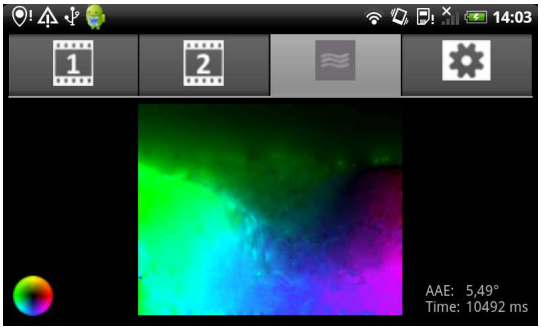


Figure 5.11: Runtime of our CFED GPU algorithm w.r.t. the image size. *Illustration*: P. Gwosdek.

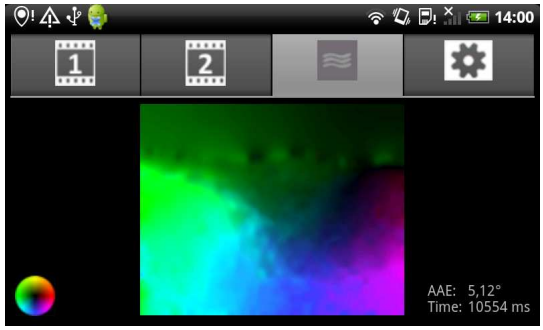
(a) First image



(b) Parameter dialog



(c) Horn and Schunck



(d) CLG

Figure 5.12: Optic flow on a mobile phone. *First row, from left to right:* **(a)** First image of the *Yosemite without clouds* sequence. **(b)** Parameter adjustment dialog. *Second row, from left to right:* **(c)** Result with the Horn and Schunck model (AAE = 5.49° in 10.5 seconds). Flow field visualised by colour code shown in bottom left corner. **(e)** Result with the CLG model (AAE = 5.12° in 10.6 seconds).

Chapter 6

Applications

> Il n'existe pas de sciences appliquées, mais seulement des applications de la science.
>
> *Louis Pasteur*

In Chapter 1, we pointed out that various applications in the fields of computer vision and image processing require to solve correspondence problems. One example is shown in the following where we consider a problem from the field of computational photography: the alignment of exposure series for HDR imaging [DM97, MP95, RWPD05, RBS99]. This comes down to estimating the displacements in the exposure series, which we realise by a modified variational optic flow approach that we presented in [9].

In the same work we leveraged the subpixel precision of the resulting flow fields to combine concepts from *super-resolution (SR)* methods [PPK03] and HDR imaging in a *joint super-resolution and HDR (SR-HDR)* technique that computes an image with enhanced dynamic range and spatial resolution.

6.1 HDR Imaging

Before presenting our alignment strategy, we first want to briefly discuss the basics of HDR imaging.

6.1.1 Why Do We Need HDR Imaging?

We usually model images as continuous functions that map from the image domain $\Omega \subset \mathbb{R}^2$ to the co-domain $\mathbb{R}$. However, digital images are inherently discrete, they are actually discrete samples of the continuous world. In Section 3.4 we discussed how the image domain Ω can be discretised. Let us for now still assume that the image domain Ω is continuous and focus on the co-domain.

Typical image formats such as JPEG, PNG or PPM reserve 1 byte, i.e. 8 bits of storage per pixel and colour channel, resulting in 256 different values for each channel. This means that the image co-domain of each channel is in fact sampled (quantised) and can be represented by the discrete set $\{0, .., 255\}$. The *dynamic range* –the ratio between smallest and largest value that can be expressed– thus spans two orders of magnitude in the discussed case. This range also correspond to the actual dynamic range of common display devices (LCD and CRT displays) as well as print media. However, the dynamic range of real world scenes is orders of magnitude larger: Outdoor scenes may exhibit a dynamic range of up to 5 orders of magnitude between shadows and highlights, or even more when a light source, e.g. the sun, is directly visible. This leads to the obvious problem that today's *low dynamic range* (*LDR*) images that have been taken with standard cameras either suffer from under-saturation (black pixels) in dark regions or from over-saturation (white pixels) in bright regions.

Note that many digital SLR cameras allow to take images in a RAW format that stores the measured data without further quantisation and allows for up to 14 bits per pixel. However, the resulting dynamic range still only spans 4 orders of magnitude which is still insufficient in difficult scenarios, as discussed above.

6.1.2 The Basics of HDR Imaging

A popular way to overcome the described problem is to consider a set of images taken with varying exposure times, a so-called *exposure series* [1]. In this way, despite using a LDR sensor, the dark parts of the scene can be captured in the long exposures and the bright parts in the short exposures, respectively. By fusing the information from a complete exposure series, HDR imaging techniques [DM97, MP95, RWPD05, RBS99] compute a single image with increased dynamic range. More precisely, the resulting image should ideally describe the actual scene radiances and thus capture the full dynamic range of the scene. This is of course only possible if the exposure series covers a sufficiently large number of exposures. To get an idea how the radiances can be computed from the exposure series, we briefly subsume in the following the HDR imaging technique of Robertson *et al.* [RBS99].

Notation and Observation Model. Assume we have taken an exposure series consisting of n_e images $f_k : \Omega \to \{0, .., 255\}$ for $k = 1, .., n_e$. The corresponding exposure times we denote by $t_k \in \mathbb{R}^+$. Our goal is now to compute a radiance map $r : \Omega \to \mathbb{R}^+$ from the given exposure series.

The basic *observation model* for HDR imaging relates the measured image intensities f_k and corresponding exposure times t_k to the unknown radiances r via

$$f_k(x, y) = \phi\Big(t_k \, r(x, y)\Big) \ , \tag{6.1}$$

[1] Also called *Wyckoff set* in honour of Charles Wyckoff who already in the 1930s combined differently exposed film layers to extend the dynamic range of photographic films.

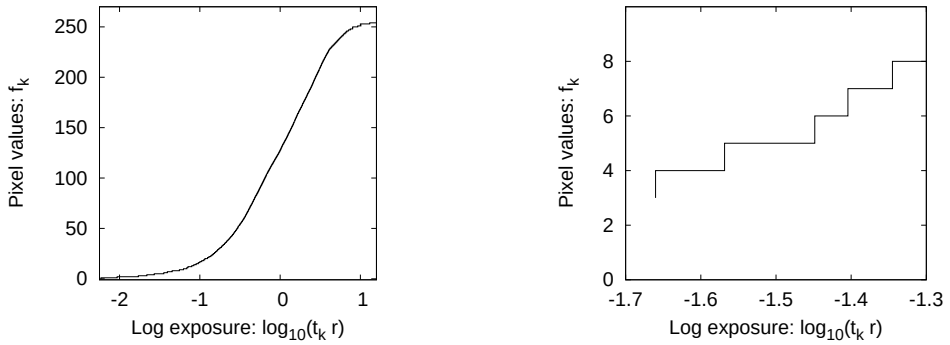


Figure 6.1: Estimated response curve for our Nikon D90 DSLR camera. It was obtained using the calibration method described in [RBS99]. We only visualise the response for the red colour channel, as it is very similar to the ones of the two other channels. *From left to right:* **(a)** Response curve (logarithmically scaled). **(b)** Close-up view.

for $k = 1, .., n_e$ and where $\phi : \mathbb{R}^+ \to \{0, .., 255\}$ denotes the *camera response function*, or *camera response curve*. The latter describes how the amount of light arriving at the image sensor (given by $t_k\, r$ and referred to as *exposure*) is mapped to a pixel value f_k. The response function can be estimated from an exposure series by a calibration procedure as for example described in [DM97, RBS99]. This has to be done only once for a specific camera and we are thus justified in assuming that the response function is known when computing the radiances.

Considering colour images, one usually assumes the same observation model for each colour channel, but with a distinct response function for each channel.

A Note on Camera Response Functions. Let us take a look at a typical response functions for digital cameras. In Figure 6.1 we show the response curve for a standard DSLR camera (Nikon D90). This camera was also used for capturing the images used in our experiments that will be shown in Sections 6.2.4 and 6.3.4.

Due to the discrete nature of digital cameras, the response curve is in fact a piecewise constant function, which means that always a whole interval of radiances is mapped to one pixel value, see Figure 6.1 (b). The size of these intervals depends on the slope of the response function. If the response function is steep, only a small interval of radiances maps to one pixel value, which is desirable. In flat regions, however, larger intervals of radiances map to the same pixel value, which reduces the reliability of the radiance computation there. If we reconsider the response function in Figure 6.1 (a), we realise that the function is rather steep in the middle range and becomes flat when approaching its extremes, 0 and 255. According to our previous argumentation, one should thus reduce the influence of rather dark and bright pixels from the radiance computation. This can easily be achieved by using a weighting function $\omega : \{0, .., 255\} \to [0, 1]$ in the final radiance computation.

In the literature, several weighting functions have been studied, see [GAW+10] for a comprehensive overview. For our experiments, we use the Gaussian-like weighting

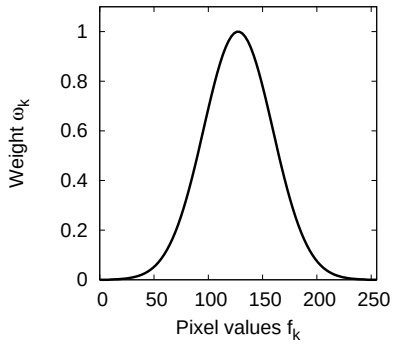


Figure 6.2: Gaussian-like weighting function.

function proposed in [RBS99]. It is given by

$$\omega\big(f_k(x,y)\big) \;=:\; \omega_k(x,y) \;:=\; \exp\left(-s\cdot\frac{\big(f_k(x,y)-127.5\big)^2}{(127.5)^2}\right) \quad , \tag{6.2}$$

with a scale parameter s. A plot of this function for $s = 8$ is shown in Figure 6.2. Furthermore, it makes sense to set $\omega = 0$ if the weight would become smaller than a threshold value, say 0.001. This completely disables the influence of less reliable pixels. Note in this context that the weight for fully saturated pixels ($f_k = 255$) should anyway be equal to zero, i.e. $\omega(255) = 0$. The reason for this is that the radiance information from saturated pixels is completely unreliable as any radiance above a certain threshold will result in a pixel value of 255.

Computing the Radiances. The sought radiances r are found by minimising an energy formulation. It solely consists of a data term that penalises deviations from the observation model (6.1) and is weighted by the discussed weight function ω. To derive the data term, let us first rewrite the observation model as

$$\begin{aligned} f_k(x,y) &= \phi\Big(t_k\, r(x,y)\Big) \\ \Rightarrow\quad \phi^{-1}\Big(f_k(x,y)\Big) &= t_k\, r(x,y) \\ \Leftrightarrow\quad \phi^{-1}\Big(f_k(x,y)\Big) - t_k\, r(x,y) &= 0 \;, \end{aligned} \tag{6.3}$$

for $k = 1, .., n_e$ and where ϕ^{-1} denotes the inverse of the camera response function. For notational convenience we further define $I_k(x,y) := \phi^{-1}(f_k(x,y))$. Using this notation and performing a quadratic penalisation of (6.3) together with the weighting yields the energy

$$E(r) = \int_\Omega \left(\sum_{k=1}^{n_e} \omega_k(x,y)\big(I_k(x,y) - t_k\, r(x,y)\big)^2\right) \mathrm{d}x\,\mathrm{d}y \;. \tag{6.4}$$

To obtain a necessary condition for a minimiser of the energy (6.4) we set its derivative w.r.t. the unknown radiance r to zero, resulting in

$$\begin{aligned} 0 \;=\; \partial_r E(r) \;&=\; -2 \sum_{k=1}^{n_e} \omega_k(x,y)\, t_k \big(I_k(x,y) - t_k\, r(x,y)\big) \\ &=\; -2 \left(\sum_{k=1}^{n_e} \omega_k(x,y)\, t_k\, I_k(x,y) - \left(\sum_{k=1}^{n_e} \omega_k(x,y)\, t_k^2 \right) r \right) \quad . \end{aligned} \tag{6.5}$$

Solving (6.5) for r finally yields

$$r(x,y) \;=\; \frac{\displaystyle\sum_{k=1}^{n_e} \omega_k(x,y)\, t_k\, I_k(x,y)}{\displaystyle\sum_{k=1}^{n_e} \omega_k(x,y)\, t_k^2} \quad . \tag{6.6}$$

If *colour images* are to be processed one applies above formula for each channel independently while considering the individual response functions of each channel.

Let us finally remark that the obtained radiances are in fact just *relative* radiances, i.e. scaled versions of the actual radiances. This scaled information is of course sufficient to solve the LDR problem as well as for most further applications of HDR images, such as image-based modelling, image processing and compositing [DM97].

Displaying HDR Images – Tone Mapping. One issue in HDR imaging is that the dynamic range of current display devices is much lower than the dynamic range of the HDR radiance maps obtained with the described method. The same holds true for any print media. Consequently, one first needs to compress the dynamic range of HDR images before displaying them, which is known as *tone mapping*.

Simple tone mapping methods apply the same transform to all pixels in the image and are thus classified as *global* operators. One of the most basic representatives of this type is a logarithmic scaling that can be written as

$$\mathcal{T}_{\log}\big(r(x,y)\big) = T_{\max} \cdot \frac{\log\left(1 - r_{\min} + r(x,y)\right)}{\log\left(1 - r_{\min} + r_{\max}\right)} \quad , \tag{6.7}$$

where $T_{\max}$ denotes the desired maximum of the tone mapped result (often $T_{\max} = 255$) and $r_{\min}$ and $r_{\max}$ are the minimum and maximum of all radiance values, respectively.

Global tone mapping algorithms are fast and easy to implement, but often discard lots of the fine scale details present in the original HDR image. To overcome this problem, more sophisticated tone mapping strategies have been proposed that try to compress the overall dynamic range, while preserving as much details as possible. One example for such a method is the *gradient domain* approach of Fattal *et al.* [FLW02].

Here, the gradients of the radiance map are considered and a scaling function is applied that compresses large gradients (reduces the dynamic range), while boosting small gradients (preserving fine details). The tone mapped result is then obtained by integrating the modified gradient field by solving the Poisson equation. We used this tone mapping operator from [FLW02] for displaying all HDR results shown in this thesis. Specifically, we use the implementation of the *pfstools* package [MKMS07] (`http://pfstools.sourceforge.net`) with the parameters $\alpha = 1.0, \beta = 0.95, s = 0.8$. Please do not confuse this parameters with the correspondent parameters in this thesis. After the tone mapping, we additionally perform a gamma correction with $\gamma = 2.0$.

Note that a comprehensive survey on other tone mapping techniques would exceed the scope of this work and we refer the interested reader to [RWPD05].

6.2 Aligning Exposure Series for HDR Imaging

Taking a closer look at equation (6.6) reveals that the HDR result at a certain position is obtained by computing a weighted average over the radiances computed from the exposure series at this very location. Such a strategy naturally requires perfectly aligned images without displacements between them. This assumption is, however, hardly fulfilled in real world scenarios due to moving objects (persons, clouds, etc.) or camera shake.

In [9] we tackled this problem by adapting a modern optic flow technique to estimate the displacements in the exposure series. This approach seems promising due to the robustness, high accuracy and efficiency of recent optic flow methods, which was also demonstrated in the previous chapters of this thesis. Note that as soon as the displacements are known, one can align the exposure series by compensating the images for the displacements.

6.2.1 Variational Optic Flow on Exposure Series

Assume we are given an exposure series $f_k : \Omega \to \mathbb{R}$ for $k = 1, \dots, n_e$. Our goal is now to estimate the displacements $\boldsymbol{u}_k := (u_k, v_k)^\top$ between a reference image f_r and the images f_k. Note that for $k = r$, we have $\boldsymbol{u}_k = \mathbf{0}$, of course. Some methods, e.g. [MG07] define the reference image f_r as the one with the highest entropy. Assuming that the exposure series is ordered w.r.t. exposure times, the image with the highest entropy will mostly be the middle image in the series. We are thus justified in defining $r := \lceil m/2 \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function. The latter is defined as $\lceil x \rceil = \min\{n \in \mathbb{Z} \mid n \geq x\}$.

To use the notation we have introduced for optic flow, we can use the same idea as in the stereo case and embed f_k and f_r in an image sequence. This results in $f_r(x, y) \equiv f(x, y, t)$ and $f_k(x, y) \equiv f(x, y, t+1)$. This allows to compute the sought displacement

(a) g_1 (1/30 s) **(b)** g_3 (1/80 s) **(c)** $\partial_x g_1$ **(d)** $\partial_x g_3$

Figure 6.3: Visualising the x-derivative for two images of an exposure series. *From left to right:* **(a)** Image 1 (exposure time 1/30 s) **(b)** Image 3 (exposure time 1/80 s) **(c)** Visualisation of the x-derivative of the red channel from the image in (a). **(d)** Same for (b).

fields $\boldsymbol{u}_k$ by minimising an energy functional of our general form (4.1):

$$E(u_k, v_k) = \int_\Omega \big(M(u_k, v_k) + \alpha\, V(\boldsymbol{\nabla}_2 u_k, \boldsymbol{\nabla}_2 v_k)\big)\, \mathrm{d}x\, \mathrm{d}y \ . \tag{6.8}$$

Data Term. An appropriate design of the data term $M(u_k, v_k)$ is mandatory for obtaining reasonable results given images with different exposures. Note for example that imposing brightness constancy does not make sense at all in this case. We propose to use the gradient constancy assumption (4.17) [BBPW04, Sch94, TP84] in our data term which results in using edge information for the alignment, see Fig. 6.3. As noted in [War03], edges are not completely invariant under exposure changes, but our experiments in Section 6.2.4 will show that they give a sufficient cue when used within a robust variational approach.

To render our approach robust against outliers caused by saturation problems, noise or occlusions, we perform a subquadratic penalisation of the gradient constancy assumption. Furthermore, we normalise the data term to prevent an undesirable overweighting that leads to artefacts in the displacement fields. Incorporating these concepts (that we discussed in detail in Section 4.2) leads to the data term

$$M(u_k, v_k) \ = \ \Psi_M \left(\begin{pmatrix} \boldsymbol{u}_k^\top, 1 \end{pmatrix} \ \bar{\boldsymbol{J}}_{xy}^{\mathcal{C}} \ \begin{pmatrix} \boldsymbol{u}_k \\ 1 \end{pmatrix} \right) \ , \tag{6.9}$$

where we use the robust penaliser $\Psi_M(s^2) := \sqrt{s^2 + \varepsilon^2}$ as proposed in [BBPW04]. The motion tensor is defined as

$$\bar{\boldsymbol{J}}_{xy}^{\mathcal{C}} \ := \ \sum_{c=1}^{3} \bar{\boldsymbol{J}}_{xy}^{c} \tag{6.10}$$

$$:= \ \sum_{c=1}^{3} \left[\theta_x^c \left(\boldsymbol{\nabla}_3 f_x^c \ (\boldsymbol{\nabla}_3 f_x^c)^\top \right) + \theta_y^c \left(\boldsymbol{\nabla}_3 f_y^c \ (\boldsymbol{\nabla}_3 f_y^c)^\top \right) \right] \ , \tag{6.11}$$

where the superscript c denotes the RGB colour channels and the normalisation factors θ are defined in (4.22).

In preliminary experiments, we also investigated data terms that impose constancy of the radiances to obtain invariance under exposure changes. This, however, requires a preceding camera calibration step and knowledge of the exposure times as the radiances are computed via applying the inverse camera response function to the pixel values and then dividing by the exposure time. Since the radiances make no sense in saturated regions, and due to possible errors in the calibration, working in the radiance domain alone produced unsatisfactory results. By additionally imposing constancy of the radiance gradients we could improve the results. However, we could not obtain better results than with our proposed approach, which is simpler and more efficient as it omits calibration and radiance computations. Note that due to the latter, our approach also allows to align image sets for exposure fusion [MKR09] where no camera response curve or exposure times are given.

Also note that using a HSV colour representation as proposed in Section 4.2 does not make sense for exposure series. Despite the fact that the exposure changes are multiplicative and that the hue as well as the saturation channel are invariant under such changes, the nonlinearity introduced by the camera response function ϕ causes problems. Thus, one would first need to apply the inverse response function which is undesirable, as discussed above.

Smoothness Term. Concerning the smoothness term, we opt for an isotropic flow-driven regulariser [Sch94, SH89]

$$V(\boldsymbol{\nabla}_2 u_k, \boldsymbol{\nabla}_2 v_k) \;=\; \Psi_V\big(|\boldsymbol{\nabla}_2\, u_k|^2 + |\boldsymbol{\nabla}_2\, v_k|^2\big) \quad , \tag{6.12}$$

using the regularised total variation (TV) penaliser [ROF92] given by $\Psi_V(s^2) = \sqrt{s^2 + \varepsilon^2}$. This discontinuity-preserving smoothness term allows for sharp edges in the flow field, which is important for a favourable alignment, see our experiment in Figure 6.8. We also experimented with more advanced smoothness terms, like our complementary regulariser [10, 11], but they could not improve the results, as will be shown in the experiment in Figure 6.9.

We further wish to note that the smoothness term plays a very important role in HDR alignment tasks: Flat image regions are omnipresent in exposure series due to over- and undersaturation problems. In this regions, however, the data term does not give any information and the smoothness term has to fill in the flow field there.

6.2.2 Energy Minimisation

As before, we minimise the energy by solving the associated Euler-Lagrange equations which read as

$$\sum_{c=1}^{3}\left(\Psi'_M\left(\left(\boldsymbol{u}_k^\top,1\right)\,\bar{\boldsymbol{J}}^{\mathcal{C}}_{xy}\begin{pmatrix}\boldsymbol{u}_k\\1\end{pmatrix}\right)\cdot\left(\left[\bar{\boldsymbol{J}}^{c}_{xy}\right]_{1,1}u_k+\left[\bar{\boldsymbol{J}}^{c}_{xy}\right]_{1,2}v_k+\left[\bar{\boldsymbol{J}}^{c}_{xy}\right]_{1,3}\right)\right) \tag{6.13}$$
$$-\alpha\,\operatorname{div}\Big(\Psi'_V\big(|\boldsymbol{\nabla}_2\,u_k|^2+|\boldsymbol{\nabla}_2\,v_k|^2\big)\;\boldsymbol{\nabla}_2 u_k\Big)=0\;,$$

$$\sum_{c=1}^{3}\left(\Psi'_M\left(\left(\boldsymbol{u}_k^\top,1\right)\,\bar{\boldsymbol{J}}^{\mathcal{C}}_{xy}\begin{pmatrix}\boldsymbol{u}_k\\1\end{pmatrix}\right)\cdot\left(\left[\bar{\boldsymbol{J}}^{c}_{xy}\right]_{1,2}u_k+\left[\bar{\boldsymbol{J}}^{c}_{xy}\right]_{2,2}v_k+\left[\bar{\boldsymbol{J}}^{c}_{xy}\right]_{2,3}\right)\right) \tag{6.14}$$
$$-\alpha\,\operatorname{div}\Big(\Psi'_V\big(|\boldsymbol{\nabla}_2\,u_k|^2+|\boldsymbol{\nabla}_2\,v_k|^2\big)\;\boldsymbol{\nabla}_2 v_k\Big)=0\;.$$

In order to handle large displacements (which often occur in HDR imaging due to significant camera shake or fast moving objects), we embed the solution of the Euler-Lagrange equations in the warping framework described in Section 5.3.1. For an efficient solution of the problem at each warping level, we propose two strategies, based on the available hardware architecture: On a sequential CPU architecture, we use the full multigrid solver of Bruhn *et al.* [BWKS06], which we briefly discussed in Section 5.3.3. For parallel GPU architectures, we use the cascadic FED solver from our paper [3], which is described in Section 5.3.5.

One problem with applying the FED scheme for solving above Euler-Lagrange equations is the following: When using a TV penaliser in the smoothness term, the eigenvalues of the diffusion tensor are given by

$$\Psi'_V(s^2)\;=\;\frac{1}{2\sqrt{s^2+\varepsilon^2}}\;\leq\;\frac{1}{2\varepsilon}\;=:\;\lambda_{\max}\;. \tag{6.15}$$

For a standard choice of $\varepsilon = 0.001$ the maximal eigenvalue $m_{\max}$ is thus equal to 500, which significantly restricts the stability limit of an explicit solver, see (5.80). As a consequence, the time steps that can be chosen by the FED scheme are tangibly smaller.

Fortunately, there is a simple remedy to this problem by using the Charbonnier penaliser function [CBFAB94] which is defined as

$$\Psi_V(s^2)=2\,\varepsilon^2\sqrt{1+\frac{s^2}{\varepsilon^2}}\;, \tag{6.16}$$

in the data and the smoothness term. As one can easily show, see Appendix A, the resulting energy is up to a constant factor equal to our proposed energy that uses the subquadratic function $\Psi_V(s^2)=\sqrt{s^2+\varepsilon^2}$ in the data and smoothness term. The minimum of both energies is thus the same, yielding the same solution. On the other hand, the derivative of the Charbonnier function is given by

$$\Psi'_V(s^2)\;=\;\frac{1}{\sqrt{1+\frac{s^2}{\varepsilon^2}}}\;\leq\;1\;, \tag{6.17}$$

Figure 6.4: A freehand exposure series for HDR imaging. Note that severe displacements due to moving clouds and camera shake. *From left to right:* **(a)** Image 1 of the exposure series. **(b)** Image 2. **(c)** Image 3 (reference image). **(d)** Image 4. **(e)** Image 5.

which results in $m_{\max} = 1$ and allows for a favourable performance of the FED scheme.

6.2.3 Actual Alignment

Having estimated the displacements $(u_k, v_k)^\top$, we can align the exposure series by performing a simple motion compensation. This comes down to replacing each image $f_k(x, y)$ by

$$f_k^{\text{align}}(x, y) \;:=\; f_k(x+u_k, y+v_k) \;. \tag{6.18}$$

6.2.4 Experiments

Comparison to Other Alignment Approaches. In a first experiment, we compare our proposed optic flow-based HDR alignment method to several competing methods. To this end, a real world exposure series was taken freehand using a standard DSLR camera (Nikon D90). The series consists of 5 images taken with exposure times between 1/250 to 1/30 seconds and the image size was rescaled to 512×340 pixels. Some images of the series are depicted in the first row of Figure 6.4 where we want to point out that the series suffers from severe displacements due to camera shake and moving clouds.

After aligning the images with the different approaches, we compute HDR reconstructions following the approach in [RBS99] that we have also briefly reviewed before. To improve the reconstruction quality, we incorporate information on the camera response curve, which we estimated beforehand by the method also described in [RBS99]. Implementations of the before mentioned algorithms can be found in the *pfstools* package [MKMS07] (`http://pfstools.sourceforge.net`).

Let us now compare our proposed method to some other alignment strategies. These are *(i)* the global method of Ward [War03], *(ii)* the `align_image_stack` algorithm from the *Hugin* toolkit which implements a homography-based approach as in [TM07], *(iii)* a variant of our method without data term normalisation, and *(iv)* the hierarchical block matching technique of Menzel and Guthe [MG07].

In Figure 6.5 (b) we show the result obtained with the global alignment strategy from [War03], where the black border marks pixels warped outside the image domain. A corresponding zoom is shown in Figure 6.6 (b). It becomes clear that the estimated global translation completely fails in describing the complex displacements between the exposures. We also tried to apply a local optic flow approach [LK81] to the globally aligned images after transferring them to the radiance domain. We found that such a global-local strategy as in [KUWS03] leads to unusable results due to the poor global alignment. Also the `align_image_stack` algorithm, see Figure 6.5 (c) and Figure 6.6 (c), fails to correctly align the images. Obviously, the homography computed from feature matches is not expressive enough to describe the displacements in the exposure series.

In Figure 6.5 (d) we show the result of our method, but without using data term normalisation. Especially when considering the zoom shown in Figure 6.6 (d), and comparing it to our final result in Figure 6.6 (f), we realise that the proposed normalisation allows to resolve any problems with unpleasant artefacts that arise due to an overweighting of the data term at image edges. This is further substantiated by the flow fields shown in Figure 6.7 (a) and (b). Of course, a larger smoothness weight α would also resolve this problem, but creates an oversmoothed displacement field that cannot capture the discontinuity between the buildings and the clouds any more, see Figure 6.7 (c).

The results obtained with the approach of Menzel and Guthe [MG07] are shown in Figure 6.5 (e). Altough the overall result looks satisfying, the zoom in Figure 6.6 (e) reveals that the hierarchical block matching technique yields unpleasant artefacts at the roof of the house since it does not impose smoothness assumptions on the displacement field.

Our final result in Figure 6.5 (f) and Figure 6.6 (f) shows a favourable HDR reconstruction without any disturbing artefacts. This becomes possible due to the robust and accurate displacement estimation of our proposed approach. The latter is demonstrated in Figure 6.8 that visualises all computed flow fields. As one can see, the motion of the clouds and the shift of the buildings due to camera shake are nicely discriminated in the first three rows. For the image in the last row, estimating the motion of the clouds is difficult, but also unimportant, because the cloud region is oversaturated and thus not considered in the HDR reconstruction, anyway.

Runtimes. Aligning the whole exposure series from Figure 6.4 that consists of 5 images with 512×340 pixels each took 75 seconds using our CPU version (full multigrid solver) with unoptimised C code on a 3.2 GHz Intel Pentium 4. An optimised parallel implementation of a cascadic FED solver on a GeForce GTX 480 reduced the runtime

(a) Reference image **(b)** Ward **(c)** `align_image_stack`

(d) No normalisation **(e)** Menzel and Guthe **(f)** Proposed method

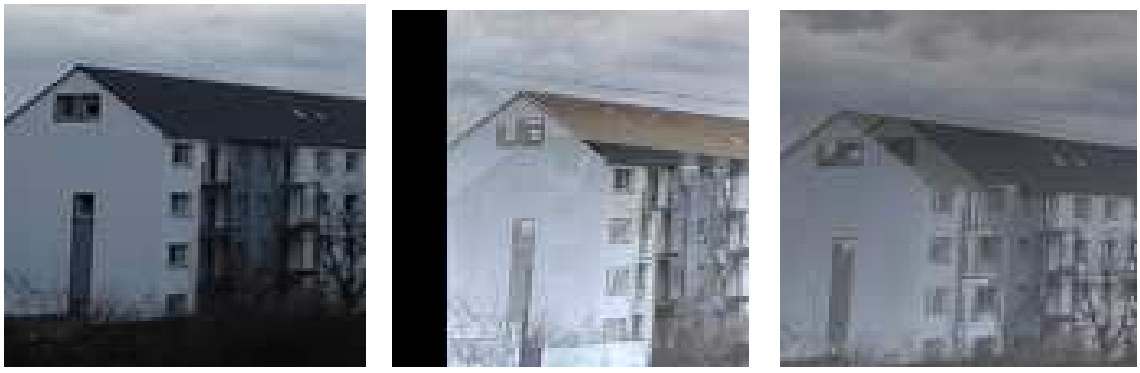
Figure 6.5: HDR imaging under severe displacements due to moving clouds and camera shake. *First row, from left to right:* **(a)** Reference image of the exposure series shown in Figure 6.4. **(b)** Tone mapped HDR image after alignment with the method of Ward [War03]. **(c)** Same using `align_image_stack`. *Second row, from left to right:* **(d)** Same using our method, but without data term normalisation. **(e)** Same using the method of Menzel and Guthe [MG07]. **(f)** Same using our proposed method.

to less than 3 seconds for the whole series.

Using More Advanced Optic Flow Methods. To see if it is possible to further improve on our results, we experimented with more sophisticated optic flow models, e.g. using our anisotropic complementary smoothness term [10, 11]. We found that using such a regulariser even deteriorates the results at occlusion boundaries, see Figure 6.9 (c) and (d) and focus on the zooms. The reason for this is that the used smoothness terms incorporates directional information from image edges. Thus, the flow estimates at occlusions are on the one hand rather sharp, see Figure 6.9 (b). This is in general a desirable property, but for alignment tasks it is better to allow for a slight over-smoothing at occlusion boundaries to avoid artefacts. This is illustrated in Figure 6.9 (a) and (c).

More Results. To demonstrate the favourable performance of our alignment strategy in general, we tested it on several freehand exposure series that are shown in Figures 6.10 to 6.13. In order to visualise the present displacements in the exposure series, we also show results without prior alignment. All results were aligned using the fixed parameters $\alpha = 2.0, \sigma = 0.5, \varepsilon = 0.001, \eta = 0.95$.

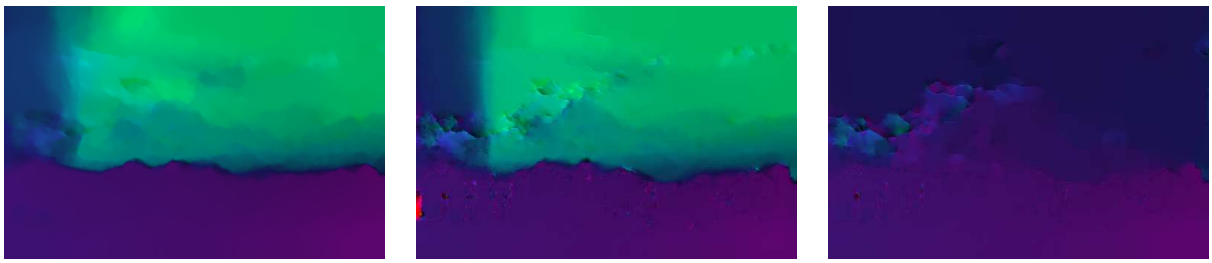
Limitations. Let us conclude our experiments with presenting the limitations of the proposed alignment strategy. As our alignment is based on a variational optic flow



(a) Reference image **(b)** Ward **(c)** `align_image_stack`

(d) No normalisation **(e)** Menzel and Guthe **(f)** Proposed method

Figure 6.6: Zoom in the results from Figure 6.5. *First row, from left to right:* **(a)** Reference image of the exposure series shown in Figure 6.4. **(b)** Tone mapped HDR image after alignment with the method of Ward [War03]. **(c)** Same using `align_image_stack`. *Second row, from left to right:* **(d)** Same using our method, but without data term normalisation. **(e)** Same using the method of Menzel and Guthe [MG07]. **(f)** Same using our proposed method.



(a) Normalisation ($\alpha = 2.0$) **(b)** No normalisation ($\alpha = 2.0$) **(c)** No normalisation ($\alpha = 2.5$)

Figure 6.7: Comparing flow fields obtained with and without data term normalisation. *From left to right:* **(a)** Flow field from image 3 (reference) to image 1 of our exposure series from Figure 6.4, computed with normalisation and $\alpha = 2.0$. **(b)** Same, but without normalisation and $\alpha = 2.0$. **(c)** No normalisation and $\alpha = 2.5$. For all results, we used the fixed parameters $\sigma = 0.5, \eta = 0.95$.

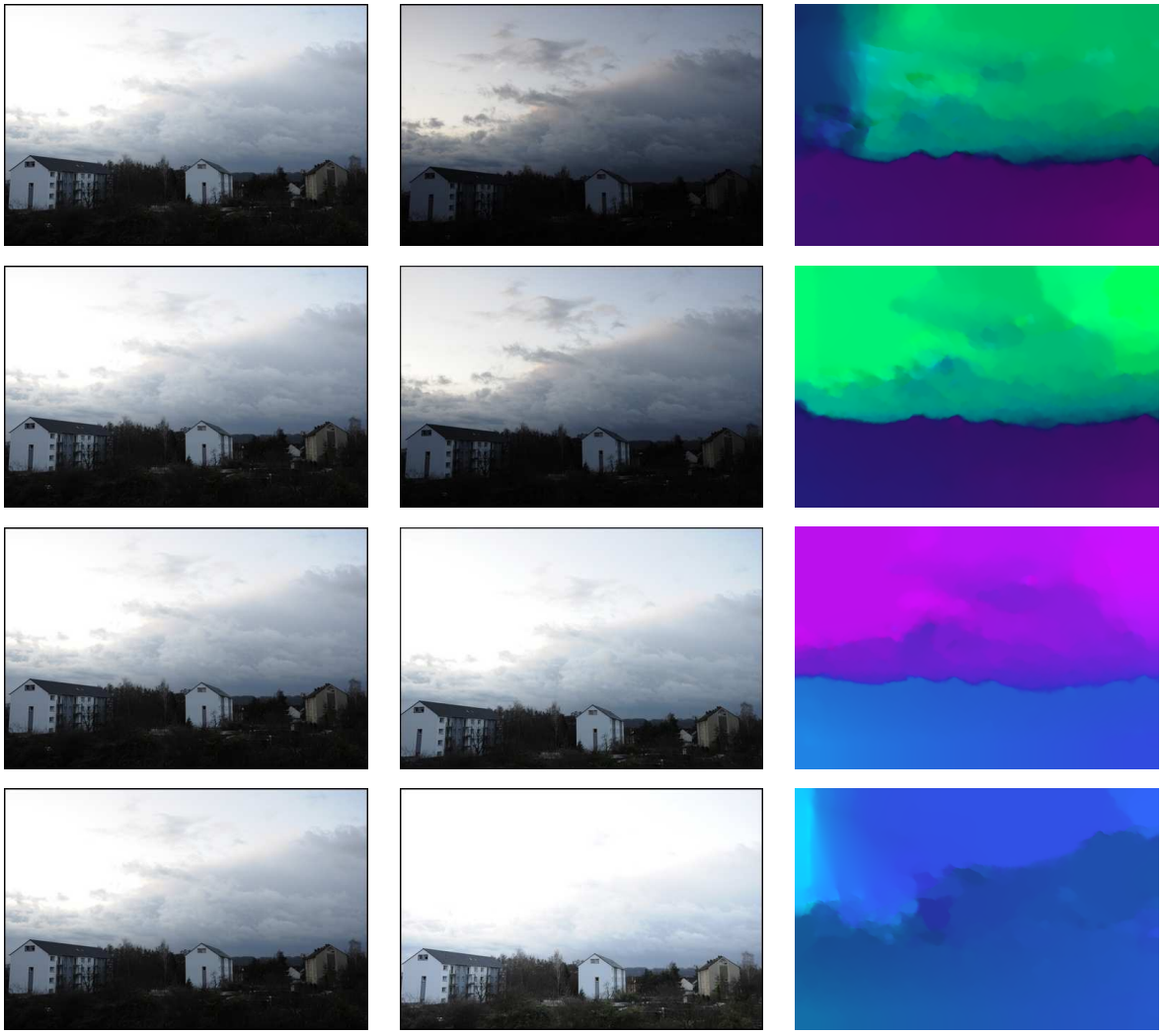
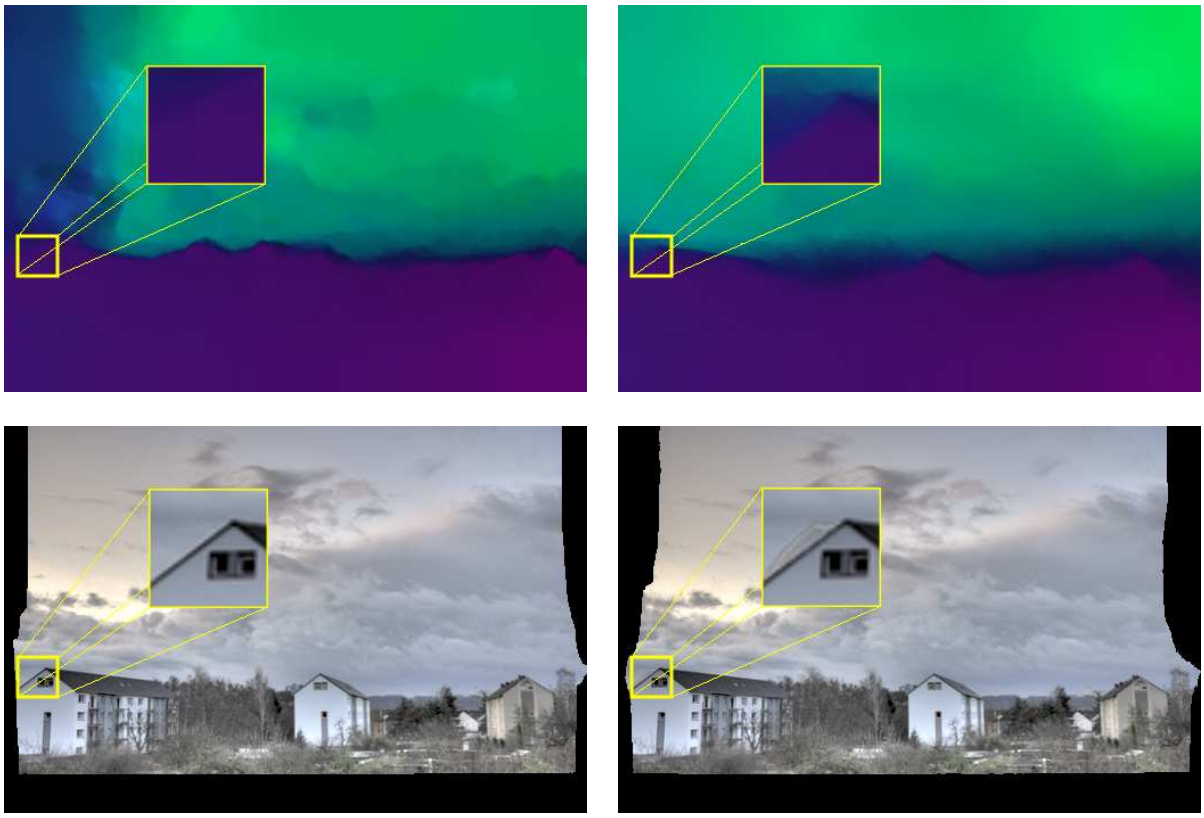


Figure 6.8: Flow fields computed on the exposure series from Figure 6.4. *First row, from left to right:* **(a)**–**(c)** Image 1. Image 3. Flow field from image 3 to image 1. *Second row, from left to right:* **(d)**–**(f)** Same for image 2 and 3. *Third row, from left to right:* **(g)**–**(i)** Same for image 4 and 3. *Fourth row, from left to right:* **(j)**–**(l)** Same for image 5 and 3. For all results, we used the fixed parameters $\alpha = 2.0, \sigma = 0.5, \varepsilon = 0.001, \eta = 0.95$.



Proposed TV smoothness term | **Complementary smoothness term**

Figure 6.9: Using a more advanced smoothness term. *First row, from left to right:* **(a)** Flow field from image 3 (reference) to image 1 of our exposure series from Figure 6.4, using the proposed TV regulariser. **(b)** Same using our complementary smoothness term from [10, 11]. *Second row, from left to right:* **(c)** Resulting tone mapped HDR after alignment using our proposed method. **(d)** Same, but using our complementary smoothness term from [10, 11] in the alignment step.

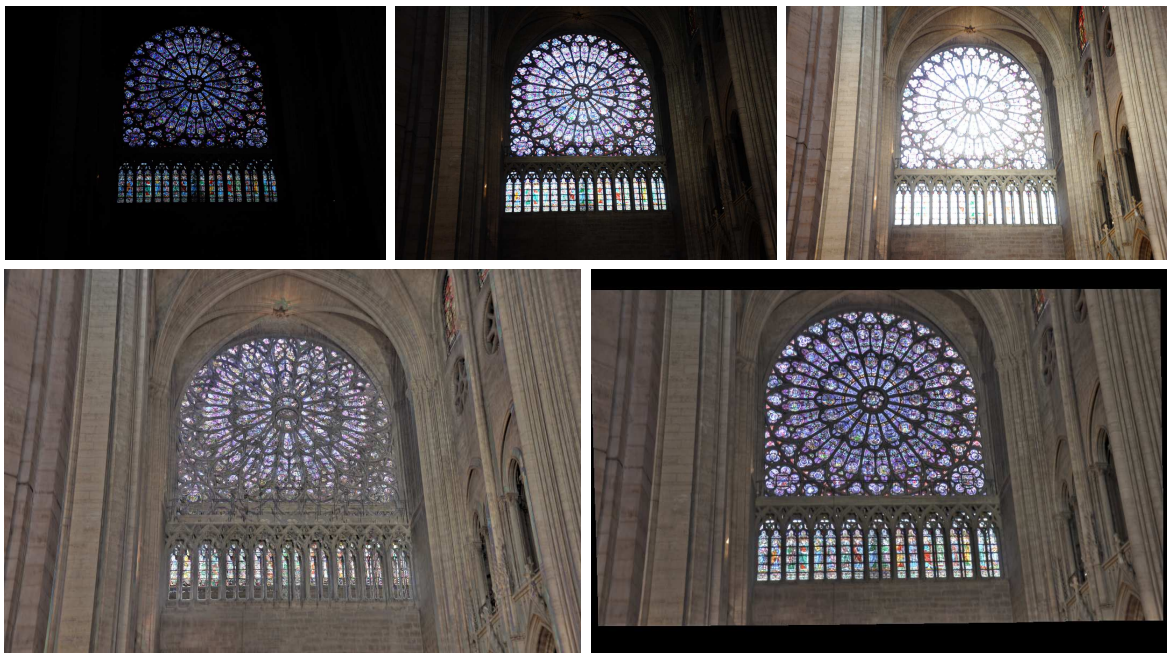


Figure 6.10: Freehand HDR imaging of a window in the cathedral Notre Dame de Paris. *First row, from left to right:* **(a)**–**(c)** Image 2, 4 and 6 of the exposure series (7 images, 1072×712 pixels, exposure times between 1/200 and 2 seconds. *Second row, from left to right:* **(c)** Resulting tone mapped HDR image without alignment. **(d)** Same using our proposed alignment method.

approach, we naturally have to use a coarse-to-fine warping strategy to handle large displacements in the exposure series, see Section 5.3.1. However, we have argued in Section 4.5 that techniques relying on a warping strategy cannot estimate large displacements of small objects. This problem can for example occur if persons or cars move relatively fast while the exposure series is taken. In such cases, our alignment strategy will fail to estimate the motion of these objects, which leads to unpleasant artefacts in the resulting HDR image, see Figure 6.14 and focus on the zooms in (d) and (e).

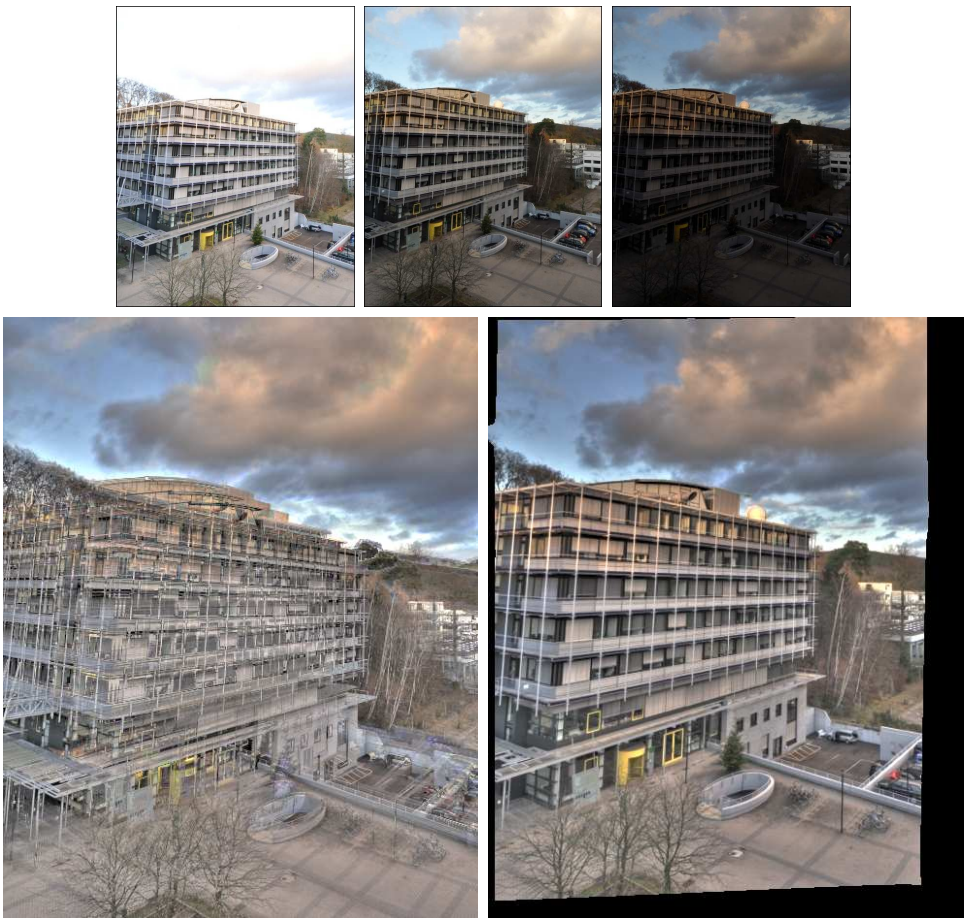


Figure 6.11: Freehand HDR imaging of the Max-Planck Institute for Informatics in Saarbrücken. *First row, from left to right:* **(a)**–**(c)** Image 1, 3 and 5 of the exposure series (5 images, 490×620 pixels, exposure times between 1/800 and 1/50 seconds. *Second row, from left to right:* **(c)** Resulting tone mapped HDR image without alignment. **(d)** Same using our proposed alignment method.

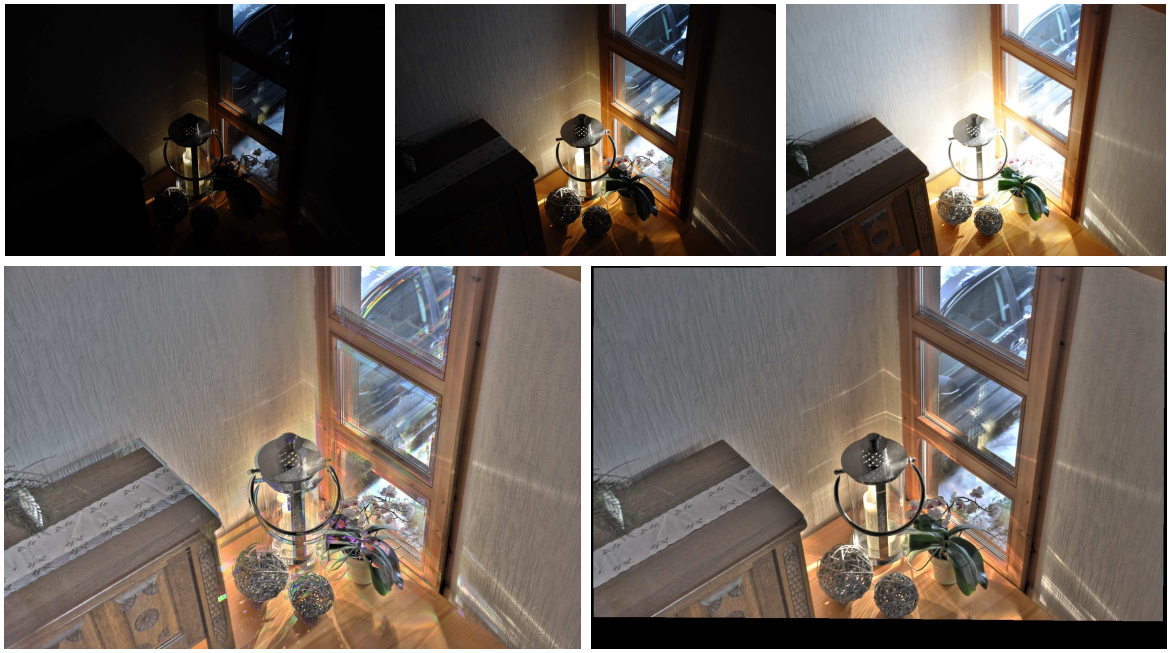
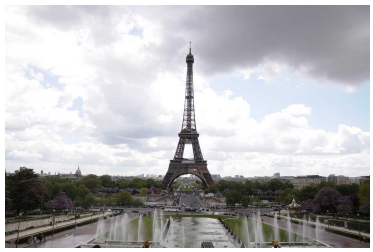


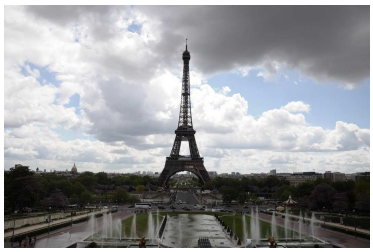
Figure 6.12: Freehand HDR imaging of an indoor scene (*Lantern*). *First row, from left to right:* **(a)**–**(c)** Image 1, 3 and 5 of the exposure series (5 images, 1268×854 pixels, exposure times between 1/320 and 1/6 seconds. *Second row, from left to right:* **(c)** Resulting tone mapped HDR image without alignment. **(d)** Same using our proposed alignment method.

Figure 6.13: Freehand HDR imaging of another indoor scene (*Room*). *First row, from left to right:* **(a)**–**(c)** Image 1, 2 and 3 of the exposure series (3 images, 1024×680 pixels, exposure times between 1/60 and 0.3 seconds. *Second row, from left to right:* **(c)** Resulting tone mapped HDR image without alignment. **(d)** Same using our proposed alignment method.

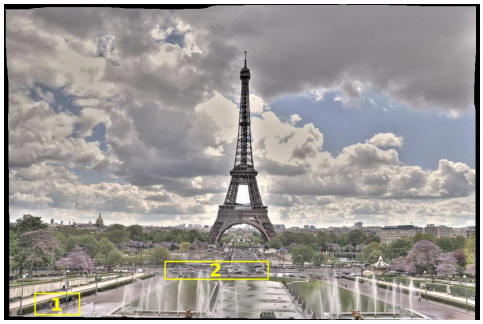
(a) Image 1



(b) Image 2



(c) Image 4



(d) Aligned HDR result

(e) Zoom in region 1 in (d)

(f) Zoom in region 2 in (d)

Figure 6.14: Limitations of our method (freehand HDR imaging of the Eiffel tower in Paris). *First row, from left to right:* **(a)**–**(c)** Image 1, 2 and 4 of the exposure series (4 images, 1072×712 pixels, exposure times between 1/4000 and 1/320 seconds. *Second row, from left to right:* **(d)** Resulting tone mapped HDR image using our proposed alignment method. **(e)** Zoom in region 1 marked in (d). *Third row:* **(f)** Zoom in region 2 marked in (d).

6.3 Joint Super-Resolution and HDR (SR-HDR) Reconstruction

Super-resolution (SR) techniques such as [EF97, FREM04, MPSC09, PPK03, TH84] combine the information from several low-resolution (LR) images of the same scene to compute an image with increased spatial resolution. Note that although megapixel image sensors are widely available nowadays, spatial resolution enhancement is still crucial in applications such as surveillance [ZZSL10] or medical imaging [Gre09], where tiny details can be of great importance. Furthermore, cameras in mobile phones still suffer from a limited spatial resolution. The prerequisites for a successful SR reconstruction are: *(i)* Some amount of *subpixel* displacement between the LR input images. Note that if there are no or only integer displacements, all images somehow contain redundant information and no resolution enhancement is possible [2]. *(ii)* The images should suffer from *aliasing*, i.e. they have been sampled below the Nyquist frequency [SW49, Uns00]. In this way the LR images contain valuable high-frequency information that helps to recover the SR result [VSVV06].

The displacement fields resulting from our previously described HDR alignment strategy (Section 6.2) are by construction of subpixel precision. If one additionally assures that the exposure series suffers from aliasing –e.g. if one uses a camera with a limited resolution like in a mobile phone– we meet all requirements for a SR reconstruction. This motivated our work in [9] where we exploited the displacements in a freehand exposure series in a *joint super-resolution and HDR* (*SR-HDR*) approach similar to [CPK09, GG06, HTO07, NYUS08, NN05, RMVS07, SSM09]. This allows to turn the usually perturbing displacements in the exposure series into an advantage: We can not only enhance the dynamic range, but also the spatial resolution of the reconstructed image. In this context, we also want to point out that SR and HDR imaging nicely complement each other: While SR methods increase the resolution of the image domain, HDR techniques increase the resolution of the image co-domain, i.e. the dynamic range.

6.3.1 Super-Resolution Basics

Before we describe our joint SR-HDR approach, we want to briefly review the basic concepts of SR reconstruction. For a more extensive survey, we refer the interested reader to [PPK03].

Notation and Observation Model. Assume we are given a set of LR input images $f_k : \Omega_{\text{disc}} \to \{0, .., 255\}$ for $k = 1, .., n_e$ and where $\Omega_{\text{disc}} := \{1, .., n_x\} \times \{1, .., n_y\}$ denotes the discrete representation of the image domain Ω. Note that we now model the images in a completely discrete manner as the SR problem stems from the discretisation and does not exist in a continuous model. Our goal is now to compute a SR image F :

[2] Combing a sequence of images with no or only integer displacements would, however, allow to obtain a denoised result by simply averaging the noisy images.

$\Omega_{\text{disc}}^{\text{SR}} \to \{0,..,255\}$, where $\Omega_{\text{disc}}^{\text{SR}} := \{1,..,z \cdot n_x\} \times \{1,..,z \cdot n_y\}$ denotes the discrete SR image domain and $z \in \mathbb{R}$ is the corresponding SR zoom factor, so usually $z > 1$.

The standard observation model for SR imaging (as for example used in [EF97]) can be written as

$$SBW_k F = f_k \ , \tag{6.19}$$

for $k = 1,..,n_e$. It models the assumption that the LR images f_k are obtained from the unknown SR image F by a backward warping (W_k), blurring (B) and a downsampling (S).

The warping W_k describes the displacements $(u_k, v_k)^\top$ between the images, which need to be estimated beforehand. This could for example be done by applying some kind of optic flow algorithm. As the warping operator acts on the SR image F, the displacements have to be transferred to the SR grid. To perform this upsampling, we use an area-based resampling strategy as proposed in [BWF+05]. Note that in the SR case, the displacements are computed from the images f_k to a reference image f_r. For the HDR alignment in Section 6.2, they were computed in the opposite direction.

The blurring operator B performs a Gaussian convolution with standard deviation σ_{SR}. In accordance to the sampling theorem [SW49, Uns00], we set $\sigma_{\text{SR}} = z \cdot \sqrt{2}/4$. For simplicity, we follow Mitzel *et al.* [MPSC09] and assume the same blurring for all input images. For implementing the downsampling operator S, we also adapt the strategy in [MPSC09] and use a simple nearest-neighbour interpolation strategy.

If *colour images* are given, it is usually assumed that the observation model holds for each of the colour channels independently.

Energy-based Super-Resolution Reconstruction. There are different basic strategies for computing the SR image from the LR input images. The survey in [PPK03] classifies SR techniques into non-uniform interpolation approaches, frequency domain approaches and energy-based approaches. We will consider the latter class that finds the SR image F by minimising an energy function of the form

$$E(F) = \sum_{(i,j) \in \Omega_{\text{disc}}^{\text{SR}}} \Big[M_{\text{SR}}(F) + \vartheta \, V_{\text{SR}}(\boldsymbol{\nabla}_2 F) \Big] \ , \tag{6.20}$$

where $\boldsymbol{\nabla}_2 F$ denotes a discrete version of the gradient operator in this case. Here, and for our further considerations on SR and joint SR-HDR reconstruction, we will intentionally misuse concepts borrowed from continuous models to facilitate notation.

The term M_{SR} denotes the data term that penalises deviations from the observation model (6.19). Using a quadratic penalisation and considering colour images one ends up with

$$M_{\text{SR}}(F) = \sum_{k=1}^{n_e} \sum_{c=1}^{3} \left(SBW_k F^c - f_k^c \right)^2 \ , \tag{6.21}$$

where c denotes the colour channels. Note that for greyscale images this summation over the channels can be omitted.

In general, SR reconstruction is an ill-posed problem, which means that we cannot determine an unique solution from the data term alone. This is due to a usually insufficient number of available LR images. To overcome this problem, energy-based SR approaches additionally use a smoothness term S_{SR} that models the assumption of a smooth SR result. As in the optic flow case, it thus allows to fill in the solution in regions where the data alone (the input images) do not provide sufficient information. The degree of smoothness is determined by a regularisation parameter $\vartheta > 0$. A basic, homogeneous regulariser is thus given by

$$V_{\mathrm{SR}}(\boldsymbol{\nabla}_2 F) = \sum_{c=1}^{3} |\boldsymbol{\nabla}_2 F^c|^2 \ . \tag{6.22}$$

Energy Minimisation. A necessary condition for a minimiser of the energy (6.20) is given by

$$0 \ = \ \partial_{F^c} E \ = \ \sum_{k=1}^{n} \left(W_k^\top B S^\top \left(S B W_k \, F^c - f_k^c \right) \right) - \ \vartheta \, \triangle F^c \ . \tag{6.23}$$

Concerning the occurring transposed operators, we implement $W_k^\top$ by a forward warping, $B = B^\top$ for Gaussian blurring [MPSC09], and $S^\top$ results in upsampling. To solve the equation (6.23), we follow the approach in [MPSC09] and apply an explicit gradient descent scheme that reads as

$$\frac{[F^c]^{(\ell+1)} - [F^c]^{(\ell)}}{\tau} \ = \ \sum_{k=1}^{n} \left(W_k^\top B S^\top \left(f_k^c - S B W_k \, [F^c]^{(\ell)} \right) \right) + \ \vartheta \, \triangle [F^c]^{(\ell)} \ . \tag{6.24}$$

Solving for the unknown at the new time level $l+1$ then gives the iterative scheme

$$[F^c]^{(\ell+1)} \ = \ [F^c]^{(\ell)} + \tau \left[\sum_{k=1}^{n} \left(W_k^\top B S^\top \left(f_k^c - S B W_k \, [F^c]^{(\ell)} \right) \right) + \ \vartheta \, \triangle [F^c]^{(\ell)} \right] \ , \tag{6.25}$$

where $[F^c]^{(\ell)}$ denotes the result at iteration ℓ and the parameter τ serves as a numerical time step size. For stability reasons it is set to the small value of $\tau = 0.01$ in accordance to [MPSC09].

Some Results and Model Extensions. Let us have a look at results obtained with the described, basic SR approach. In Figure 6.15 (a) we show one frame of an image sequence depicting a car. Obviously, the resolution of the used camera was very low and we cannot recognise any details, as for example the license plate, which is shown in the zoom in Figure 6.15 (d). Considering the SR result in Figure 6.15 (b) and (e), a drastic improvement becomes visible. However, the SR result looks rather blurry, which can be explained as follows: The quadratic data term is rather sensitive to outliers caused by noise, occlusions or misregistrations. This means that a relatively large smoothness

weight ϑ has to be chosen, which in turn blurs the result due to the homogeneous smoothing of the result.

These problems can be diminished by following Mitzel *et al.* [MPSC09] who proposed to use a subquadratic data term (L_1-norm) in combination with a TV regulariser. This results in the energy

$$E(F) = \sum_{(i,j)\,\in\,\Omega_{\text{disc}}^{\text{SR}}} \left[\sum_{k=1}^{n_e} \Psi_M \left(\sum_{c=1}^{3} \left(SBW_k\, F^c - f_k^c \right)^2 \right) \,+\, \vartheta\, \Psi_V \left(\sum_{c=1}^{3} |\boldsymbol{\nabla}_2 F^c|^2 \right) \right] \,, \tag{6.26}$$

with $\Psi_M(s^2) = \Psi_V(s^2) = \sqrt{s^2 + \varepsilon^2}$. With above modifications, the corresponding gradient descent scheme is given by

$$\begin{aligned} [F^c]^{(\ell+1)} \quad &= \quad [F^c]^{(\ell)} \\ &+ \tau \left[\sum_{k=1}^{n} \left(W_k^\top B S^\top\, \Psi_M' \left(\sum_{c=1}^{3} \left(SBW_k\, [F^c]^{(\ell)} - f_k^c \right)^2 \right) \left(f_k^c - SBW_k\, [F^c]^{(\ell)} \right) \right) \right. \\ &\qquad \left. + \vartheta\ \operatorname{div} \left(\Psi_V' \left(\sum_{c=1}^{3} \left| \boldsymbol{\nabla}_2 [F^c]^{(\ell)} \right|^2 \right) \boldsymbol{\nabla}_2 [F^c]^{(\ell)} \right) \right] \,. \end{aligned} \tag{6.27}$$

When considering the corresponding results in Figure 6.15 (c) and (f), we see that the improved model allows to obtain a much sharper reconstruction, which makes it easier to recognise the license plate.

6.3.2 A Novel Joint Super-Resolution and HDR (SR-HDR) Model

In the following we show how to extend the presented energy-based SR approach (6.26) to a novel joint super-resolution and HDR (SR-HDR) method that we presented in [9].

We assume to be given a freehand exposure series $f_k : \Omega_{\text{disc}} \to \{0, .., 255\}$ for $k = 1, .., n_e$ with corresponding exposure times t_k. Furthermore, a zoom factor z has to be specified and we assume that the displacements between the images of the exposure series has been estimated. The latter is realised by applying our method described in Section 6.2. From these inputs, we compute a SR-HDR reconstruction, i.e. a super-resolved radiance map $R : \Omega_{\text{disc}}^{\text{SR}} \to \mathbb{R}^+$.

Energy Formulation

We compute the SR-HDR image R by minimising a suitable energy function of the general form [CPK09, GG06]

$$E(R) = \sum_{(i,j)\in\Omega_{\text{disc}}^{\text{SR}}} \Big[M(R) + \vartheta\, V(\boldsymbol{\nabla}_2 R) \Big] \,, \tag{6.28}$$

with a smoothness weight $\vartheta > 0$.

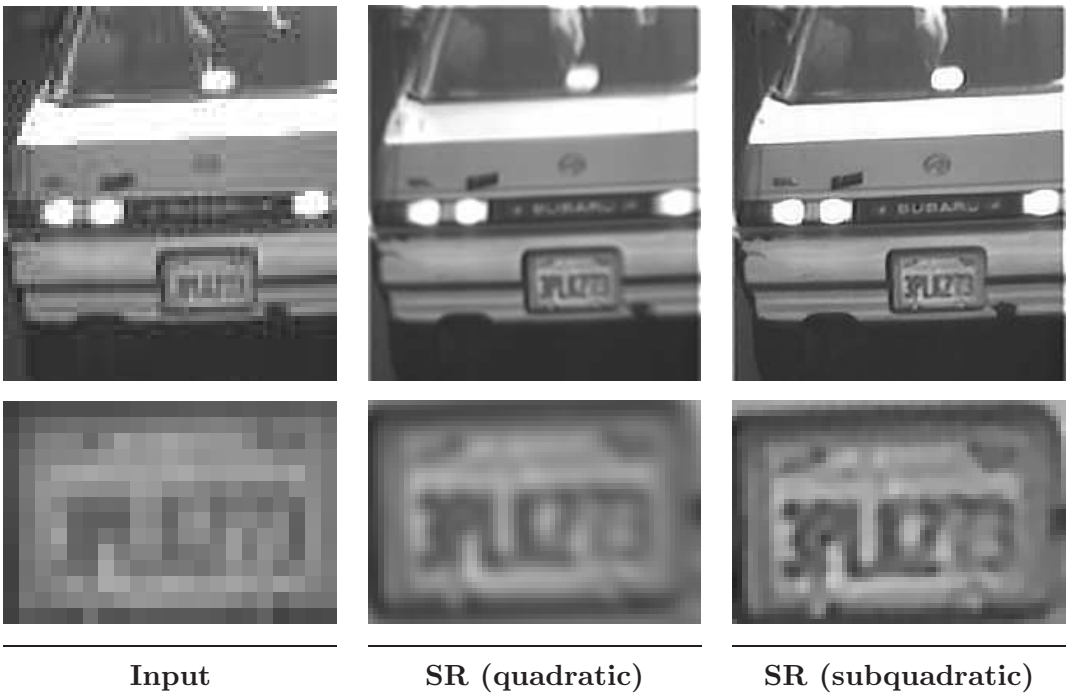


Figure 6.15: Super-resolution example. *First row, from left to right:* **(a)** One of the input images (64 images, 72×121 pixels, taken from `http://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html`). **(b)** Super-resolution result using a quadratic data term and a homogeneous smoothness term ($\vartheta = 1.0$). The used zoom factor was $z = 3$. **(c)** Same using a L_1 data term and a TV smoothness term ($\vartheta = 0.7$). *Second row, from left to right:* **(d)–(f)** Zooms in (a)–(c).

Data Term. The data term $M(R)$ combines concepts from the SR data term (6.26) and the HDR data term (6.4), which results in

$$M(R) = \sum_{k=1}^{n_e} \omega_k \, \Psi_M\left(\sum_{c=1}^{3} \left(SBW_k \, R^c - \frac{I_k^c}{t_k} \right)^2 \right) \quad , \tag{6.29}$$

where red colour marks SR concepts and blue colour marks HDR concepts. Further note that we reformulated the HDR data term $(I_k^c - t_k \, r^c)^2$ as $(r^c - I_k^c/t_k)^2$ where we recall that $I_k^c := \phi^{-1}(f_k^c)$ denotes the result of applying the inverse camera response function ϕ^{-1} to the input images f_k^c.

As a novelty compared to existing energy-based SR-HDR approaches [CPK09, GG06], we use the subquadratic L_1 penaliser function $\Psi_M(s^2) = \sqrt{s^2 + \varepsilon^2}$ to render our approach robust against outliers due to noise or incorrect displacement estimates.

Smoothness Term. The smoothness term plays a major role in the SR-HDR context as it has to fill in the result R in regions where no information is available. This can either be due to saturation problems (where $\omega_k \approx 0$), or in pixels where the warping by the displacements does not provide information.

For our SR-HDR approach, we propose a novel anisotropic smoothness term that adapts its smoothing direction to local image structures. To obtain the required directional information, we consider the structure tensor [FG87] of an image $\widetilde{R}$ that is obtained by upsampling a pure HDR reconstruction computed from the aligned input images. This leads to the following structure tensor definition:

$$\widetilde{\boldsymbol{S}} = \sum_{c=1}^{3} K_{\rho_1} * \left(\boldsymbol{\nabla}_2 \widetilde{R}_{\rho_2}^c \left(\boldsymbol{\nabla}_2 \widetilde{R}_{\rho_2}^c \right)^\top \right) \quad , \tag{6.30}$$

with

$$\widetilde{R}_{\rho_2}^c := K_{\rho_2} * \widetilde{R}^c \quad , \tag{6.31}$$

and where ρ_1 and ρ_2 serve as neighbourhood and smoothing scale, respectively. As we already noted earlier in Section 4.3.1, the eigenvectors $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ of the structure tensor give the desired direction of local image structures: The vector $\boldsymbol{s}_1$ points across image structures, and $\boldsymbol{s}_2$ points along them.

Following the design concept of our complementary regulariser [10, 11] (see Section 4.3.2), our smoothness term then penalises the projections of the image gradients onto $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ in different manners: Along image edges, we perform a quadratic penalisation to obtain a strong smoothing that pronounces the edges. In the orthogonal direction across edges, we use a robust penaliser function that reduces the smoothing and helps to preserve edges. This results in the smoothness term

$$V(\boldsymbol{\nabla}_2 R) = \Psi_V\left(\sum_{c=1}^{3} \left(\boldsymbol{s}_1^\top \boldsymbol{\nabla}_2 R^c \right)^2 \right) + \sum_{c=1}^{3} \left(\boldsymbol{s}_2^\top \boldsymbol{\nabla}_2 R^c \right)^2 \quad . \tag{6.32}$$

For the robust penaliser Ψ_V, we use the Charbonnier function (6.16) with $\varepsilon = 0.1$.

6.3.3 Energy Minimisation

We compute the SR-HDR image R by minimising the energy (6.28). To this end, we use a *semi-implicit* gradient descent scheme given by

$$\begin{aligned}\frac{[R^c]^{(\ell+1)} - [R^c]^{(\ell)}}{\tau} &= \\ \sum_{k=1}^{n_e} \left(W_k^\top B S^\top \omega_k \, \Psi'_M \left(\sum_{c=1}^{3} \left(SBW_k \, [R^c]^{(\ell)} - \frac{I_k^c}{t_k} \right)^2 \right) \left(\frac{I_k^c}{t_k} - SBW_k \, [R^c]^{(\ell)} \right) \right) \\ + \vartheta \, \mathrm{div} \left(D\left(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{\nabla}_2 R^{(\ell)} \right) \boldsymbol{\nabla}_2 [R^c]^{(\ell,\ell+1)} \right) \ ,\end{aligned} \tag{6.33}$$

The occurring diffusion tensor D is given by

$$D\left(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{\nabla}_2 R^{(\ell)} \right) = \Psi'_V \left(\sum_{c=1}^{3} \left(\boldsymbol{s}_1^\top \boldsymbol{\nabla}_2 [R^c]^{(\ell)} \right)^2 \right) \boldsymbol{s}_1 \boldsymbol{s}_1^\top + \boldsymbol{s}_2 \boldsymbol{s}_2^\top \ . \tag{6.34}$$

Further note that term $\boldsymbol{\nabla}_2 [R^c]^{(\ell,\ell+1)}$ occurring in (6.33) has two indices in the superscript. By this, we denote that when discretising the divergence expression, the old value at iteration l is used for the neighbouring pixels, whereas the new result at iteration $l+1$ is used for the central pixel. The gradient descent scheme (6.33) can thus be rewritten as

$$\frac{[R^c]_{i,j}^{(\ell+1)} - [R^c]_{i,j}^{(\ell)}}{\tau} = [m^c]_{i,j}^{(\ell)} + \vartheta \left(\sum_{\tilde{i},\tilde{j} \in \mathcal{N}_{i,j}} [w]_{\tilde{i},\tilde{j}} \left([R^c]_{\tilde{i},\tilde{j}}^{(\ell)} - [R^c]_{i,j}^{(\ell+1)} \right) \right) \ , \tag{6.35}$$

using the abbreviation

$$\begin{aligned}[m^c]_{i,j}^{(\ell)} \ &:= \\ \sum_{k=1}^{n_e} \left(W_k^\top B S^\top [\omega_k]_{i,j} \, \Psi'_M \left(\sum_{c=1}^{3} \left(SBW_k \, [R^c]_{i,j}^{(\ell)} - \frac{[I_k^c]_{i,j}}{t_k} \right)^2 \right) \left(\frac{[I_k^c]_{i,j}}{t_k} - SBW_k \, [R^c]^{(\ell)} \right) \right) \ ,\end{aligned} \tag{6.36}$$

and where the weights $[w]_{\tilde{i},\tilde{j}}$ can be derived from the diffusion tensor (6.34) as described in Section 5.1.1. Solving (6.35) for the unknown update $[R^c]^{(\ell+1)}$ results in the final iterative scheme given by

$$[R^c]_{i,j}^{(\ell+1)} = \frac{[R^c]_{i,j}^{(\ell)} + \tau \left([m^c]_{i,j}^{(\ell)} + \vartheta \sum_{\tilde{i},\tilde{j} \in \mathcal{N}_{i,j}} [w]_{\tilde{i},\tilde{j}} \, [R^c]_{i,j}^{(\ell)} \right)}{1 + \vartheta \sum_{\tilde{i},\tilde{j} \in \mathcal{N}_{i,j}} [w]_{\tilde{i},\tilde{j}}} \ . \tag{6.37}$$

Concerning the numerical time step size, we experimentally found that setting $\tau = 0.5$ is feasible, which is more than one order of magnitude larger than the step size used in the explicit scheme (6.25) which was taken from [MPSC09]. We further note that due to the complicated structure of the data term (6.29), more efficient solution strategies as in the optic flow case (see Section 5.3.2) are difficult to realise here.

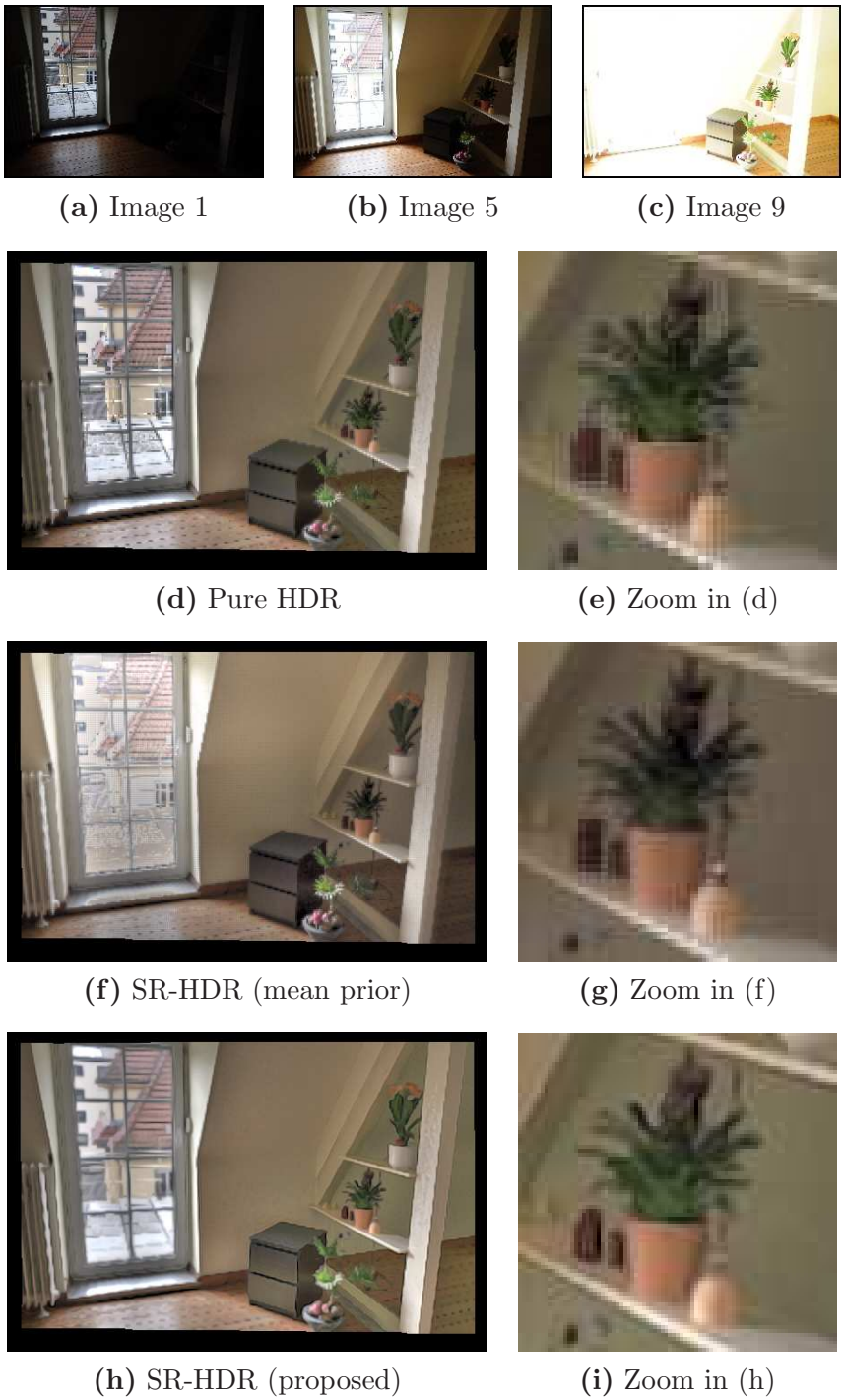
6.3.4 Experiments

For our first experiments, we took a freehand, real world exposure series consisting of 9 images with a resolution of 268×178 pixels and with exposure times between 1/800 and 1/8 seconds. Some of the input images are shown in the first row of Figure 6.16. As zoom factor we set $z = 2$ for all our experiments. To estimate the displacements between the images of the exposure series, we use our technique described before in Section 6.2. The model parameters were set to $\alpha = 2.0, \sigma = 0.5, \eta = 0.7$. If not stated otherwise, these parameters are also used for the other SR-HDR experiments in this section.

Our SR-HDR Method versus Pure HDR and Existing Approaches. In a first experiment (see Figure 6.16) we show two benefits of our proposed SR-HDR method: *(i)* We illustrate that our obtained SR-HDR reconstruction shown in Figure 6.16 (h) gives a remarkable improvement over a pure HDR reconstruction (Figure 6.16 (d)). Especially when comparing the corresponding zooms in Figure 6.16 (e) and (i) we see that our SR-HDR result gives a much less pixelated result. *(ii)* We also want to show that our proposed model (L_1-data term with anisotropic smoothness term) gives much more appealing results compared to existing energy-based SR-HDR approaches [CPK09, GG06] that use a quadratic data term and a mean prior. Recall that the latter enforces the result to be close to a mean image that is obtained by averaging the radiances computed from all input images. Although this prior stabilises the minimisation, it does *not* allow to fill in missing information and to smooth the resulting image. Consequently, the result of such a method (see Figure 6.16 (f) and (g)) hardly improves over a pure HDR reconstruction and is inferior to our result.

Runtimes. As we have noted before, it is much more difficult to come up with an efficient numerical solver for our SR-HDR approach. By using a semi-implicit scheme, we could improve the performance by more than one order of magnitude, as it now becomes possible to use a time step size of 0.5 instead of the 0.01 which was reported for an explicit scheme in [MPSC09]. However, computing the final SR-HDR result for the exposure series from Figure 6.16 still takes about half an hour on a standard PC (3.2 GHz Intel Pentium 4).

Different Data and Smoothness Terms. Our next experiment compares the influence of using different data and smoothness terms in energy-based SR-HDR approaches.



(a) Image 1 **(b)** Image 5 **(c)** Image 9

(d) Pure HDR **(e)** Zoom in (d)

(f) SR-HDR (mean prior) **(g)** Zoom in (f)

(h) SR-HDR (proposed) **(i)** Zoom in (h)

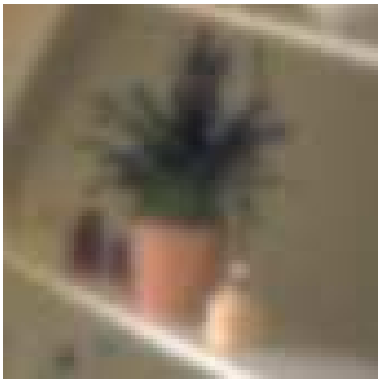
Figure 6.16: Joint SR-HDR reconstruction on a real world exposure series. *First row, from left to right:* **(a)** – **(c)** Images 1, 5 and 9. *Second row, from left to right:* **(d)** Upsampled HDR reconstruction. **(e)** Zoom in (d). *Third row, from left to right:* **(f)** SR-HDR result of [CPK09, GG06] (quadratic data term, mean prior, $\vartheta = 5.0, \tau = 0.1$, 25 iterations). **(g)** Zoom in (f). *Fourth row, from left to right:* **(h)** Result of proposed method ($\vartheta = 0.3, s = 8, \rho_1 = 0.4, \rho_2 = 0.2, \tau = 0.5$, 1000 iterations). **(i)** Zoom in (h).

As input images, we use the same exposure series as before in Figure 6.16. In Figure 6.17 (d) and (e) we show the result when using a quadratic data and smoothness term. As the quadratic data term is not robust against outliers we have to use a relatively large smoothness weight ϑ to avoid unpleasant artefacts. Consequently, the result looks rather blurry and lacks detail as the quadratic regulariser leads to a homogeneous diffusion of the image. When using a robust L_1 data term in combination with a TV smoothness term as proposed for pure SR by Mitzel *et al.* [MPSC09], the results look already much sharper, see Figure 6.17 (f) and (g). However, the edges in the image are rather jagged, which becomes obvious at the planks in the zoom in Figure 6.17 (g). Replacing the TV regulariser by our proposed anisotropic smoothness term (Figure 6.17 (h) and (i)) allows to obtain a much better reconstruction of edges due to the pronounced smoothing in edge direction. The reduced smoothing in the orthogonal direction allows to preserve fine details, as can be seen at the plant in the zoom in Figure 6.17 (i).

More Results. To conclude our experimental section, we compare our SR-HDR results against pure HDR reconstructions for further real world exposure series. The exposure series together with the corresponding results are shown in Figures 6.18 to 6.20. In our opinion, the SR-HDR results always look more appealing compared to a pure HDR reconstruction. As fixed parameters we set $s = 8, \tau = 0.5$ and performed 1000 iterations. The only remaining tuning parameter is the smoothness weight ϑ. Its values are given in the captions of the respective figures.

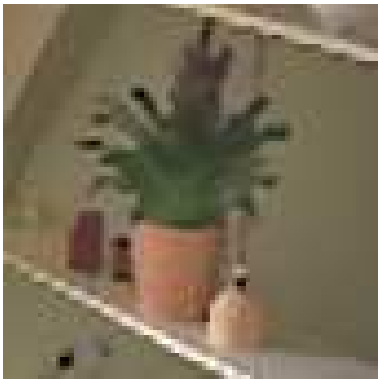
The astute reader may also notice that the black border for the HDR and SR-HDR results are different. This is explained by the fact that in the SR-HDR context, the displacements are computed in the opposite direction compared to HDR alignment, as noted before.

(a) Quadratic data and smoothness term



(b) Zoom in (d)

(c) L_1 data term, TV smoothness term



(d) Zoom in (d)

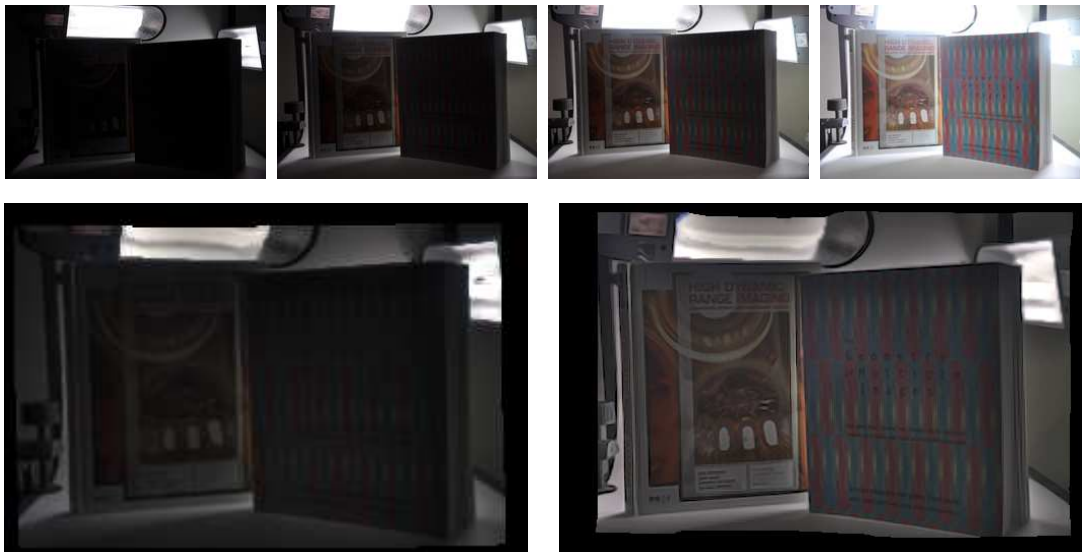
(e) L_1 data term, anisotropic smoothness term

(f) Zoom in (d)

Figure 6.17: Comparing data and smoothness terms in energy-based SR-HDR methods (using the exposure series from Figure 6.16). *First row, from left to right:* **(a)** Result with a quadratic data and smoothness term ($\vartheta = 0.75, s = 8, \tau = 0.5$, 1000 it.). **(b)** Zoom in (a). *Second row, from left to right:* **(c)** Result with a L_1 data and a TV smoothness term. ($\vartheta = 0.4, \varepsilon = 0.001$, other parameters as in (a)). **(d)** Zoom in (c). *Third row, from left to right:* **(e)** Result with our proposed method with a L_1 data term and an anisotropic smoothness term ($\vartheta = 0.3, \mu = 0.1, \rho_1 = 0.4, \rho_2 = 0.2$, others as in (c)). **(f)** Zoom in (e).

Pure HDR **Joint SR-HDR**

Figure 6.18: Comparing a pure HDR to a joint SR-HDR result on an indoor scene. *First row, from left to right:* **(a)** – **(d)** Images 1, 3, 6 and 8 of the exposure series (8 images, 268×178 pixels, exposure times between 1/200 and and 1/8 seconds.) *Second row, from left to right:* **(d)** HDR result after aligning the exposure series with our proposed technique. **(e)** SR-HDR result using our proposed approach. (same alignment, $\vartheta = 0.6$).



Pure HDR **Joint SR-HDR**

Figure 6.19: Comparing a pure HDR to a joint SR-HDR result on a lab-made scene. *First row, from left to right:* **(a)** – **(d)** Images 2, 5, 7 and 11 of the exposure series (12 images, 257×171 pixels, exposure times between 1/2500 and and 1/4 seconds.) *Second row, from left to right:* **(d)** HDR result after aligning the exposure series with our proposed technique. **(e)** SR-HDR result using our proposed approach. (same alignment, $\vartheta = 0.4$).

Pure HDR **Joint SR-HDR**

Figure 6.20: Comparing a pure HDR to a joint SR-HDR result on an outdoor scene. *First row, from left to right:* **(a)** – **(d)** Images 1, 4, 9 and 12 of the exposure series (12 images, 268×178 pixels, exposure times between 1/2500 and and 1/20 seconds.) *Second row, from left to right:* **(d)** HDR result after aligning the exposure series with our proposed technique. **(e)** SR-HDR result using our proposed approach. (same alignment, $\vartheta = 0.2$).

Chapter 7

Summary and Future Work

Nothing shocks me. I'm a scientist.

from the motion picture
"Indiana Jones and the Temple of Doom"

7.1 Summary

In the previous chapters we presented contributions to various areas of research on correspondence problems. Thereby, we mainly focused on optic flow as a representative and rather general example. At some points we also discussed the more specific stereo matching problem. Moreover, we only considered energy-based (variational) approaches that find the optic flow field as minimiser of a suitable energy formulation. In recent years these strategies have shown to give the most accurate results, and also efficient solvers have been proposed.

Modelling

Data Term. In Chapter 4 we started by reviewing previous ideas for the design of robust data terms, such as subquadratic penaliser functions and higher order constancy assumptions. These concepts were combined into a novel robust data term that offers a high degree of invariance under illumination changes by considering a *HSV colour representation* and applying robust penaliser functions to each channel separately. In order to avoid an undesirable overweighting of the data term at large image gradient locations, we incorporated the idea of *data term normalisation* [LV98, SC06, SAH91] into our data term, thereby utilising this concept for modern data terms.

Smoothness Term. After discussing the data term, we turned to the design of an appropriate smoothness term. This resulted in a *complementary smoothness term* that

adapts its anisotropic smoothing directions to the imposed data constraints: In the direction where the data terms provides information, we reduce the smoothing, but opt for a strong smoothing in the orthogonal direction where the data constraints do not provide information. Moreover, our smoothness term can be classified as being *joint image- and flow-driven* as it adapts the direction of smoothing to image edges, whereas the amount of smoothing depends on the flow contrast. This allows to obtain sharp flow edges as known from image-driven methods, but remedies their oversegmentation problems. Finally, we presented a *general framework* that introduces a common notation of several existing as well as our novel smoothness term and thus eases their comparison.

We also discussed smoothness terms for variational stereo approaches where e completed the taxonomy by presenting the first disparity-driven regularisation strategy.

Parameter Selection. Having adapted the smoothness term and the data term, what remains is to appropriately balance the two terms, i.e. determine the optimal value of the smoothness weight for the considered image sequence. To this end, we introduced a novel concept that we call *optimal prediction principle (OPP)*. It states that an optimal flow field (obtained by using an optimal smoothness weight) allows for the best possible prediction of the next images in the sequence. Assuming constant speed and linear trajectories of objects, we judge the prediction quality by evaluating the data constraints between the first and the third frame using the doubled flow. Our parameter selection approach then simply computes flow fields for several values of the smoothness weight and selects the weight corresponding to the flow field with the best prediction quality.

Combining this parameter selection with our complementary smoothness term yields a variational approach where all components (data term, smoothness term and smoothness weight) are in harmony, motivating the name *optic flow in harmony* method.

Other Terms. We also briefly investigated an additional predictor term in the energy that incorporates prior knowledge by enforcing the resulting flow field to be close to a given, *predictor flow field*, e.g. a (possibly sparse) predictor flow field obtained by a feature matching technique. Altough these methods give in general worse results than variational methods, they do not require a coarse-to-fine warping strategy to handle large displacements and can thus capture small flow details. As the latter are problematic for variational methods, a feature matching predictor can help to remedy these problems.

Numerics

For minimising the proposed energy functionals we deduce and solve the associated Euler-Lagrange equations. Discretising the Euler-Lagrange equations requires to approximate occurring derivatives, which is mostly done by symmetric (central) finite difference approximations. We could show that using correctly-oriented, one-sided differences may lead to improved results which constitutes a novel finding for the computer vision community. However, such *upwind discretisations* are a classic concepts for the solution of hyperbolic partial differential equations. Following further ideas from this

context, we also showed that combining one-sided and central differences based on a smoothness measure leads to even better results. Here, we could draw an interesting conclusion from our extensive experiments: The more sophisticated the continuous models become, the less influential is an appropriate numerical approximation of derivative. Nevertheless, all results illustrated the importance of an appropriate numerical approximation of derivatives to some extent and eventually showed that advanced numerics might be an alternative to sophisticated models.

After discretisation, the Euler-Lagrange equations constitute a nonlinear system of equations. For an efficient solution of the latter, we used two strategies, depending on the underlying hardware architecture: On the CPU, we used the full multigrid solver proposed in [BWKS06]. To exploit the possibilities of modern graphics hardware (GPUs), we implemented an explicit gradient descent scheme that was speeded up by a coarse-to-fine computation and using variable *fast explicit diffusion (FED)* [GWB10] time step sizes. Due to its simple nature, we could easily implement our solver on a parallel GPU architecture using the NVidia CUDA framework and achieved speedup factors of up to 55 compared to the multigrid scheme on the CPU. Additionally, we could port this simple solver to a modern smartphone which allowed us to implement basic optic flow methods with an interactive performance [4].

Applications

As an example of a novel application area for modern optic flow methods we considered *high dynamic range (HDR)* imaging. Here, a set of images taken with varying exposure times (exposure series) is fused to compute an image with increased dynamic range, which allows to capture that details in dark as well as bright regions. However, the practical applicability of HDR techniques is limited as they assume that the exposure series is perfectly aligned, which is often violated due to camera shake or moving objects. By carefully modifying a modern optic flow approach, we managed to robustly estimate the displacements and thus to align even challenging real world exposure series. The main modification was to only impose gradient constancy in the data term as the common brightness constancy assumption will not hold due to the varying exposure times.

Exploiting the subpixel precision of the estimated displacement fields, we combined ideas from super-resolution and HDR techniques into an energy-based *joint super-resolution and HDR (SR-HDR)* approach that computes an image with increased spatial resolution and dynamic range. In this context, we introduced sophisticated regularisation strategies in the SR-HDR context, which we could show to be a key aspect for obtaining visually appealing results.

7.2 Future Work

Our contributions have shown that despite three decades of research on correspondence problems, significant improvements in the accuracy as well as in the efficiency are still

possible. Moreover, growing new research areas such as computational photography give rise to interesting novel applications for correspondence problems. Having this in mind, there are several future directions that seem promising to be explored.

Data Term

We proposed to use an HSV colour representation to obtain robustness under difficult illumination conditions. We would thus like to investigate the usefulness of alternative colour spaces, e.g. HSL, YUV, or CIELAB. Moreover, we still did not consider extreme distortions in the images, like heavy noise or strong compression artefacts caused by common image and video codecs like JPEG or MPEG.

Smoothness Term

Our experiments have shown that joint image- and flow-driven regularisers like our complementary smoothness term mostly remedy oversegmentation problems. However, we found that small artefacts might still be visible in some sequences; see our remarks in Section 4.3.5. We are thus interested in analysing and resolving these remaining issues. Concerning extensions of our smoothness term, we would like to see if the incorporation of recent non-local regularisation ideas [SRB10, WPB10] allows to further improve the performance of our smoothness term.

Parameter Selection

For our proposed parameter selection approach based on the optimal prediction principle (OPP), we plan to investigate several points:

(i) We would like to improve the efficiency, for example by finding a modification that yields similar results but requires to compute less flow fields during the estimation process. *(ii)* It also seems very interesting to apply our approach in cases where only two images are available. This would also remedy the problems with non-constant speed of objects that we have mentioned in Section 4.4.3. *(iii)* Two interesting extensions of our method could be to apply the parameter estimation on every level of the warping scheme, or for subsequent image pairs in an image sequence, respectively. In both cases, we hope that our method could profit from an initialisation with the previous results, i.e. from the previous warping level or from the previous image pair, respectively. *(iv)* Finally, a deeper theoretical underpinning of the optimal prediction principle may give further ideas for improvements and extensions. For example, we hope to explain why our method seems to underestimate the smoothness weight in general while still giving favourable results in terms of quality measures.

Joint Super-Resolution and HDR (SR-HDR) Reconstruction

The major limitation of our SR-HDR approach is that its optic flow-based alignment step cannot handle large displacements of small objects in the given exposures series.

This leads to unpleasant artefacts in the reconstruction (see Figure 6.14) and is caused by the coarse-to-fine warping strategy. One possible solution to this issue could thus be to incorporate a feature-matching prior as discussed in Section 4.5. Alternatively, if we do not want, or if we are not able to correct the flow estimate, applying an anti-ghosting technique [KAR06] will reduce the disturbing artefacts.

Furthermore, a joint estimation of displacements *and* SR-HDR reconstruction seems interesting. However, it is questionable if the more difficult minimisation of such a model pays off in terms of quality. As a starting point, we refer to the SR method in [EM09] where a joint estimation of displacements and super-resolved image has been proposed.

An Alternative Modelling and Minimisation Framework

We used the Euler-Lagrange framework for minimising our continuous energy functionals. While this framework allows to minimise rather general energies in an efficient way, it may still restrict the design of the energy terms. For example, non-differentiable penaliser functions like the Potts model [Pot52] cannot be used although this might be useful for obtaining extremely sharp discontinuities. Furthermore, although the applied coarse-to-fine warping strategy often allows to compute a good local minimum of the energy, there is no guarantee on how well the desired global minimum is approximated.

These problems can be alleviated by formulating the correspondence problem as a multi-label problem where each pixel is assigned two labels from a discrete set of labels to represent the two components of the displacement vector. A globally optimal solution of such multi-label problems can be computed using graph cut optimisation in a discrete setting [BVZ01]. In a continuous setting, e.g. [PSG$^+$08], one transforms the problem into a continuous convex problem, computes the global minimum and then projects the continuous solution back onto the discrete label space. Until recently, these solvers featured a high computational complexity and an immense memory consumption, which rendered them practically unusable for large label spaces like the ones typically encountered in optic flow. However, recently discrete [SKH08] as well continuous solvers [GC10] for multi-label problems have been proposed that *approximate* a globally optimal solution while drastically reducing the computational complexity as well as the memory requirements. As the approximation is fairly close to the global optimum (less than 5% in [GC10]), adapting these methods to our framework may not only give more freedom in modelling, but in the long run also a further boost in quality and efficiency.

Further Ideas

We have shown that a carefully modified optic flow method is well-suited for aligning exposure series in the context of HDR imaging. Building on this finding, we would like to see if modern optic flow approaches can also help to align the input images for other multi-shot techniques such as extended depth of field (focus stacking) [ADA$^+$04], flash/no-flash photography [PSA$^+$04], image stitching [Sze06] or day and night fusion [RIY05]. However, it is not obvious how the optic flow approaches have to be

adapted to the specific tasks.

Building on our first work in [4], we are interested in porting more of our algorithms to modern smartphones (like the iPhone 4 or Android phones). Nowadays, these platforms offer reasonable built-in cameras, a remarkable computational power and also flexible programming environments that support native programming languages like C/C++. This opens the door for interesting mobile image processing and computer vision applications that can additionally profit from features like broadband internet access, GPS positioning and gyroscopic sensors that can be found in almost every modern smartphone.

Appendix A

Proofs

A.1 Reformulation of the Nagel and Enkelmann Regulariser

We want to show the correctness of our reformulation of the Nagel and Enkelmann regulariser from (4.56):

$$\begin{aligned} V_{\mathrm{AI}}(\nabla_2 u, \nabla_2 v) &= (\nabla_2 u)^\top \boldsymbol{P}(\nabla_2 f)\ \nabla_2 u + (\nabla_2 v)^\top \boldsymbol{P}(\nabla_2 f)\ \nabla_2 v \\ &= \frac{\kappa^2}{\operatorname{tr} \boldsymbol{S}_0 + 2\kappa^2}\left(u^2_{\boldsymbol{s}_1^0} + v^2_{\boldsymbol{s}_1^0}\right) + \frac{\operatorname{tr} \boldsymbol{S}_0 + \kappa^2}{\operatorname{tr} \boldsymbol{S}_0 + 2\kappa^2}\left(u^2_{\boldsymbol{s}_2^0} + v^2_{\boldsymbol{s}_2^0}\right) \quad , \end{aligned} \tag{A.1}$$

We proceed in two steps: First, we show that $\boldsymbol{p}_1 := \boldsymbol{s}_1^0 = \nabla_2 f / |\nabla_2 f|$ and $\boldsymbol{p}_2 := \boldsymbol{s}_2^0 = \nabla_2 f^\perp / |\nabla_2 f|$ are the eigenvectors of the matrix $\boldsymbol{P}$ as defined in (4.55). The corresponding eigenvalues that we denote by q_1 and q_2 shall be given by the factors in front of $(u^2_{\boldsymbol{s}_1^0} + v^2_{\boldsymbol{s}_1^0})$ and $(u^2_{\boldsymbol{s}_2^0} + v^2_{\boldsymbol{s}_2^0})$, respectively. To show that $\boldsymbol{p}_1$ is an eigenvector of $\boldsymbol{P}$ with corresponding eigenvalue q_1, it suffices to prove that $\boldsymbol{P}\,\boldsymbol{p}_1 = q_1\,\boldsymbol{p}_1$:

$$\boldsymbol{P}\,\boldsymbol{p}_1 = \frac{1}{|\nabla_2 f|^2 + 2\kappa^2}\left(\nabla_2 f^\perp \left(\nabla_2 f^\perp\right)^\top + \kappa^2 \boldsymbol{I}\right)\frac{\nabla_2 f}{|\nabla_2 f|} \tag{A.2}$$

$$= \frac{1}{|\nabla_2 f|^2 + 2\kappa^2}\left(\frac{\nabla_2 f^\perp \left(\nabla_2 f^\perp\right)^\top \nabla_2 f}{|\nabla_2 f|} + \frac{\kappa^2}{|\nabla_2 f|}\nabla_2 f\right) \tag{A.3}$$

$$\overset{(*)}{=} \frac{\kappa^2}{|\nabla_2 f|^2 + 2\kappa^2}\,\frac{\nabla_2 f}{|\nabla_2 f|} \tag{A.4}$$

$$= \frac{\kappa^2}{|\nabla_2 f|^2 + 2\kappa^2}\,\boldsymbol{s}_1^0 = q_1\,\boldsymbol{p}_1 \quad , \tag{A.5}$$

where $(*)$ holds as $\left(\boldsymbol{\nabla}_2 f^\perp\right)^\top \boldsymbol{\nabla}_2 f = 0$. For the second eigenvector and eigenvalue, one obtains

$$\begin{aligned}
\boldsymbol{P}\,\boldsymbol{p}_2 &= \frac{1}{|\boldsymbol{\nabla}_2 f|^2 + 2\kappa^2}\left(\boldsymbol{\nabla}_2 f^\perp \left(\boldsymbol{\nabla}_2 f^\perp\right)^\top + \kappa^2 \boldsymbol{I}\right)\frac{\boldsymbol{\nabla}_2 f^\perp}{|\boldsymbol{\nabla}_2 f|} && \text{(A.6)}\\
&= \frac{1}{|\boldsymbol{\nabla}_2 f|^2 + 2\kappa^2}\left(\frac{\boldsymbol{\nabla}_2 f^\perp \left(\boldsymbol{\nabla}_2 f^\perp\right)^\top \boldsymbol{\nabla}_2 f^\perp}{|\boldsymbol{\nabla}_2 f|} + \frac{\kappa^2}{|\boldsymbol{\nabla}_2 f|}\boldsymbol{\nabla}_2 f^\perp\right) && \text{(A.7)}\\
&\overset{(*)}{=} \frac{1}{|\boldsymbol{\nabla}_2 f|^2 + 2\kappa^2}\left(\boldsymbol{\nabla}_2 f^\perp |\boldsymbol{\nabla}_2 f| + \frac{\kappa^2}{|\boldsymbol{\nabla}_2 f|}\boldsymbol{\nabla}_2 f^\perp\right) && \text{(A.8)}\\
&= \frac{1}{|\boldsymbol{\nabla}_2 f|^2 + 2\kappa^2}\left(\frac{\boldsymbol{\nabla}_2 f^\perp |\boldsymbol{\nabla}_2 f|^2 + \kappa^2 \boldsymbol{\nabla}_2 f^\perp}{|\boldsymbol{\nabla}_2 f|}\right) && \text{(A.9)}\\
&= \frac{1}{|\boldsymbol{\nabla}_2 f|^2 + 2\kappa^2}\left(\frac{\boldsymbol{\nabla}_2 f^\perp \left(|\boldsymbol{\nabla}_2 f|^2 + \kappa^2\right)}{|\boldsymbol{\nabla}_2 f|}\right) && \text{(A.10)}\\
&= \frac{|\boldsymbol{\nabla}_2 f|^2 + \kappa^2}{|\boldsymbol{\nabla}_2 f|^2 + 2\kappa^2}\,\frac{\boldsymbol{\nabla}_2 f^\perp}{|\boldsymbol{\nabla}_2 f|} && \text{(A.11)}\\
&= \frac{|\boldsymbol{\nabla}_2 f|^2 + \kappa^2}{|\boldsymbol{\nabla}_2 f|^2 + 2\kappa^2}\,\boldsymbol{s}_2^0 = q_2\,\boldsymbol{p}_2\ , && \text{(A.12)}
\end{aligned}$$

where $(*)$ holds as $\left(\boldsymbol{\nabla}_2 f^\perp\right)^\top \boldsymbol{\nabla}_2 f^\perp = |\boldsymbol{\nabla}_2 f|^2$.

What remains to show is the validity of the following equation:

$$\boldsymbol{\nabla}_2 u^\top \boldsymbol{P}\,\boldsymbol{\nabla}_2 u + \boldsymbol{\nabla}_2 v^\top \boldsymbol{P}\,\boldsymbol{\nabla}_2 v = q_1(u_{\boldsymbol{p}_1^2} + v_{\boldsymbol{p}_1^2}) + q_2(u_{\boldsymbol{p}_2^2} + v_{\boldsymbol{p}_2^2})\ . \tag{A.13}$$

The proof is a straightforward rewriting, which we exemplify for the u-part:

$$\begin{aligned}
\boldsymbol{\nabla}_2 u^\top \boldsymbol{P}\,\boldsymbol{\nabla}_2 u &= \boldsymbol{\nabla}_2 u^\top \left(q_1\,\boldsymbol{p}_1\boldsymbol{p}_1^\top + q_2\,\boldsymbol{p}_2\boldsymbol{p}_2^\top\right)\,\boldsymbol{\nabla}_2 u && \text{(A.14)}\\
&= q_1\left(\boldsymbol{\nabla}_2 u^\top \boldsymbol{p}_1\boldsymbol{p}_1^\top \boldsymbol{\nabla}_2 u\right) + q_2\left(\boldsymbol{\nabla}_2 u^\top \boldsymbol{p}_2\boldsymbol{p}_2^\top \boldsymbol{\nabla}_2 u\right) && \text{(A.15)}\\
&= q_1\left(\boldsymbol{p}_1^\top \boldsymbol{\nabla}_2 u\right)^2 + q_2\left(\boldsymbol{p}_2^\top \boldsymbol{\nabla}_2 u\right)^2 && \text{(A.16)}\\
&= q_1 u_{\boldsymbol{p}_1}^2 + q_2 u_{\boldsymbol{p}_2}^2\ . && \text{(A.17)}
\end{aligned}$$

Equivalently one obtains for the v-part

$$\boldsymbol{\nabla}_2 v^\top \boldsymbol{P}\,\boldsymbol{\nabla}_2 v = q_1 v_{\boldsymbol{p}_1}^2 + q_2 v_{\boldsymbol{p}_2}^2\ . \tag{A.18}$$

Combining u- and v-part finally shows the correctness of (A.13):

$$\begin{aligned}
\boldsymbol{\nabla}_2 u^\top \boldsymbol{P}\,\boldsymbol{\nabla}_2 u + \boldsymbol{\nabla}_2 v^\top \boldsymbol{P}\,\boldsymbol{\nabla}_2 v &= q_1 u_{\boldsymbol{p}_1}^2 + q_2 u_{\boldsymbol{p}_2}^2 + q_1 v_{\boldsymbol{p}_1}^2 + q_2 v_{\boldsymbol{p}_2}^2 && \text{(A.19)}\\
&= q_1\left(u_{\boldsymbol{p}_1}^2 + v_{\boldsymbol{p}_1}^2\right) + q_2\left(u_{\boldsymbol{p}_2}^2 + v_{\boldsymbol{p}_2}^2\right)\ . && \text{(A.20)}
\end{aligned}$$

□

A.2 Diffusion Tensor of the Complementary Regulariser

Consider the complementary regulariser from (4.82)

$$\begin{aligned} V(\boldsymbol{\nabla}_2 u, \boldsymbol{\nabla}_2 v) &= \Psi_V\Big(u_{\boldsymbol{r}_1}^2 + v_{\boldsymbol{r}_1}^2\Big) + u_{\boldsymbol{r}_2}^2 + v_{\boldsymbol{r}_2}^2 && \text{(A.21)} \\ &= \Psi_V\Big((\boldsymbol{r}_1^\top \boldsymbol{\nabla}_2 u)^2 + (\boldsymbol{r}_1^\top \boldsymbol{\nabla}_2 v)^2\Big) + (\boldsymbol{r}_2^\top \boldsymbol{\nabla}_2 u)^2 + (\boldsymbol{r}_2^\top \boldsymbol{\nabla}_2 v)^2 && \text{(A.22)} \\ &= \Psi_V\Big((r_{11} u_x + r_{12} u_y)^2 + (r_{11} v_x + r_{12} v_y)^2\Big) && \text{(A.23)} \\ &\quad + \ (r_{21} u_x + r_{22} u_y)^2 + (r_{21} v_x + r_{22} v_y)^2 \ , \end{aligned}$$

with $\boldsymbol{r}_1 := (r_{11}, r_{12})^\top$ and $\boldsymbol{r}_2 := (r_{21}, r_{22})^\top$. Its contributions to the Euler-Lagrange equations are given by

$$\operatorname{div} \begin{pmatrix} \partial_{u_x} V \\ \partial_{u_y} V \end{pmatrix} \quad \text{and} \quad \operatorname{div} \begin{pmatrix} \partial_{v_x} V \\ \partial_{v_y} V \end{pmatrix} \ , \tag{A.24}$$

respectively. We exemplify the computation of the first expression (A.24). The second one then follows analogously.

With the abbreviation

$$\psi_V'(\boldsymbol{r}_i) := \Psi_V'\Big(u_{\boldsymbol{r}_i}^2 + v_{\boldsymbol{r}_i}^2\Big) \ , \tag{A.25}$$

we start by computing the expressions

$$\begin{aligned} \partial_{u_x} V &= 2\left(\psi_V'(\boldsymbol{r}_1)\, u_{\boldsymbol{r}_1} r_{11} + u_{\boldsymbol{r}_2} r_{21}\right) \ , && \text{(A.26)} \\ \partial_{u_y} V &= 2\left(\psi_V'(\boldsymbol{r}_1)\, u_{\boldsymbol{r}_1} r_{12} + u_{\boldsymbol{r}_2}\, r_{22}\right) \ . && \text{(A.27)} \end{aligned}$$

Plugging (A.26) and (A.27) into (A.24) gives

$$\operatorname{div} \begin{pmatrix} \partial_{u_x} V \\ \partial_{u_y} V \end{pmatrix} = 2 \operatorname{div} \begin{pmatrix} \psi_V'(\boldsymbol{r}_1)\, u_{\boldsymbol{r}_1} r_{11} + u_{\boldsymbol{r}_2} r_{21} \\ \psi_V'(\boldsymbol{r}_1)\, u_{\boldsymbol{r}_1} r_{12} + u_{\boldsymbol{r}_2} r_{22} \end{pmatrix} \ . \tag{A.28}$$

By multiplying out the expressions inside the divergence expressions one ends up with

$$\operatorname{div} \begin{pmatrix} \partial_{u_x} V \\ \partial_{u_y} V \end{pmatrix} = 2 \operatorname{div} \begin{pmatrix} (\psi_V'(\boldsymbol{r}_1)\, r_{11}^2 + r_{21}^2)\, u_x + (\psi_V'(\boldsymbol{r}_1)\, r_{11} r_{12} + r_{21} r_{22})\, u_y \\ (\psi_V'(\boldsymbol{r}_1)\, r_{11} r_{12} + r_{21} r_{22})\, u_x + (\psi_V'(\boldsymbol{r}_1)\, r_{12}^2 + r_{22}^2)\, u_y \end{pmatrix} \ . \tag{A.29}$$

Note that the divergence expression has the form

$$\operatorname{div} \begin{pmatrix} a\, u_x + b\, u_y \\ b\, u_x + c\, u_y \end{pmatrix} = \operatorname{div} \left(\begin{pmatrix} a & b \\ b & c \end{pmatrix} \boldsymbol{\nabla}_2 u \right) \ , \tag{A.30}$$

using a diffusion tensor

$$\boldsymbol{D} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \ . \tag{A.31}$$

Hence we can write equation (A.29) in diffusion tensor notation as

$$\operatorname{div}\begin{pmatrix}\partial_{u_x} V\\ \partial_{u_y} V\end{pmatrix} = 2\operatorname{div}\left(\boldsymbol{D}\,\boldsymbol{\nabla}_2 u\right) \quad , \tag{A.32}$$

with the diffusion tensor

$$\boldsymbol{D} := \begin{pmatrix}\psi_V'(\boldsymbol{r}_1)\cdot r_{11}^2 + 1\cdot r_{21}^2 & \psi_V'(\boldsymbol{r}_1)\cdot r_{11}r_{12} + 1\cdot r_{21}r_{22}\\ \psi_V'(\boldsymbol{r}_1)\cdot r_{11}r_{12} + 1\cdot r_{21}r_{22} & \psi_V'(\boldsymbol{r}_1)\cdot r_{12}^2 + 1\cdot r_{22}^2\end{pmatrix} \quad . \tag{A.33}$$

We multiplied the second term of each sum by a factor of 1 to clarify that the eigenvalues of $\boldsymbol{D}$ are $\psi_V'(\boldsymbol{r}_1)$ and 1, respectively. The corresponding eigenvectors are $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. We thus can rewrite the tensor $\boldsymbol{D}$ as

$$\begin{aligned}\boldsymbol{D} &= \begin{pmatrix}r_{11} & r_{21}\\ r_{12} & r_{22}\end{pmatrix}\begin{pmatrix}\psi_V'(\boldsymbol{r}_1)\cdot r_{11} & \psi_V'(\boldsymbol{r}_1)\cdot r_{12}\\ 1\cdot r_{21} & 1\cdot r_{22}\end{pmatrix}\\ &= (\boldsymbol{r}_1 \,|\, \boldsymbol{r}_2)\begin{pmatrix}\psi_V'(\boldsymbol{r}_1) & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\boldsymbol{r}_1^\top\\ \boldsymbol{r}_2^\top\end{pmatrix} \quad .\end{aligned} \tag{A.34}$$

This shows that $\boldsymbol{D}$ is identical to $\boldsymbol{D}_{\mathrm{CR}}$ from (4.83), as it can be written as

$$\boldsymbol{D} \;=\; \psi_V'(\boldsymbol{r}_1)\,\boldsymbol{r}_1\,\boldsymbol{r}_1^\top + \boldsymbol{r}_2\,\boldsymbol{r}_2^\top \;=\; \Psi_V'\Big(u_{\boldsymbol{r}_1}^2 + v_{\boldsymbol{r}_1}^2\Big)\cdot \boldsymbol{r}_1\,\boldsymbol{r}_1^\top + 1\cdot \boldsymbol{r}_2\,\boldsymbol{r}_2^\top \quad . \tag{A.35}$$

□

A.3 Equivalence between Two Energy Functionals

Consider the two energy functionals

$$E_1 \;=\; \int_\Omega \left(\sqrt{m^2+\varepsilon^2} \;+\; \alpha\sqrt{v^2+\varepsilon^2}\,\right)\mathrm{d}x\,\mathrm{d}y \quad , \tag{A.36}$$

$$E_2 \;=\; \int_\Omega \left(2\varepsilon^2\sqrt{1+\frac{m^2}{\varepsilon^2}} \;+\; \alpha\,2\varepsilon^2\sqrt{1+\frac{v^2}{\varepsilon^2}}\,\right)\mathrm{d}x\,\mathrm{d}y \quad , \tag{A.37}$$

where m^2 denotes some quadratic data term and v^2 some quadratic smoothness term, respectively. Note that both energies use the same penaliser in the data and smoothness term. The energy E_1 uses subquadratic penaliser $\Psi(s^2) = \sqrt{s^2+\varepsilon^2}$ that leads to modified L_1 penalisation in the data term and to TV regularisation in the smoothness term. The energy E_2 uses the Charbonnier function in both terms.

We now wish to prove that E_2 is (up to a constant factor) equivalent to E_1. This in turn shows that minimising both energies leads to the same solution. Let us denote the

integrand of E_2 by F_2. The latter, we rewrite as

$$F_2 = 2\varepsilon^2\sqrt{1+\frac{m^2}{\varepsilon^2}} + \alpha\, 2\varepsilon^2\sqrt{1+\frac{v^2}{\varepsilon^2}} \tag{A.38}$$

$$= 2\varepsilon^2\left(\sqrt{1+\frac{m^2}{\varepsilon^2}} + \alpha\sqrt{1+\frac{v^2}{\varepsilon^2}}\right) \tag{A.39}$$

$$= 2\varepsilon^2\left(\sqrt{\frac{\varepsilon^2+m^2}{\varepsilon^2}} + \alpha\sqrt{\frac{\varepsilon^2+v^2}{\varepsilon^2}}\right) \tag{A.40}$$

$$= 2\varepsilon^2\left(\frac{1}{\varepsilon}\sqrt{m^2+\varepsilon^2} + \alpha\frac{1}{\varepsilon}\sqrt{v^2+\varepsilon^2}\right) \tag{A.41}$$

$$= 2\varepsilon\left(\sqrt{m^2+\varepsilon^2} + \alpha\sqrt{v^2+\varepsilon^2}\right) \quad . \tag{A.42}$$

This shows that $E_2 = 2\varepsilon\, E_1$ and concludes our proof.

□

Own Publications

[1] M. Breuß, H. Zimmer, and J. Weickert. Can variational models for correspondence problems benefit from upwind discretisations? *Journal of Mathematical Imaging and Vision*, 39(3):230–244, 2011.

[2] O. Demetz, J. Weickert, A. Bruhn, and H. Zimmer. Optic flow scale space. In A. M. Bruckstein et al., editor, *Scale Space and Variational Methods in Computer Vision (SSVM)*, volume 6667 of *Lecture Notes in Computer Science*, pages 713–724. Springer, Berlin, 2011.

[3] P. Gwosdek, H. Zimmer, S. Grewenig, A. Bruhn, and J. Weickert. A highly efficient GPU implementation for variational optic flow based on the Euler-Lagrange framework. In *Proc. 2010 ECCV Workshop on Computer Vision with GPUs*, Heraklion, Greece, Sept. 2010. Web page: `http://www.mia.uni-saarland.de/Research/ComplOF-GPU/`.

[4] A. Luxenburger, H. Zimmer, P. Gwosdek, and J. Weickert. Fast PDE-based image analysis in your pocket. In A. M. Bruckstein et al., editor, *Scale Space and Variational Methods in Computer Vision (SSVM)*, volume 6667 of *Lecture Notes in Computer Science*, pages 544–555. Springer, Berlin, 2011.

[5] L. Valgaerts, A. Bruhn, H. Zimmer, J. Weickert, C. Stoll, and C. Theobalt. Joint estimation of motion, structure and geometry from stereo sequences. In *Computer Vision – ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, pages 568–581. Springer, Berlin, 2010. Web page: `http://www.mia.uni-saarland.de/valgaerts/eccv10/sceneflow/`.

[6] S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer. Modeling temporal coherence for optical flow. In *Proc. Thirteenth International Conference on Computer Vision (ICCV)*, Barcelona, Spain, Nov. 2011. IEEE Computer Society Press.

[7] H. Zimmer, M. Breuß, J. Weickert, and H.-P. Seidel. Hyperbolic numerics for variational approaches to correspondence problems. In X.-C. Tai et al., editor, *Scale Space and Variational Methods in Computer Vision (SSVM)*, volume 5567 of *Lecture Notes in Computer Science*, pages 636–647. Springer, Berlin, 2009.

[8] H. Zimmer, A. Bruhn, L. Valgaerts, M. Breuß, J. Weickert, B. Rosenhahn, and H.-P. Seidel. PDE-based anisotropic disparity-driven stereo vision. In O. Deussen, D. Keim, and D. Saupe, editors, *Proceedings of Vision, Modeling, and Visualization (VMV)*, pages 263–272. AKA Heidelberg, Oct. 2008.

[9] H. Zimmer, A. Bruhn, and J. Weickert. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. *Computer Graphics Forum (Proceedings of Eurographics)*, 30(2):405–414, Apr. 2011. Web page: `http://www.mia.uni-saarland.de/Research/SR-HDR/`.

[10] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, Apr. 2011.

[11] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H.-P. Seidel. Complementary optic flow. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 5681 of *Lecture Notes in Computer Science*, pages 207–220. Springer, Berlin, 2009.

Bibliography

[ADA+04] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. In *Proc. ACM SIGGRAPH*, pages 294–302, 2004.

[ADK99] G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal on Applied Mathematics*, 60(1):156–182, 1999.

[ADSW00] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert. Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach. In *Proc. IAPR Workshop on Machine Vision*, pages 423–427, Tokyo, Japan, November 2000.

[ADSW02] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert. Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach. *Journal of Visual Communication and Image Representation*, 13(1/2):3–21, 2002.

[AELS99] L. Alvarez, J. Esclarín, M. Lefébure, and J. Sánchez. A PDE model for computing the optical flow. In *Proc. XVI Congreso de Ecuaciones Diferenciales y Aplicaciones*, pages 1349–1356, Las Palmas de Gran Canaria, Spain, September 1999.

[Ana89] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[AWS00] L. Alvarez, J. Weickert, and J. Sánchez. Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39(1):41–56, August 2000.

[BA96] M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, January 1996.

[BAS07] R. Ben-Ari and N.A. Sochen. Variational stereo vision with sharp discontinuities and occlusion handling. In *Proc. 2007 IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007. IEEE Computer Society Press.

[BBH03] M.Z. Brown, D. Burschka, and G.D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, August 2003.

[BBPW04] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004, Part IV*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, 2004.

[BFB94] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, February 1994.

[BGW91] J. Bigün, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):775–790, August 1991.

[BM04] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(1):221–255, 2004.

[BM11] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):500–513, March 2011.

[BPT88] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, August 1988.

[Bra77] A. Brandt. Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation*, 31(138):333–390, April 1977.

[Bru06] A. Bruhn. *Variational Optic Flow Computation: Accurate Modelling and Efficient Numerics*. PhD thesis, Department of Mathematics and Computer Science, Saarland University, Saarbrücken, Germany, July 2006.

[BSL+10] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, September 2010. Web page: `http://vision.middlebury.edu/flow/`.

[BVZ01] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, November 2001.

[BW05] A. Bruhn and J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proc. Tenth International Conference on Computer Vision*, volume 1, pages 749–755, Beijing, China, October 2005. IEEE Computer Society Press.

[BWF+05] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr. Variational optical flow computation in real-time. *IEEE Transactions on Image Processing*, 14(5):608–615, May 2005.

[BWKS06] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision*, 70(3):257–277, December 2006.

[BWS05] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.

[BZ87] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.

[CBFAB94] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proc. 1994 IEEE International Conference on Image Processing*, volume 2, pages 168–172, Austin, TX, November 1994. IEEE Computer Society Press.

[Cha04] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1–2):89–97, 2004.

[CLMC92] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 32:1895–1909, 1992.

[Coh93] I. Cohen. Nonlinear variational method for optical flow computation. In *Proc. Eighth Scandinavian Conference on Image Analysis*, volume 1, pages 523–530, Tromsø, Norway, May 1993.

[Coo67] S. A. Coons. Surfaces for computer aided design of space forms. Technical Report MIT/LCS/TR-41, Massachusetts Institute of Technology, Cambridge, MA, June 1967.

[Coo08] T. Cooke. Two applications of graph-cuts to image processing. In *Proc. International Conference on Digital Image Computing: Techniques and Applications, DICTA 2008*, pages 498–504, Canberra, Australia, December 2008. IEEE Computer Society.

[CPK09] J. Choi, M.K. Park, and M.G. Kang. High dynamic range image reconstruction with spatial resolution enhancement. *The Computer Journal*, 52(1):114–125, 2009.

[DM97] P.E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. ACM SIGGRAPH*, pages 369–378, 1997.

[dSK02] G.N. de Souza and A.C. Kak. Vision for mobile robot navigation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):237–267, 2002.

[EF97] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6(12):1646–1658, December 1997.

[Els62] L. E. Elsgolc. *Calculus of Variations*. Pergamon, London, England, 1962.

[EM09] M. Ebrahimi and A.L. Martel. A PDE approach to coupled super-resolution with non-parametric motion. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 5681 of *Lecture Notes in Computer Science*, pages 112–125. Springer, Berlin, 2009.

[FG87] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, Switzerland, June 1987.

[FKN73] S. Fučik, A. Kratochvil, and J. Nečas. Kačanov–Galerkin method. *Commentationes Mathematicae Universitatis Carolinae*, 14(4):651–659, 1973.

[FLP01] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, Cambridge, MA, 2001.

[FLW02] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. In *Proc. ACM SIGGRAPH*, pages 249–256, 2002.

[FM08] B. Fischer and J. Modersitzki. Ill-posed medicine – an introduction to image registration. *Problems*, 24(3):1–19, 2008.

[FREM04] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):1327–1344, October 2004.

[GAW+10] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H.P.A. Lensch. Optimal hdr reconstruction with linear digital cameras. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 215–222, an Francisco, CA, USA, June 2010. IEEE Computer Society Press.

[GB97] P. Golland and A. M. Bruckstein. Motion from color. *Computer Vision and Image Understanding*, 68(3):346–362, December 1997.

[GC10] B. Goldluecke and D. Cremers. Convex relaxation for multilabel problems with product label spaces. In *Computer Vision – ECCV 2010*, volume 6315 of *Lecture Notes in Computer Science*, pages 225–238. Springer, Berlin, 2010.

[GG06] B.K. Gunturk and M. Gevrekci. High-resolution image reconstruction from multiple differently exposed images. *IEEE Signal Processing Letters*, 13(4):197–200, 2006.

[GG09] J. Gulbins and R. Gulbins. *Photographic Multishot Techniques*. Rocky Nook, Santa Barbara, CA, USA, 2009.

[Gla84] F. Glazer. Multilevel relaxation in low-level computer vision. In A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis*, pages 312–330. Springer, Berlin, 1984.

[Gre09] H. Greenspan. Super-resolution in medical imaging. *The Computer Journal*, 52(1):43–63, 2009.

[Gro06] T. Grosch. Fast and robust high dynamic range image generation with camera and object movement. In L. Kobbelt, T. Kuhlen, T. Aach, and R. Westermann, editors, *Proceedings of Vision, Modeling, and Visualization (VMV) 2006*, pages 277–284. AKA Heidelberg, 2006.

[GS78] W. Gentzsch and A. Schlüter. Über ein Einschrittverfahren mit zyklischer Schrittweitenänderung zur Lösung parabolischer Differentialgleichungen. *Zeitschrift für angewandte Mathematik und Mechanik*, 58:T415–T416, 1978.

[GT08] H. Grossauer and P. Thoman. GPU-based multigrid: Real-time performance in high resolution nonlinear image processing. In A. Gasteratos, M. Vincze, and J. K. Tsotsos, editors, *Computer Vision Systems*, volume 5008 of *Lecture Notes in Computer Science*, pages 141–150. Springer, Berlin, 2008.

[GvV96] S. Ghosal and P. Č. Vaněk. Scalable algorithm for discontinuous optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2):181–194, February 1996.

[GWB10] S. Grewenig, J. Weickert, and A. Bruhn. From box filtering to fast explicit diffusion. In M. Goesele, S. Roth, A. Kuijper, B. Schiele, and K. Schindler, editors, *Pattern Recognition*, volume 6376 of *Lecture Notes in Computer Science*, pages 533–542. Springer, Berlin, 2010.

[Hac85] W. Hackbusch. *Multigrid Methods and Applications*. Springer, New York, 1985.

[Hor86] B. K. P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

[HS81] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[HTO07] H.B. Haraldsson, M. Tanaka, and M. Okutomi. Reconstruction of a high dynamic range and high resolution image from a multisampled image sequence. In *Proc. 14th International Conference on Image Analysis and Processing (ICIAP)*, pages 303–310, Modena, Italy, September 2007.

[Hub81] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[HZ00] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[JLW08] K. Jacobs, C. Loscos, and G. Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications*, 28:84–93, 2008.

[JO08] T. Jinno and M. Okuda. Motion blur free HDR image acquisition using multiple exposures. In *Proc. 2008 IEEE International Conference on Image Processing*, pages 1304–1307, San Diego, CA, USA, October 2008.

[Joh87] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Studentlitteratur, Lund, 1987.

[KAR06] E.A. Khan, A.O. Akyuz, and E. Reinhard. Ghost removal in high dynamic range images. In *Proc. 2006 IEEE International Conference on Image Processing*, pages 2005–2008, Atlanta, USA, August 2006.

[KDN93] D. Koller, K. Daniilidis, and H.H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10(3):257–281, 1993.

[KER95] W. Kruger, W. Enkelman, and S. Rossle. Real-time estimation and tracking of optical flow vectors for obstacle detection. In *Proc. IEEE Computer Society Intelligent Vehicles '95 Symposium*, pages 204–209, Detroit, MI, USA, September 1995. IEEE Computer Society.

[KKR07] E.M. Kalmoun, H. Köstler, and U. Rüde. 3D optical flow computation using a parallel variational multigrid scheme with application to cardiac C-arm CT motion. *Image and Vision Computing*, 25(9):1482–1494, 2007.

[KM07] K. Krajsek and R. Mester. Bayesian model selection for optical flow estimation. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 142–151. Springer, Berlin, 2007.

[KP04] S.J. Kim and M. Pollefeys. Radiometric alignment of image sequences. In *Proc. 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 645–651, Washington, DC, June 2004. IEEE Computer Society Press.

[KR03] E.M. Kalmoun and U. Rüde. A variational multigrid method for computing the optical flow. In T. Ertl, B. Girod, G. Greiner, H. Niemann, H.-P. Seidel, E. Steinbach, and R. Westermann, editors, *Proceedings of Vision, Modeling, and Visualization (VMV)*, pages 577–584. Akademische Verlagsgesellschaft Berlin, 2003.

[KS03] H. Kim and K. Sohn. Hierarchical disparity estimation with energy-based regularization. In *Proc. 2003 IEEE International Conference on Image Processing*, volume 1, pages 373–376, Barcelona, Spain, September 2003.

[KUWS03] S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. In *Proc. ACM SIGGRAPH*, pages 319–325, 2003.

[LeV92] R. J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser, Basel, 1992.

[LeV02] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, Cambridge, UK, 2002.

[Lin94] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer, Boston, 1994.

[LK81] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, Canada, August 1981.

[LLB+10] C. Lipski, C. Linz, K. Berger, A. Sellent, and M. Magnor. Virtual Video Camera: Image-based viewpoint navigation through space and time. *Computer Graphics Forum*, 29(8):2555–2568, 2010.

[Low99] D.G. Lowe. Object recognition from local scale-invariant features. In *Proc. Seventh International Conference on Computer Vision*, pages 1150–1157, Corfu, Greece, September 1999.

[Low04] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[LV98] S.-H. Lai and B. C. Vemuri. Reliable and efficient computation of optical flow. *International Journal of Computer Vision*, 29(2):87–105, October 1998.

[LY09] C. Lei and Y. H. Yang. Optical flow estimation on coarse-to-fine region-trees using discrete optimization. In *Proc. 2009 IEEE International Conference on Computer Vision*, Kyoto, Japan, October 2009. IEEE Computer Society Press.

[LYT+08] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman. SIFT flow: Dense correspondence across different scenes. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part III*, volume 5304 of *Lecture Notes in Computer Science*, pages 28–42, Berlin, 2008. Springer.

[MB87] D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2):220–228, March 1987.

[MBW07] Y. Mileva, A. Bruhn, and J. Weickert. Illumination-robust variational optical flow with photometric invariants. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 152–162. Springer, Berlin, 2007.

[MG07] N. Menzel and M. Guthe. Freehand HDR photography with motion compensation. In H.P.A. Lensch, B. Rosenhahn, H.-P. Seidel, P. Slusallek, and J. Weickert, editors, *Proceedings of Vision, Modeling, and Visualization (VMV)*, pages 127–134. AKA Heidelberg, 2007.

[MKMS07] R. Mantiuk, G. Krawczyk, R. Mantiuk, and H.-P. Seidel. High dynamic range imaging pipeline: Perception-motivated representation of visual content. In B.E. Rogowitz, T.N. Pappas, and S.J. Daly, editors, *Human Vision and Electronic Imaging XII*, volume 6492 of *Proceedings of SPIE*, San Jose, USA, February 2007. SPIE.

[MKR09] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum*, 28(1):161–171, 2009.

[MKW97] P. Moulin, R. Krishnamurth, and J.W. Woods. Multiscale modeling and estimation of motion fields for video coding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(12):1606–1620, 1997.

[MM94] K. W. Morton and L. M. Mayers. *Numerical Solution of Partial Differential Equations*. Cambridge University Press, Cambridge, UK, 1994.

[MMK98] A.-R. Mansouri, A. Mitiche, and J. Konrad. Selective image diffusion: application to disparity estimation. In *Proc. 1998 IEEE International Conference on Image Processing*, volume 3, pages 284–288, Chicago, IL, October 1998.

[MN03] P. Mrázek and M. Navara. Selection of optimal stopping time for nonlinear diffusion filtering. *International Journal of Computer Vision*, 52(2/3):189–203, 2003. stopping time.

[MO99] A. Marquina and S. Osher. Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal. *SIAM Journal on Scientific Computing*, 22(2):387–405, 1999.

[MP95] S. Mann and R.W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. 48th IS&T Annual Conference*, pages 442–448, Washington DC, USA, May 1995.

[MP98] E. Mémin and P. Pérez. A multigrid approach for hierarchical motion estimation. In *Proc. Sixth International Conference on Computer Vision*, pages 933–938, Bombay, India, January 1998. Narosa Publishing House.

[MPG85] H.G. Musmann, P. Pirsch, and H.J. Grallet. Advances in picture coding. *Proceedings of the IEEE*, 73(4):523–547, 1985.

[MPSC09] D. Mitzel, T. Pock, T. Schoenemann, and D. Cremers. Video super resolution using duality based TV-L_1 optical flow. In J. Denzler, G. Notni, and H. Süße, editors, *Pattern Recognition*, volume 5748 of *Lecture Notes in Computer Science*, pages 432–441. Springer, Berlin, 2009.

[Nag90] H.-H. Nagel. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow. In O. Faugeras, editor, *Computer Vision – ECCV '90*, volume 427 of *Lecture Notes in Computer Science*, pages 139–148. Springer, Berlin, 1990.

[NBK08] T. Nir, A. M. Bruckstein, and R. Kimmel. Over-parameterized variational optical flow. *International Journal of Computer Vision*, 76(2):205–216, 2008.

[NE86] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.

[NN05] S.G. Narasimhan and S.K. Nayar. Enhancing resolution along multiple imaging dimensions using assorted pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):518–530, 2005.

[NS97] L. Ng and V. Solo. A data-driven method for choosing smoothing parameters in optical flow problems. In *Proc. 1997 IEEE International Conference on Image Processing*, volume 3, pages 360–363, Los Alamitos, CA, USA, 1997. IEEE Computer Society.

[NVI10] NVIDIA Corporation. *NVIDIA CUDA Programming Guide*, 3rd edition, February 2010. `http://developer.download.nvidia.com/compute/cuda/3_0/toolkit/docs/NVIDIA_CUDA_ProgrammingGuide.pdf`.

[NYUS08] H. Nakai, S. Yamamoto, Y. Ueda, and Y. Shigeyama. High resolution and high dynamic range image reconstruction from differently exposed images. In G. Bebis and et al., editors, *Advances in Visual Computing*, volume 5359 of *Lecture Notes in Computer Science*, pages 713–722. Springer, Berlin, 2008.

[PAT96] I. Patras, N. Alvertos, and G. Tziritas. Joint disparity and motion field estimation in steroscopic image sequences. In *Proc. 13th International Conference on Pattern Recognition*, volume 1, pages 359–362, Vienna, Austria, August 1996.

[PBB+06] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, April 2006.

[PM90] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[Pot52] R.B. Potts. Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society*, 48:106–109, 1952.

[PPK03] S. Park, M. Park, and M. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20:21–36, 2003.

[PSA+04] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. In *Proc. ACM SIGGRAPH*, pages 664–672, 2004.

[PSG+08] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part III*, volume 5304 of *Lecture Notes in Computer Science*, pages 792–805. Springer, Berlin, 2008.

[PTK85] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, September 1985.

[RBS99] M.A. Robertson, S. Borman, and R.L. Stevenson. Dynamic range improvement through multiple exposures. In *Proc. Sixth International Conference on Image Processing*, volume III, pages 159–163, Kobe, Japan, October 1999.

[RD96] L. Robert and R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In B. Buxton and R. Cipolla, editors, *Computer Vision – ECCV '96*, volume 1064 of *Lecture Notes in Computer Science*, pages 439–451. Springer, Berlin, 1996.

[RIY05] R. Raskar, A. Ilie, and J. Yu. Image fusion for context enhancement and video surrealism. In *Proc. ACM SIGGRAPH Courses*, pages 85–93, 2005.

[RMVS07] A.A. Rad, L. Meylan, P. Vandewalle, and S. Süsstrunk. Multidimensional image enhancement from a set of unregistered differently exposed images. In *Proc. IS&T/SPIE Electronic Imaging: Computational Imaging V*, volume 6498, 2007.

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[RSS07] P. Ruhnau, A. Stahl, and C. Schnörr. Variational estimation of experimental fluids flows with physics-based spatio-temporal regularization. *Measurement Science and Technology*, 18:755–763, 2007.

[RWPD05] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann Publishers, 2005.

[RWWK07] M. Raffel, C.E. Willert, S.T. Wereley, and J. Kompenhans. *Particle Image Velocimetry: A Practical Guide*. Springer, Berlin, second edition, 2007.

[SAH91] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proc. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 310–315, Maui, HI, June 1991. IEEE Computer Society Press.

[SBK10] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by GPU-accelerated large displacement optical flow. In *Computer Vision – ECCV 2010*, volume 6311 of *Lecture Notes in Computer Science*, pages 438–451. Springer, Berlin, 2010. number I.

[SBW05] N. Slesareva, A. Bruhn, and J. Weickert. Optic flow goes stereo: A variational method for estimating discontinuity-preserving dense disparity maps. In W. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 33–40. Springer, Berlin, 2005.

[SC06] T. Schoenemann and D. Cremers. Near real-time motion segmentation using graph cuts. In K. Franke, K.-R. Müller, B. Nickolay, and R. Schäfer, editors, *Pattern Recognition*, volume 4174 of *Lecture Notes in Computer Science*, pages 455–464. Springer, Berlin, 2006.

[Sch93] C. Schnörr. On functionals with greyvalue-controlled smoothness terms for determining optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1074–1079, 1993.

[Sch94] C. Schnörr. Segmentation of visual motion by minimizing convex non-quadratic functionals. In *Proc. Twelfth International Conference on Pattern Recognition*, volume A, pages 661–663, Jerusalem, Israel, October 1994. IEEE Computer Society Press.

[SH89] D. Shulman and J. Hervé. Regularization of discontinuous flow fields. In *Proc. Workshop on Visual Motion*, pages 81–86, Irvine, CA, March 1989. IEEE Computer Society Press.

[SKH08] A. Shekhovtsov, I. Kovtun, and V. Hlaváč. Efficient MRF deformation model for non-rigid image matching. *Computer Vision and Image Understanding*, 112:91–99, October 2008.

[SPC09a] F. Steinbrücker, T. Pock, and D. Cremers. Advanced data terms for variational optic flow estimation. In M.A. Magnor, B. Rosenhahn, and H. Theisel, editors, *Proceedings of the Vision, Modeling, and Visualization Workshop (VMV)*, pages 155–164. DNB, November 2009.

[SPC09b] F. Steinbrücker, T. Pock, and D. Cremers. Large displacement optical flow computation without warping. In *Proc. Twelfth International Conference on Computer Vision*, Kyoto, October 2009. IEEE Computer Society Press.

[SRB10] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 2010. IEEE Computer Society Press.

[SRLB08] D. Sun, S. Roth, J. P. Lewis, and M. J. Black. Learning optical flow. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part III*, volume 5304 of *Lecture Notes in Computer Science*, pages 83–97. Springer, Berlin, 2008.

[SS02] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002.

[SSB10] D. Sun, E. Sudderth, and M. J. Black. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23, pages 2226–2234. MIT Press, 2010.

[SSM09] F. Schubert, K. Schertler, and K. Mikolajczyk. A hands-on approach to high-dynamic-range and superresolution fusion. In *Proc. 2009 IEEE Workshop on Applications of Computer Vision (WACV)*, Snowbird, Utah, December 2009. IEEE Computer Society Press.

[ST94] J. Shi and C. Tomasi. Good features to track. In *Proc. 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 593–600, Seattle, WA, USA, June 1994. IEEE Computer Society Press.

[ST04] P. Sand and S.J. Teller. Video matching. In *Proc. ACM SIGGRAPH*, pages 592–599, 2004.

[SW49] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, 1949.

[Sze06] R. Szeliski. Image alignment and stitching: a tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1):1–104, 2006.

[Ter86] D. Terzopoulos. Image analysis using multigrid relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):129–139, March 1986.

[TH84] R. Tsai and T. Huang. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*, 1(2):317–339, 1984.

[Tho49] L. H. Thomas. Elliptic problems in linear difference equations over a network. Technical report, Watson Scientific Computing Laboratory, Columbia University, New York, NJ, 1949.

[TM07] A. Tomaszewska and R. Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *Proc. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, pages 49–56, Plzen-Bory, CZ, January 2007.

[Tor99] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*. Springer, Berlin, second edition, 1999.

[TP84] O. Tretiak and L. Pastor. Velocity estimation from image sequences with second order differential operators. In *Proc. Seventh International Conference on Pattern Recognition*, pages 16–19, Montreal, Canada, July 1984.

[Uns00] M. Unser. Sampling–50 years after Shannon. *Proceedings of the IEEE*, 88(4):569–587, April 2000.

[VBR+05] S. Vedula, S. Baker, P. Rander, R.T. Collins, and T. Kanade. Three-dimensional scene flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):475–480, March 2005.

[vG04] J. van de Weijer and T. Gevers. Robust optical flow from photometric invariants. In *Proc. 2004 IEEE International Conference on Image Processing*, volume 3, pages 1835–1838, Singapore, October 2004. IEEE Signal Processing Society.

[VSVV06] P. Vandewalle, L. Sbaiz, J. Vandewalle, and M. Vetterli. Aliasing is good for you: Joint registration and reconstruction for super-resolution. Technical Report LCAV-2006-001, School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, 2006.

[War03] G. Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics, GPU, and Game Tools*, 8(2):17–30, 2003.

[WCPB09] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure- and motion-adaptive regularization for high accuracy optic flow. In *Proc. 2009 IEEE International Conference on Computer Vision*, Kyoto, Japan, October 2009. IEEE Computer Society Press.

[Wei94] J. Weickert. Scale-space properties of nonlinear diffusion filtering with a diffusion tensor. Technical Report 110, Laboratory of Technomathematics, University of Kaiserslautern, Germany, October 1994.

[Wei96] J. Weickert. Theoretical foundations of anisotropic diffusion in image processing. *Computing Supplement*, 11:221–236, 1996.

[Wei98] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.

[WMR+09] A. Wedel, A. Meißner, C. Rabe, U. Franke, and D. Cremers. Detection and segmentation of independently moving objects from dense scene flow. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition – EMMCVPR*, volume 5681 of *Lecture Notes in Computer Science*, pages 14–27. Springer, Berlin, 2009.

[WPB10] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 2010. IEEE Computer Society Press.

[WPZ+08] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for TV-L^1 optical flow computation. In D. Cremers, B. Rosenhahn, A. L. Yuille, and F. R. Schmidt, editors, *Statistical and Geometrical Approaches to Visual Motion Analysis*, volume 5604 of *Lecture Notes in Computer Science*, pages 23–45. Springer, Berlin, September 2008.

[WRV+08] A. Wedel, C. Rabe, T. Vaudrey, T. Brox, U. Franke, and D. Cremers. Efficient dense scene flow from sparse or dense stereo data. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008*, volume 5302 of *Lecture Notes in Computer Science*, pages 739–751. Springer, Berlin, 2008.

[WS01a] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, December 2001.

[WS01b] J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14(3):245–255, May 2001.

[WTK87] A. Witkin, D. Terzopoulos, and M. Kass. Signal matching through scale space. *International Journal of Computer Vision*, 1(2):133–144, 1987.

[WTP+09] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic Huber-L^1 optical flow. In *Proc. 20th British Machine Vision Conference*, London, UK, September 2009. British Machine Vision Association.

[XCJ08] L. Xu, J. Chen, and J. Jia. A segmentation based variational model for accurate optical flow estimation. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part I*, volume 5302 of *Lecture Notes in Computer Science*, pages 671–684. Springer, Berlin, 2008.

[XJM10] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 2010. IEEE Computer Society Press.

[Yar85] L. P. Yaroslavsky. *Digital Picture Processing: An Introduction*. Springer, Berlin, 1985.

[YK06] K.-J. Yoon and I.-S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.

[You03] D. M. Young. *Iterative Solution of Large Linear Systems*. Dover, New York, 2003.

[ZPB07] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L^1 optical flow. In F.A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 214–223. Springer, Berlin, 2007.

[ZZSL10] L. Zhang, H. Zhang, H. Shen, and P. Li. A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, 90(3):848–859, February 2010.