Feature Invariance versus Change Estimation in Variational Motion Estimation

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

vorgelegt von

Oliver Demetz

Saarbrücken, 2015

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

Tag des Kolloquiums:
25.09.2015

Dekan:
Prof. Dr. Markus Bläser

Prüfungsausschuss:

Prof. Dr. Rainer Schulze-Pillot-Ziemen (Vorsitzender)

Prof. Dr. Joachim Weickert (1. Gutachter)

Prof. Dr. Andrés Bruhn (2. Gutachter)
Universität Stuttgart

Dr. Sven Gehring (akademischer Mitarbeiter)

Copyright

Kurzzusammenfassung

Die robuste Schätzung von Korrespondenzen in Bildfolgen ist eines der grundlegenden Probleme des Maschinensehens. Eine der großen Herausforderungen in der Praxis sind hierbei Aussehensänderungen von Objekten, da die traditionelle Helligkeitskonstanzannahme nur unter idealisierten Bedingungen gilt. Zwei Kapitel dieser Dissertation befassen sich mit diesem Problem im Kontext der Schätzung dichten optischen Flusses. Wir tragen zwei Lösungen bei die sehr verschiedenen Strategien folgen. Zuerst betrachten wir Invarianzen, also Eigenschaften des gegebenen Bildmaterials, die unter bestimmen Beleuchtungs- und Aussehensänderungen unverändert bleiben. Wir geben eine systematische Übersicht bestehender Invarianzen und stellen unsere *Complete Rank Transform* vor. Unser prototypisches variationelles Framework kann jede der besprochenen Invarianzen verwenden und macht so einen fairen Vergleich möglich. Unser zweiter Beitrag ist von der Erkenntnis motiviert, dass schwierige Aussehensänderungen, wie z.B. Schlagschatten, oft lokale Erscheinungen sind, Invarianzen jedoch global wirken. Wir entwickeln daher ein Modell das Beleuchtungsänderungen explizit einbezieht und die Konstanzannahme um diese Änderungen *lokal kompensiert*. Der dritte Beitrag dieser Dissertation betrifft das Skalenraum-Verhalten variationeller Optischer-Fluss-Methoden. Das Vorhandensein eines solchen Verhaltens im Regularisierungsparameter ist unstrittig, jedoch liegen bislang keine gründlichen Untersuchungen hierzu vor. Diese sind der Beitrag des vierten Kapitels dieser Dissertation.

Short Abstract

The robust estimation of correspondences in image sequences belongs to the fundamental problems in computer vision. One of the big practical challenges are appearance changes of objects, because the traditional brightness constancy assumption only holds under idealised conditions. Two chapters of this thesis approach this problem in the context of dense optic flow estimation. We contribute two solutions that follow very different strategies. First, we consider invariances, i.e. features of the input images that remain unaffected under certain changes in illumination and appearance. We present a systematic overview of available invariances and introduce our *complete rank transform*. Our prototypical variational framework can incorporate any of the discussed invariant features and facilities a fair comparison. Our second contribution is motivated by the insight that challenging appearance changes such as drop shadows often are local phenomena, but invariances operate globally. Thus, we develop a method that integrates illumination changes explicitly into our model and *compensates* the constancy assumption locally for them. The third contribution of this thesis considers the scale space behaviour of variational optic flow methods. The existence of such a behaviour in the regularisation parameter is unquestionable, however, there is no through analysis in this direction up to now. We close this gap as the third contribution of this thesis.

Zusammenfassung

Die robuste Schätzung von Korrespondenzen in Bildsequenzen ist eines der fundamentalen Probleme des Maschinensehens. Hierbei sind Änderungen des Aussehens von Objekten eine der größten Herausforderungen, da die traditionelle Helligkeitskonstanzannahme lediglich unter idealisierten Voraussetzungen gilt. Unter realistischen Bedingungen sind weitergehende Konzepte notwendig um verlässliche Korrespondenzen schätzen zu können. Im Zusammenhang mit der dichten Schätzung von optischem Fluss ist das Hauptziel dieser Dissertation, Vorgehensweisen und Konstanzannahmen zu entwickeln, zu analysieren und zu vergleichen, die robust mit unkontrollierten Beleuchtungssituationen umgehen können. In dieser Hinsicht leisten wir zwei Beiträge zur Lösung dieses Problems, die sehr verschiedenen Strategien folgen. Zuerst betrachten wir Invarianzen, die den weit-verbreitetsten Lösungsansatz in der Literatur heute darstellen. Dies sind Eigenschaften die vom gegebenen Bildmaterial abgeleitet werden können, und die unter bestimmen Beleuchtungs- und Aussehensänderungen unverändert bleiben. Wir geben eine systematische Übersicht bestehender Invarianzen und stellen zwei neue Reihenfolgen-basierte Features vor, die *Complete Rank Transform* und die *Complete Census Transform*. Wir untersuchen wichtige Eigenschaften und geeignete Metriken für diese Signaturen, und stellen ein generisches variationelles Framework vor. Jede der besprochenen Invarianzen kann ohne weitere Anpassungen in diesem Modell verwandt werden. Unser zweiter Beitrag ist vom großen Nachteil aller invarianzbasierten Konstanzannahmen motiviert, nämlich dass schwierige Aussehensänderungen lokale Erscheinungen sind, Invarianzen jedoch global wirken. Zum Beispiel im Falle von Schlagschatten ändern sich nur Teile des Bildes. Dies bedeutet im Umkehrschluss, dass Invarianzen in Bereichen ohne Änderung potentiell wertvolle Bildinformation blind verwerfen. Aus dieser Erkenntnis heraus entwickeln wir daher ein Modell das Beleuchtungsänderungen explizit einbezieht und das die Konstanzannahme um diese Änderungen *lokal* kompensiert. Schließlich betrachten wir einen weiteren Aspekt variationeller Optischer-Fluss-Methoden die die heute vorherrschenden Methoden zur Schätzung von optischem Fluss darstellen. All solche Methoden haben einen Regulariserungsterm gemein, dessen Gewicht typischerweise der entscheidende Parameter der gesamten Methode ist. Je größer sein Wert, desto glatter wird das Ergebnis sein. Offensichtlich besteht ein Skalenraum-Verhalten in diesem Parameter. Dieses ist jedoch nur im Kontext von Signalregularisierung und Bild-Skalenräumen wohlverstanden. Aus diesem Grund sind diese fehlenden Untersuchungen des *Optic Flow Scale Space* der dritte Beitrag dieser Dissertation.

Abstract

The robust estimation of correspondences in image sequences belongs to the fundamental problems in computer vision. One of the biggest challenges in this context are appearance changes, because the traditional brightness constancy assumption only holds under idealised conditions. In realistic scenarios, more advanced concepts are necessary to estimate correspondence reliably. Thus, in the context of dense optic flow estimation, the main goal of this thesis is to develop, analyse and compare strategies and constancy assumptions that are able to handle uncontrolled lighting situations robustly. In that respect, we contribute two solutions to the mentioned problem, that, however, follow very different strategies. First, we consider invariances, which are the dominating concept in literature to tackle appearance changes and uncontrolled lighting conditions today. Invariant features are properties that can be derived from the input images that remain unaffected under certain classes of intensity rescalings. We present a systematic and broad overview of available invariances from the literature and introduce two novel ordering-based features, the *complete rank transform* and the *complete census transform.* We analyse important properties and suitable metrics for these signatures and present a generic variational framework that allows to incorporate any of the discussed signatures. Our second contribution is motivated by the biggest disadvantage of invariance-based constancy assumptions: the most challenging appearance changes are *local* phenomena, but invariances act globally. For instance in case of drop shadows, only parts of the image change. Thus, in regions without appearance changes, the invariance blindly discards potentially valuable information. This fact motivates us to develop a method that integrates illumination changes explicitly into the data model and compensates the constancy assumption *locally* for occurring changes. Finally, we consider another aspect of optic flow, where variational methods are prevailing today. All such functionals have some type of regularisation term in common, whose weight is the crucial parameter in practice. The larger its value, the smoother the solution will be. Obviously, there exists scale space behaviour in this parameter, which is, however, only well understood in the context of signal regularisation and image scale spaces. Thus, we perform this missing analysis of the *optic flow scale space* as the third contribution of this thesis.

Acknowledgements

Let me begin by expressing my gratitude to Prof. Dr. Joachim Weickert for being the supervisor of my dissertation and for giving me the chance to join his Mathematical Image Analysis (MIA) group. He sparked my scientific interest for the field of computer vision and made this dissertation possible. I also want to thank Prof. Dr. Andrés Bruhn for agreeing to be the second reviewer of this dissertation. His kindliness, support, advice, expertise, and cleverness have always been a role model for me. I also want to thank my numerous collaborators for good and efficient teamwork, the friendly and positive atmosphere and a lot of motivation and fun. These are David Hafner, Christopher Schroers, Henning Zimmer, Michael Stoll, Vladislav Kramarev, Nico Persch, Levi Valgaerts, and Sebastian Volz. Moreover, I want to thank all other current members of the MIA group that provide this good working atmosphere every day. Especially, I want to mention and thank the heart and soul of our chair, Mrs. Ellen Wintringer, for keeping all organisatorial issues and bureaucracy as far as possible away from us. Also, there are a number of past group members I want to thank and remember: My former room mates Sven Grewenig and Kai Uwe Hagenburg, the fruitful discussions with Prof. Dr. Martin Welk and Prof. Dr. Michael Breuss, Marcus Hargarter for his Linux expertise and patience in explaining me all the bells and whistles of system administration, Simon Setzer, Markus Mainberger and Pascal Gwosdek. Many of the aforementioned people have become friends during my time in the MIA group, and I am very grateful for that. I further want to acknowledge the funding of my last three month that was made possible by the Graduate School of Computer Science of Saarland University. Generally, I want to mention very positively the almost ideal research conditions and environment here at Saarland University in which I was allowed to conduct my studies.

Finally, my family deserves my biggest gratitude. My parents Bertilla and Jürgen have always and unconditionally supported me with everything they could. I cannot say "Thank you" enough. Last but not least, I want to thank my girlfriend Kristina for her patience, support, warmth and love.

Contents

Chapter 1

Introduction

Given two images, the establishment of correspondence relations between positions in these two images is one of the fundamental problems in computer vision. Usually, these two images are consecutive frames of an image sequence – a movie –, and the task is to establish this relation for every pixel *densely*. This means, we want to determine for every position (or pixel) in the first frame to which position in the second frame it corresponds. We quantify this relation by a displacement vector field pointing from each position in the first frame to the corresponding new position in the second image. This is what we call *optic flow* field.

Since more than three decades, many researchers have addressed this problem. Interestingly, variational energy functionals for optic flow, whose history dates back to the very first dense optic flow method proposed by Horn and Schunck [1981], are still the prevailing methodology today, c.f. various accuracy benchmark platforms [Baker et al., 2011; Geiger et al., 2012; Butler et al., 2012]. Also the present thesis focusses completely on variational strategies to compute optic flow. The core of such methods is a global cost functional that integrates a cost function over the image domain, and the goal is to find a flow field that minimises the cost. The modeling within such an energy functional can be performed in a completely transparent and elegant way, as all assumptions just need to be encoded in cost terms that are added together: All desired properties of the solution can be rewarded with low costs, and, vice versa, unmeant configurations in the solution can be penalised severely. Typically, an energy functional for optic flow comprises at least two types of cost terms: a data term relates the input image sequence with the optic flow field, and ensures that the optic flow is appropriate for the given images. The other term is a smoothness term that usually demands some kind of spatial regularity of the flow, for instance that neighbouring pixels have a similar displacement vector.

1.1 Applications

Being able to reliably and accurately estimate optic flow fields is an important prerequisite for an enormous variety of applications.

Optic flow is probably the most intuitive and straightforward strategy for tracking [Shi and Tomasi, 1994] objects from frame to frame. The trajectory of an object over time can be determined by computing optic flow fields between all consecutive frames and following the displacement vectors at the actual object position. Ochs et al. [2013] uses optic flow as input data for his long-term trajectory analysis. Also in the context of video compression, motion information can be useful if the appearance of a moving object does not change. The MPEG video coding standard incorporates a block motion compensation feature, where frames can be predicted from reference frames with motion information about blocks of size 16×16 pixel. The more recent High Efficient Video Coding (HEVC) standard exploits accurate motion information even more [Sullivan et al., 2012]. Not only for video compression purposes, also in order to interpolate movies in temporal direction [Rakêt et al., 2012] optic flow is helpful. For instance, if the frame rate of a video should be changed [Wang et al., 2014], additional frames are necessary. Optic flow also plays an important role for visual effects in movies [Radke, 2012; Sadek et al., 2013]. Another example is the so-called bullet time effect, where a scene is captured from a couple of different view points. Using optic flow, these different view angles can be interpolated and a version of the movie can be computed where a virtual camera seems to fly around the scene. Another application of optic flow in the graphics domain is high dynamic range (HDR) reconstruction. Whenever differently exposed images of a scene should be reconstructed to a HDR composite and these exposures were subject to any motion, the necessary alignment can be achieved with optic flow [Zimmer et al., 2011a][Hafner et al., 2014]. In automotive driver assistance systems, optic flow plays an increasingly important role [Onkarappa and Sappa, 2014], for instance for the detection of obstacles. In the same context, Stein [2004] proposes the usage of correspondences computed using the census transform [Zabih and Woodfill, 1994]. Generally, visual correspondence algorithms are important for the whole research field of robotics and are used for various purposes, e.g. for optic flow based obstacle avoidance [Santos-Victor et al., 1993]. The research field on structure-from-motion relies on the assumption that the captured scene did not contain individually moving objects such as cars or persons. Then, any motion between consecutive frames can be attributed to camera motion, and the geometry of the scene can be deduced. This property is also exploited in the field of photometric 3-D reconstruction [Schroers et al., 2012], [Schneevoigt et al., 2014] and visual odometry [Steinbrücker et al., 2011]. Garrido et al. [2013] use optic flow to improve face reconstructions. In case of a stereoscopic setting, i.e. if two cameras capture a picture at the same time, the static-scene assumption is always fulfilled and the correspondence problem becomes much easier [Hartley and Zisserman, 2004]. Another research field where dense correspondence plays an important role is registration. In this setting, typically the two images to be related are of different kind, for instance, one could register the CT scan of a patient's hand to the CT scan of a *normal* (or standardised) hand and then compute an image of the patients hand in *normalised* shape and pose. Such applications are not exclusively in the medical domain, optic flow-based registration can also be applied for face registration [Demetz et al., 2007]. Finally, many denoising algorithms try to remove noise by averaging similar pixels. If more than one image of the same (moving) object are available, Nir et al. [2005] average in the spatial and temporal domain by considering the corresponding pixels from the other images.

1.2 Main Challenges in Optic Flow

The seminal work of Horn and Schunck [1981] constitutes the starting point of the research on optic flow. Since then, the main challenges and major problems in the context of optic flow estimation have become evident and have been approached.

Noise. Almost all methods for optic flow are vulnerable against noise in the input image sequence. To this end, Bruhn et al. [2005] combined the local and noise-robust method of Lucas and Kanade [1981] with the global variational ansatz of Horn and Schunck [1981]. Another way to approach this problem is proposed by Nir et al. [2005], who additionally minimise for a denoised version of the input images and thus incorporate the problematic noise explicitly into their model.

Occlusions. Another major problem of dense correspondence estimation algorithms are occluded image regions, because there no corresponding structure exists at all. This happens if a structure enters or leaves the image, or if one object moves behind another and becomes invisible by that. Also in this respect, several ways to tackle this problem have been proposed. As one example, Alvarez et al. [2007] propose a method to estimate such regions explicitly and their model ignores them in further computations.

Outliers. In general, all types of phenomena that violate our assumptions about the scene or input images can be considered as outliers. For such cases, Black and Anandan [1991] incorporated the concepts of robust statistics into variational optic flow.

Appropriate regularisation. Horn and Schunck [1981] proposed the first variational formulation for optic flow, where they incorporated a quadratic regularisation of the flow derivatives [Tikhonov, 1963]. This type of regularisation however penalises any discontinuities in the solution, although image sequences as they are captured in real world do contain sharp discontinuities. Many researchers have addressed this issue. One of the earliest attempts in this direction is made by Nagel and Enkelmann [1986]. They propose an anisotropic regularisation term where smoothness is demanded mainly along images edges, across them the smoothness assumption is weighted down. The works of Shulman and Herve [1989] and Cohen [1993] propose sub-quadratic regularisation terms for optic flow, such that discontinuities can be preserved.

Large displacements. The typical data terms for optic flow are highly non-convex expressions, as the sought unknown flow appears in the argument of the image function. Thus, most variational methods for optic flow make use of a linearised data term, i.e. approximate the image sequence locally by a function that is linear in space and time. At the one hand, this linearisation is very important as it makes the minimisation of the functional tractable and converts the formerly non-convex problem into a convex one. On the other hand, this linearisation has serious problems with large displacements. To this end, multi-scale strategies have been proposed by Witkin et al. [1987] and Anandan [1989]. Alvarez et al. [1999b] refrain from the linearisation and instead perform a gradient descent on the non-convex functional. To not be trapped in local minima, they embedded also their descent in a coarse-to-fine strategy. A more efficient solution to this problem is proposed by Brox et al. [2004], who solve on each level only for a flow increment.

Varying illumination. In principle all motion estimation methods rely on some sort of constancy assumption. The most basic assumption that the colour of an object does not change over time is intuitively correct. However, there is a difference between the intuitive understanding of the diffuse reflectance colour of a surface and the actually perceived light spectrum that is reflected from an object surface into the direction of the observer in reality. In practice, this depends on the surface orientation, the light source, the surface roughness, and many more factors. There exist various different mathematical reflectance models, see e.g. the book of [Pharr and Humphreys, 2010]. Thus, the intuitive assumption that the grey value of corresponding pixels remains constant over time unfortunately only holds under idealised conditions, and further concepts must be employed to handle realistic phenomena such as automatic camera re-adjustments, changeable weather conditions or physical effects such as shadows and highlights reliably.

1.3 Scope of this Thesis

The first two main chapters of this thesis are devoted to the last-mentioned problem, the adequate treatment of appearance changes in variational optic flow methods. The goal is to analyse different strategies to tackle such changes and to perform an objective comparison of these concepts. In the end, the reader should have a broad overview over existing alternative strategies as well as their individual advantages and shortcomings. The different strategies to approach this problem can be roughly divided into two classes:

First, the predominant concept to tackle this problem today is through *invariances.* Approaches of this type are covered in Chapter 2 of this dissertation. Depending on the situation, a feature, i.e. a property which can be computed from the input images, is chosen that is not affected by the type of illumination change one expects to occur. As an example, for colour image material Zimmer et al. [2011b] choose the hue channel of the hue-saturation-value (HSV) colour space representation of the images and argue that this hue channel does not change when entering a shadow region. However, there is no free lunch, and every invariance discards information. Especially for the class of morphologically invariant transform that we will consider especially in this thesis, the problem of discarding too much information is an issue. Throughout our analysis, we will see that in designing invariance-based data terms, the big issue is to find features that discard as little information as possible.

This inevitable loss of information is the main motivation for the second class of strategies being discussed in Chapter 3 of this thesis. Their idea is – instead of ignoring parts of the available information – to explicitly *estimate* the change. That way, the change becomes an additional unknown, and assumptions about this change, e.g. smoothness and plausibility, can be incorporated formally. Moreover, absolutely no image information is discarded when following this strategy. However, we will see that an appropriate parametrisation of the illumination change is important, and we will show how to learn the parametrisation from training data.

Chapter 4 of this dissertation is devoted to the *optic flow scale space.* The central quantity being analysed in this context is the weight of the flow regularisation term, which is usually the crucial parameter of all variational optic flow methods. The larger its value, the smoother the minimising flow field will be. The existence of a scale space behaviour in potentially all variational optic flow methods is thus undeniable. However, formal scale space evolution equations as they are well known and understood for traditional image scale spaces have not been found before. This is the contribution of Chapter 4. For the seminal model of Horn and Schunck [1981], we derive the evolution equations and analyse important properties and differences to established concepts. Practical experiments underline our theoretical findings.

1.4 Formal Problem Definition

Let us define the optic flow correspondence problem formally. We assume to be given two consecutive frames of an input image sequence

$$\boldsymbol{f}_i : \Omega \to \mathcal{R}^{n_c}, \quad i \in \{1, 2\} \tag{1.1}$$

where $\Omega \subset \mathbb{R}^2$ is the rectangular image domain and $\mathcal{R} \subset \mathbb{R}$ denotes the range of intensity values for each channel. The dimensionality n_c of the co-domain of $\boldsymbol{f}$ depends on the type of image material; for grey value imagery $n_c = 1$, and for typical colour images (e.g. RGB imagery) $n_c = 3$. The range of intensity values $\mathcal{R}$ in each dimension is a matter of definition. Typically, images are stored on disk as integer values with 8 bit per channel which suggests the interval $[0, 255]$.

With the term *optic flow field*, we denote a dense 2-D vector field. This vector field describes the relative displacement for each position between the first frame $\boldsymbol{f}_1$ and the second frame $\boldsymbol{f}_2$ of the image sequence. More formally, the optic flow field $\boldsymbol{u} = (u, v)^\top : \Omega \to \mathbb{R}^2$ parametrises at each position $\boldsymbol{x} = (x, y)^\top \in \Omega$ the displacement by the horizontal offset $u(\boldsymbol{x})$ and the vertical offset $v(\boldsymbol{x})$. Hence, the flow $(u, v)^\top$ expresses the correspondence relation

$$(x, y)^\top \multimap\!\bullet \; (x + u(x, y), y + v(x, y))^\top . \tag{1.2}$$

From a continuous-time point of view, the two frames $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ can be understood as slices of a spatio-temporal image sequence

$$\hat{\boldsymbol{f}} : \Omega \times [0, T] \to \mathcal{R}^{n_c} , \tag{1.3}$$

where the third dimension corresponds to time. Thus, for an arbitrary time instant $t \in [0, T - 1]$, we would have that

$$\boldsymbol{f}_1(x, y) = \hat{\boldsymbol{f}}(x, y, t) \quad \text{and} \quad \boldsymbol{f}_2(x, y) = \hat{\boldsymbol{f}}(x, y, t + 1) , \tag{1.4}$$

and the flow establishes a correspondence for a time span of length 1

$$(x, y, t)^\top \rightsquigarrow (x + u(x, y), y + v(x, y), t + 1)^\top . \tag{1.5}$$

For convenience, we will use the abbreviation $\boldsymbol{w} = (u, v, 1)^\top$ to denote the flow vector with one variable that includes the temporal offset.

Discrete vs. Continuous Images and Differentiability

In practice, all image acquisition devices convert the observed scene into discrete data. Thereby, the image sensor performs a spatial integration of the incident light over each pixel element. Neglecting the spatial extent of each pixel element, this can also be interpreted as a multiplication of the continuous incident light function with a regular Dirac comb. Throughout this thesis, we will however model almost all concepts with continuous functions. Such continuous counterparts of the recorded discrete samples can be reconstructed by convolving the discrete samples with a small Gaussian kernel. As a side effect, any this way reconstructed functions inherit the property of being infinitely many times differentiable from the Gaussian. Thus, without further mention, all continuous image functions treated in this thesis are assumed to be the result of this acquisition pipeline.

1.4.1 Flow Field Visualisation

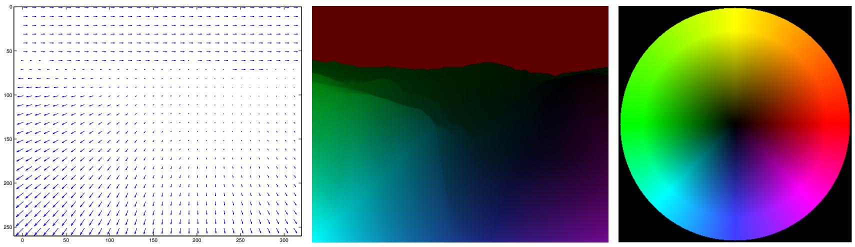


Figure 1.1: Visualisation schemes for optic flow fields. **Left:** Arrow plot of the ground truth flow field of the *Yosemite* [2] sequence. **Center:** Corresponding colour visualisation. **Right:** Colour circle. The colours associated to each direction can be seen from this image: For instance, a vector pointing in right direction ($(u, v)^\top = (1, 0)^\top$) is visualised with red colour, cf. the *sky* region in the exemplary flow field.

To visualise optic flow fields, colour visualisations are well suited. In contrast to the classical arrow plots, a colour visualisation allows to visualise

[2] The Yosemite image sequence was originally created by Lynn Quam at SRI.

the vector field densely and with high angular and radial resolution for the human observer. The direction of a flow vector is used to defines its colour, and its length determines the brightness. Thus, each displacement vector (u, v) is decomposed into polar coordinates, i.e. angle $\phi = \operatorname{atan2}(u, v)$ and length $r = \sqrt{u^2 + v^2}$. The visualisation colour is computed in HSV colour space, where the angle of the flow vector defines the hue component, the length is set to the value component and the saturation is kept fixed at the maximal value, see also Figure 1.1

1.4.2 Error Measures

To quantify the accuracy of an estimated flow field, its distance to the so-called *ground truth* flow field must be computed. Ground truth flow information is only available in special cases, e.g. if the image sequence is synthetic and has been rendered, or in special scenarios like for the data acquisition car of the KITTI Vision Benchmark suite [Geiger et al., 2012].

Average Angular Error

The classical error measure introduced by Barron et al. [1994] is the *average angular error* (AAE), which measures the average angular deviation of the estimated optic flow $\boldsymbol{w}_\mathrm{e} = (u_\mathrm{e}, v_\mathrm{e}, 1)^\top$ from the ground truth flow field $\boldsymbol{w}_\mathrm{g} = (u_\mathrm{g}, v_\mathrm{g}, 1)^\top$ in a spatio-temporal sense

$$\begin{aligned} \mathrm{AAE}(\boldsymbol{w}_\mathrm{e}, \boldsymbol{w}_\mathrm{g}) &= \frac{1}{\Omega} \int_\Omega \arccos\left(\frac{\boldsymbol{w}_\mathrm{e}^\top \boldsymbol{w}_\mathrm{g}}{|\boldsymbol{w}_\mathrm{e}||\boldsymbol{w}_\mathrm{g}|} \right) \mathrm{d}\boldsymbol{x} \\ &= \frac{1}{\Omega} \int_\Omega \arccos\left(\frac{u_\mathrm{e} u_\mathrm{g} + v_\mathrm{e} v_\mathrm{g} + 1}{(u_\mathrm{e}^2 + v_\mathrm{e}^2 + 1)^{1/2} \cdot (u_\mathrm{g}^2 + v_\mathrm{g}^2 + 1)^{1/2}} \right) \mathrm{d}\boldsymbol{x} \,. \end{aligned} \tag{1.6}$$

Average Endpoint Error

A slightly different error measure is the *average endpoint error* (AEE) which computes the average Euclidean difference of two flow fields

$$\begin{aligned} \mathrm{AEE}(\boldsymbol{u}_\mathrm{e}, \boldsymbol{u}_\mathrm{g}) &= \frac{1}{\Omega} \int_\Omega \sqrt{(u_\mathrm{e} - u_\mathrm{g})^2 + (v_\mathrm{e} - v_\mathrm{g})^2} \, \mathrm{d}\boldsymbol{x} \\ &= \frac{1}{\Omega} \int_\Omega ||\boldsymbol{u}_\mathrm{e} - \boldsymbol{u}_\mathrm{g}||_2 \, \mathrm{d}\boldsymbol{x} \,. \end{aligned} \tag{1.7}$$

Bad Pixel Measure

Recently, Geiger et al. [2012] released the KITTI Vision Benchmark Suite that offers a huge amount of image material from urban driving scenarios. The ground truth optic flow data of this benchmark is recorded using a Velodyne laser scanner device that is mounted besides the image sensors. To account for the small scale imprecision contained in the measurements of this laser scanner, Geiger et al. [2012] adopted the so-called *bad pixel* (BP) error measure from the stereo vision community [Scharstein and Szeliski, 2002]

$$\mathrm{BP}K(\boldsymbol{u}_\mathrm{e}, \boldsymbol{u}_\mathrm{gt}) = \frac{1}{\Omega} \int_\Omega \mathbb{1}_{(\|\boldsymbol{u}_\mathrm{e}(\boldsymbol{x}) - \boldsymbol{u}_\mathrm{gt}(\boldsymbol{x})\|_2 < K)} \, \mathrm{d}\boldsymbol{x} \quad K \in \mathbb{N}\,, \tag{1.8}$$

where $\mathbb{1}_{(\cdot)}$ is 1 if its argument is true, and 0 else. For instance, the BP3 error, which we will always consider, expresses the percentage of estimated flow vectors that differ by more than 3 pixels form the ground truth solution, i.e. the percentage of pixels with an Euclidean endpoint error above 3 pixels.

1.5 Thesis Organisation

This dissertation is organised in three main chapters.

Chapter 2 is devoted to invariances. After giving clarifying notations in Section 2.1, a broad overview of available invariant features is presented in Section 2.2. After that, Section 2.3 discusses appropriate metrics for the various features, and in Section 2.4 we develop a generic variational energy functional that allows to incorporate all previously discussed invariances. After that, we evaluate our findings experimentally in Section 2.5.

Chapter 3 discusses our change estimation and compensation strategy. We begin by presenting a variational model that estimates appearance changes in Section 3.2. In the following Section 3.3, we describe how to estimate a suitable basis from training data, and we evaluate our method in Section 3.4 with experiments.

In Chapter 4, we turn our attention to the scale space behavior of optic flow. First, in Section 4.1, we rewrite the energy functional of Horn and Schunck in terms of a signal regularisation like approach. After that, we generalise the functional by introducing two additional degrees of freedom in Section 4.2. After that we derive the evolution equations and discuss their implementation in Sections 4.3 and 4.4, respectively. Finally, we discuss how to select an optimal scale in Section 4.5 and analyse the scale space behaviour experimentally in Section 4.6.

Finally, in Chapter 5 we conclude this dissertation and briefly touch on possible future research directions.

Chapter 2

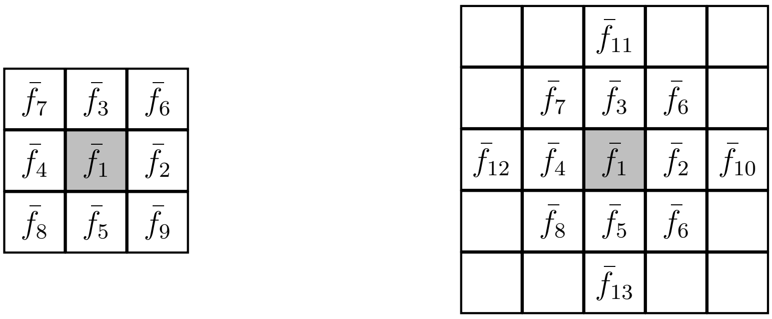
Invariance

This chapter is dedicated to invariance-based constancy assumptions. Thus, in the first section, we will start by giving a broad and structured overview over available features and the type of invariance they provide. After that, we will analyse in Section 2.3 what suitable distance metrics for the discussed descriptors are. Then, we will develop a general variational framework for optic flow in Section 2.4 which allows to incorporate each of these assumptions in a generic manner. Finally, we will present extensive experiments that analyse the properties and performance of the discussed concepts in practice. This chapter bases on the BMVC paper [Demetz et al., 2013] and the follow-up IJCV article [Demetz et al., 2015].

In the upcoming section, the goal will be to define a transformation for each possible feature. Such a transformation takes as input a smooth image $f \in C^\infty(\Omega, \mathcal{R})$, where $C^\infty(\Omega, \mathcal{R})$ denotes the set of infinitely often differentiable image functions that map from the 2-D spatial image domain Ω to image intensity values $\mathcal{R}$. The transformed feature $\mathcal{T}(f)$ is a generic m-dimensional feature vector field. For colour images $\boldsymbol{f} \in C^\infty(\Omega, \mathcal{R}^{n_c})$ and transformations that do not depend on all n_c colour channels, we apply the transformation to each channel and concatenate the transformation results:

$$\mathcal{T}(\boldsymbol{f}) := (\mathcal{T}(f_1)^\top, \ldots, \mathcal{T}(f_{n_c})^\top)^\top . \tag{2.1}$$

Regarding the differentiability of the output, some of the inherently discrete transforms that we will introduce later do not lead to differentiable transformation results. However, as mentioned in Section 1.4, in this case we choose the same countermeasure to fix this theoretical problem and perform a Gaussian post-smoothing after the transformation has been applied; see also Section 2.4.3.



$\bar{f}_7$	$\bar{f}_3$	$\bar{f}_6$
$\bar{f}_4$	$\bar{f}_1$	$\bar{f}_2$
$\bar{f}_8$	$\bar{f}_5$	$\bar{f}_9$

		$\bar{f}_{11}$		
	$\bar{f}_7$	$\bar{f}_3$	$\bar{f}_6$	
$\bar{f}_{12}$	$\bar{f}_4$	$\bar{f}_1$	$\bar{f}_2$	$\bar{f}_{10}$
	$\bar{f}_8$	$\bar{f}_5$	$\bar{f}_6$	
		$\bar{f}_{13}$		

$$\bar{\boldsymbol{f}} := \mathcal{N}^9(f) = (\bar{f}_1, \dots, \bar{f}_9)^\top \qquad \bar{\boldsymbol{f}} := \mathcal{N}^{13}(f) = (\bar{f}_1, \dots, \bar{f}_{13})^\top$$

Figure 2.1: Illustration of the neighbourhood operator for two different neighbourhood sizes. **Left:** $\mathcal{N}^9$ on a 3×3 patch. The grey-shaded cell denotes the centre reference pixel. The four direct neighbours are closer than the diagonal neighbours and thus have lower indices. **Right:** Example for $\mathcal{N}^{13}$.

2.1 Point- vs. Patch-based Descriptors

Many of the common standard features that we are now going to discuss are *point-based*, i.e. the features are computed for each point only by exploiting image information in that point. In the discrete case, a small fixed set of neighbouring pixels around a point is sometimes necessary to approximate the derivatives, however we still consider such features as point-based.

If we give up this property and consider a neighbourhood around each point, we will see that a variety of other features with special properties and invariances becomes available. Of course, the size of this neighbourhood is a free parameter for which we will also have to choose a suitable value.

Formally, with the notion *image patch* we mean a discrete and finite set of neighbouring pixels. To this end we have to introduce a neighbourhood operator $\mathcal{N}^k$ that formally converts the scalar image function $f \in C^\infty(\Omega, \mathcal{R})$ into a k-dimensional vector field. This is done by sampling the signal on a regular grid and stacking the neighbours of each pixel into a vector. Another intuition how the k-dimensional signal is built is that each of the k components contains a copy of original signal, however, each component is spatially shifted by a shift corresponding to the offset to the neighbour.

At this point, the fully discrete nature of patch-based concepts enters the game, because this local neighbourhood is sampled on a regular grid. A continuous counterpart for some of the upcoming concepts is not available in all cases. However, there are exceptions for specifically shaped neighbourhoods, see e.g. the study of Hafner et al. [2013] about the continuous implications of the circular census transform. Thus, we define the local neighbourhood to contain the k closest grid positions with respect to Euclidean distance, i.e. the neighbourhood has an approximately circular shape. As a convention, we order the values in the vector by increasing spatial Euclidean distance from the centre reference position, c.f. Figure 2.1. To ease notation, we will abbreviate the neighbourhood operator by over-lining and bold-face font to indicate that the result is vector-valued:

$$\bar{\boldsymbol{f}} := \mathcal{N}^k(f)\,. \tag{2.2}$$

The neighbourhood operator includes the centre pixel, thus, the first component always coincides with the input image:

$$\bar{f}_1 = f\,. \tag{2.3}$$

2.2 Invariant Descriptors

Let us now give an overview over features that exhibit invariances, structured by their class of invariance.

The most basic of all features is the input grey value or colour image sequence itself. Formally one could say that these raw intensity values exhibit the weakest form of invariance, namely *no invariance* since any rescaling that is not the identity changes the function.

2.2.1 Additive Rescalings

The second-weakest class of invariance comprises all features that stay unchanged under additive rescalings. Formally, this invariance is fulfilled if it holds that

$$\mathcal{T}(f+a) = \mathcal{T}(f) \quad \forall\, a \in \mathcal{R}, f \in C^\infty(\Omega, \mathcal{R})\,. \tag{2.4}$$

Here, a denotes a globally constant intensity offset.

It is obvious that if the additive term from Equation (2.4) is constant in space, then it vanishes by differentiation. A large variety of feature transformations based on local derivative information is available. A good overview on this class is given by Papenberg et al. [2006] and is summarised here for the sake of completeness.

Image Gradient

The image gradient has a long tradition as constancy assumption in variational optic flow [Nagel, 1983b; Tretiak and Pastor, 1984; Uras et al., 1988; Schnörr, 1993]. It leads to a two-dimensional feature per colour channel and represents a vector that points into the direction of steepest ascent in the current colour channel. Its length can be interpreted as a measure of the local image contrast. Formally, the sought transform is simply defined as the nabla operator

$$\mathcal{T}_{\mathrm{grad}}(\cdot) := \boldsymbol{\nabla}(\cdot) = \begin{pmatrix} \partial_x(\cdot) \\ \partial_y(\cdot) \end{pmatrix} , \tag{2.5}$$

which is a linear operator. One drawback of the gradient is its missing rotation invariance, i.e. if a structure undergoes rotational motion, the direction of the gradient changes.

Gradient Magnitude

The magnitude of the gradient, however, does possess the property of rotational invariance:

$$\mathcal{T}_{\mathrm{gradmag}}(\cdot) := |\boldsymbol{\nabla}(\cdot)| = \sqrt{(\partial_x \cdot)^2 + (\partial_y \cdot)^2} \, , \tag{2.6}$$

because the directional component of the gradient is discarded, and only the local contrast, which is rotationally invariant, counts.

Gradient Orientation

For the sake of completeness, we note that the orientation of the gradient plays an important role in the context of image registration [Modersitzki, 2009]. Such constancy assumptions are often realised via the scalar product of the normalised image gradients at corresponding locations. However, this scalar product does not fit into our framework and cannot be represented with our type of transformation operator [Haber and Modersitzki, 2007].

Hessian

Of course, also higher-order derivatives can serve as feature to match. The Hessian

$$\boldsymbol{\mathcal{H}}(\cdot) := \begin{pmatrix} \partial_{xx}(\cdot) & \partial_{xy}(\cdot) \\ \partial_{xy}(\cdot) & \partial_{yy}(\cdot) \end{pmatrix} , \tag{2.7}$$

summarises all second-order derivatives and offers several possibilities for deriving constancy assumptions.

Generally, all derivative based features are members of the same invariance class, they are only invariant under additive rescalings. However, for second-order features, the additive rescaling might be spatially affine:

$$\mathcal{T}(f(\boldsymbol{x}) + a(\boldsymbol{x})) = \mathcal{T}(f) \quad \forall\, a : \Omega \to \mathcal{R} \text{ with } \boldsymbol{\mathcal{H}}(a) = \mathbf{0}\,, \tag{2.8}$$

and where $f \in C^\infty(\Omega, \mathcal{R})$ denotes the image function. In contrast, for first-order features, the additive rescaling must be constant.

A basic assumption would be about the entries of the Hessian, leading to the 3-dimensional feature transform

$$\mathcal{T}_{\text{hess}}(\cdot) := \begin{pmatrix} \partial_{xx}(\cdot) \\ \partial_{xy}(\cdot) \\ \partial_{yy}(\cdot) \end{pmatrix}\,, \tag{2.9}$$

where the second component holding the mixed derivatives might be double-weighted since this entry appears twice in the full Hessian (2.7).

As for the raw gradient feature (2.5), the latter feature is not rotationally invariant. However, for instance the trace of the Hessian, the so-called Laplace operator, fulfills this requirement

$$\mathcal{T}_{\text{laplace}}(\cdot) := \Delta(\cdot) = \partial_{xx}(\cdot) + \partial_{yy}(\cdot)\,, \tag{2.10}$$

and leads to a one-dimensional feature per colour channel.

Structure-Texture Decomposition

Under the assumption that illumination changes vary smoothly in space, Wedel et al. [2008] have used the *total variation* (TV) denoising model of Rudin et al. [1992] to separate the input images into a so-called *structure* and a *texture* part. The texture part is the difference of original input image and its smoothed version (also called *method noise* in other contexts). Wedel et al. [2008] argue that smooth illumination changes will only affect the structure part and the texture part remains unchanged. Hence, the feature to impose a constancy assumption on is the texture part. This is a reasonable assumption for smooth additive illumination changes, however, multiplicative rescalings do not fit into this model. In practice, Wedel et al. [2008] perform a blending of structure and texture part

$$s(f) = \operatorname*{argmin}_u \Big((u - f)^2 + \lambda_{\text{ROF}} |\boldsymbol{\nabla} u| \Big)\,, \tag{2.11}$$

$$t(f) = f - \alpha_{\text{blend}} \cdot s(f)\,, \tag{2.12}$$

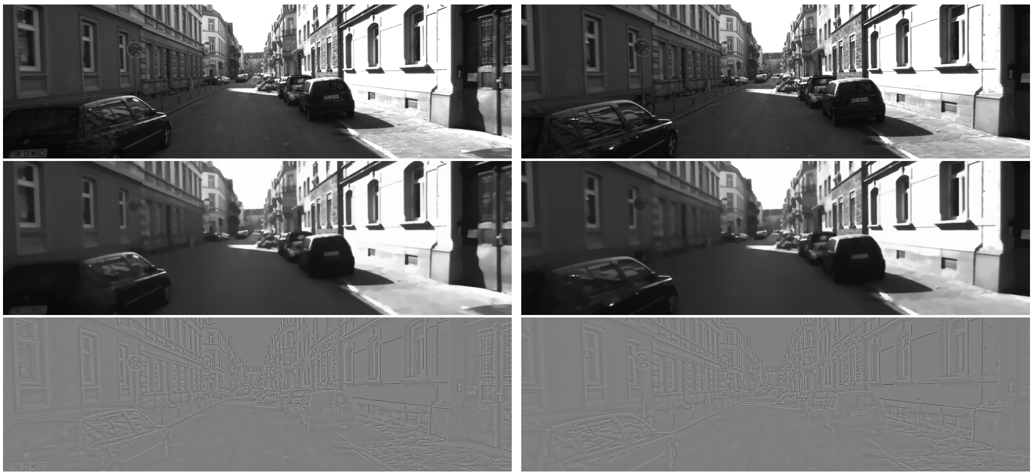


Figure 2.2: Structure-texture decomposition. **Left column:** Frame 1 of KITTI training sequence #15. **Right column:** Corresponding frame 2. **Top row:** Input images. **Middle row:** Structure part. **Bottom row:** Texture part.

where the smoothness parameter $\lambda_{\text{ROF}} > 0$ has to be chosen according to Aujol et al. [2006], and the optimal value for the blending parameter is experimentally found as $\alpha_{\text{blend}} = 0.95$ [Wedel et al., 2008]. The texture part is then considered as feature

$$\mathcal{T}_{\text{texture}}(\cdot) := t(\cdot)\,. \tag{2.13}$$

For an exemplary decomposition, see Figure 2.2. This feature however should not be applied to colour data in the straightforward channel-by-channel manner, because the original ROF model is not well suited for multivariate inputs. An appropriate smoothing process should align the edges of all channels, as proposed and applied for instance by Gerig et al. [1992]; Blomgren and Chan [1998]; Chambolle [1994], and Kramarev et al. [2013].

Centralised Differences

The first patch-based concept we consider in this thesis are centralised differences that Vogel et al. [2013] introduce as a sum of absolute differences-based approximation of the census transform, c.f. Section 2.2.4. Its invariance versus additive changes originates from considering the difference between each pixel value of the patch and the centre value. For a patch size k, the resulting feature is $k-1$-dimensional and reads

$$\mathcal{T}_{\mathrm{centr-diff}}(\bar{\boldsymbol{f}}) := \begin{pmatrix} \bar{f}_2 - \bar{f}_1 \\ \vdots \\ \bar{f}_k - \bar{f}_1 \end{pmatrix} , \tag{2.14}$$

where, as defined before, $\bar{f}_1$ corresponds to the grey value of the central pixel of the patch $\bar{\boldsymbol{f}}$.

2.2.2 Multiplicative Rescalings

Let us now extend the invariance class to include multiplicative rescalings, i.e. we consider transformations that remain unchanged under the following type of changes:

$$\mathcal{T}(a \cdot f) = \mathcal{T}(f) \quad \forall\, a \in \mathcal{R}, f \in C^\infty(\Omega, \mathcal{R}) , \tag{2.15}$$

where $a > 0$ is a globally constant, positive number. A more extensive study on the features that we will discuss in this subsection is presented by Mileva et al. [2007].

Derivatives of Logarithms

For multiplicative changes, we can exploit the general logarithm rule

$$\log(a \cdot b) = \log(a) + \log(b) , \tag{2.16}$$

and that it holds for space variant f and constant a that

$$\partial_x(\log(a \cdot f)) = \partial_x(\log a) + \partial_x(\log f) . \tag{2.17}$$

The first summand, i.e. the former constant factor, vanishes. Hence, the according two-dimensional feature reads

$$\mathcal{T}_{\mathrm{logderiv}}(\cdot) := \begin{pmatrix} \partial_x(\log \cdot) \\ \partial_y(\log \cdot) \end{pmatrix} . \tag{2.18}$$

Analogously to the gradient, this feature is not rotationally invariant.

Normalisation

Let us consider the special case of being given vector-valued (colour) images, i.e. $\boldsymbol{f} \in C^\infty(\Omega, \mathcal{R}^{n_c})$, and let the special assumption hold that the multiplicative factor $a \in \mathcal{R}$ is the same for all channels but can vary spatially:

$$f_{\text{changed},i} = a \cdot f_{\text{orig},i} \qquad \text{for all channels } i\,. \tag{2.19}$$

For this scenario, Golland and Bruckstein [1997] propose a strategy that eliminates the multiplicative change by means of normalisation. After computing either the arithmetic or geometric mean of each colour vector

$$\tilde{f}_{\text{a}} = \frac{1}{n_c}(f_1 + \ldots + f_{n_c})\,, \quad \text{or} \quad \tilde{f}_{\text{g}} = \sqrt[n_c]{f_1 \cdot \ldots \cdot f_{n_c}}\,, \tag{2.20}$$

each channel is normalised by dividing though the computed mean. Thus, we obtain one normalised feature for each input channel which finally amounts to

$$\mathcal{T}_{\text{norm}-\text{a}}(\boldsymbol{f}) := \frac{1}{\tilde{f}_{\text{a}}}\boldsymbol{f}\,, \quad \text{and} \quad \mathcal{T}_{\text{norm}-\text{g}}(\boldsymbol{f}) := \frac{1}{\tilde{f}_{\text{g}}}\boldsymbol{f}\,. \tag{2.21}$$

2.2.3 Affine Rescalings

In this section we will discuss features that are invariant under affine rescalings, i.e. transformations for which holds

$$\mathcal{T}(a \cdot f + b) = \mathcal{T}(f) \quad \forall\, a, b \in \mathcal{R}, f \in C^\infty(\Omega, \mathcal{R})\,, \tag{2.22}$$

where $a > 0$ and $b \in \mathcal{R}$ are constants.

Hue Saturation Value (HSV) Colour Space

In this colour space, a colour is characterised in a cylindrical coordinate representation, rather than in a Cartesian space like the standard RGB colour space. Instead of red, green and blue value, the hue angle H, the saturation S and the value V specify a colour. The hue component $H \in [0, 2\pi]$ defines an angle on the colour wheel which selects the pure colour. The saturation $S \in [0, 1]$ represents the pureness of the colour, i.e. a pure red has saturation equal to one. If the saturation of a colour is zero, then only shades of grey can be represented. Finally, the value $V \in [0, 1]$ defines the brightness of a colour. Figure 2.2.3 illustrates this concept. For the sake of completeness, we

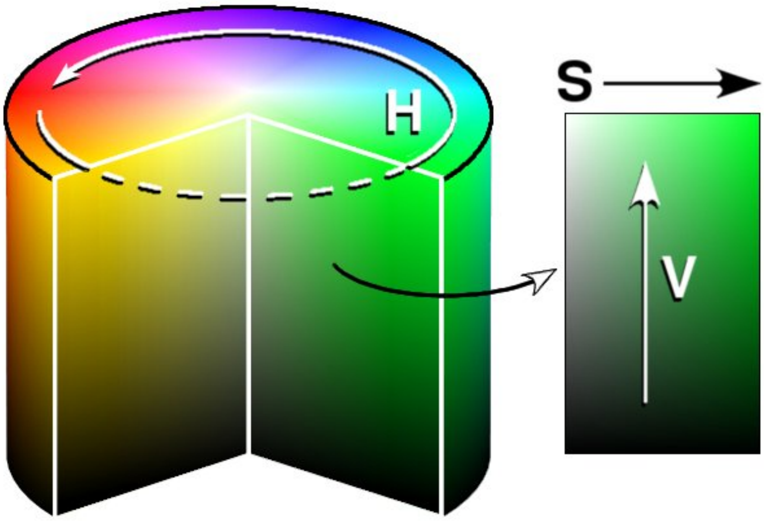


Figure 2.3: HSV colour space.[1]

depict the general transformation rules from RGB to HSV space, c.f. Mileva et al. [2007],

$$\begin{aligned} H\,(r,g,b) &= \begin{cases} \frac{g-b}{M-m} \quad\quad\quad \cdot {}^{\pi}\!/_{3} & \text{if } R = M\,, \\ (\frac{b-r}{M-m} + 2) \cdot {}^{\pi}\!/_{3} & \text{if } G = M\,, \\ (\frac{r-g}{M-m} + 4) \cdot {}^{\pi}\!/_{3} & \text{if } B = M\,, \end{cases} \\ S\,(r,g,b) &= \frac{M-m}{M}\,, \\ V\,(r,g,b) &= M\,, \end{aligned} \tag{2.23}$$

where $m = \min(r,g,b)$ and $M = \max(r,g,b)$ are abbreviations.

To gain an invariance in the HSV colour space, we need the assumption that both, the multiplicative and the additive change are the same for all colour channels, i.e.

$$f_{\text{changed},i} = a \cdot f_{\text{orig},i} + b \qquad \text{for all channels } i\,. \tag{2.24}$$

In this setting, the hue component H is invariant against such affine rescalings [Geusebroek et al., 2001]. This invariance is expolited by van de Weijer

[1] Image source: Wikipedia `https://de.wikipedia.org/wiki/HSV-Farbraum`

and Gevers [2004] by embedding a constancy assumption on the hue component into a local optic flow approach. Mileva et al. [2007] proposes a variational optic flow method with a constancy assumption on the hue component. In the work of Zimmer et al. [2011b], the ambiguity in the hue channel is solved by expressing the angle in terms of its sine and cosine. That way, the angles 0 and 2π, which correspond to the same hue colour angle, are expressed by the same vector. The feature finally reads

$$\mathcal{T}_{\text{hue}}(\boldsymbol{f}) := \begin{pmatrix} \cos H(\boldsymbol{f}) \\ \sin H(\boldsymbol{f}) \end{pmatrix} . \tag{2.25}$$

Correlation Transform

The *correlation transform* of Drulea and Nedevschi [2013] uses image statistics of the local neighbourhood $\mathcal{N}$ such as the mean

$$\mu_f(\boldsymbol{x}) = \frac{1}{|\mathcal{N}|} \int_{\mathcal{N}} f(\boldsymbol{x}+\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \, , \tag{2.26}$$

and standard deviation

$$\sigma_f(\boldsymbol{x}) = \sqrt{\frac{1}{|\mathcal{N}|} \int_{\mathcal{N}} (f(\boldsymbol{x}+\boldsymbol{y}) - \mu_f(\boldsymbol{x}))^2 \, \mathrm{d}\boldsymbol{y}} \, , \tag{2.27}$$

and assumes the scalar feature

$$\mathcal{T}_{\text{correlation}}(f) := \frac{f - \mu_f}{\sigma_f} \, , \tag{2.28}$$

to remain unchanged over time. This is related – but not equivalent – to assuming directly in a data term that two patches correspond if their normalised cross correlation is large [Hermosillo et al., 2002; Werlberger et al., 2010; Molnár et al., 2010]. However, such a direct normalised cross correlation term does not fit into our transformation framework. Moreover, the linearisation and minimisation of this assumption also differs in important points from the scheme that we will discuss in the next section.

Keypoint Descriptors for Large Neighbourhoods

There is a large variety of features available in the literature that computes very discriminative high-dimensional feature vectors on larger neighbourhoods. Among those, the *Scale Invariant Feature Transform* (SIFT) [Lowe, 2004], *Speeded Up Robust Features* (SURF) [Bay et al., 2008], or *Histogram* *of Gradients* (HOG) [Dalal and Triggs, 2005], are most popular. There exist many more such features and an extensive review is out of the scope of this dissertation. For a comparative study we refer to Mikolajczyk and Schmid [2005]. Almost all of these descriptors are invariant w.r.t. affine rescalings, since they mostly operate on grey value differences and include some kind of normalisation step. However, due to their large neighbourhood window, such descriptors are mainly suitable for sparse feature matching methods and are less appropriate for our dense correspondence estimation setting. Moreover, these descriptors are typically very high dimensional, for instance the SIFT descriptor is 128-dimensional. This renders their use in a data term of a variational energy functional difficult. One remarkable direct integration of SIFT descriptors in an optic flow method is *SIFT-Flow* [Liu et al., 2011], where the authors concentrate on matching images across different scenes (similar to general registration tasks, c.f. Subsection 2.2.5), and employ discrete optimisation strategies to compute a locally optimal solution. Moreover, an integration of HOG features into a variational model is presented by Rashwan et al. [2013]. An alternative way to exploit the remarkably good sparse matching quality of descriptors from this section in variational optic flow methods is to follow a two step strategy: First, estimate sparse matches with one of the mentioned key point descriptors, then integrate them as soft constraint into a variational model. Such ideas have been proposed e.g. by Brox and Malik [2011], Stoll et al. [2013], and Braux-Zin et al. [2013].

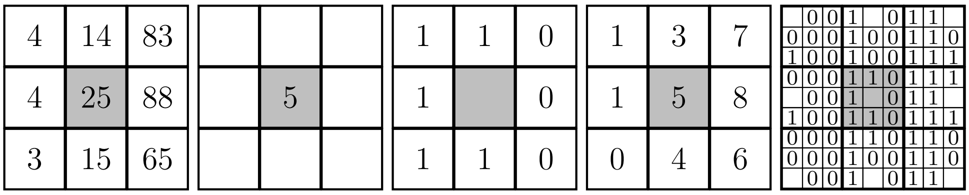
2.2.4 Monotonically Increasing Rescalings

Let us now discuss transforms that are invariant under any global monotonically increasing rescalings of the input signal. Transforms of this class are also called *morphologically invariant*, in the sense of Alvarez et al. [1993] and Chambolle [1994]. An illustration of the important transforms in this class is given in Figure 2.4. All morphologically invariant transforms have to fulfill formally:

$$\mathcal{T}(g(f)) = \mathcal{T}(f) \quad \forall g \in C^1(\mathcal{R},\mathcal{R}),\, g' > 0,\, f \in C^\infty(\Omega,\mathcal{R})\,. \tag{2.29}$$

For instance, if a video camera adjusts its gain to adapt to brightness changes in the captured scene, the effect on the recorded images can be approximated by an exponential rescaling (also called γ-correction), which is monotonically increasing and thus does not fit into any of the previous invariance classes.

Transforms of the current class of morphologically invariant descriptors are closely connected to the *order* of intensity values, because a monotonically increasing function can change anything – except the order of pixel intensities.



(a) Intensities. (b) Rank. (c) Census. (d) Complete rank. (e) Complete census.

Figure 2.4: Illustration of the presented intensity order transforms *(b)–(d)* with a 3×3 neighbourhood patch *(a)*, where the reference pixel is marked in grey.

Rank

The *rank transform* (RT) was proposed by Zabih and Woodfill [1994], and encodes for each pixel the position of its grey value in the ranking of all grey values in the neighbourhood. In other words, it is the number of neighbours with a smaller grey value than the central one. Formally, for a patch of size k the rank transform maps each pixel to its scalar rank signature $\mathcal{T}_{\text{rank}}\left(\bar{\boldsymbol{f}}\right) \in \{0, \dots, k-1\}$. It can be computed as

$$\mathcal{T}_{\text{rank}}\left(\bar{\boldsymbol{f}}\right) := \sum_{i=2}^{k} \mathbb{1}_{(\bar{f}_i < \bar{f}_1)}\,, \tag{2.30}$$

where $\mathbb{1}_{(x)}$ denotes the indicator function

$$\mathbb{1}_{(x)} := \begin{cases} 1 & \text{if } x \text{ is true,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.31}$$

In Figure 2.4, the image patch

$$\bar{\boldsymbol{f}}_{\text{example}} = (25, 88, 14, 4, 15, 83, 4, 3, 65)^\top\,, \tag{2.32}$$

is depicted, and its rank transform is

$$\mathcal{T}_{\text{rank}}(\bar{\boldsymbol{f}}_{\text{example}}) = 5\,, \tag{2.33}$$

since 5 of the 8 neighbouring intensities in the patch are smaller than the reference value.

Census

In the same paper, Zabih and Woodfill [1994] also introduced another descriptor, the so-called *census transform* (CT). It has attracted a lot of attention in recent years and can be seen as an extension of the rank transform: Besides encoding the rank, it adds a spatial component by expressing the relationship between the central pixel and each of its neighbours explicitly. Specifically, one bit of information is stored for each pixel of the neighbourhood: If the neighbour is smaller than the reference pixel the bit is 1, and 0 otherwise. In the final binary signature, all bits are concatenated. While the order of this concatenation is in general arbitrary, it has to be consistent such that each bit can be uniquely associated with one neighbour. Throughout this thesis, we will stick to our definition from Section 2.1. In mathematical terms, each image patch of size k is mapped to a binary signature $\mathcal{T}_{\text{census}}\left(\bar{\boldsymbol{f}}\right) \in \{0,1\}^{k-1}$ of length $k-1$. We choose the following formal representation to compute a census signature:

$$\mathcal{T}_{\text{census}}\left(\bar{\boldsymbol{f}}\right) := (\mathbb{1}_{(\bar{f}_2<\bar{f}_1)}, \ldots, \mathbb{1}_{(\bar{f}_k<\bar{f}_1)})^\top . \tag{2.34}$$

Hence, every neighbouring pixel is compared to the central one. For the exemplary patch form Figure 2.4, we have:

$$\mathcal{T}_{\text{census}}(\bar{\boldsymbol{f}}_{\text{example}}) = (0,1,1,1,0,1,1,0)^\top . \tag{2.35}$$

Furthermore, the sum of the digits of a census signature coincides with the rank of that pixel:

$$\mathcal{T}_{\text{rank}}\left(\bar{\boldsymbol{f}}\right) \;=\; \sum_{i=1}^{k-1} \left(\mathcal{T}_{\text{census}}\left(\bar{\boldsymbol{f}}\right)\right)_i . \tag{2.36}$$

There are several recent publications that incorporate the census transform in variational optical flow or stereo methods: Müller et al. [2011] propose a census-based data term for optical flow, and Ranftl et al. [2012] as well as Mei et al. [2011] present census-based stereo methods. Braux-Zin et al. [2013] combine census and grey value constancy assumption in a data term for optic flow and additionally integrate sparse feature matches. The theoretical study of Hafner et al. [2013] explains the reasons why census-based data terms for variational optic flow are successful.

Variants. The original census concept of Zabih and Woodfill [1994] has been extended in various respects.

Especially in homogeneous regions of an image, the decision between 0 and 1 for each bit of a census signature can be influenced severely even by very low amounts of noise. Stein [2004] proposes one possibility to address this problem by extending the concept to a ternary signature and introducing a parameter that represents the noise level. Thus, if one grey value is clearly smaller or larger than the other one, a 1 or -1 is stored, respectively. However, the case when two grey values are very similar has its own representation and is encoded by the signature value 0. The novel transform can formally be written as

$$\mathcal{T}_{\text{tern-census}}\left(\bar{\boldsymbol{f}}\right) := \left(\mathbb{1}^{\epsilon}_{(\bar{f}_2,\bar{f}_1)}, \ldots, \mathbb{1}^{\epsilon}_{(\bar{f}_k,\bar{f}_1)}\right)^{\top}, \tag{2.37}$$

where the ternary indicator $\mathbb{1}^{\epsilon}_{(\cdot)}$ is defined as

$$\mathbb{1}^{\epsilon}_{(x,y)} := \begin{cases} 1 & \text{if } x - y < -\epsilon\,, \\ -1 & \text{if } x - y > \epsilon\,, \\ 0 & \text{otherwise.} \end{cases} \tag{2.38}$$

Note that, strictly speaking, the ternary census transform is only invariant versus additive rescalings, since for any difference of two grey values, there exists a multiplicative factor that can scale this difference below or above the noise threshold ϵ.

Fröba and Ernst [2004] address the vulnerability of the census transform against noise in another way: The idea of their so-called *modified census transform* is to exchange the centre grey value by the mean grey value of the patch as comparison target. This concept formally reads

$$\mathcal{T}_{\text{mod-census}}\left(\bar{\boldsymbol{f}}\right) := \left(\ \mathbb{1}_{(\bar{f}_1<\bar{f}_{\text{mean}})}, \ldots, \mathbb{1}_{(\bar{f}_k<\bar{f}_{\text{mean}})}\right)^{\top}, \tag{2.39}$$

where $\bar{f}_{\text{mean}} = (\bar{f}_1 + \ldots + \bar{f}_k)/k$ represents the mean grey value of the patch. Note that the modified census signature also contains one digit for a comparison of the centre pixel value with the mean, thus leads to a k-dimensional feature. Moreover, by comparing against the mean, the morphological invariance of the modified transform is destroyed, and only an invariance under affine rescalings is left.

Mohamed and Mertsching [2012] combine the two latter concepts and use sparse matches from the *ternary modified* census transform as soft constraints in an otherwise standard variational optic flow approach of $TV - L_1$ type [Zach et al., 2007].

In the spirit of local binary patterns [Pietikäinen et al., 2011; Calonder et al., 2012], where larger and sparse stencils are common, Ranftl et al. [2014] recently proposed a scale-adaptive census transform. Depending on the scale of the underlying structure, the census signature is computed on a circular stencil of varying radii. The sampling of this circular stencil involves bi-linear interpolation. Thus, unfortunately also this census variant formally looses the morphological invariance.

Complete Rank

Although the two signatures by Zabih and Woodfill [1994] exhibit the same morphological invariance, the census transform obviously encodes by construction more information than the pure rank.

However, there is still some more information that can be used without losing the desired invariance. To this end, let us now introduce an extension of Zabih and Woodfill's basic transform: the *complete rank transform* (CRT). We will see that the resulting signature carries much more information than its predecessors.

Given the census signature of an image patch, we know which pixels in the patch are smaller than the *central* one. However, the relationships among *all* neighbours cannot be determined by the pure census information. For instance, if two neighbouring pixels are both smaller than the central one, it is still unclear which of the two neighbours is smallest.

To also encode this information, we propose the *complete rank transform.* We compute the rank of each pixel of the patch and store this information in a k-dimensional signature $\mathcal{T}_{\mathrm{comp-rank}}\left(\bar{\boldsymbol{f}}\right) \in \{0, \dots, k-1\}^k$:

$$\begin{aligned} \mathcal{T}_{\mathrm{comp-rank}}\left(\bar{\boldsymbol{f}}\right) &:= \left(s^1_{\mathrm{rank}}, \dots, s^k_{\mathrm{rank}}\right)^\top, \\ s^j_{\mathrm{rank}} := \mathcal{T}^j_{\mathrm{rank}}\left(\bar{\boldsymbol{f}}\right) &:= \sum_{\substack{i=1\\ i\neq j}}^{k} \mathbb{1}_{(f_i < f_j)} \,. \end{aligned} \tag{2.40}$$

With this CRT signature, the whole intensity order is represented. From the viewpoint of morphological invariance, this is the maximal amount of information that can be extracted without leaving this class of invariance. Regarding the exemplary patch from Figure 2.4, we have:

$$\mathcal{T}_{\mathrm{comp-rank}}(\bar{\boldsymbol{f}}_{\mathrm{example}}) = (5, 8, 3, 1, 4, 7, 1, 0, 6)^\top, \tag{2.41}$$

where the rank 1 is tied, which means two pixels in the patch have the same intensity, so the corresponding rank 1 occurs twice. This scenario is not uncommon in practice, since most images are stored after a quantisation to 256 intensity levels. By this, tied ranks occur frequently especially in homogeneous image regions. Of course, by Gaussian pre-smoothing, this phenomenon can be largely suppressed.

The computation rule for CRT signatures as shown in Equation 2.40 is demonstrative and intuitively understandable, but also inefficient (quadratic complexity in k). Essentially what has to be done is to sort the intensities, and afterwards note their position. Thus, we propose to use an efficient sorting algorithm such as *Quicksort* for this task (asymptotical complexity $\mathcal{O}(k \log k)$); see e.g. Press et al. [2007].

The first appearance of ordinal measures of full patches in the literature goes back to work on block matching based stereo correspondence of Bhat and Nayar [1998]. Related to that, also more recently, several sparse interest point descriptors building on intensity order-based ideas have been proposed: With their chained circular neighbourhoods, Chan et al. [2012] make a first step towards representing neighbourhood ordinal information. The *Local Intensity Order Pattern* (LIOP) descriptor of Wang et al. [2011] describes the intensity order of a very large neighbourhood and is specifically tailored for sparse interest point matching. A similar idea of matching order distributions is proposed by Tang et al. [2009]. Mittal and Ramesh [2006] combine order and intensity information to increase the robustness against Gaussian noise.

Complete Census

After motivating the complete rank transform via the missing information about ordering relationships between all pixels in the patch, another transform comes naturally into mind, namely an analogue extension of the census transform: the *complete census transform* (CCT).

Instead of storing all k ranks, it stores for each pixel of the patch its own census transform, i.e. it is smaller or larger than *any other* pixel in the patch. Thus, we obtain a signature $\mathcal{T}_{\mathrm{comp-census}}\left(\bar{\boldsymbol{f}}\right) \in \{0,1\}^{k\cdot(k-1)}$ which contains all census signatures with each of the pixels as reference:

$$\mathcal{T}_{\mathrm{comp-census}}\left(\bar{\boldsymbol{f}}\right) := \left({\boldsymbol{s}^{1}_{\mathrm{census}}}^{\top}, \ldots, {\boldsymbol{s}^{k}_{\mathrm{census}}}^{\top}\right)^{\top} , \tag{2.42}$$

with

$$\begin{aligned} \boldsymbol{s}^{j}_{\mathrm{census}} := \mathcal{T}^{j}_{\mathrm{census}}\left(\bar{\boldsymbol{f}}\right) := (&\mathbb{1}_{(f_1<f_j)}, \ldots, \mathbb{1}_{(f_{j-1}<f_j)}, \\ &\mathbb{1}_{(f_{j+1}<f_j)}, \ldots, \mathbb{1}_{(f_k<f_j)})^{\top} . \end{aligned} \tag{2.43}$$

Evidently, the original census signature from Equation 2.34 coincides with this novel definition for $j = 1$: $\mathcal{T}_{\text{census}}(\cdot) = \mathcal{T}^1_{\text{census}}(\cdot)$. The information contained in complete rank and complete census is equivalent. This can be seen from the bijection between them: Instead of computing the CCT signature from the original intensity patch, we could as well compute it from the CRT signature of the patch. The result would be the same:

$$\mathcal{T}_{\text{comp-census}}\Big(\bar{\boldsymbol{f}}\Big) = \mathcal{T}_{\text{comp-census}}\Big(\mathcal{T}_{\text{comp-rank}}\Big(\bar{\boldsymbol{f}}\Big)\Big)\,. \tag{2.44}$$

Vice versa, the complete rank digits are just the sums of corresponding CCT bits:

$$(\mathcal{T}_{\text{comp-rank}}\Big(\bar{\boldsymbol{f}}\Big))_j = \sum_{i=1}^{k-1} (\boldsymbol{s}^j_{\text{census}})_i\,. \tag{2.45}$$

Both signatures, complete rank and complete census, can also represent tied ranks, i.e. if pixels in the patch have the same intensity. In this case the same rank occurs multiple times in a CRT signature, and corresponding CCT digits are both 0. Thus, the number of possible signatures for a patch with k pixels is the k-th *ordered Bell number* $\mathrm{OBN}(k)$ [Sloane and Plouffe, 1995] (also called k-th Fubini number), which is defined by

$$\mathrm{OBN}(k) \;=\; \sum_{i=0}^{k}\sum_{j=0}^{i}(-1)^{i-j}\binom{i}{j}j^k\,. \tag{2.46}$$

It expresses the maximally possible number of weak orderings of a set of k elements. The fact that tied ranks can be represented increases the number of possibilities drastically, see Table 2.1.

Vulnerabilities

When dealing with morphologically invariant transforms, one has to be very careful not to accidentally destroy this invariance by performing operations on the signals before the transformation: For instance any averaging, interpolation or smoothing of the intensity values before performing the transformation can cause a change of the intensity order. Let us illustrate this with an example: We consider an exemplary patch of three intensities

$$\boldsymbol{p}_1 = (3,1,4)^\top\,, \quad \mathcal{T}_{\text{comp-rank}}(\boldsymbol{p}_1) = (1,0,2)^\top\,, \tag{2.47}$$

and as non-affine, monotonically increasing rescaling we consider the cubic function $g(f) = f^3$. Morphological invariance means that the order of intensities of the original patch and the rescaled one remains unaltered:

$$\boldsymbol{p}_2 = (27,1,64)\,, \quad \mathcal{T}_{\text{comp-rank}}(\boldsymbol{p}_2) = (1,0,2)^\top\,, \tag{2.48}$$

Table 2.1: Ordered Bell Numbers in comparison to factorial k, the number of possible rankings without tied ranks.

k	OBN(k)	k!
0	1	1
1	1	1
2	3	2
3	13	6
4	75	24
5	541	120
6	4683	720
7	47293	5040
8	545835	40320
9	7087261	362880
10	102.247.563	3.628.800

which is obviously the case. If we now compute the arithmetic mean of neighbouring pixels on the original patch we obtain:

$$\boldsymbol{p}_3 = (3,\, \underbrace{2}_{=\frac{3+1}{2}}, 1,\, \underbrace{2.5}_{=\frac{1+4}{2}}, 5)\,, \quad \mathcal{T}_{\text{comp}-\text{rank}}(\boldsymbol{p}_3) = (3, 1, 0, 2, 4)^\top\,, \tag{2.49}$$

and if we first apply the cubic rescaling and average the result, we obtain:

$$\boldsymbol{p}_4 = (27,\, 14,\, 1,\, 32.5,\, 64)\,, \quad \mathcal{T}_{\text{comp}-\text{rank}}(\boldsymbol{p}_4) = (2, 1, 0, 3, 4)^\top\,. \tag{2.50}$$

In the non-transformed patch, the right averaged value of 2.5 was smaller than the first entry 3. This relation is not true anymore after the rescaling: now, 32.5 is larger than 27.

This example shows that the invariance of a transformation is a very fragile property that can easily be affected by pre-processing steps or the like.

2.2.5 Monotonically Decreasing Rescalings

The class of monotonically decreasing rescalings is not in the focus of this thesis. For instance, a grey value inversion would belong to this class of rescalings, i.e. if bright structures in the first frame correspond to dark parts in the second image, and vice-versa. Such phenomena typically do not occur in real image sequences. However, this class of features is in fact relevant in general image registration contexts, such as medical image registration [Modersitzki, 2009].

2.2.6 Discussion

We have revisited a big variety of descriptors in this section, and we have ordered them systematically by their class of invariance. For the sake of clarity, we summarise all discussed transforms and compare their essential properties in Table 2.2.

Table 2.2: Summary of the proposed intensity order transforms. The number of pixels in the considered neighbourhood is given by k.

transform	range $\mathcal{D}$ of one digit	signature length m	spatial information	rotation invariance	invariance class
$\mathcal{T}_{\text{intensity}}$	$\mathbb{R}$	1	–	–	none
$\mathcal{T}_{\text{grad}}$	$\mathbb{R}$	2	✓	–	additive
$\mathcal{T}_{\text{gradmag}}$	$\mathbb{R}$	1	✓	✓	additive
$\mathcal{T}_{\text{hess}}$	$\mathbb{R}$	3	✓	✓	additive
$\mathcal{T}_{\text{laplace}}$	$\mathbb{R}$	1	✓	✓	additive
$\mathcal{T}_{\text{texture}}$	$\mathbb{R}$	1	–	✓	(additive)
$\mathcal{T}_{\text{centr−diff}}$	$\mathbb{R}$	k	✓	–	additive
$\mathcal{T}_{\text{logderiv}}$	$\mathbb{R}$	1	–	✓	multiplicative
$\mathcal{T}_{\text{norm−g}}$	$[0,1]$	3	–	✓	multiplicative
$\mathcal{T}_{\text{norm−a}}$	$[0,1]$	3	–	✓	multiplicative
$\mathcal{T}_{\text{hue}}$	$[-1,1]$	2	–	✓	affine
$\mathcal{T}_{\text{correlation}}$	$\mathbb{R}$	1	–	✓	affine
$\mathcal{T}_{\text{rank}}$	$\{0,\ldots,k–1\}$	1	–	(–)	morphological
$\mathcal{T}_{\text{census}}$	$\{0,1\}$	$k-1$	✓	(–)	morphological
$\mathcal{T}_{\text{tern−census}}$	$\{-1,0,1\}$	$k-1$	✓	(–)	additive
$\mathcal{T}_{\text{mod−census}}$	$\{0,1\}$	k	✓	(–)	affine
$\mathcal{T}_{\text{comp−rank}}$	$\{0,\ldots,k–1\}$	k	✓	(–)	morphological
$\mathcal{T}_{\text{comp−census}}$	$\{0,1\}$	$k(k-1)$	✓	(–)	morphological

However, still the question *Which transform is best?* has not yet been answered. Unfortunately, a universal answer to this question *cannot be given*, because the big trade-off between invariance and accuracy persists: On the one hand we would like to be as invariant as possible, because more invariance means more robustness. On the other hand, with every invariance we discard information, i.e. we disregard the property we are invariant against. So, from the analytical point of view, the only valid answer to the question is that the *best* feature is situation-dependent and that no universally best feature can be determined. It is up to our experiments to find appropriate solutions for exemplary scenarios, and to identify features that perform well in many situations.

However, among all features exhibiting one particular invariance, in fact there do exist differences in terms of the information that is contained in the different signatures, or – the other way round – in terms of the information that is discarded. For instance, we have seen that, in each pixel, our complete rank and census signatures contain the full local image intensity order. Obviously this is much more information than the rank or census signatures carry. In particular, it is not even possible to encode any more local image information without leaving the class of morphologically invariant descriptors. The reason for this is as follows: All monotonic functions have in common that they cannot alter the order of inputs. This means if one intensity is smaller than another, this relation will still be valid after any monotonic function has been applied. Any further information that goes beyond the order of intensities, for instance differences, sums or quotients of intensities, can thus be altered by monotonic functions. As consequence, since our complete signatures carry the ordering relations between every possible pair of intensities, it is not possible to encode more morphologically invariant information. The same maximal amount of image information is also contained in the complete census transform. However, the reason to prefer our proposed complete rank signature is its much more compact representation and lower dimensionality, compared to the complete census signature.

Nevertheless, this alternative census-inspired perspective offers an unexpected insight: Hafner et al. [2013] points out that each binary digit of a census signature can be regarded as the sign of the corresponding directional derivative (in a finite difference sense). Thus, from this point of view, one can conclude that the complete rank transform inherently contains rich local differential information. In this regard, dealing with derivatives of such signatures as in [Puxbaum and Ambrosch, 2010] actually corresponds to second order image derivative information. This fact is not obvious from just considering the rank representation and should be kept in mind.

Let us further note that our notion of *image information* does not coincide with the classical notion of information content in terms of coding length. The work by Soatto [2009] goes more in our direction, as his notion of *actionable information* only includes the information content of image (sequences) that is relevant for the task at hand. In particular, this means that so-called *nuissances* – illumination and viewpoint changes – are discarded and do not count as actionable information. Interestingly, Soatto [2009] models illumination changes with monotonically increasing continuous functions. In that sense, our morphologically invariant descriptors do only discount nuisances and no actionable information.

2.3 Signature Distance Metrics

Besides the question which signature to chose, an equally important decision to take is the metric in which to compare the chosen signatures.

For the real- and vector-valued transforms, such as $\mathcal{T}_{\text{grad}}$, $\mathcal{T}_{\text{grad}-\text{mag}}$, etc., considering an L_p-norm of the signature difference most often is already a suitable solution. However, for the ordering-based descriptors, an appropriate metric is not immediately obvious.

For the classical rank and census transform there are suitable solutions available: For ranks, the absolute value of their differences is an appropriate metric because smaller rank difference means higher patch similarity. Let, analogously to Section 2.2.4, $\bar{\boldsymbol{f}}$ and $\bar{\boldsymbol{g}}$ denote two patches to compare. Then, the corresponding metric for the pure rank reads

$$d(\mathcal{T}_{\text{rank}}(\bar{\boldsymbol{f}}), \mathcal{T}_{\text{rank}}(\bar{\boldsymbol{g}})) = |\mathcal{T}_{\text{rank}}(\bar{\boldsymbol{f}}) - \mathcal{T}_{\text{rank}}(\bar{\boldsymbol{g}})| \,. \tag{2.51}$$

In case of census signatures, the Hamming distance is a natural choice since it reflects the number of pixel comparisons that are in agreement:

$$\begin{aligned} d(\mathcal{T}_{\text{census}}(\bar{\boldsymbol{f}}), \mathcal{T}_{\text{census}}(\bar{\boldsymbol{g}})) &= \sum_{i=1}^{k-1} \mathbb{1}_{((\mathcal{T}_{\text{census}}(\bar{\boldsymbol{f}}))_i \neq (\mathcal{T}_{\text{census}}(\bar{\boldsymbol{g}}))_i)} \,, \\ &= |\mathcal{T}_{\text{census}}(\bar{\boldsymbol{f}}) - \mathcal{T}_{\text{census}}(\bar{\boldsymbol{g}})|_0 \,. \end{aligned} \tag{2.52}$$

In the latter equation, we make use of the l_0 norm, i.e. the total number of non-zero elements of its argument vector. Vogel et al. [2013] propose the so-called *Centralised Sum of Absolute Distances* (CSAD) as a convex approximation of the Hamming distance for Census signatures. However, this approximation looses many invariances, in fact even the invariance under multiplicative rescalings is lost. Thus, this is not an option for our framework.

The Hamming distance is a concept for binary signatures, thus for ternary census signatures an appropriate metric has to be found. However, in terms of comparison similarity, an L_p-norm of the difference vector of two ternary signatures makes sense:

$$\begin{aligned} & d(\mathcal{T}_{\mathrm{tern-census}}(\bar{\boldsymbol{f}}), \mathcal{T}_{\mathrm{tern-census}}(\bar{\boldsymbol{g}})) \\ & \quad = \|\mathcal{T}_{\mathrm{tern-census}}(\bar{\boldsymbol{f}}) - \mathcal{T}_{\mathrm{tern-census}}(\bar{\boldsymbol{g}})\|_p \,. \end{aligned} \tag{2.53}$$

This can be explained by regarding the possible values that each component of the difference vector can attain: If it is zero, then the ternary digits of the two signatures to compare coincide in the considered component. If it is 2 or -2, then the two signatures differ completely, one digit is 1, the other -1. Accordingly, the contribution to the norm is large. The interesting case is if the difference is 1 or -1. Then, one of the signatures must have a zero in the respective component, the other signature however must be one or minus one. One can thus argue that the two signatures are *less* different than in the second case and consequently, the contribution to the norm is smaller.

The straightforward generalisation of the absolute rank difference to its complete counterpart is the Euclidean norm of the difference vector ($p = 2$) or the sum of absolute component differences ($p = 1$):

$$\begin{aligned} & d(\mathcal{T}_{\mathrm{comp-rank}}(\bar{\boldsymbol{f}}), \mathcal{T}_{\mathrm{comp-rank}}(\bar{\boldsymbol{g}})) \\ & \quad = \left(\sum_{j=1}^{k} |(\mathcal{T}_{\mathrm{comp-rank}}(\bar{\boldsymbol{f}}))_j - (\mathcal{T}_{\mathrm{comp-rank}}(\bar{\boldsymbol{g}}))_j|^p \right)^{1/p} . \end{aligned} \tag{2.54}$$

However, we are actually interested in the number of pixel comparisons in the patch not being in agreement. In this regard, the desired dissimilarity measure can be obtained by applying the Hamming distance to the complete census signatures:

$$\begin{aligned} & d(\mathcal{T}_{\mathrm{comp-census}}(\bar{\boldsymbol{f}}), \mathcal{T}_{\mathrm{comp-census}}(\bar{\boldsymbol{g}})) \\ & \quad = \sum_{i=1}^{k(k-1)} \mathbb{1}_{((\mathcal{T}_{\mathrm{comp-census}}(\bar{\boldsymbol{f}}))_i = (\mathcal{T}_{\mathrm{comp-census}}(\bar{\boldsymbol{g}}))_i\ )} \,, \end{aligned} \tag{2.55}$$

because this metric measures exactly what we would like. Interestingly, the two latter metrics exhibit a close relationship that becomes clear by plugging Equation 2.45 into the former one:

$$\begin{aligned} & d(\mathcal{T}_{\mathrm{comp-rank}}(\bar{\boldsymbol{f}}), \mathcal{T}_{\mathrm{comp-rank}}(\bar{\boldsymbol{g}})) \\ & \quad = \left(\sum_{j=1}^{k} \left| \sum_{i=1}^{k-1} (\boldsymbol{s}^j_{\mathrm{census}}(\boldsymbol{f}))_i - \sum_{i=1}^{k-1} (\boldsymbol{s}^j_{\mathrm{census}}(\boldsymbol{g}))_i \right|^p \right)^{1/p} . \end{aligned} \tag{2.56}$$

Comparing Equation 2.55 and 2.56, one can see that instead of counting each individual disagreeing pixel comparison, the disagreements are accumulated for each of the k CRT components. This accumulation performs a *best case estimate*, i.e. as many census digits as possible are assumed to coincide. In other words, the best possible case for each CRT component is assumed.

In the case of the sum of absolute differences ($p = 1$), the metric exactly represents the lowest possible bound on the Hamming distance of CCT signatures. For the Euclidean distance ($p = 2$), the individual rank differences are amplified by the square function. Thus, more disagreeing pixel comparisons than the least possible are presumed.

To conclude, CCT in combination with the Hamming distance might be the most intuitive measure. However, the CCT introduces a large computational overhead compared to CRT; signature length $k(k-1)$ versus k. Nevertheless, we have seen that the CRT metrics approximate this CCT metric in a meaningful way. It is up to our experiments to show the difference of both signature-metric combinations in terms of accuracy.

2.4 Variational Optic Flow Model

Let us now develop a general variational framework for optic flow in which all the previously discussed transforms can be embedded.

The seminal work of Horn and Schunck [1981] constitutes the starting point of the whole research branch on variational methods for optic flow. They were the first to formulate the problem of determining the optic flow as that of finding the minimiser of an energy functional of type

$$E(\boldsymbol{u}) = \int_{\Omega} \Big(D(\boldsymbol{u}) + \alpha \cdot S(\boldsymbol{u}) \Big) \, \mathrm{d}\boldsymbol{x} \,, \tag{2.57}$$

where the terms D and S denote the so-called *data* term and *smoothness* term, respectively. In such an energy minimisation setting, both terms can be understood as penalising functions, i.e. the terms should attain small values for a good solution. Here, a *good solution* means a solution that fits well to the model. The relative weight of data and smoothness term can be influenced with the positive smoothness parameter α.

In the following sections, we will discuss generic choices for both terms of the latter functional.

2.4.1 Data Term

The purpose of the data term is to establish a relationship between the input images and the sought optic flow field. This relationship usually consists of the assumption that some quantity that can be deduced from the images – a so-called *feature* – remains constant between the two images. For instance, the constancy assumption behind the classical work of Horn and Schunck [1981] is that the intensity of corresponding pixels stays unchanged over time. Thus, here the mentioned feature would be the intensity, which does not exhibit any invariance c.f. Section 2.2. Formally, this assumption can be expressed with the equation

$$\begin{aligned} & f_2(x+u(x,y), y+v(x,y)) = f_1(x,y) && (2.58)\\ \Leftrightarrow\quad & f_2(x+u(x,y), y+v(x,y)) - f_1(x,y) = 0\,, && (2.59) \end{aligned}$$

where $f(x, y, t)$ is the intensity of a pixel in the first frame and $f(x + u(x, y), y+v(x, y), t+1)$ represents the grey value of the pixel at the displaced corresponding position in the second frame. For the sake of readability, we will omit from now on the spatial arguments (x, y) of u and v.

Any non-zero value of the difference in Equation (2.59) indicates a violation of the constancy assumption. The larger this violation is, the larger will be the absolute value of the difference. Thus, to incorporate this constancy assumption into an energy functional we square it and obtain the basic data term

$$D_f(\boldsymbol{u}) = (f_2(x+u, y+v) - f_1(x,y))^2\,. \tag{2.60}$$

Vector-Valued Features

Many of the previously presented features are vector-valued, c.f. the overview of the previous Section 2.2. As discussed in Section 2.3, in most of these cases, the L_2 norm is a suitable metric. Thus, let us assume that for $j \in \{1, 2\}$, the feature under consideration $\mathcal{T}(f_j) =: \boldsymbol{p}_j$ is m-dimensional. The task is then to set up a data term reflecting the assumption that the value of each component remains constant for corresponding points. Analogously to 2.59, we can then transform each single feature constancy assumption into a data term by squaring and adding all squared data terms together:

$$\begin{aligned} D_{\boldsymbol{p}}(\boldsymbol{u}) &= \frac{1}{m}\Big|\, \boldsymbol{p}_2(x+u, y+v) - \boldsymbol{p}_1(x,y)\,\Big|_2^2 \\ &= \frac{1}{m}\sum_{i=1}^{m}\Big(p_{2,i}(x+u, y+v) - p_{1,i}(x,y)\Big)^2\,, \end{aligned} \tag{2.61}$$

and where the division by the number of components m normalises the whole expression w.r.t. the number of features. Note that in the same way also different constancy assumptions can be imposed jointly. For instance the model of Brox et al. [2004], which assumes grey value and gradient constancy, can be represented via $\mathcal{T}_{\text{Brox}}(f) := (f, f_x, f_y)^\top$.

Linearisation

The representation of our constancy assumptions as difference of grey values has one big practical drawback: The unknowns u and v show up in the argument of the image sequence. As consequence, Equations (2.60) and (2.61) are not convex and far from linear. Thus, the direct minimisation of a functional with this data term is very difficult: Due to the non-convexity, there might be many local minimisers and even several distinct globally optimal solutions. Thus, if we would find a minimiser with some strategy, e.g. some kind of descent scheme, we might just have been trapped in a local minimum and we cannot to decide if it is the globally optimal solution. Non-convex optimisation is a very difficult problem in general and the way we approach this issue will be discussed in Section 2.4.3 later in this thesis.

Nevertheless, Horn and Schunck [1981] and Lucas and Kanade [1981] approximated the non-linear expression of the data term with a first-order Taylor expansion. This linear approximation leads to

$$\begin{aligned} & f_2(x+u, y+v) && - f_1(x,y) \\ \approx\; & f_2(x,y) + f_{2,x}(x,y)\, u + f_{2,y}(x,y)\, v && - f_1(x,y) \\ =\; & f_{2,x}(x,y)\, u + f_{2,y}(x,y)\, v + f_t(x,y)\,, \end{aligned} \tag{2.62}$$

where the abbreviation $f_t := f_2 - f_1$ can be seen as a finite temporal derivative and the notion $f_{i,*} := \partial_*(f_i)$ abbreviates a partial derivative of the i-th frame. Due to this linearisation, we obtain linear (and thus also convex) minimality conditions when minimising the functional later on. Such a linear approximation, however, only holds if the signal f is indeed sufficiently linear, and if the displacements u and v are small.

We note that each constancy assumption leads to one constraint, but depends on two unknowns u and v. In the linearised case (2.62), this underdeterminedness shows up in the fact that there exists one line in the (u,v)-space of which all points are solutions of the constancy assumption. This line is perpendicular to the so-called *normal flow*

$$\boldsymbol{u}_{\text{normal}} = \frac{-f_t \boldsymbol{\nabla}_2 f}{|\boldsymbol{\nabla}_2 f|^2}\,, \tag{2.63}$$

which is the shortest possible of all solutions on that line. A visualisation of this solution is depicted in Figure 2.5. In Chapter 4, we will analyse the role of the normal flow in further detail.

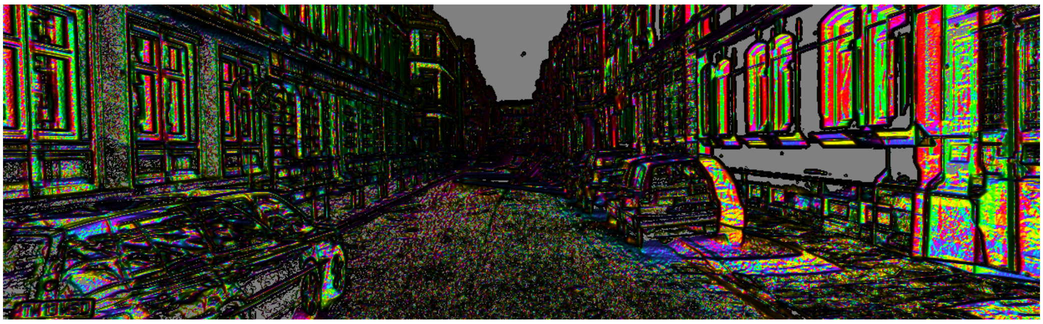


Figure 2.5: Visualisation of the normal flow of the linearised grey value constancy assumption. Colours encode direction, and regions where the image gradient is zero are depicted in grey. Image sequence: sequence #15 of the KITTI benchmark [Geiger et al., 2012].

Motion Tensor Notation

Especially for the case of multiple features, the motion tensor notation of Bruhn [2006] allows a very elegant formulation of the energy components. This tensor notation results in a data term consisting of one quadratic form – the so-called *motion tensor* – which is independent of the number of incorporated features. For the vector-valued feature $\boldsymbol{p}$, it can be deduced as follows:

$$\begin{aligned} D_{\boldsymbol{p},\mathrm{lin}}(\boldsymbol{u}) &= \sum_{i=1}^{m} \Big(p_{i,x} u + p_{i,y} v + p_{i,t} \Big)^2 \\ &= \sum_{i=1}^{m} \Big(\boldsymbol{\nabla}_3 p_i^\top \boldsymbol{w} \Big)^2 = \sum_{i=1}^{m} \boldsymbol{w}^\top \boldsymbol{\nabla}_3 p_i \boldsymbol{\nabla}_3 p_i^\top \boldsymbol{w} \\ &= \boldsymbol{w}^\top \underbrace{\sum_{i=1}^{m} \boldsymbol{\nabla}_3 p_i \boldsymbol{\nabla}_3 p_i^\top}_{\text{motion tensor } \boldsymbol{J}} \boldsymbol{w} \,. \end{aligned} \tag{2.64}$$

Here, the spatio-temporal offset vector $\boldsymbol{w} = (u, v, 1)^\top$ contains the flow components, and $\boldsymbol{\nabla}_3 = (\partial_x, \partial_y, \partial_t)^\top$ denotes the spatio-temporal gradient, c.f. Equation (2.62).

Normalisation

Linearised data terms as in Equation (2.62) penalise the distance of a flow vector to the line defined by the normal flow. However, as deduced by Lai and Vemuri [1998], Schönemann and Cremers [2006] and Zimmer et al. [2011b], the true distance of the flow to that line is locally re-weighted by the squared magnitude of the image gradient, as can be seen from the following computation:

$$\begin{aligned}(\boldsymbol{\nabla}_3 f^\top \boldsymbol{w})^2 &= (\boldsymbol{\nabla}_2 f^\top \boldsymbol{u} + f_t)^2 \\ &= |\boldsymbol{\nabla}_2 f|^2 \left(\frac{(\boldsymbol{\nabla}_2 f^\top \boldsymbol{u} + f_t)}{|\boldsymbol{\nabla}_2 f|} \right)^2 \\ &= |\boldsymbol{\nabla}_2 f|^2 \left(\frac{\boldsymbol{\nabla}_2 f^\top}{|\boldsymbol{\nabla}_2 f|} \Big(\boldsymbol{u} - \frac{-\boldsymbol{\nabla}_2 f f_t}{|\boldsymbol{\nabla}_2 f|^2} \Big) \right)^2 \\ &\overset{(2.63)}{=} |\boldsymbol{\nabla}_2 f|^2 \left(\frac{\boldsymbol{\nabla}_2 f^\top}{|\boldsymbol{\nabla}_2 f|} \Big(\boldsymbol{u} - \boldsymbol{u}_{\text{normal}} \Big) \right)^2 .\end{aligned} \tag{2.65}$$

The consequence of this re-weighting is that high-contrast positions have more influence on the minimiser than positions where the image gradient is small. To eliminate this behaviour, Lai and Vemuri [1998] proposed to normalise the constancy assumption. Zimmer et al. [2011b] applied this strategy by dividing the motion tensor of each feature by the trace of its upper left 2 × 2 sub-tensor:

$$\bar{\boldsymbol{J}}_{\boldsymbol{p}} = \sum_{i=1}^{n} \frac{1}{|\boldsymbol{\nabla}_2 p_i|^2 + \xi^2} \boldsymbol{\nabla}_3 p_i \boldsymbol{\nabla}_3 p_i^\top , \tag{2.66}$$

where ξ is a small constant to avoid division by zero. Throughout all experiments, its value is kept constant at $\xi = 10^{-2}$. For the rest of this thesis, we will indicate normalised motion tensors with bars, e.g. $\bar{\boldsymbol{J}}$. Generally, the concept of constraint normalisation can be applied to any linearised constancy assumption [Valgaerts et al., 2010]. In motion tensor notation, the factor to normalise any constraint is given by $\theta := (J_{11} + J_{22} + \xi^2)^{-1}$, where $\boldsymbol{J}$ represents the motion tensor:

$$\bar{\boldsymbol{J}} = \underbrace{\frac{1}{J_{11} + J_{22} + \xi^2}}_{=:\theta} \cdot \boldsymbol{J} . \tag{2.67}$$

On the one hand, normalising the data constraints removes the dependency on the local feature contrast. On the other hand, however, from a global energy point of view, this changes the interplay of data and smoothness term completely: Without normalisation, the data term only gives a contribution if reliable image information is available, i.e. if the image gradient is large. Otherwise, the energy is determined solely by the smoothness constraint, this creates the so-called filling-in effect. Contrarily, with normalisation, small image gradients are amplified and each position finally has the same weight.

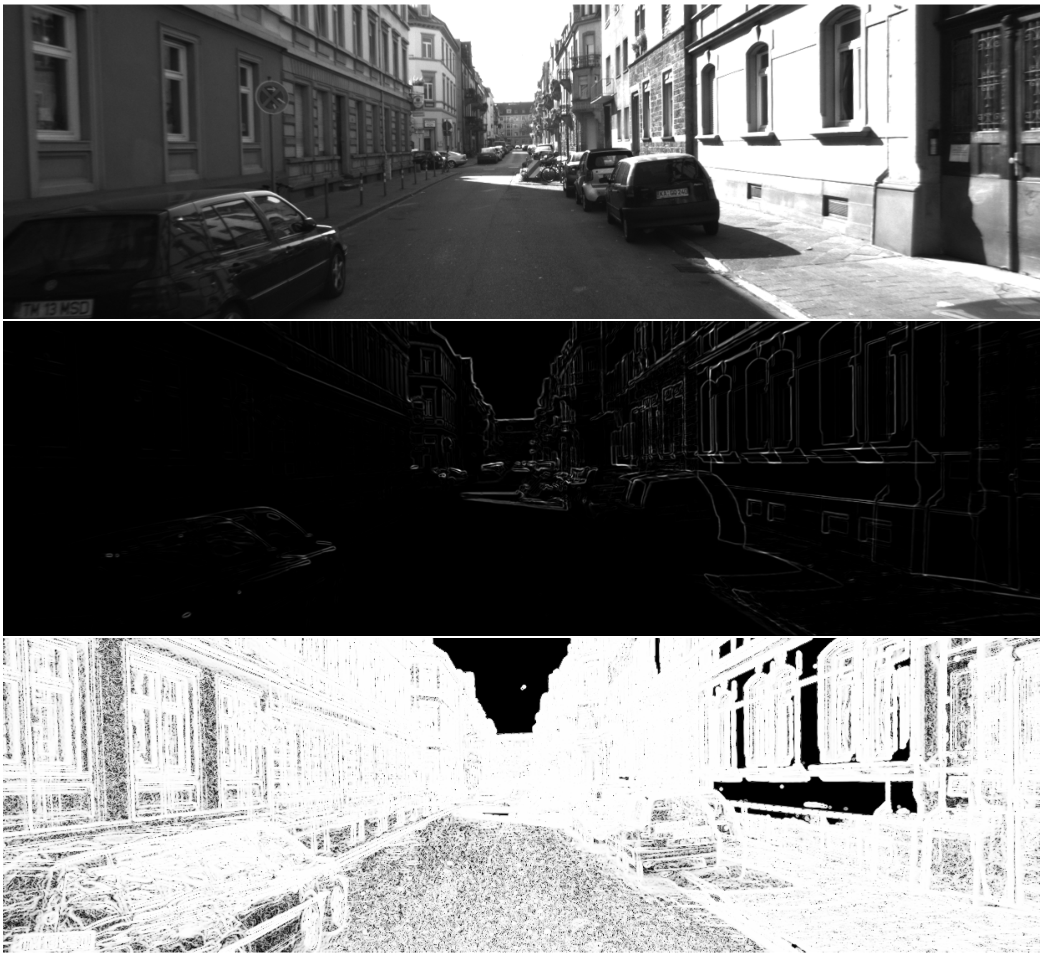


Figure 2.6: Data term normalisation. **Top:** First frame of KITTI training sequence *#15* [Geiger et al., 2012]. Grey values are in $[0, 255]$. **Middle:** Trace of motion tensor without normalisation. Edge pixels have a clearly larger influence than pixels in smooth regions. Values are in the range $[0, 25]$. **Bottom:** Trace of motion tensor after normalisation. Values in range $[0, 1]$. Positions with non-vanishing image gradient all have similar weight. However, regions where the image gradient vanishes still have no influence (e.g. the sky).

This can be problematic for instance if noise is present. In that respect, the small constant ξ plays an important role, as it can also be seen as a parameter allowing to influence the noise sensitivity of the normalisation. Any image noise below ξ is not overamplified.

Figure 2.6 juxtaposes the weightings of the linearised grey value constancy assumption for one sequence of the KITTI Vision benchmark with and without motion tensor normalisation. One can see clearly that without normalisation, the few image regions with high local contrast dominate all other areas completely. Moreover, the regions where the local contrast is really vanishes are mainly saturated image parts.

Robust Data Terms

So far, we have analysed and developed a global data term, in the sense that it couples the flow and input image data *in every pixel*. There are, however, cases, where such a coupling is not appropriate everywhere. For instance, if the image sequence is perturbed by a stochastic process like noise, we cannot expect that corresponding pixels in both frames are affected exactly in the same way. Moreover, *any* feature constancy assumption is conceptually wrong in occlusions, i.e. regions that are – due to the apparent motion in the scene – only visible in one of the two frames.

In any of the described cases, we can expect the intensity or feature difference (c.f. Equation (2.59)) to attain large values. Consequently, with quadratic penalisation, the data term (2.60) will nevertheless try to match such unmatchable regions. This will have negative effects, and, because of the global coupling of the regularisation term, this will spoil the result globally in the whole image domain.

As a remedy, the concept of *sub-quadratic penalisation* from robust statistics can help [Huber, 1981]: By surrounding the quadratic expression with a function

$$\begin{aligned} \Psi_\varepsilon(z^2) &= 2\varepsilon\sqrt{z^2+\varepsilon^2} - 2\varepsilon^2 \\ &= 2\varepsilon^2\sqrt{z^2/\varepsilon^2+1} - 2\varepsilon^2\,, \end{aligned} \tag{2.68}$$

where $\varepsilon > 0$ is a small positive constant, we can limit the influence of outliers [Black and Anandan, 1991] while ensuring the convexity of the expression. Thus, a generic robustified data term would read:

$$D_{f,\Psi}(\boldsymbol{u}) = \Psi\Big(\,(f_2(x+u,y+v) - f_1(x,y))^2\,\Big)\,. \tag{2.69}$$

We note that our choice of the function Ψ is a regularised and differentiable version of the absolute value function (re-scaled by ε). Its derivative Ψ' is the so-called Charbonnier diffusivity [Charbonnier et al., 1994]

$$\Psi'_\varepsilon(z^2) \;=\; \frac{\varepsilon}{\sqrt{z^2+\varepsilon^2}} \;=\; \frac{1}{\sqrt{1+\frac{z^2}{\varepsilon^2}}}\,, \tag{2.70}$$

which is a decreasing function. If its argument is zero, i.e. if the constancy assumption is fulfilled, it has the value 1. On the contrary, if the constancy assumption is violated its value goes to zero.

Separate vs. Joint Robustification. This robustification step offers additional degrees of freedom if more than one constancy assumption is present, or if the feature has multiple channels.

Depending on the situation, it can be beneficial to either impose a *joint* robustification, i.e.

$$\begin{aligned} D_{\boldsymbol{p},\Psi,\text{joint}}(\boldsymbol{u}) &= \Psi\Big(\frac{1}{m}\,\Big|\,\boldsymbol{p}_2(x+u,y+v)-\boldsymbol{p}_1(x,y)\,\Big|^2\Big) \\ &= \Psi\left(\frac{1}{m}\sum_{i=1}^{m}\Big(p_{2,i}(x+u,y+v)-p_{1,i}(x,y)\Big)^2\right), \end{aligned} \tag{2.71}$$

or a separate one, which reads:

$$D_{\boldsymbol{p},\Psi,\text{separate}}(\boldsymbol{u}) = \frac{1}{m}\sum_{i=1}^{m}\Psi\left(\big(\,p_{2,i}(x+u,y+v)-p_{1,i}(x,y)\,\big)^2\right)\,. \tag{2.72}$$

Since our choice of the sub-quadratic function $\Psi_\varepsilon(z^2)$ approximates the absolute value function $|z|$ in the limiting case $\varepsilon \to 0$, joint robustification approximates the L_2-norm of the feature vector difference. Ignoring the influence of the regularisation term for a moment, we find that a jointly robustified data term thus demands the solution to be close to the least squares solution in each point. However, in case of separate robustification, the L_1-norm is approximated, which is closely related to the median.

As a rule of thumb, joint robustification is appropriate if the features should hold and fail together; for instance it is unlikely that the red channel of a colour image sequence matches, but the green and blue channel do not fit well. On the contrary, if one feature can match independently of another, it makes sense to robustify them separately. Of course, also hybrid robustification schemes might make sense, i.e. mixed schemes where several vector-valued features are separately robustified, c.f. [Bruhn and Weickert, 2005].

With the latter robustification strategies for multiple features, an assumption is considered reliable if its cost is low. Thus it is possible that multiple separately robustified feature constancy assumptions are active at same time. There is work by Xu et al. [2010] and Kim et al. [2013] where this situation is avoided, and in which configurations where only one assumption is active at a time are preferred. However, we do not consider such strategies in this work.

2.4.2 Smoothness Term

Let us now turn our attention to the smoothness term. Its purpose is to resolve the ambiguities of the data constraints and to ensure the existence of a dense solution. It is important to choose the smoothness term carefully, because each possible type of smoothness term prefers solutions with different properties. Thus, with a certain choice, we express our prior knowledge that the solution shall have its associated properties and thus influence the resulting flow field.

There is a big variety of possible smoothness terms available in the literature, however the discussion of those is not in the scope of this thesis. Instead, let us now consider a general smoothness term which allows to represent several popular terms from the literature. This general regulariser reads:

$$S(\boldsymbol{u},\boldsymbol{a},\boldsymbol{b}) = R_1(\boldsymbol{u},\boldsymbol{a},\boldsymbol{b}) + \beta R_2(\boldsymbol{a},\boldsymbol{b})\,, \tag{2.73}$$

and consists of two components: R_1 is a *coupling term*:

$$R_1(\boldsymbol{u},\boldsymbol{a},\boldsymbol{b}) \;=\; \Psi\Big(|\boldsymbol{\nabla} u - \boldsymbol{a}|^2 + |\boldsymbol{\nabla} v - \boldsymbol{b}|^2\Big)\,, \tag{2.74}$$

which demands the gradients of the flow field to be similar to the auxiliary vector fields $\boldsymbol{a}$ and $\boldsymbol{b}$. The second term R_2 is a classical smoothness term that penalises changes in the auxiliary functions:

$$R_2(\boldsymbol{a},\boldsymbol{b}) \;=\; \Psi\Big(|\boldsymbol{\mathcal{J}}\boldsymbol{a}|_F^2 + |\boldsymbol{\mathcal{J}}\boldsymbol{b}|_F^2\Big)\,. \tag{2.75}$$

In the latter equation,

$$\boldsymbol{\mathcal{J}}\boldsymbol{v} := \begin{pmatrix} \partial_x v_1 & \partial_y v_1 \\ \partial_x v_2 & \partial_y v_2 \end{pmatrix} \tag{2.76}$$

denotes the Jacobian matrix that captures the first-order derivatives, and $|\cdot|_F$ denotes the Frobenius norm.

Our general smoothness term (2.73) has several well-known special cases that we now want to review briefly.

Tikhonov Regularisation. For the choice $\Psi(z^2) = z^2$, and the constraint $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{0}$, the second term R_2 vanishes naturally and the first term can be simplified to

$$R_1^{\text{Tikhonov}}(\boldsymbol{u}) = |\boldsymbol{\nabla} u|^2 + |\boldsymbol{\nabla} v|^2 . \tag{2.77}$$

By restricting the minimisation to the unknown flow $\boldsymbol{u}$ and setting $\boldsymbol{a}$ and $\boldsymbol{b}$ to zero, we revert the role of the coupling term to a traditional smoothness term that demands the flow derivatives to be small. As result, this term coincides with a standard Tikhonov regularisation [Tikhonov, 1963] term as used by Horn and Schunck [1981].

First-Order Flow-Driven Regularisation. By only setting $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{0}$, again R_2 drops out naturally, and the term

$$R_1^{\text{non-lin}}(\boldsymbol{u}) = \Psi\left(|\boldsymbol{\nabla} u|^2 + |\boldsymbol{\nabla} v|^2\right) . \tag{2.78}$$

realises the popular non-linear isotropic regularisation. This concept has a successful and long tradition in the context of computer vision [Blake and Zisserman, 1987; Shulman and Herve, 1989; Black and Anandan, 1991; Cohen, 1993; Schnörr, 1994; Brox et al., 2004].

Depending on the choice of Ψ, a discontinuity preserving or even discontinuity enhancing behaviour can be achieved. Our choice of a regularised absolute value function from Equation (2.68) approximates total variation (TV) regularisation [Rudin et al., 1992; Acar and Vogel, 1994].

Second-Order Flow-Driven Regularisation

When exploiting its full potential, the general framework from (2.73) can realise a second-order regularisation. To this end, the set of unknowns has to be extended by the vector fields $\boldsymbol{a}, \boldsymbol{b} : \Omega \to \mathbb{R}^2$. The task of the auxiliary variables is then to *approximate* the first order derivatives of the optic flow field:

$$\boldsymbol{a} \approx \boldsymbol{\nabla} u \quad , \quad \boldsymbol{b} \approx \boldsymbol{\nabla} v . \tag{2.79}$$

Consider for a moment the case that the coupling term would be perfectly fulfilled, i.e. equality would be enforced to hold everywhere in terms of a hard constraint, such that $\boldsymbol{a} = \boldsymbol{\nabla} u$ and $\boldsymbol{b} = \boldsymbol{\nabla} v$. Consequently, in this case, R_1 would be satisfied and evaluate to zero. However, the term R_2 would penalise *second-order* derivatives of the flow in this case. More precisely, because of $\boldsymbol{\mathcal{J}}\boldsymbol{\nabla}(\cdot) = \boldsymbol{\mathcal{H}}(\cdot)$, exactly all entries of the Hessian matrices of u and v would be penalised, and an equivalent second-order regularisation term would read:

$$R_2^{\text{direct}}(\boldsymbol{u}) \;=\; \Psi\Big(|\boldsymbol{\mathcal{H}} u|_F^2 + |\boldsymbol{\mathcal{H}} v|_F^2\Big) . \tag{2.80}$$

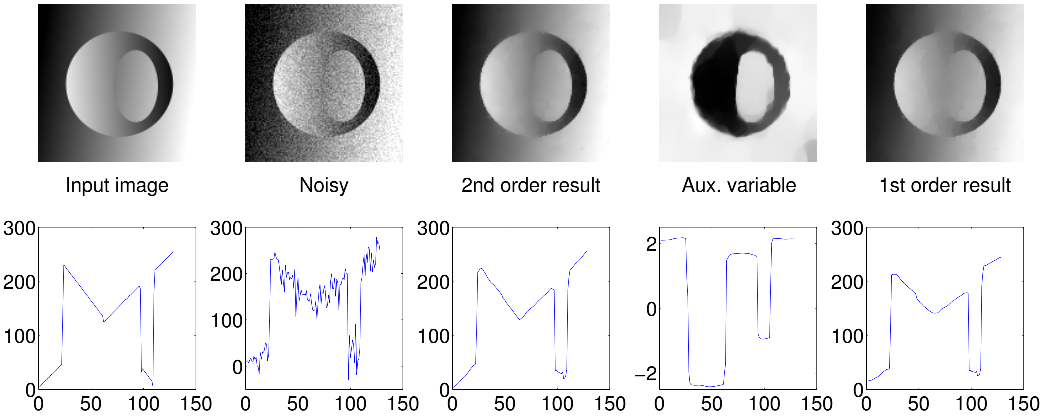


Figure 2.7: Regularisation of a piecewise affine test image. **Top row, from left to right:** Input image, perturbed version with additive Gaussian noise ($\sigma = 10$), denoised version with coupled regulariser ($\alpha = 40$, $\beta = 125$, $\varepsilon = 10^{-2}$), visualisation of the resulting auxiliary coupling variable for the x-derivative, and a corresponding first order regularisation as reference (c.f. Eq. (2.78), $\alpha = 60$). **Bottom row:** Plots of respective horizontal scan lines of the centre row of the images above.

However, such a direct second-order term prefers *continuous* piecewise affine solutions [Lysaker et al., 2003; Didas, 2008]. This means the latter direct second-order regularizer favors continuous solutions, because only the first order derivatives exhibit discontinuities. This corresponds to bends in the solution. Another direct second-order regularisation term for optic flow is presented by Trobin et al. [2008] that penalises the deviation of the solution from an affine function in each position. Also the work of Yuan et al. [2007] considers second-order regularisation term for variational optic flow estimation in the context of fluid dynamics.

In contrast, in the model used in this thesis, the derivatives are allowed to deviate from the auxiliaries (c.f. (2.79)). As consequence, the solution can exhibit both: discontinuities and bends. This is illustrated in Figure 2.7 in the image regularisation setting: Especially when comparing the resulting images with and without coupling, the typical staircasing artefacts of the first-order regulariser become obvious. For the second-oder model this can be avoided. In general, the regularisation scheme from Equation (2.73) can be seen as a regularised and differentiable approximation of the second order *total generalised variation* (TGV^2) smoothness term of Bredies et al. [2010]. The TGV-norm, in turn, has a close relationship to infimal convolution [Chambolle and Lions, 1997]. Setzer et al. [2011] proposes a modified infimal convolution regulariser for the discrete setting that coincides with the TGV-norm. This modification clearly shows that the difference between the original infimal convolution and TGV-norm is one additinal degree of freedom in the latter framework. Hewer et al. [2013] use a linear variant of the coupling scheme for the estimation of Lagrangian strain tensors from optic flow field derivatives in mechanical engineering applications. The TGV^2 norm has also recently been used in several stereo [Ranftl et al., 2012] and optic flow methods, e.g. [Braux-Zin et al., 2013; Vogel et al., 2013].

Other Regularisation Schemes

Our regularisation term cannot cover several other important regularisation concepts. Among those, the isotropic image-driven ideas of Alvarez et al. [1999a] align discontinuities of the solution with edges in the input images. Furthermore, the anisotropic image regularisation term of Nagel and Enkelmann [1986] also considers edge information of the input images, and demands the solution to be smooth along image edges but not across them. The joint image- and flow-driven regularisation of Zimmer et al. [2011b] also performs an anisotropic smoothing, but rather considers directional derivatives of the flow along directions given by the data term. A purely flow-driven anisotropic flow regularisation scheme is presented by Weickert and Schnörr [2001]. The non-local total variation concept is used by Werlberger et al. [2010] and allows to incorporate tonal weights. The work of Sun et al. [2010] shows the relationship between such non-local terms and median filtering.

2.4.3 Multi-Scale Technique

As we have argued before, from an optimisation point of view, the main problem with the constancy assumptions so far is that they consist of highly non-linear and non-convex terms that cannot be minimised without further ado. As one first countermeasure, we already discussed a generic linearisation strategy that leads to a locally convexified and linear (sub-)problem. In the field of nonlinear least squares minimisation, such a linearisation can be interpreted as one step of the Gauss-Newton method [Nocedal and Wright, 2006]. However, this linearisation is merely a linear approximation to the original constraints that obviously only holds if the underlying signal is in fact linear (i.e. smooth), and if the approximation distance, i.e. the length of the flow vector, is sufficiently small. Both these requirements do not hold for general real world image sequences: We have to cope with very large displacements and arbitrary image material. Figure 2.8 illustrates this problem. This exemplary sequence has an image resolution of 1241 × 376

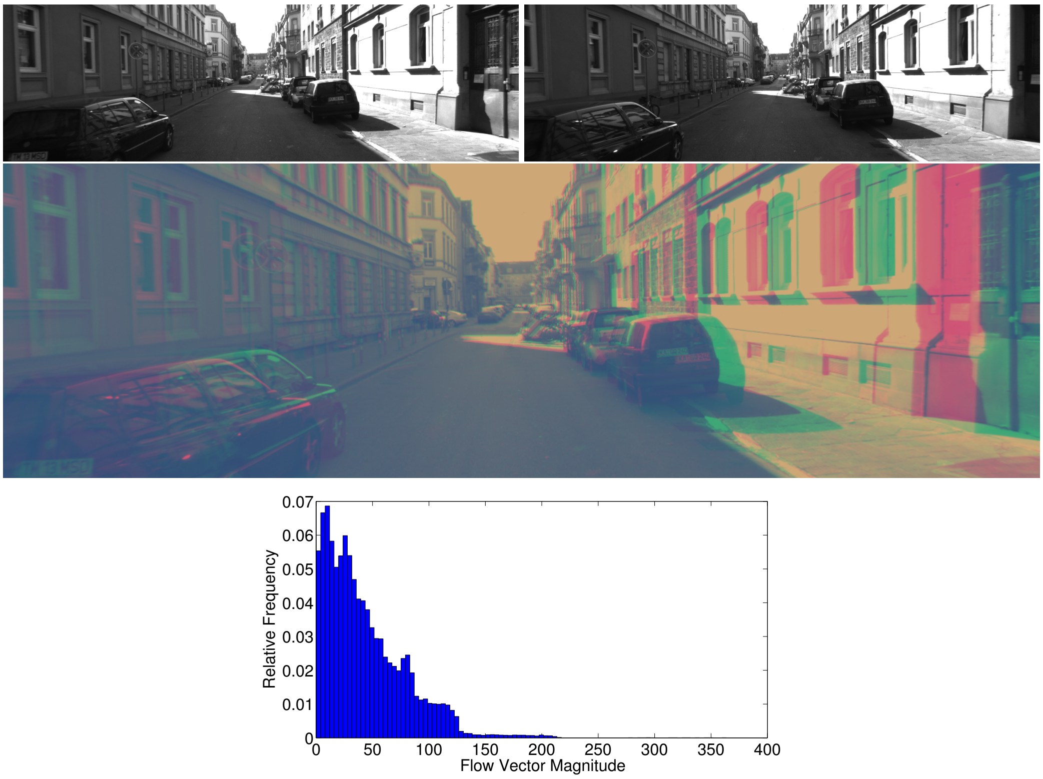


Figure 2.8: Illustration of large displacements in real world image sequence. **Top row:** first and second frame of KITTI training sequence #15. **Middle:** Both frames overlaid as green and red channel, respectively. **Plot:** Distribution of flow vector magnitudes in this image sequence.

pixels and 8.2% of all flow vectors are longer than 100 pixels.

The common solution for this problem in the literature is to perform more than one step of the Gauss-Newton method [Nagel, 1983a], and to embed these steps in a multi-scale strategy [Witkin et al., 1987; Enkelmann, 1988; Anandan, 1989; Brox et al., 2004]. Then, the original non-linear problem is approached step-by-step on different levels of a scale space representation of the input images [Iijima, 1962; Witkin, 1983]. Beginning on a very smooth scale, the solution is successively refined on sharper levels. This strategy makes the estimation of large displacements possible: On very smooth scales, the linearisation holds since the error of a linear approximation becomes less significant. Practically, the refinement of a previous solution is realised by computing an additive increment on each scale, such that the sum of the solution from the previous scale and the increment approximates the solution of the non-linear problem at the current scale. Alternatively to the Gauss-

Newton method, Alvarez et al. [2002] performs a gradient descent on the original non-linear functional which is embedded into a scale space strategy.

In detail, our scheme acts on a finite set of scales indicated with the variable ℓ. We split the overall estimation of the final flow into a series of estimations of flow increments $\boldsymbol{du}^\ell = (du^\ell, dv^\ell)^\top$, starting from the smoothest scale $\ell = 0$ up to the finest scale $\ell_{\max}$. Each incremental flow allows to compute the refined flow for the next finer scale $\ell + 1$ as:

$$u^{\ell+1} = u^\ell + du^\ell \qquad \text{and} \qquad v^{\ell+1} = v^\ell + dv^\ell \,. \tag{2.81}$$

At the coarsest scale, the flow $\boldsymbol{u}^0$ is initialised with a zero flow. On each scale, we minimise the following *incremental* functional:

$$E^\ell(\boldsymbol{du}^\ell) = \int_\Omega \Big(D^\ell(\boldsymbol{u}^\ell + \boldsymbol{du}^\ell) + \alpha \; S(\boldsymbol{u}^\ell + \boldsymbol{du}^\ell) \Big) \; \mathrm{d}\boldsymbol{x} \,. \tag{2.82}$$

In principle, the integrand of the incremental energy coincides with our original functional (2.57). However, the functional (2.82) is minimised only w.r.t. the flow increment and only the smoothed versions of the input images are considered.

Concerning the regularisation term, this means that the estimated increment can be non-smooth. The functional only demands smoothness of the overall refined flow:

$$S(\boldsymbol{u}^\ell + \boldsymbol{du}^\ell, \boldsymbol{a}, \boldsymbol{b}) \;=\; R_1(\boldsymbol{u}^\ell + \boldsymbol{du}^\ell, \boldsymbol{a}, \boldsymbol{b}) + \beta R_2(\boldsymbol{a}, \boldsymbol{b}) \,, \tag{2.83}$$

where

$$\begin{aligned} & R_1(\boldsymbol{u}^\ell + \boldsymbol{du}^\ell, \boldsymbol{a}, \boldsymbol{b}) \\ & \quad = \; \Psi\Big(|\boldsymbol{\nabla} u^\ell + \boldsymbol{\nabla} du^\ell - \boldsymbol{a}|^2 + |\boldsymbol{\nabla} v^\ell + \boldsymbol{\nabla} dv^\ell - \boldsymbol{b}|^2 \Big) \,. \end{aligned} \tag{2.84}$$

The main difference between the original functional and the incremental energy lies in its data term that acts on images at the scale ℓ; for instance a basic robustified version reads:

$$D^\ell_{f,\Psi,\text{joint}}(\boldsymbol{u}^\ell + \boldsymbol{du}^\ell) = \Psi\Big(\Big(\, f_2^\ell(x + u^\ell + du^\ell, y + v^\ell + dv^\ell) - f_1^\ell(x, y) \, \Big)^2 \Big) \,. \tag{2.85}$$

In the latter definition, the smoothed versions f^ℓ are computed by:

$$f^\ell = \mathcal{G}_{\sigma_\ell} * f \quad , \quad \sigma_\ell = \frac{\nu}{\eta^{\ell_{\max} - \ell}} \,, \tag{2.86}$$

where the 2-D Gaussian kernel $\mathcal{G}_\sigma$ with standard deviation σ reads

$$\mathcal{G}_\sigma(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right) . \tag{2.87}$$

Thus, the standard deviation of the Gaussian (i.e. the scale) is determined from the scale index ℓ and the parameter $\eta \in (0,1)$ that steers how far apart two consecutive scales are. Moreover, the parameter $\nu > 0$ scales the size of the Gaussian linearly. For efficiency reasons, we can sub-sample the (discretised) images after smoothing: When going from one scale to the next smoother one, we divide the grid size by the factor η and resample. Thus, the standard deviation of the Gaussian increases on smoother scales to avoid aliasing artifacts.

Once a solution is found on a scale, the transition to the next finer scale is performed as follows: First, the coarse optic flow field is resampled to the finer scale. There, we perform the linearisation around the displaced position: We *warp* the second frame as well as its derivatives by the current overall flow field $\boldsymbol{u}^\ell$. This means we compensate for this motion and consider

$$f_{2,*}^{\ell,\boldsymbol{u}^\ell}(x,y) := f_{2,*}^\ell(x+u^\ell(x,y),\ y+v^\ell(x,y)) . \tag{2.88}$$

Thus, the linearisation around the the point $(\boldsymbol{x}+\boldsymbol{u}^\ell)$ (instead of $\boldsymbol{x}$) reads:

$$\begin{aligned}
&f_2^\ell(x+u^{\ell+1}, y+v^{\ell+1}) && - f_1^\ell(x,y)\\
&= f_2^\ell(x+u^\ell+du^\ell, y+v^\ell+dv^\ell) && - f_1^\ell(x,y)\\
&\approx f_2^{\ell,\boldsymbol{u}^\ell}(x,y) + f_{2,x}^{\ell,\boldsymbol{u}^\ell}(x,y)\,du^\ell + f_{2,y}^{\ell,\boldsymbol{u}^\ell}(x,y)\,dv^\ell && - f_1^\ell(x,y)\\
&= f_{2,x}^{\ell,\boldsymbol{u}^\ell}(x,y)\,du^\ell + f_{2,y}^{\ell,\boldsymbol{u}^\ell}(x,y)\,dv^\ell + f_t^{\ell,\boldsymbol{u}^\ell}(x,y) .
\end{aligned} \tag{2.89}$$

Again, in the last line, the difference between the warped second frame and the first frame is interpreted as a temporal derivative (along the overall flow trajectory). Thus, also the incremental linearised data term can conveniently be transformed into motion tensor notation. We extend our notation from Equation (2.66), and denote with $\bar{\boldsymbol{J}}_{\boldsymbol{p}}^{\ell,\boldsymbol{u}^\ell}$ a normalised motion tensor of features $\boldsymbol{p}$ on scale ℓ, which have been compensated with the flow field $\boldsymbol{u}^\ell$. Thus, our linearised data term (analogously to Equation (2.64)) reads:

$$D_{\boldsymbol{p},\Psi,\text{joint,linearised}}^{\ell,\boldsymbol{u}^\ell}(\boldsymbol{du}^\ell) = \Psi\left(\boldsymbol{dw}^{\ell\top} \cdot \bar{\boldsymbol{J}}_{\boldsymbol{p}}^{\ell,\boldsymbol{u}^\ell} \cdot \boldsymbol{dw}^\ell\right) \tag{2.90}$$

Note that, formally, this multi-resolution strategy differs slightly from the strategy of Brox et al. [2004]: We solve a series of incremental energy functionals, whereas Brox et al. [2004] embed the minimality conditions of the original functional into a fixed point iteration. Our series of functionals makes the use of images from different scales explicit. Moreover, the series of incremental energies is an explicit and deterministic convexification of the non-convex original problem: after linearisation, the convex problem that is solved throughout each incremental step is clearly defined, c.f. the Gauss-Newton method [Nocedal and Wright, 2006].

Further, please note that for the special case $\ell_{\max} = 0$, the multi-scale strategy exactly falls back to the original scheme without warping: the pre-smoothing of the images at scale 0 with a Gaussian of standard deviation $\sigma_0 = \nu$ ensures their differentiability, cf. Section 1.4. The back-registration with a zero flow field leaves the second frame unchanged.

Let us also remark that it is possible to iterate the computation of incremental flows on each scale [Sun et al., 2010]. This means, instead of going to the next finer scale as soon as the increment has been found, one can stay at the current scale, add overall flow $\boldsymbol{u}^\ell$ and increment $\boldsymbol{du}^\ell$ together, back-register the second frame by the sum, and recompute a new flow increment. Sun et al. [2010] report that these intra-scale iterations allow to increase the accuracy of the resulting flow slightly. However, we could not reproduce this behavior unless we decrease the number of iterations of the linear system solver significantly.

Large Displacements of Small Objects

The linearised data term in Equation (2.82) only allows to estimate small displacements – relative to the current scale. This means that on very smooth scales the displacements being estimated can indeed be large with respect to the original scale. However, the displacements being estimated there are optimal for the smoothed images, where each pixel represents a local average or the original imagery. Let us now consider small objects with high velocity. On a scale that is smooth enough to capture such high velocities, small objects are *smoothed out*: due to their small size, such objects do not have enough weight in the local average and disappear. As soon as as the multi-scale strategy reaches a finer scale where the small object survives, however, such large motions cannot be recovered anymore.

Thus, the problem that small objects undergoing large displacements cannot be estimated is an inherent property of the multi-scale algorithm. Moreover, seen from the viewpoint of stability and robustness, this property is even desired: If it would be possible to alter the strategy such that on smooth scales, small objects could dominate their surrounding, then the algorithm would effectively trust in outliers. This would weaken the overall strategy and cannot lead to a reliable estimation scheme.

The recent work of Sevilla-Lara et al. [2014] partly aims in that direction. By *exploding* the input images to multiple channels with a different intensity intervals for each channel, the phenomenon that small objects disappear on smoother scales shall be suppressed.

Another possibility to capture small objects with long motion vectors is via the integration of sparse feature matches. Such methods are proposed by Brox and Malik [2011], Stoll et al. [2013] and Braux-Zin et al. [2013].

2.4.4 Minimisation

Until now, we have only introduced and explained our model assumptions from an energy point of view. All our assumptions have lead to penalty terms that attain small values if the model is fulfilled and large values if it is violated. Thus, the solution we are searching for minimises the developed energy functionals. Let us now come to the question how to find such a minimiser.

The calculus of variations [Gelfand and Fomin, 2000] tells us that any minimiser of a generic energy functional of type

$$E(u) = \int_\Omega F(\boldsymbol{x}, u, \boldsymbol{\nabla} u)\, \mathrm{d}\boldsymbol{x}\,, \tag{2.91}$$

has to fulfill the Euler-Lagrange equation:

$$F_u - \mathrm{div}\, F_{\boldsymbol{\nabla} u} = 0\,, \tag{2.92}$$

as well as the boundary conditions

$$\boldsymbol{n}^\top F_{\boldsymbol{\nabla} u} = 0 \qquad \forall \boldsymbol{x} \in \partial(\Omega)\,, \tag{2.93}$$

where $\mathrm{div} \;= \boldsymbol{\nabla}^\top := (\partial_x, \partial_y)$ is the divergence operator, $\boldsymbol{n}$ the unit outer normal vector, and the boundary of the image domain is denoted with $\partial(\Omega)$.

For convenience, we repeat our functional (2.82) with linearised data term (2.90), whose minimisation will be discussed in the following:

$$\begin{aligned} E^\ell(\boldsymbol{du}^\ell, \boldsymbol{a}^\ell, \boldsymbol{b}^\ell) = \int_\Omega \Big(& D^\ell(\boldsymbol{u}^\ell + \boldsymbol{du}^\ell) \\ & + \alpha\Big(R_1(\boldsymbol{u}^\ell + \boldsymbol{du}^\ell, \boldsymbol{a}^\ell, \boldsymbol{b}^\ell) + \beta R_2(\boldsymbol{a}^\ell, \boldsymbol{b}^\ell)\Big)\Big)\, \mathrm{d}\boldsymbol{x}\,. \end{aligned} \tag{2.94}$$

For the sake of readability, we omit the scale ℓ for the rest of this section, and obtain a system of six coupled Euler-Lagrange equations:

$$\begin{aligned}
&D_{du} && -\alpha\,\mathrm{div}\;R_{1,\boldsymbol{\nabla} du} && = 0\,, && (2.95)\\
&D_{dv} && -\alpha\,\mathrm{div}\;R_{1,\boldsymbol{\nabla} dv} && = 0\,, && (2.96)\\
&\alpha\;(R_{1,a_1} && -\beta\,\mathrm{div}\;R_{2,\boldsymbol{\nabla} a_1}) && = 0\,, && (2.97)\\
&\alpha\;(R_{1,a_2} && -\beta\,\mathrm{div}\;R_{2,\boldsymbol{\nabla} a_2}) && = 0\,, && (2.98)\\
&\alpha\;(R_{1,b_1} && -\beta\,\mathrm{div}\;R_{2,\boldsymbol{\nabla} b_1}) && = 0\,, && (2.99)\\
&\alpha\;(R_{1,b_2} && -\beta\,\mathrm{div}\;R_{2,\boldsymbol{\nabla} b_2}) && = 0\,. && (2.100)
\end{aligned}$$

We introduce the following abbreviations:

$$\Psi'_D := \Psi'\Big(\boldsymbol{dw}^\top \cdot \bar{\boldsymbol{J}}_{\boldsymbol{p}}^{\ell,\boldsymbol{u}^\ell} \cdot \boldsymbol{dw}\Big)\,, \tag{2.101}$$

$$\Psi'_{R_1} := \Psi'\Big(|\boldsymbol{\nabla} u + \boldsymbol{\nabla} du - \boldsymbol{a}|^2 + |\boldsymbol{\nabla} v + \boldsymbol{\nabla} dv - \boldsymbol{b}|^2\Big)\,, \tag{2.102}$$

$$\Psi'_{R_2} := \Psi'\Big(|\boldsymbol{\mathcal{J}} \boldsymbol{a}|^2 + |\boldsymbol{\mathcal{J}} \boldsymbol{b}|^2\Big)\,. \tag{2.103}$$

With these abbreviations, the Euler-Lagrange equations that stem from the optic flow increments du and dv (2.95) and (2.96) can be written as

$$\Psi'_D \cdot (J_{11} du + J_{12} dv + J_{13}) - \alpha\;\mathrm{div}\left(\Psi'_{R_1} \cdot (\boldsymbol{\nabla} u + \boldsymbol{\nabla} du - \boldsymbol{a})\right) \;=\; 0\,, \tag{2.104}$$

$$\Psi'_D \cdot (J_{12} du + J_{22} dv + J_{23}) - \alpha\;\mathrm{div}\left(\Psi'_{R_1} \cdot (\boldsymbol{\nabla} v + \boldsymbol{\nabla} dv - \boldsymbol{b})\right) \;=\; 0\,, \tag{2.105}$$

where J_{ij} denote the entries of the motion tensor, cf. (2.101). The latter two equations are associated to natural boundary conditions, which read:

$$\boldsymbol{n}^\top(\boldsymbol{\nabla} u + \boldsymbol{\nabla} du - \boldsymbol{a}) = 0\,, \tag{2.106}$$

$$\boldsymbol{n}^\top(\boldsymbol{\nabla} v + \boldsymbol{\nabla} dv - \boldsymbol{b}) = 0\,. \tag{2.107}$$

As one can see, the first auxiliary variable $\boldsymbol{a}$ couples to the gradient of the horizontal flow component $u + du$, and the second variable $\boldsymbol{b}$ belongs to the vertical part $v + dv$. Moreover, concerning the boundary conditions, if the the auxiliaries do not vanish at the boundary, the given boundary conditions ensure that also the outer normal derivatives of the flow do not vanish there.

Let us now come to the remaining four coupling variables. Having a closer look at the functional, one can recognise that the term R_1 demands the auxiliary variables to be similar to the derivatives of the flow and R_2 demands them to be smooth. This bears analogy to variational approaches for signal or image regularisation. In such methods, a noisy input image corresponds to the flow derivatives in our functional. This similarity is also obvious in the minimality conditions:

$$
\begin{aligned}
-\Psi'_{R_1} \cdot (u_x + du_x - a_1) \; - \; \beta \text{ div } (\Psi'_{R_2} \boldsymbol{\nabla} a_1) \qquad &= 0\,, && (2.108)\\
-\Psi'_{R_1} \cdot (u_y + du_y - a_2) \; - \; \beta \text{ div } (\Psi'_{R_2} \boldsymbol{\nabla} a_2) \qquad &= 0\,, && (2.109)\\
-\Psi'_{R_1} \cdot (v_x + dv_x - b_1) \; - \; \beta \text{ div } (\Psi'_{R_2} \boldsymbol{\nabla} b_1) \qquad &= 0\,, && (2.110)\\
-\Psi'_{R_1} \cdot (v_y + dv_y - b_2) \; - \; \beta \text{ div } (\Psi'_{R_2} \boldsymbol{\nabla} b_2) \qquad &= 0\,. && (2.111)
\end{aligned}
$$

The boundary conditions for the latter four equations are homogeneous Neumann conditions:

$$
\boldsymbol{n}^\top \boldsymbol{\nabla} c = 0 \quad , \qquad c \in \{a_1, a_2, b_1, b_2\}\,. \tag{2.112}
$$

Lagged Nonlinearity Algorithm

In total, we have deduced six coupled PDEs that determine the minimiser of our energy. These PDEs are in general non-linear due to the Ψ'-terms (2.101)–(2.103). In order to resolve this source of non-linearity, we apply the Kačanov-method [Fučik et al., 1973; Zeidler, 1990; Vogel and Oman, 1996] which introduces a fixed point iteration with index k. In each iteration, the nonlinear Ψ'-terms are converted into linear factors by fixing them to the old iterate. What remains is a system of six *linear* PDEs which reads:

$$
\begin{aligned}
&\Psi'^k_D \cdot (J_{11} du^{k+1} + J_{12} dv^{k+1} + J_{13})\\
&\qquad\qquad - \alpha \text{ div } \Big(\Psi'^k_{R_1} \cdot (\boldsymbol{\nabla} u + \boldsymbol{\nabla} du^{k+1} - \boldsymbol{a}^{k+1}) \Big) = 0\,,
\end{aligned} \tag{2.113}
$$

$$
\begin{aligned}
&\Psi'^k_D \cdot (J_{12} du^{k+1} + J_{22} dv^{k+1} + J_{23})\\
&\qquad\qquad - \alpha \text{ div } \Big(\Psi'^k_{R_1} \cdot (\boldsymbol{\nabla} v + \boldsymbol{\nabla} dv^{k+1} - \boldsymbol{b}^{k+1}) \Big) = 0\,,
\end{aligned} \tag{2.114}
$$

$$\begin{aligned}-\Psi'^{k}_{R_1} \cdot (u_x + du_x^{k+1} - a_1^{k+1}) \\ - \beta \operatorname{div} (\Psi'^{k}_{R_2} \boldsymbol{\nabla} a_1^{k+1}) = 0\,,\end{aligned} \tag{2.115}$$

$$\begin{aligned}-\Psi'^{k}_{R_1} \cdot (u_y + du_y^{k+1} - a_2^{k+1}) \\ - \beta \operatorname{div} (\Psi'^{k}_{R_2} \boldsymbol{\nabla} a_2^{k+1}) = 0\,,\end{aligned} \tag{2.116}$$

$$\begin{aligned}-\Psi'^{k}_{R_1} \cdot (v_x + dv_x^{k+1} - b_1^{k+1}) \\ - \beta \operatorname{div} (\Psi'^{k}_{R_2} \boldsymbol{\nabla} b_1^{k+1}) = 0\,,\end{aligned} \tag{2.117}$$

$$\begin{aligned}-\Psi'^{k}_{R_1} \cdot (v_y + dv_y^{k+1} - b_2^{k+1}) \\ - \beta \operatorname{div} (\Psi'^{k}_{R_2} \boldsymbol{\nabla} b_2^{k+1}) = 0\,,\end{aligned} \tag{2.118}$$

where the abbreviation Ψ'^{k}_{*} means that the respective nonlinearity is evaluated with the unknowns from iteration k. This lagging non-linearity is the reason why the Kačanov method is also known under the name *lagged-nonlinearity algorithm*. We initialise the flow increment $\boldsymbol{du}^0$ with zero and the auxiliary coupling variables with the (resampled) solution from the previous scale.

Reinterpretation. The lagged nonlinearity method allows an alternative interpretation: Under the name *iteratively reweighed least squares*, an advanced least-squares technique is known, where weights g^k play an important role. These weights pop up in the associated normal equations in terms of a diagonal weight matrix, and the weighted least squares problem is solved several times, always with updated weights. Let us now consider the following series of functionals:

$$\begin{aligned}E(\boldsymbol{du}^{k+1}, \boldsymbol{a}^{k+1}, \boldsymbol{b}^{k+1}) = \int_\Omega \Bigg(g_D^k \cdot \left(\boldsymbol{dw}^{k+1\top} \cdot \bar{\boldsymbol{J}}_{\boldsymbol{p}}^{\ell,\boldsymbol{u}^\ell} \cdot \boldsymbol{dw}^{k+1} \right) \\ + \alpha \cdot \Bigg(g_{R_1}^k \cdot \left(|\boldsymbol{\nabla} u + \boldsymbol{\nabla} du^{k+1} - \boldsymbol{a}^{k+1}|^2 + |\boldsymbol{\nabla} v + \boldsymbol{\nabla} dv^{k+1} - \boldsymbol{b}^{k+1}|^2 \right) \\ + \beta \cdot g_{R_2}^k \cdot \left(|\boldsymbol{\mathcal{J}} \boldsymbol{a}^{k+1}|^2 + |\boldsymbol{\mathcal{J}} \boldsymbol{b}^{k+1}|^2 \right) \Bigg) \Bigg) \, \mathrm{d}\boldsymbol{x}\,.\end{aligned} \tag{2.119}$$

The interesting point about these functionals is that they are linear in all unknowns. The nonlinearities are hidden in the terms $g_*^k = \Psi'^{k}_{*}$. However, the associated Euler-Lagrange equations coincide with the those of the lagged nonlinearity method, cf. Equations (2.113)–(2.118). Thus, for our type of energy functional, both non-linear optimisation methods, the lagged nonlinearity scheme as well as the iteratively reweighted least squares technique, coincide since they lead to the same minimisation procedure.

2.4.5 Numerical Algorithm and Implementation

At this point, we have developed two nested fixed point iterations to solve the original functional: The outer loop (iteration variable ℓ) realises the multi-scale strategy and makes the estimation of large displacements feasible. The nested inner iterations (iteration variable k) tackle the non-linear terms that stem from the robust functions. At this stage, however, we are still left with a linear system of equations with 6 unknowns per pixel to solve.

Discretisation

We assume the images to be sampled on a regular grid with horizontal and vertical grid size h_1 and h_2, respectively. We use finite differences to approximate derivatives in discrete images: The spatial derivatives of the input image data appear in the motion tensor in (2.90). They are computed with a 4th-order stencil $(1, -8, 0, 8, -1)/(24h_d)$, $d = 1, 2$. The temporal derivative f_t is determined by a simple forward difference, c.f. Equation (2.62). Moreover, during the backward-registration step from Equation (2.88), any interpolation is performed using bicubic interpolation [Keys, 1981]. A discrete linear system of equations corresponding to the system of partial differential equations (2.113)–(2.118) can be deduced in two ways: either by discretising the mentioned PDEs directly, or by developing a discrete version of the corresponding linearised functional given in Equation (2.119). Due to the non-standard boundary conditions (2.107), we choose the latter alternative and discretise the functional. This discretisation is postponed to Appendix A.

Iterative Solution

As a fast and easily implementable linear system solver, we use the *Fast Jacobi* method of Grewenig et al. [2013]. It is perfectly suited for an implementation on parallel hardware architectures such as modern GPUs. Basically, it is based on a standard Jacobi solver. However, varying cyclic under- and over-relaxations ω, where even half of them may violate the stability limit allow an enormous speed-up. More precisely, one iteration step at the pyramid level ℓ with pixel index i and iteration index k is for the flow increments

du^ℓ and dv^ℓ given by

$$\begin{aligned} du_i^{\ell,k+1} = (1-\omega)\cdot du_i^{\ell,k} \; + \; \omega \; \cdot \; \Bigg(& - {\Psi'}_{Di} \cdot (J_{12,i} \cdot dv_i^{\ell,k} + J_{13,i}) \\ & + \sum_{d=1}^{2} \sum_{j \in \mathcal{N}_d(i)} \alpha \cdot \frac{{\Psi'}_{R_1 i} + {\Psi'}_{R_1 j}}{2h_d} \cdot \left(\frac{u_j^{\ell,k} - u_i^{\ell,k} + du_j^{\ell,k}}{h_d} + a_{di}^{\ell,k} - a_{dj}^{\ell,k} \right) \Bigg) \\ & \Bigg/ \left({\Psi'}_{Di} \cdot J_{11,i} + \sum_{d=1}^{2} \sum_{j \in \mathcal{N}_d(i)} \alpha \cdot \frac{{\Psi'}_{R_1 i} + {\Psi'}_{R_1 j}}{2h_d^2} \right), \end{aligned} \tag{2.120}$$

and

$$\begin{aligned} dv_i^{\ell,k+1} = (1-\omega)\cdot dv_i^{\ell,k} \; + \; \omega \; \cdot \; \Bigg(& - {\Psi'}_{Di} \cdot (J_{12,i} \cdot du_i^{\ell,k} + J_{23,i}) \\ & + \sum_{d=1}^{2} \sum_{j \in \mathcal{N}_d(i)} \alpha \cdot \frac{{\Psi'}_{R_1 i} + {\Psi'}_{R_1 j}}{2h_d} \cdot \left(\frac{v_j^{\ell,k} - v_i^{\ell,k} + dv_j^{\ell,k}}{h_d} + b_{di}^{\ell,k} - b_{dj}^{\ell,k} \right) \Bigg) \\ & \Bigg/ \left({\Psi'}_{Di} \cdot J_{22,i} + \sum_{d=1}^{2} \sum_{j \in \mathcal{N}_d(i)} \alpha \cdot \frac{{\Psi'}_{R_1 i} + {\Psi'}_{R_1 j}}{2h_d^2} \right), \end{aligned} \tag{2.121}$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ describe the neighbouring pixels in horizontal and vertical direction, respectively. In an analogous way, the iteration step for $\boldsymbol{a} = (a_1, a_2)^\top$ and $\boldsymbol{b} = (b_1, b_2)^\top$ for $p = 1, 2$ reads

$$\begin{aligned} a_{pi}^{k+1} = (1-\omega)\cdot a_{pi}^{k} \; + \; \omega \; \cdot \; \Bigg(& {\Psi'}_{R_1 i} \cdot \frac{u_{n_p^+}^{\ell,k} - u_{n_p^-}^{\ell,k} + du_{n_p^+}^{\ell,k} - du_{n_p^-}^{\ell,k}}{2h_p} \\ & + \sum_{d=1}^{2} \sum_{j \in \mathcal{N}_d(i)} \beta \cdot \frac{{\Psi'}_{R_2 i} + {\Psi'}_{R_2 j}}{2h_d^2} \cdot a_{pj}^{k} \Bigg) \\ & \Bigg/ \left({\Psi'}_{R_1 i} + \sum_{d=1}^{2} \sum_{j \in \mathcal{N}_d(i)} \beta \cdot \frac{{\Psi'}_{R_2 i} + {\Psi'}_{R_2 j}}{2h_d^2} \right), \end{aligned} \tag{2.122}$$

and

$$
\begin{aligned}
b_{pi}^{k+1} = (1-\omega)\cdot b_{pi}^{k} \;+\; \omega \;\cdot\; \Bigg(& {\Psi'}_{R_1 i} \cdot \frac{v^{\ell,k}_{n_p^+} - v^{\ell,k}_{n_p^-} + dv^{\ell,k}_{n_p^+} - dv^{\ell,k}_{n_p^-}}{2h_p} \\
& + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \beta \cdot \frac{{\Psi'}_{R_2 i} + {\Psi'}_{R_2 j}}{2h_d^2} \cdot b_{pj}^{k} \Bigg) \\
& \Bigg/ \Bigg({\Psi'}_{R_1 i} + \sum_{d=1}^{2} \sum_{j\in\mathcal{N}_d(i)} \beta \cdot \frac{{\Psi'}_{R_2 i} + {\Psi'}_{R_2 j}}{2h_d^2} \Bigg) \,,
\end{aligned}
\tag{2.123}
$$

where n_1^- and n_1^+ describe the left and right neighbouring pixels in horizontal direction. In a similar way, the vertical neighbours are denoted by n_2^- and n_2^+.

Additionally, we embed the iterative Fast Jacobi scheme that we perform on each scale into a cascadic coarse-to-fine scheme, c.f. Bruhn [2006] and Meister [2008]. This means that the solution on each level of the *outer warping pyramid* comprises another *inner solver pyramid*. To this end, after having warped the input images and having computed all motion tensor entries at each scale of the outer pyramid, we resample all tensors and other quantities to the coarsest scale of the inner pyramid. There, we initialise the unknowns with zero and perform a number of Fast Jacobi cycles. After that, we resample the intermediate solution to the next finer scale of the inner pyramid where we perform again a number of cycles. This algorithm is continued until we arrive at the finest inner scale. The outer warping pyramid is typically very steep, i.e. the resolution only changes by a factor of 0.95 between subsequent scales. However, the purpose of the inner pyramid is merely to provide a good initialisation. So, this pyramid can be less steep, and we half the resolution between two scales.

2.5 Experiments

In this section, let us evaluate the discussed concepts. Before we actually present the experiments, we are going to detail on our experimental setup in Section 2.5.1. After that, we perform a general comparison of all invariant descriptors that have been considered in this chapter (Section 2.5.2). Next, we demonstrate that order-based descriptors do also perform well on perturbed image material by analysing their performance under synthetic perturbations (Section 2.5.3). Then, we turn to the evaluation of the components of the presented framework in Section 2.5.4 and present a real world multi-modal image matching example in Section 2.5.5. Finally, we compare our framework to the state-of-the-art in Section 2.5.6.

2.5.1 Experimental Setup

Before starting with the actual evaluation, let us detail on our general experimental setup.

Optimal Parameter Estimation

When trying to find the best descriptor or the best strategy, it is important to optimise the crucial regularisation parameters of all competing methods equally well. Thus, in order to perform such a fair comparison, a clear parameter optimisation strategy has to be defined. In this thesis we will always use the following fixed iterative strategy: We pick a set of image sequences, e.g. all Middlebury training image sequences [Baker et al., 2011], for which ground truth optic flow fields are available, and our objective is to minimise the average error over all sequences. To this end, we follow an iterative strategy that shortens in each step an open search interval S_k in which the optimal parameter value is contained. Initially, we have to define the search interval $S_0 = (a_0, b_0)$, that definitely contains the optimal value. Usually, the lower bound of this interval is $a_0 = 0$. Throughout the experiments for this thesis, all optimal parameters have shown to be in the interval $(0, 1000)$.

In each iteration of our strategy, we sample the current interval at positions s_i that lie in different orders of magnitudes:

$$s_i = a_k + \frac{b_k - a_k}{d_k^i}\,, \quad i = 1, ..., n_s\,. \tag{2.124}$$

At each position, we compute the average error over the chosen set of image sequences. Typically, we sample $n_s = 5$ orders of magnitude, and we choose the divisor $d_0 = 10$ initially. Assume the j-th sample s_j of the interval $S_k = (a_k, b_k)$ leads to the smallest mean error. Then we define the restricted search interval for the next iteration to be the order of magnitude above and the one below the best sample:

$$S_{k+1} = \left(a_k + \frac{b_k - a_k}{d_k^{j+1}},\ a_k + \frac{b_k - a_k}{d^{j-1}} \right) \tag{2.125}$$

It is easy to show that $S_{k+1} \subset S_k$. Moreover, we decrease the step size to achieve a finer sampling with every iteration:

$$d_{k+1} = \sqrt{d_k}\,. \tag{2.126}$$

Table 2.3: First three iterations of an exemplary optimisation sequence. Optimal samples are marked by an asterisk.

k	a_k	b_k	d_k	s_5	s_4	s_3	s_2	s_1
0	0	1000	10	0.01	0.1	1^*	10	100
1	0.1	10	3.16	0.131	0.199	0.413	1.09	3.23^*
2	1.09	10	1.77	1.59	1.981^*	2.674	3.90	6.1

In each iteration, the search interval shrinks and thus also the minimal and maximal measured average errors come closer. Thus, we stop after 4 iterations or if the maximal error is less than 1% larger than the minimal one.

Example. Let us consider the initial interval $S_0 = (0, 1000)$, $n_s = 5$ and $d_0 = 10$. For a hypothetical optimisation, we summarise the computed samples and intervals in Table 2.3. In this exemplary case, during the initial iteration the sample s_3 leads to the least error, in the second iteration s_1 and in the last iteration the sample s_4 are optimal.

If more than one parameter has to be optimised, we perform this logarithmic sampling strategy in N-D, leading to n_s^N mean error evaluations per iteration.

Instead of choosing logarithmically distributed sampling positions, one could also perform an equidistant sampling. However, if the interval is chosen too large in the beginning, much more iterations of the scheme are necessary. There do exist alternative methods, such as Golden Section Search [Press et al., 2007] or standard binary search schemes. However, in our practical experiments, we found out that the initial search interval has to be chosen much more carefully for binary schemes than for the described one. Any search scheme performing binary decisions runs a high risk of turning into a wrong direction if the interval is very large. Due to the $n_s > 2$ samplings, this risk is much lower for our scheme. However, the most important requirement to any of these schemes is to perform a fair comparison of competing strategies. This is fulfilled by our scheme.

It is obvious that in order to execute the described optimisation strategy, large compute resources are indispensable. For instance, to perform three iterations of the 2-D scheme with an image set consisting of 10 images with 5 samplings per dimension, 750 flow fields have to be computed. Fortunately, we can access the compute cluster of the Cluster of Excellence “Multimodal Computing and Interaction” (MMCI) at Saarland University to master the workload. In detail, this cluster consists of 78 *DELL PowerEdge 1950* servers, each being equipped with two *Intel XEON E5430* CPUs (2.66GHz quad core) and 16 GB RAM. Without this cluster, our upcoming experiments would not have been feasible in this elaborateness.

Implementation Details

Our main implementation is written in C and basic parallelisation is realised via OpenMP directives. Although our numerical scheme is described in detail in Section 2.4.5, the multi-scale technique comprises several resampling and interpolation operations that we realise as follows: The backward-registration as well as the resampling of smoothed images to a coarser grid is performed using bicubic interpolation [Keys, 1981]. However, to avoid over- and undershoots, we only use bilinear interpolation for the prolongation operator that is used to transfer flow fields as well as auxiliary variables from a coarse scale to the next finer one. Moreover, we perform 10 inner and 5 outer iteration steps of our cascadic numerical solver scheme on each level. Compared to 5 inner iterations, hardly any changes of the solution can be observed.

Available Image Material

Since our main objective is to achieve the highest possible accuracy, we have to be able to assess the quality of estimated flow fields objectively, by means of average error measures, cf. Section 1.4.2. However, all such measures can only be evaluated if a ground truth optic flow field is available.

For synthetic image sequences depicting rendered scenes, the acquisition of ground truth flow is not an issue. Unfortunately, often the resulting rendered images look unrealistic and do not reproduce well certain characteristics of real photographs such as lighting, noise, motion blur, atmospheric effects, etc.. Nevertheless, rendered scenes offer very reliable ground truth flow fields and occlusion masks. The meanwhile classical Middlebury benchmark [Baker et al., 2011] is partly based on such synthetic sequences. The more recent *MPI Sintel benchmark* of Butler et al. [2012] is an large scale source of such synthetic image sequences. Here, parts of the open source short film *Sintel*[2] build the basis of the benchmark that comprises around 1000 frames with ground truth flow fields.

However, the reliable measurement of ground truth flow information in real world scenarios is still an open problem, even if huge technical effort is taken. The state-of-the-art in this respect is shown by Geiger et al. [2012], who equip a car with many different kinds of sensors to allow a flow estimation while the car is driving. Nonetheless, the acquired measurements contain small scale imprecisions, compelling the authors to propose a non-standard

[2] Official webpage `https://durian.blender.org`

Table 2.4: Uncritical model parameters and chosen values.

Parameter	ε_c	ε_s	ε_d	η	ν	k
Value	10^{-2}	10^{-2}	10^{-2}	0.95	0.75	13

error metric to account for this problem. Moreover, another severe restriction of the laser scans of Geiger et al. [2012] is that the captured scene must be static. This is very difficult to guarantee, and in some of the published training sequences independently moving cars are visible that violate this requirement. We further remark that the authors of this benchmark have published two different ground truth flow fields for each image sequence, one excluding and one including regions that leave the field of view. Throughout this thesis we will always refer to the latter one (abbreviated with *occ*).

Finally, the recent benchmark of Scharstein et al. [2014] represents the state-of-the-art for the real-world stereo setting. For a set of 11 static scenes in calibrated stereo camera setups, very accurate disparity maps are acquired with structured lighting techniques. Moreover, the pictures of the scenes are available under very different lighting situations such that very challenging test problems are available. Although these image sequences are provided in a rectified ortho-parallel stereo setting, where we can be sure that the vertical flow component v must vanish, it is still possible to apply our methods test-wise to such image sequences without enforcing v to be zero in any way. Since all discussed descriptors are used within the same variational framework and implementation, the results are well comparable.

To obtain experimental results that are as representative as possible, we select a subset of the mentioned benchmarks. The images of this set are depicted in Figures 2.9 and 2.10. The MPI Sintel sequences and the classic Middlebury problems exhibit only very few illumination changes. For the new Middlebury stereo sequences, we have exchanged the original second frame with the second frame taken in the alternative lighting situation. By that, the three selected image sequences exhibit drastic appearance changes.

Choice of Parameters

Our model comprises a number of *uncritical* parameters that we fix to one value throughout all our experiments. These parameters are summarised in Table 2.4. In this table, ε_c, ε_s, and ε_d are the parameters of the robust functions around the coupling, auxiliary variable smoothness, and data term, respectively, c.f. Equation (2.68). The parameters η and ν affect the multi-

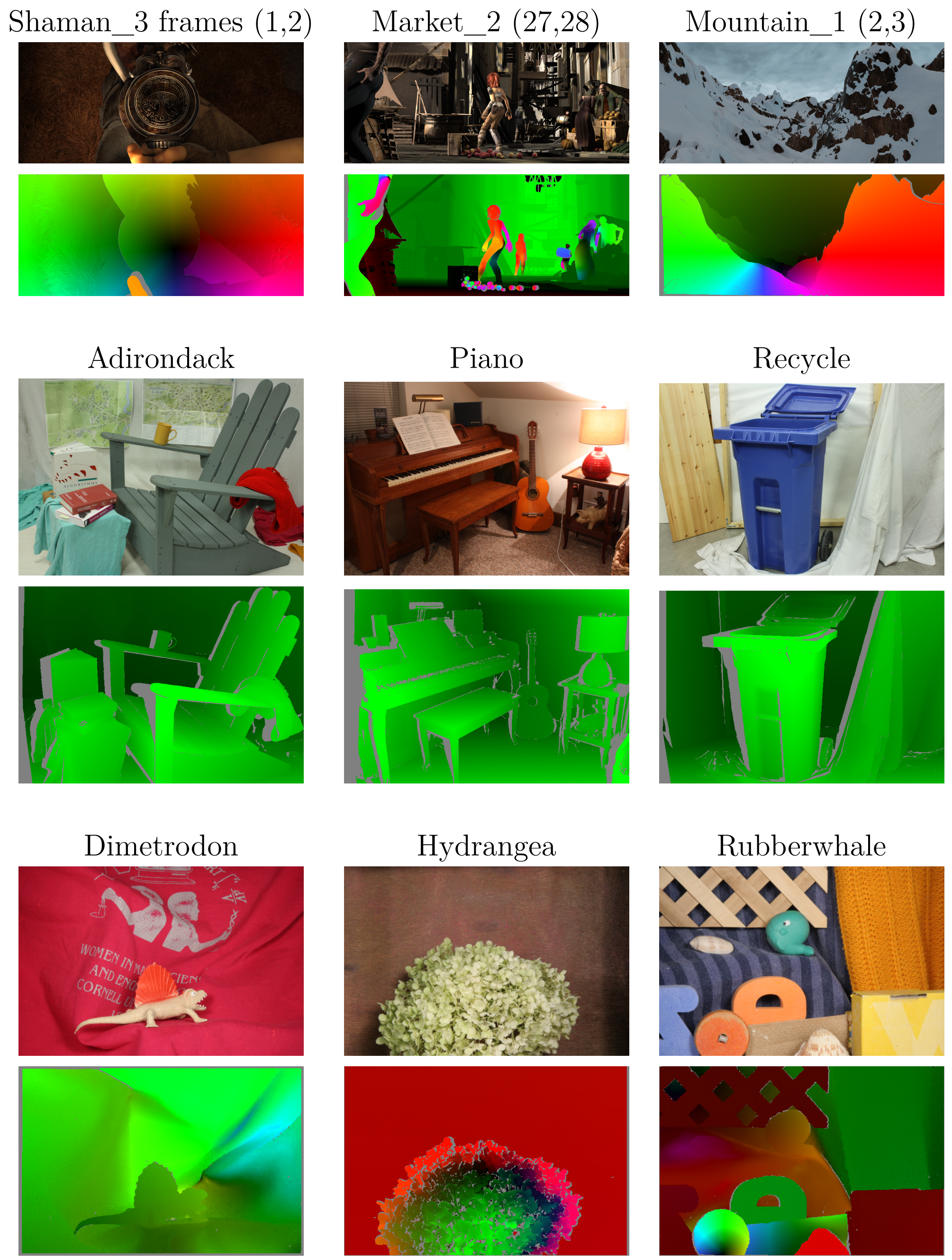


Figure 2.9: Image sequences selected for the experiments of this thesis (abbreviation *subset9*). First frame of each image sequence with corresponding flow visualisation below it. Grey regions in the flow fields indicate occlusions. **Top:** MPI Sintel subset [Butler et al., 2012], abbreviation *subset3sintel*. **Middle:** New Middlebury stereo subset [Scharstein et al., 2014], abbreviation *subset3newmiddle*. **Bottom:** Classic Middlebury flow subset [Baker et al., 2011], abbreviation *subset3middle*.

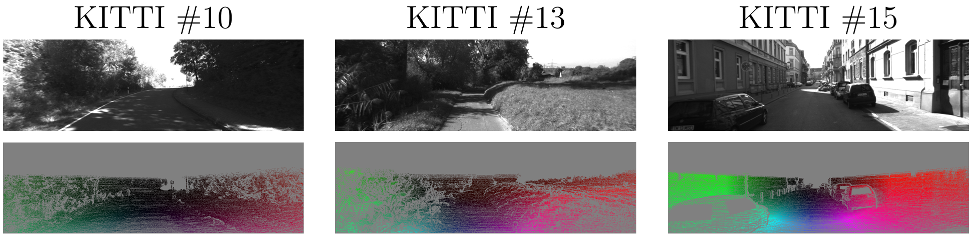


Figure 2.10: KITTI image sequences selected for the experiments of this thesis (abbreviation *subset3kitti*). Unfortunately, the KITTI sequences are only available as grey value images. First frame of each image sequence with corresponding flow visualisation below it. Grey regions in the flow fields indicate occlusions or missing measurements.

scale strategy, and k represents the size of the neighbourhood of patch-based descriptors. Effectively, this means that there are two important parameters to be chosen: the weight α of the coupling term, and the weight β of the auxiliary variable regularisation term. In Table 2.5, some optimal combinations of α and β are shown. Smaller values for ε_c lead to sharper discontinuities, which is reflected by more accurate flow fields on three of four image sets. The reason that the KITTI imagery does not profit from sharp flow discontinuities are the sparse ground truth measurements, where smooth and affine solutions are more important than realistically reproduced flow discontinuities. Thus, we will perform all further experiments with a standard choice $\varepsilon_c = 0.01$, except for KITTI sequences where we choose $\varepsilon_c = 0.5$.

2.5.2 General Comparison

Let us start our evaluation with a general comparison of all discussed features from this Section 2.2. To this end, we choose the mixed set of 9 image sequences depicted in Figure 2.9 (*subset9*). Since the KITTI benchmark does not provide colour imagery, we do not consider it for this experiment, as several of the discussed features such as the normalisation methods or the hue channel cannot be applied for scalar valued images. We apply the discussed optimisation procedure in 2-D to find optimal values for the coupling weight α as well as the smoothness weight β for each descriptor, and we present the results Table 2.6. Please note that we do not consider combinations of features, as there would be too many possible combinations.

As a result of this general comparison, we can see that the order-based descriptors do perform quite well on the selected images. But also the classical gradient constancy can still keep up - especially in view of the fact that also

Table 2.5: Optimal combinations of the weight of the coupling term α and the regularisation parameter of the auxiliary coupling variables β with varying ε_c. When changing ε_c, both regularisation parameters have to be adapted. The applied optimisation strategy is described in Section 2.5.1.

Image sequence	ε_c	α	β	avg. error
kitti10	0.5	0.003	243.787	**11.82 %**
	0.01	0.063	4.601	13.06 %
	0.0001	4.326	0.073	13.52 %
subset3middle	0.5	0.044	4.119	0.14 px
	0.01	0.084	847.625	0.12 px
	0.0001	7.642	151.623	**0.11 px**
subset3newmiddle	0.5	0.065	375.785	4.15 px
	0.01	0.072	535.233	3.38 px
	0.0001	7.16	2.813	**3.32 px**
subset3sintel	0.5	0.027	19.133	0.25 px
	0.01	0.126	847.625	0.19 px
	0.0001	11.517	6.466	**0.18 px**
subset9	0.5	0.019	847.625	1.53 px
	0.01	0.095	348.688	1.24 px
	0.0001	7.16	2.813	**1.22 px**

its computational requirements are very low (number of features per colour channel $m = 2$). Furthermore, the colour-based features *norm-a*, *norm-g* and *hue* perform clearly below average. This indicates that the occurring type of changes cannot be well described by a multiplicative function, which these features are invariant against. Moreover, the hue-feature only seems to work well in a combination with other constraints, as demonstrated by Zimmer et al. [2011b].

2.5.3 Behaviour under Synthetic Perturbations

Concerning synthetic perturbations, we consider again the mixed set of 9 image sequences.

Invariance to γ Changes

Our first experiment examines the behaviour of the proposed features under monotonically increasing intensity changes. Assuming the red, green and blue values of the input images to lie in the interval $[0, 255]$, we apply a

Table 2.6: General signature comparison. The table depicts for each of the discussed features the average endpoint error (AEE). The optimal value of the coupling weight α and the smoothness parameter β has been obtained with the strategy described in Section 2.5.1. For the patch-based descriptors, the patch size was fixed to 13 for this experiment. Smallest errors are highlighted in boldface font.

Signature type	AEE
rank	1.852
census	1.326
comp-rank	**1.246**
comp-census	1.269
intensity	3.760
centr-diff	1.694
correlation	7.367
tern-census	1.463
mod-census	1.561
hue	10.790
norm-a	9.258
norm-g	8.192
logderiv	1.697
texture	1.609
laplace	1.780
hess	1.398
gradmag	1.600
gradient	1.556

γ-correction to the second frame of each sequence:

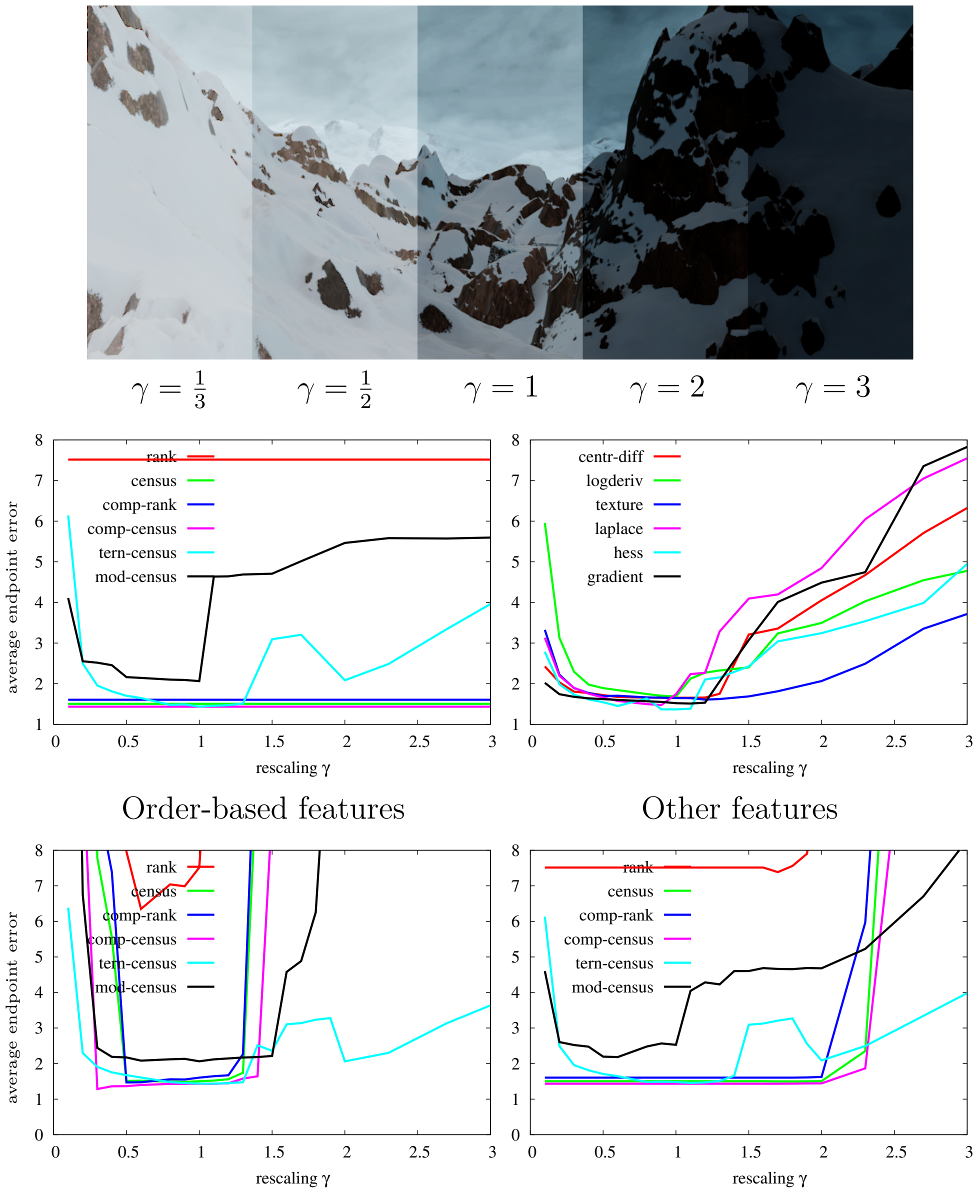
$$f_\gamma(\boldsymbol{x}) := 255 \cdot \left(\tfrac{1}{255} f(\boldsymbol{x})\right)^\gamma \,. \qquad (2.127)$$

The top part of Figure 2.11 shows the result of such rescalings for one of the used images. In this experiment, we optimise for each descriptor the coupling and smoothness weight for $\gamma = 1$ (c.f. Table 2.6), and use this value to compute the average error for the 9 test images. The corresponding plots are depicted in the middle row of Figure 2.11, where two plots are shown. In the left plot, the performance of the ordering-based descriptors is depicted, and the right plot shows the behavior of the classical features. As one can see, the classical ones, which are also in theory not invariant against such γ-rescalings are severely affected. However, for instance the texture part of the structure-texture decomposition does perform clearly better than its competitors – especially for severe rescalings. Furthermore, the ordering-based signatures also behave as expected. The rank, census, complete rank and complete census transforms are *fully invariant* against γ-rescalings (they compute perfectly the same results), whereas the modified and ternary census transforms loose accuracy for drastic rescalings as expected. Considering the behavior of the ternary census transform if further detail, another disadvantageous property becomes visible. As soon as the value of γ becomes larger than 1, the error jumps up. By this rescaling, many intensities come closer to each other. This can change the signatures, because 1 and -1 digits can be changed into 0's. Hence, the modified census transform is highly vulnerable against rescalings of this type.

The gamma corrections in the mentioned middle row of Figure 2.11 are performed with floating point accuracy. However, to simulate the image acquisition process in a digital camera more realistically, we also want to take the subsequent quantisation into account. Such a quantisation is a highly nonlinear post-processing step, which, in particular, can affect the intensity order. In our next experiment, we thus simulate the quantisation at two different bit depths: Most often, digital images are quantised with 8 bit. As can be seen in Figure 2.11, the theoretically unconditional invariance does not hold for any transform in this case. However, many cameras offer RAW sensor data that is quantised with 12 bit. Also many CMOS sensors and high-quality webcams offer a capture mode with such an increased dynamic range. Thus, we have also requantised the adjusted images with 12 bit, and analysed those results. From Figure 2.11, one can see that these 4 bit more tonal resolution are in practice enough to shift the point at which results deteriorate considerably to higher γ-values.

Sensitivity to Noise

It is clear that noise can have a severe effect on all order-based signatures. In fact only one changed pixel value can change the rank of each pixel in a patch. Thus, the question how vulnerable the signatures are in practice is interesting. To analyse this, we again consider the set of 9 mixed image sequences (*subset9*) and perturb them with additive zero-mean Gaussian noise. We consider two cases: a relatively low amount of noise ($\sigma_n = 2$), as well as one case of severe noise ($\sigma_n = 20$). The outcome of this experiment is shown in Table 2.7. Each single absolute error measurement in this table has been optimised separately and depicts the average endpoint error of the nine sequences. Moreover, we have also computed the relative error of each measurement compared to the noise-free result (abs-column with $\sigma = 0$). This relative value allows to compare how severely different features



Order-based features

Other features

8 bit quantisation

12bit quantisation

Figure 2.11: Behaviour under γ changes. **Top:** Exemplary γ-rescalings of the *mountain_1* test image [Butler et al., 2012]. For the centre field, $\gamma = 1$ means no change. As one can see, the interval $[\frac{1}{3}, 3]$ covers the whole range of realistic γ-rescalings. **Bottom:** The plots show the average accuracy of the features under γ variations of the second frames. **Centre left plot:** Behaviour of ordering-based features. **Centre right plot:** Other descriptors. All features that have not been included in any plot here failed completely under γ-rescalings. **Bottom left:** γ change followed by 8-bit quantisation. Practically all transforms loose their invariance. **Bottom right:** With 12-bit quantisation.

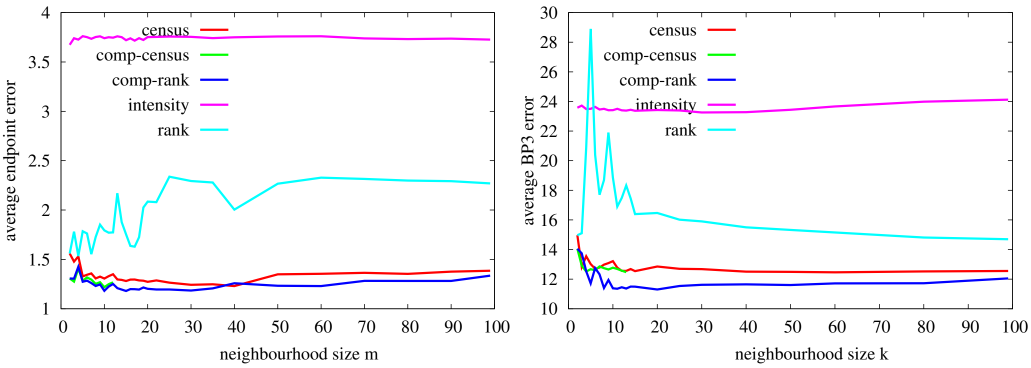
Table 2.7: Behaviour of the discussed features under additive white Gaussian noise of varying standard deviations σ_n. The relative error of each descriptor describes how much the accuracy changes if noise is present.

Signature	Average Endpoint Error (AEE)					
	$\sigma_n = 0$		$\sigma_n = 2$		$\sigma_n = 20$	
	abs	rel	abs	rel	abs	rel
rank	9.55	1.00	6.59	0.69	10.78	1.13
census	1.54	1.00	2.29	1.48	10.76	6.98
comp-rank	**1.36**	1.00	2.12	1.56	3.14	2.31
comp-census	1.46	1.00	2.23	1.52	4.50	3.08
intensity-pure	5.07	1.00	5.10	**1.01**	6.03	1.19
centr-diff	1.85	1.00	1.91	1.03	2.89	1.56
tern-census	2.83	1.00	2.07	0.73	5.65	2.00
mod-census	4.58	1.00	2.21	0.48	4.44	**0.97**
logderiv	2.17	1.00	2.26	1.04	3.38	1.56
texture	3.68	1.00	5.14	1.40	10.72	2.91
laplace	1.88	1.00	1.91	1.02	6.14	3.26
gradmag	1.90	1.00	2.11	1.11	7.20	3.79
gradient	1.54	1.00	**1.63**	1.06	**2.83**	1.84

are affected under noise: As could be expected, the pure rank is very robust especially under zero-mean noise. Apart from that, this experiment proves that the complete rank and complete census signatures do not fail under noise. On the other hand, the classical census transform looses accuracy for the case $\sigma = 20$.

2.5.4 Component Evaluation

In the following, the goal will be to evaluate single components of our framework.



subset9 KITTI training images #10-#19

Figure 2.12: Behaviour of the average error under varying neighbourhood sizes. Due to its high dimensionality, we did not test larger neighbourhoods for the complete census transform.

Neighbourhood Size

Our next experiments are devoted to appropriate neighbourhood sizes for the proposed patch-based descriptors, and are depicted in Figure 2.12. Note that all depicted measurements here are the result of the described parameter optimisation strategy. First, we optimise the parameters of our method for the set of 9 mixed training image sequences for many neighbourhood sizes (*subset9*). The results of this experiment are shown in the left part of Figure 2.12. First, we consider patch-based intensity matching, which is very similar to the Combined Local-Global method of Bruhn et al. [2005] (with a *hard* window instead of a Gaussian). Here, the best patch size is 1, i.e. the standard point-based grey value constancy is best. This is in line with results reported by Zimmer et al. [2011b]. Next, the pure rank transform attains the best error for the relatively small patch size of 4. For such small sizes the rank seems to be a significant information. With growing size, results get worse for the rank. The error plot for the original census transform attains its minimal value at a neighbourhood size of 40. This is also the only position where its average error is below the error of the complete rank transform. For all other measurements, the complete rank and complete census transforms outperform the original CT. In detail, the CRT performs best for a patch size of 10, with only little variation over the whole tested range of neighbourhood sizes. For the CCT, a 15-pixel neighbourhood is optimal. We repeated this experiment on a different set of images to test if these findings are also valid in other situations. The results for the a selection of ten KITTI image sequences are shown in the right plot of Figure 2.12. Again, our complete

Table 2.8: Comparison of different regularisation strategies on different image sets. The table depicts the average accuracy of different configurations of our framework for different image sets in terms of the $BP3$ or AEE. The subquadratic coupling scheme leads to very good result consistently.

Benchmark	Coupling	1st order	2nd order
KITTI	subquadratic	21.3%	**11.9%**
	quadratic	21.4%	12.2%
Middlebury	subquadratic	**0.116 px**	0.117 px
	quadratic	0.158 px	0.151 px
Sintel	subquadratic	0.215 px	**0.189 px**
	quadratic	0.306 px	0.281 px
New Middlebury	subquadratic	**3.226 px**	3.405 px
	quadratic	3.505 px	21.113 px

signatures outperform the classical census consistently.

Regularisation Term

As described in Section 2.4.2, the discussed coupling framework allows to resemble several different regularisation strategies. Our upcoming experiments will evaluate how different regularisation schemes affect the accuracy of the computed flow. To this end, we consider 4 regularisation settings: On the one hand it is possible to enable or disable the coupling scheme, and on the other hand we can have quadratic or subquadratic coupling and coefficient regularisation terms. In Table 2.8 we present the results of these configurations on four different image sets. First of all, one can clearly state that the sub-quadratic schemes outperform the quadratic ones consistently. For both Middlebury benchmarks, the sub-quadratic 1st order scheme performs best, but the 2nd order scheme is only slightly worse. This can be explained by the dominating type of motion in those benchmarks, which is mainly fronto-parallel and requires piecewise constant flow fields. Thus, the second-order regulariser offers a degree of freedom that is not necessary here. Regarding the KITTI benchmark, the opposite is the case. Here, the divergent flow fields resulting from the driving scenario are mainly affine and not piecewise constant. As one can see from the results in Table 2.8, the accuracy of the second-order scheme is much higher.

Also our next experiment substantiates that the second-order coupling model is superior to the first-order model. To this end we choose the same set of images as selected for the *GCPR 2013 - Special Session on Robust Optic*

Table 2.9: Behaviour in real-world scenarios. Errors are given in terms of the *BP3* measure, i.e. the percentage of pixels having a Euclidean error larger than 3.

KITTI image sequence:	#11	#15	#44	#74	average
Zimmer et al. [2011b]	37.3	32.3	23.2	62.9	38.9
Bruhn and Weickert [2005]	33.9	47.7	32.4	71.4	46.7
census(1st order)	36.5	28.6	28.5	63.8	39.4
comp-rank (1st order)	29.8	22.8	22.6	61.5	34.2
comp-rank (2nd order)	**22.9**	**13.5**	**15.2**	**56.3**	**27.0**

Table 2.10: Runtimes of our framework using the complete rank transform. The time per pixel-value is approximately the same for all images.

Image	Resolution	Runtime
Middlebury Rubberwale	584 × 388 × 3	80.6 s
Sintel ambush2	1024 × 436 × 3	162.8 s
Kitti #15	1241 × 376 × 1	75.9 s
New Middlebury Adirondack	718 × 496 × 3	138.3 s

Flow [3]. Table 2.9 summarises the obtained results. As reference, the numbers for the method of Zimmer et al. [2011b] and Bruhn and Weickert [2005] are taken from the website of this special session. The method of Bruhn and Weickert [2005] is particularly interesting to compare, since its regularisation term coincides with our first-order sub-quadratic strategy. Additionally, we depict optimised results with the census transform for comparison. As one can see, the complete rank transform consistently outperforms the competing methods.

Further Model Evaluation

Runtimes. Table 2.10 summarises the runtime of our method on various image sequences. As evaluation system we use a Macbook Air with Intel Core i5 processor (1.3 GHz). This experiment substantiates that in practice the runtime of our framework is linear in the number of pixels to be processed.

Normalisation. In Section 2.4.1, we discussed the data term normalisation. Here, we want to evaluate if this concept is also beneficial for our order-based complete rank feature. The results are depicted in Table 2.11.

[3] `http://www.dagm.de/symposien/special-sessions/`

Table 2.11: Influence of data term normalisation. As can be seen, this concept increases the accuracy on all test sequences consistently.

Image set	Without Normalisation	With Normalisation
subset3middle	0.117 px	**0.116 px**
subset3newmiddle	3.405 px	**3.135 px**
subset3sintel	0.189 px	**0.174 px**
kitti10	13.217 %	**11.470 %**

Table 2.12: Effect of different robustification strategies.

	Joint Rob.	Separate Rob.
comp-rank	1.268 px	**1.248 px**
comp-census	**1.275 px**	1.309 px

Obviously, the data term normalisation is also beneficial in the case of matching complete rank signatures.

Joint vs. Separate Robustification. We have discussed the different metrics that are applicable to the complete rank and complete census signature differences and their implications in Section 2.3. Let us now analyse how the different robustification type affect the accuracy of flow computations. To this end we have optimised the average endpoint error for the set 9 mixed image sequences (*subset9*) with separate and joint robustification, corresponding to the L_1 and L_2 norm, respectively. Since the complete census signatures become very large with increasing neighbourhood size, we have fixed the patch size to $k = 5$ for this experiment. The results are shown in Table 2.12. These results indicate that the computationally much more efficient joint robustification performs comparably well in practice.

2.5.5 Real-world Example

In this experiment we demonstrate the matching capabilities of the complete rank transform in a scenario with extreme appearance changes. The task is to recover the optic flow between the two raw images provided by a Microsoft Kinect[4]. This device is equipped with two sensors side-by-side: an infra red (IR) camera captures a structured light pattern that a built-in IR-projector casts into the scene. Simultaneously, a standard CMOS sensor for the visible light spectrum provides a colour image of the scene. In this experimental

[4] `http://www.xbox.com/en-US/xbox-360/accessories/kinect`

Table 2.13: Quantitative comparison of the rank (RT), census (CT) and complete rank transform (CRT) on the Middlebury training images. Numbers are average endpoint errors ×10^{-1}.

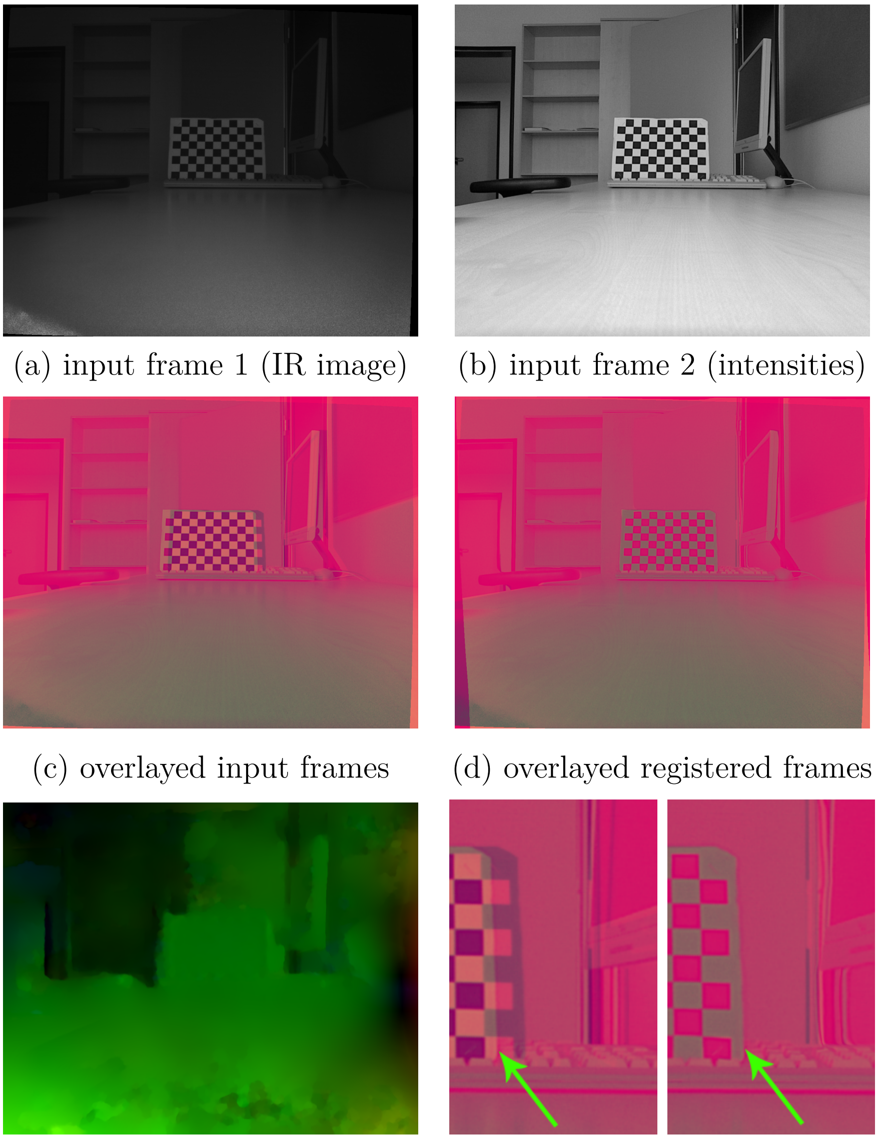
	rw	dime.	gr2	gr3	hydr.	urb2	urb3	yos	**avg**
RT	1.11	0.92	1.91	7.64	1.91	4.57	10.3	2.11	3.81
CT	1.02	0.90	1.69	6.46	**1.47**	3.78	8.19	1.69	3.16
CRT	**1.00**	**0.76**	**1.54**	**5.85**	1.58	**3.24**	**5.29**	**1.50**	**2.60**

setup, we switch off the IR-projector such that the IR-camera records an unperturbed (but dark) image of the IR light in the scene. The two top images of Figure 2.13 show a pair of two such captured images. The IR image contains quite large amounts of noise. Nevertheless, the results in this figure show that a reliable matching is possible using the complete rank transform.

2.5.6 Comparison to State-of-the-Art

After having evaluated various strategies inside our framework, let us now come to the question how our framework performs in comparison to other state-of-the-art methods. Luckily, there are several public benchmark systems for optic flow available that make an objective and fair comparison possible.

Middlebury Benchmark. First, we assess the error rates on the Middlebury training images in Table 2.13, where the results from our conference contribution [Demetz et al., 2013] are reproduced. As also noted by Vogel et al. [2013], the image sequences of that benchmark exhibit mainly frontoparallel motion, so use our first order regularisation term here. Furthermore, note that the Middlebury sequences are also less demanding with respect to illumination changes. Hence, the goal of this experiment is to show that also under normal lighting conditions reasonable flow fields can be obtained with our CRT-based data term. Furthermore, we prove with this experiment that our CRT is also in this setting generally preferable over the rank and census transform. Again, for each signature type, the regularisation parameter α has been optimised and then kept constant over all images. For the sake of completeness, we also evaluated our method with TV regularisation on the testing images of the public Middlebury benchmark [Baker et al., 2011]. Also these test sequences exhibit almost no illumination changes or other scenarios that our highly invariant descriptor is designed for, so we cannot expect



(a) input frame 1 (IR image) (b) input frame 2 (intensities)

(c) overlayed input frames (d) overlayed registered frames

(e) reconstructed flow zooms into (c) and (d)

Figure 2.13: Kinect registration example. **Tow row:** Input images captured with a Kinect RGB-D camera. **Centre row:** Colour Visualisation of misalignments. The original input frames have been combined into different channels (image (c)). One can clearly see the large misalignments. Image (d) shows the same visualisation after back-registration of the second frame. The matching was successful. **Bottom row:** Recovered flow (image (e)) and magnifications of (c) and (d). The arrows highlight the alignment before and after the back-registration.

Table 2.14: Quantitiative Middlebury results in terms of the average endpoint error (AEE) $\times 10^{-1}$.

Army	Mequon	Schefflera	Wooden	Grove	Urban	Yosemite	Teddy
1.1	2.4	5.0	2.3	8.6	6.0	1.2	7.9

Table 2.15: Detailed results of our method on the KITTI benchmark.

Error	**Out-Noc**	**Out-All**	**Avg-Noc**	**Avg-All**
2 pixels	8.84 %	15.38 %	2.0 px	3.9 px
3 pixels	6.71 %	12.09 %	2.0 px	3.9 px
4 pixels	5.68 %	10.23 %	2.0 px	3.9 px
5 pixels	5.01 %	8.97 %	2.0 px	3.9 px

top-ranking results on this benchmark. Nevertheless, it turns out that our prototypical variational model can in fact keep up with its nearest competitors: Our method ranks on 59.4th average rank. That is between the method of Brox et al. [2004] (avg. rank 67.5) and the much more advanced method by Zimmer et al. [2009] (avg. rank 40.7)[5]. These results are remarkable in the sense that they prove our invariant data term to include hardly less information than the combined grey value and gradient information of [Brox et al., 2004; Zimmer et al., 2009].

KITTI Benchmark. Definitely the most interesting benchmark for our method is the KITTI Vision Benchmark Suite [Geiger et al., 2012]. It provides a huge amount of image sequences captured from a driving car, along with corresponding ground truth flow fields that are acquired with a laser scanning technique. These image sequence contain challenging illumination conditions, such as camera adjustments, light inter-reflections in the windshield, driving into or out of shadows, etc., and are thus a perfect testbed for our needs. We used our described optimisation strategy on the 194 training sequences with a neighbourhood size of 13 to find the optimal value of the coupling weight as $\alpha = 0.01$, and submitted the flow fields for all 195 test sequences to the benchmark website[6]. Detailed results are shown in Table 2.15 where the bad pixel error measure is depicted for various thresholds and for ground truth information in *all* and *non-occluded* regions.

Additionally, we present in Table 2.16 a comparison of our method to the other participants of the benchmark. In this table, we only include published

[5] Rankings as of September 28th, 2015

[6] `http://www.cvlibs.net/datasets/kitti/`

competing methods that provide a 100% dense ground truth and that consider the pure two-frame optic flow setup without stereoscopic assumptions. Methods that exploit such additional assumptions loose general applicability, because they are likely to fail e.g. in the presence of independently moving objects. The table shows the average error of each method with respect to the bad pixel 3 error (BP3) and the average endpoint error (AEE). Moreover, the evaluation is performed on all pixels (occ), or only on pixels that are visible in both frames (noc). The small numbers to the right of the measured errors indicate the ranking of the corresponding method among all competitors according to this measure.

As one can see, two entries in this table belong to methods discussed in this chapter. With the complete rank transform and second order regulariser, our method [Demetz et al., 2015] (CRT w. TGV) clearly belongs to the top-ranking ones on this benchmark. Particularly when considering the ground truth information in all image regions (occ), our method ranks third in both error measures. One reason for this is the second order regulariser that is well suited for the typical divergent motion patterns of the KITTI benchmark. However, regarding the performance of the method of Ranftl et al. [2012], which also incorporates a TGV-based regulariser, the benefits of our descriptor become apparent. The other highlighted list entry [Demetz et al., 2013] (CRT w. TV) shows the performance of our earlier conference publication with only first order regularisation.

2.6 Summary

This chapter was devoted completely to invariance-based descriptors for pattern matching. We approached this topic from the theoretical as well as the practical point of view.

First, we structured all invariant features into different mathematical classes of increasing invariance and gave a broad overview of existing invariant features from the literature. Then we concentrated on the extreme class of morphological invariance, in which all descriptors fall that remain unchanged under any monotonically increasing rescaling. Here, we introduced two novel descriptors: the complete rank transform (CRT) and the complete census transform (CCT). The unique property of these two descriptors is that they carry as much image information as possible in this class.

Second, we discussed appropriate distance metrics for the discussed features. In most cases the Euclidean distance is a natural and well-suited choice. For the binary and order-based descriptors however, this was not directly obvious.

Table 2.16: Top KITTI benchmark results as of March 31st, 2015. Only pure two-frame dense optic flow methods are shown. All our methods are highlighted in red.

Method	**BP3** [%]				**AEE** [px]			
	noc		**occ**		**noc**		**occ**	
[Ranftl et al., 2014]	**5.93**	**1**	11.96	2	1.6	4	3.8	3
[Wei et al., 2014]	6.03	2	13.08	5	1.6	4	4.2	5
[Braux-Zin et al., 2013]	6.20	3	15.15	7	**1.5**	**1**	4.5	6
[Demetz et al., 2014]	6.52	4	**11.03**	**1**	**1.5**	**1**	**2.8**	**1**
[Demetz et al., 2015] (CRT w. TGV)	6.71	5	12.09	3	2.0	8	3.9	4
[Vogel et al., 2013]	7.11	6	14.57	6	1.9	7	5.5	8
[Weinzaepfel et al., 2013]	7.22	7	17.79	8	**1.5**	**1**	5.8	9
[Rashwan et al., 2013]	7.91	9	18.90	13	2.0	8	6.1	10
[Mohamed et al., 2014]	8.67	10	18.78	12	2.4	11	6.7	13
[Timofte and Gool, 2015]	9.09	11	19.32	14	2.6	12	7.6	16
[Demetz et al., 2013] (CRT w. TV)	9.43	12	18.72	11	2.7	14	6.5	11
[Sun et al., 2014]	10.04	13	20.26	15	2.6	12	7.1	14
[Kennedy and Taylor, 2015]	10.22	14	18.46	10	2.0	8	5.0	7
[Sun et al., 2014]	10.49	15	20.64	16	2.8	15	7.2	15
[Hermann and Klette, 2013]	10.74	16	22.66	17	3.2	17	12.2	17
[Ranftl et al., 2012]	11.03	17	18.37	9	2.9	16	6.6	12

Thirdly, in order to use the discussed features for pattern matching in practice, we presented a very general variational framework for optic flow computation, in which all discussed descriptors can be used without further modifications. The data term of this functional penalises deviations of the features at corresponding positions. The regularisation term demands second-order regularity. Due to a coupling strategy, we could realise this with first-order derivatives of the unknowns only. Our basic model is non-convex and non-linear. To find a minimiser, we discussed a multi-scale minimisation strategy and explained our final numerical scheme.

However, our theoretical classification could not answer one crucial question: *Which feature is best suited in practical situations?* To find its answer, we have performed an extensive evaluation of all discussed descriptors in the final part of this chapter. With our evaluation we tried to compare the various descriptors on a maximally fair basis. As could be expected, classical features such as the gradient that have proven their usefulness in many methods do in fact perform quite well. Moreover, the experiments showed that many features lead to inferior results because they either discard too much (for instance the rank), or not enough information (e.g. normalisation strategies). Thus, the best tradeoff is difficult to find. However, overall, the theoretically maximal information content of our complete signatures is also reflected experimentally in highly accurate results. In fact, we achieved the highest accuracy with the complete rank transform, even if only minor illumination changes were present.

Chapter 3

Change Estimation

In the previous Chapter 2, the main paradigm was to *discard* fragile information through invariances. The idea was to find features that remain constant under changing illumination, and that *survive* all changes of appearance. We have seen that depending on the expected type of changes, this strategy can be successful and allows to estimate correspondence accurately. However, we also have seen that it is easily possible to employ too strong invariances: If too much image information is discarded, the accuracy of the estimated flow is affected. This is a general problem, for instance in the simplest case that no illumination changes are present in an image sequence, any invariance can potentially deteriorate the results.

For this reason, we want to approach the problem of uncontrolled lighting from a completely different direction in this chapter. Now, the main paradigm will be *compensation*: We want to compensate the observed intensities for the occurring appearance changes, such that after the compensation, the grey value constancy holds again. In order to compensate, we have to *estimate* appearance changes explicitly, and our idea is to do this *jointly* with the optic flow. Hence, we will consider the changes as an additional unknown to be estimated. The basis for this chapter is the publication at the European Conference on Computer Vision [Demetz et al., 2014].

As we have seen, already the original optic flow problem is ill-posed, and regularisation is necessary. With this additional degree of freedom, things don not get easier. However, we exploit the following observation: in the real world, appearance changes usually are *not arbitrary.* For instance, if one object causes a drop shadow onto another one, a certain larger region will be affected smoothly by this shadow. In this region, the darkening of the image content will be similar everywhere, see also Figure 3.1.

Practically, we have to estimate the so-called *brightness transfer function* (BTF) that allows to rescale the intensity of pixels in the first frame

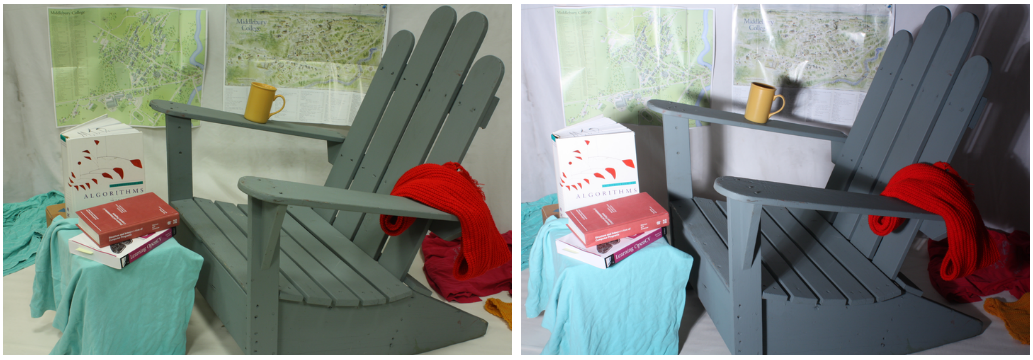


Figure 3.1: Exemplary image sequence [Scharstein et al., 2014]. Between the first and second frame, the lighting in the scene changes drastically by an additional spotlight. The Adirondack chair and the cup cast a drop shadow with sharp shadow boundary. Outside the shadow region, the full brightening illumination change is visible. One can expect that all pixels within the shadow region are less affected by the spotlight.

to the illumination setting of the second one. However, instead of the ad-hoc parametrisations proposed by Mukawa [1990] and Ayvaci et al. [2012] (additive), and Gennert and Negahdaripour [1987] (affine), we aim for more general and suitable representation of the BTF in terms of a specifically learned basis. We obtain this basis by performing a principal component analysis (PCA) of the appearance changes in training data. Contrary to the invariance-based data terms from the previous chapter, where the invariance can only be imposed globally, our ansatz will be to perform the compensation locally. Thus, the BTF we estimate will vary in space. However, as already mentioned, this variation in space cannot be arbitrary, thus, we develop a suitable regularisation strategy and include it into our variational model. This allows to separate real illumination changes from motion-induced brightness variations.

3.1 Related Work

The idea of estimating illumination changes jointly with optic flow is not new. Several – mainly older – methods in the literature follow this idea. On the one hand, there are approaches that seek to estimate a single *global* brightness transfer function to identify problematic image regions [Mann et al., 2003; Dederscheck et al., 2012]. On the other hand, there are techniques that embed the classical brightness constancy assumption into a parametrised *local* illumination model for which the coefficients are jointly estimated. Such local models include simple additive terms [Cornelius and Kanade, 1984; Mukawa, 1990], affine illumination models [Gennert and Negahdaripour, 1987; HaCohen et al., 2011; Negahdaripour and Yu, 1993; Fouad et al., 2009], as well as more complex brightness models derived from physics [Haussecker and Fleet, 2001]. Recently, HaCohen et al. [2011] combined local and global ideas: While a local affine model allows to estimate the correspondences in a PatchMatch-like approach [Barnes et al., 2010], the information is finally condensed to a single global transfer function.

Also the research field on so-called *intrinsic images* is related to the present idea. In this context, a captured image is usually considered to be the product of surface reflectance (albedo, colour) and illumination. The task is to estimate both factors. Such representations are valuable for many applications. For instance, the illumination part plays an important role for shape-from-shading methods [Horn and Brooks, 1989]. The reflectance part is free of illumination and a well suited input for instance for segmentation methods. The research on this topic has very long tradition. It goes back to contributions of Horn [1974] and Barrow and Tenenbaum [1978] who coined the expression *intrinsic image*. The two latter methods closely refer to the Retinex theory of Land and McCann [1971]. A slightly easier problem is approached by Weiss [2001] who assume to be given a sequence of images in which only the illumination changes. The recent work of Rother et al. [2011] addresses the intrinsic image decomposition problem with the assumption that the reflectance part only contains colours from a sparse set of so-called *basis colours*. The work of Chen and Koltun [2013] incorporates depth measurements as an additional cue for the accurate estimation of the illumination part. In general, the goal of intrinsic image methods is to separate the reflectance from illumination. In contrast, we are only interested in illumination *changes*, i.e. we are not interested in removing all drop shadows, but only in those regions where the shadow changes.

Related Work on Basis Learning. Apart from the aforementioned techniques that jointly estimate illumination changes and optical flow, a few more related works are worth mentioning. On the one hand, there are methods that address the estimation of camera response or brightness transfer functions, mainly in the context of HDR imaging. Such approaches include the work of Grossberg and Nayar [2002] who propose to compute the brightness transfer function via histogram specification. Also the papers of Debevec and Malik [1997] and Grossberg and Nayar [2004] consider the problem of estimating the camera response function, the latter one using learned basis functions. On the other hand, there are approaches that represent appearance changes with basis functions for illumination changes. First, Belhumeur and Kriegman [1998] neglect viewpoint changes and consider the question how a convex object can look under varying illuminations. Using PCA, they show that the space of possible images is a cone with limited dimension. Mainly in the context of face tracking, Hager and Belhumeur [1998] compute a basis of a template image to be tracked and incorporate this basis in a local optic flow method to account for changes in appearance. Similarly, the work on iconic changes by Black et al. [2000] also computes a basis of images. The work of Tieu and Miller [2002] comes closest to our formulation. There, a 3-D basis is estimated via PCA to represent so-called *colour eigenflows* that allow to transfer RGB colour vectors from the first to the second image. Finally, there exist a few optical flow methods that make use of spatial or temporal basis functions to model the flow. This applies to the approaches by Nir et al. [2008] on over-parametrised optical flow and Garg et al. [2013] on temporal tracking of non-rigid objects with subspace constraints.

3.2 Variational Model

We stick to our notation from the previous chapter, and consider a sequence of two images $\boldsymbol{f}_i : \Omega \to \mathcal{R}^{n_c}$, $i \in \{1, 2\}$, defined on the rectangular image domain $\Omega \subset \mathbb{R}^2$. Furthermore, we denote the optic flow field by $\boldsymbol{u} = (u, v)^\top : \Omega \to \mathbb{R}^2$ and parametrise the illumination changes by a coefficient field

$$\boldsymbol{c} : \Omega \to \mathbb{R}^{n_b} . \tag{3.1}$$

Then, inspired by the basic approach of Cornelius and Kanade [1984], we propose to jointly compute the optical flow and the illumination changes as minimiser of an energy functional with the following structure:

$$E(\boldsymbol{u}, \boldsymbol{c}, \boldsymbol{a}, \boldsymbol{b}) = \int_\Omega \Big(D(\boldsymbol{u}, \boldsymbol{c}) + \alpha \cdot R_{\text{flow}}(\boldsymbol{u}, \boldsymbol{a}, \boldsymbol{b}) + \lambda \cdot R_{\text{illum}}(\boldsymbol{c}) \Big) \, \mathrm{d}\boldsymbol{x} . \tag{3.2}$$

This functional is closely related to our functional of Equation (2.57) that was the central quantity to be minimised in the previous chapter. However the functional (3.2) now consists of three terms: As before, the data term D relates the two input images via the optical flow and the parametrised illumination changes (in terms of coefficients). The flow regularisation term R_{flow} coincides with the regularisation term (2.73) from the previous chapter and encourages a piecewise affine flow field. The novel coefficient regularisation term R_{illum} assumes the coefficient fields to be piecewise smooth. The two positive parameters α and λ allow to adjust the influence of the two smoothness terms. Let us now discuss these terms in detail.

3.2.1 Data Term

Unlike traditional data terms for optical flow estimation that explain brightness changes in the image sequence exclusively by motion, our data term considers changes in illumination as an additional source for brightness variations. In order to estimate these changes jointly with the motion, we make use of a *parametrised* brightness transfer function (BTF) which was originally proposed by Grossberg and Nayar [2004] in the context of photometric calibration for HDR imaging.

Parametrised Brightness Transfer Function

This function maps intensities of the first frame to the intensity level of the second frame. Given a set of n_b basis functions $\phi_j : \mathbb{R} \to \mathbb{R}$, we parametrise the corresponding brightness transfer function as follows:

$$\begin{aligned}\Phi(\boldsymbol{c}, f) &= \bar{\phi}(f) + \sum_{j=1}^{n_b} c_j \cdot \phi_j(f)\,, &(3.3)\\ &= \bar{\phi}(f) + \boldsymbol{c}^\top \boldsymbol{\phi}(f)\,, &(3.4)\end{aligned}$$

where $\bar{\phi} : \mathbb{R} \to \mathbb{R}$ is the mean brightness transfer function, and $\boldsymbol{\phi} : \mathbb{R} \to \mathbb{R}^{n_b}$ resembles all basis functions in one vector. As can be seen from Equation (3.4), independently of the shape of the basis functions, this parametrisation is linear in the weights $\boldsymbol{c} = (c_1, \ldots, c_{n_b})^\top$. Note that our concept of brightness transfer functions is closely related to the colour flows of Tieu and Miller [2002]. Such a colour flow, however, is a 3-D vector field describing the offset in RGB colour space that is added to the RGB colour value of the first image. Moreover, in the work of Tieu and Miller [2002], the linear coefficients corresponding to the weights $\boldsymbol{c}$ of our model are assumed constant over the whole image. This is not the case in our model.

Compensated Brightness Constancy Assumption

Let us now discuss how to embed this general model for brightness changes into a data term. To this end, we formulate the assumption that after the BTF has been applied to a grey value of the first frame, the resulting intensity should be equal to the corresponding intensity in the second frame at the displaced position. Such a data term thus reads:

$$D_{\text{bright}}(\boldsymbol{w}, \boldsymbol{c}) = \Psi\Big(\big(f_2\left(\boldsymbol{x} + \boldsymbol{w}\right) - \Phi\left(\boldsymbol{c}(\boldsymbol{x}), f_1(\boldsymbol{x})\right)\big)^2\Big)\,. \tag{3.5}$$

It is important to notice that the coefficient field $\boldsymbol{c} : \Omega \to \mathbb{R}^{n_b}$ is space-variant, thus this model is able to estimate different illumination changes in each position. If no appearance change has happened, the coefficients should parametrise the identity function. Note that depending on the particular choice of a basis, the identity is not represented by a zero coefficient vector. A related data term has been proposed by Hafner et al. [2014], where the camera response function is applied to the unknown radiances to establish a correspondence to the observed exposures.

Compensated Gradient Constancy Assumption

We can go further, and apply the compensation idea to the gradient constancy assumption. Consequently, the corresponding data term reads

$$D_{\text{grad}}(\boldsymbol{w}, \boldsymbol{c}) = \Psi\Big(\left\| \boldsymbol{\nabla} f_2(\boldsymbol{x}+\boldsymbol{w}) - \boldsymbol{\nabla}\Phi(\boldsymbol{c}(\boldsymbol{x}), f_1(\boldsymbol{x})) \right\|_2^2 \Big)\,, \tag{3.6}$$

and our final data term is a convex combination of the two assumptions with separate robustification with the weight $\nu \in [0, 1]$:

$$D(\boldsymbol{w}, \boldsymbol{c}) = \nu D_{\text{bright}}(\boldsymbol{w}, \boldsymbol{c}) + (1 - \nu) D_{\text{grad}}(\boldsymbol{w}, \boldsymbol{c})\,. \tag{3.7}$$

This combination is similar to the data term of Brox et al. [2004] and the separate robustification of grey value and gradient constancy assumption was proposed by Bruhn and Weickert [2005].

In both assumptions, the flow variables and illumination coefficients are intentionally distributed to different frames:

$$\underbrace{f_2\left(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x})\right)}_{\text{flow}} = \underbrace{\Phi\left(\boldsymbol{c}(\boldsymbol{x}), f_1(\boldsymbol{x})\right)}_{\text{illumination}}\,. \tag{3.8}$$

Because of this, the typical linearisation in the flow variables can be performed without having to apply the chain rule, and the resulting minimality conditions do not contain products of unknowns.

Moreover, at first glance, it may seem counter-intuitive to combine our explicit estimation strategy with a gradient constancy assumption that is invariant under additive illumination changes. However, the additional gradient constancy term supports the estimation at those locations where the coefficients cannot adapt or have not yet adapted perfectly to the illumination changes. This is for instance the case at the beginning of the estimation, when neither the flow nor the coefficients have converged to their final values yet.

3.2.2 Regularisation Terms

Compared to traditional variational optic flow methods, where no illumination coefficients have to be estimated, the present method has a lot more degrees of freedom. Especially the data term is highly under-determined: it is not clear how to distribute observed brightness changes between the motion field and the illumination compensation. Thus, spatial regularisation of both the flow variables and the illumination coefficients is indispensable. While the parametrisation in terms of basis functions already provides a meaningful representation given by the coefficient fields, the concrete modeling of both regularisation terms plays an important role in resolving this ambiguity. Let us now discuss how we model the two regularisers.

Flow Regularisation

In order to ensure a fair comparison between the invariance and estimation-based ideas we will stick to the second-order regularisation as presented in Section 2.4.2:

$$R_{\text{flow}}(\boldsymbol{u},\boldsymbol{a},\boldsymbol{b}) = \Psi\Big(|\boldsymbol{\nabla} u - \boldsymbol{a}|^2 + |\boldsymbol{\nabla} v - \boldsymbol{b}|^2\Big) \,+\, \beta \cdot \Psi\Big(|\boldsymbol{\mathcal{J}}\boldsymbol{a}|_F^2 + |\boldsymbol{\mathcal{J}}\boldsymbol{b}|_F^2\Big)\,. \quad (3.9)$$

This term can be seen as a regularised variant of the TGV^2 smoothness term [Bredies et al., 2010; Ranftl et al., 2012], and favours piecewise affine solutions. In our conference publication [Demetz et al., 2014], we employed a direct second-order regularisation term, namely a sub-quadratic penalisation of the Frobenius norm of the Hessian of both flow components u and v. For the sake of comparability with the results from the previous chapter, we refrain from this regulariser here.

Coefficient Regularisation

The smoothness assumption on the illumination coefficients is essential for our model. Without this term, the displacement field of the global minimiser of our energy would be the trivial *zero* displacement field (with energy zero): In each pixel a coefficient vector could be chosen that compensates the brightness change in that pixel (without motion) perfectly. It is only the smoothness constraint on the coefficients that prevents this degeneration, and is thus absolutely vital for this method to work.

In contrast to the flow regulariser that models a piecewise affine flow field, we basically assume that neighbouring pixels are subject to similar illumination changes, i.e. that the coefficients of the basis functions are piecewise constant. Additionally, discontinuities in the coefficient fields are assumed to be aligned with edges in the input images (e.g. shadow edges) [Nagel and Enkelmann, 1986]. Consequently, we follow the idea of Zimmer et al. [2011b] and employ the following anisotropic complementary regularisation term:

$$R_{\text{illum}}(\boldsymbol{c}) = \sum_{i=1}^{2} \Psi^i \Big(\sum_{j=1}^{n_b} \gamma_j \, (\boldsymbol{r}_i^\top \boldsymbol{\nabla} c_j)^2 \Big) \,, \tag{3.10}$$

where the two directions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2 = \boldsymbol{r}_1^\perp$ allow to adapt the smoothing direction locally across and along image edges, respectively. As proposed in Zimmer et al. [2011b], these directions can be derived as the eigenvectors of the so-called regularisation tensor. In our case, this tensor must be computed from the *photometrically uncompensated* first frame f_1 to ensure that brightness information related to illumination changes is *not* discarded. Moreover, all coefficient fields are regularised jointly with a single penaliser function per direction, since spatial changes of the brightness transfer function typically result in discontinuities in all coefficient fields. In this context, the derivatives of the coefficients have to be balanced with weights γ_j to reflect the different magnitude ranges of the coefficient fields. How we can estimate these weights together with the basis functions is discussed in Section 3.3. Finally, we have to define the penaliser functions. As suggested in Volz et al. [2011], we use the edge-enhancing Perona-Malik regulariser

$$\Psi^1_{\text{illum}}(s^2) = \varepsilon_c^2 \log(1 + s^2/\varepsilon_c^2) \tag{3.11}$$

as penaliser across edges (in $\boldsymbol{r}_1$-direction) [Perona and Malik, 1990], while we apply the edge-preserving Charbonnier regulariser denoted by $\Psi^2_{\text{illum}}(s^2)$ along them (in $\boldsymbol{r}_2$-direction), c.f. Equation (2.68).

3.2.3 Multi-Scale Minimisation

As in the previous chapter, we have developed an energy functional that is non-convex since the unknown optic flow field appears in the argument of the image function. Consequently, also in this chapter, our minimisation follows a very similar Gauss-Newton-type [Nocedal and Wright, 2006] multiscale strategy as described in detail previously (c.f. Section 2.4.3). Thus, let us now only highlight the main extensions and differences to the already presented strategy.

Although the illumination coefficients already now only appear linearly inside the data term from Equation (3.5), we handle them analogously to the flow variables, and apply a splitting into a known part $\boldsymbol{c}^\ell$ and an unknown part $\boldsymbol{dc}^\ell$ on each scale. Essentially, we minimise the following functional on each scale ℓ:

$$\begin{aligned} E^\ell(\boldsymbol{du}^\ell, \boldsymbol{dc}^\ell, \boldsymbol{a}, \boldsymbol{b}) \;=\; \int_\Omega \Big(\, & D^\ell(\boldsymbol{u}^\ell + \boldsymbol{du}^\ell, \boldsymbol{c} + \boldsymbol{dc}^\ell) \\ & + \alpha \cdot R_{\text{flow}}(\boldsymbol{u} + \boldsymbol{du}^\ell, \boldsymbol{a}, \boldsymbol{b}) \\ & \quad + \, \lambda \cdot R_{\text{illum}}(\boldsymbol{c} + \boldsymbol{dc}^\ell) \,\Big) \, \mathrm{d}\boldsymbol{x} \, . \end{aligned} \tag{3.12}$$

The minimisation takes place only w.r.t. the flow increments, thus we linearise the data term around the positions displaced by the known flow $\boldsymbol{x}+\boldsymbol{u}^\ell$. After that, also the optic flow unknowns only appear linearly in the components of the latter functional. As in the previous chapter the only remaining non-linear terms are the sub-quadratic penaliser functions Ψ. In detail, the linearised brightness constancy assumption reads

$$\begin{aligned} D^\ell_{\text{bright}} = \Psi\Big(\theta \cdot \Big(& f_{2,x}^{\ell,\boldsymbol{u}^\ell} du^\ell + f_{2,y}^{\ell,\boldsymbol{u}^\ell} dv^\ell + f_2^{\ell,\boldsymbol{u}^\ell} \\ & - \bar{\phi}(f_1^\ell) - (\boldsymbol{c}^\ell + \boldsymbol{dc}^{\boldsymbol{\ell}})^\top \boldsymbol{\phi}(f_1^\ell)\Big)^2 \,\Big) \, , \end{aligned} \tag{3.13}$$

where, as before, we use the abbreviation $f_2^{\ell,\boldsymbol{u}^\ell}(\boldsymbol{x}) := f_2^\ell(\boldsymbol{x}+\boldsymbol{u}^\ell)$ to denote the back-registered second frame. The compensated gradient constancy would require the application of the chain rule in the BTF term:

$$\begin{aligned} \boldsymbol{\nabla}\Phi(\boldsymbol{c}^\ell + \boldsymbol{dc}^\ell, f_1) = & \sum_{j=1}^{n_b} \phi_j(f_1^\ell) \cdot \boldsymbol{\nabla}(c_j^\ell + dc_j^\ell) \\ & + \Big(\bar{\phi}'(f_1^\ell) + \sum_{j=1}^{n_b} (c_j^\ell + dc_j^\ell) \cdot \phi_j'(f_1^\ell)\Big) \cdot \boldsymbol{\nabla} f_1^\ell \, . \end{aligned} \tag{3.14}$$

Since derivatives of the unknown in the data term would render the minimisation of the functional much more difficult, we omit the increments $\boldsymbol{dc}^\ell$ for the gradient constancy and obtain the modified constancy assumption:

$$\begin{aligned} D^\ell_{\text{grad}} = \Psi\Bigg(\Bigg\| \begin{pmatrix} \theta_x & 0 \\ 0 & \theta_y \end{pmatrix} \Big(& \boldsymbol{\nabla} f_{2,x}^{\ell,\boldsymbol{u}^\ell} du^\ell + \boldsymbol{\nabla} f_{2,y}^{\ell,\boldsymbol{u}^\ell} dv^\ell + \boldsymbol{\nabla} f_2^{\ell,\boldsymbol{u}^\ell} \\ & - \sum_{j=1}^{n_b} \phi_j(f_1^\ell) \cdot \boldsymbol{\nabla} c_j^\ell \\ & - \Big(\bar{\phi}'(f_1^\ell) + \sum_{j=1}^{n_b} c_j^\ell \cdot \phi_j'(f_1^\ell)\Big) \cdot \boldsymbol{\nabla} f_1^\ell \Big) \Bigg\|_2^2 \,\Bigg) \, . \end{aligned} \tag{3.15}$$

Especially if multiple warps per scale are applied, this choice should approximate original constancy assumption sufficiently well. Moreover, in the latter two data terms, constraint normalisation has already been applied to all constancy assumptions in terms of the weights θ, θ_x, and θ_y, as proposed in Valgaerts et al. [2010] for linearised constraints with more than two variables.

We omit the flow regulariser as it coincides with the one from the previous chapter (c.f. Equation (2.83)). Similarly to the flow regularisation, also the novel coefficient regularisation term demands the smoothness of the overall coefficient field, thus it penalises directional derivatives of the sum of overall and incremental coefficients:

$$R^{\ell}_{\text{illum}} = \sum_{i=1}^{2} \Psi^{i}_{\text{illum}}\left(\sum_{j=1}^{n_b} \gamma_j \left(\boldsymbol{r}_i^{\top} \boldsymbol{\nabla}(c^{\ell}_j + dc^{\ell}_j)\right)^2 \right). \tag{3.16}$$

3.2.4 Minimality Conditions

The presented energy functional has to be minimised w.r.t. a number of unknowns: two flow components du^{ℓ} and dv^{ℓ}, the auxiliary 2-D vector fields $\boldsymbol{a}$, $\boldsymbol{b}$, and the n_b illumination coefficients that parametrise the brightness transfer function. In total, this amounts to a system of $6 + n_b$ partial differential equations (PDEs). Its structure is:

$$\begin{aligned}
&D_{du} && - \alpha \,\text{div}\, R_{\text{flow},1,\boldsymbol{\nabla} du} && = 0\,, && && (3.17)\\
&D_{dv} && - \alpha \,\text{div}\, R_{\text{flow},1,\boldsymbol{\nabla} dv} && = 0\,, && && (3.18)\\
&\alpha\,(R_{\text{flow},1,a_1} && - \beta \,\text{div}\, R_{\text{flow},2,\boldsymbol{\nabla} a_1}) && = 0\,, && && (3.19)\\
&\alpha\,(R_{\text{flow},1,a_2} && - \beta \,\text{div}\, R_{\text{flow},2,\boldsymbol{\nabla} a_2}) && = 0\,, && && (3.20)\\
&\alpha\,(R_{\text{flow},1,b_1} && - \beta \,\text{div}\, R_{\text{flow},2,\boldsymbol{\nabla} b_1}) && = 0\,, && && (3.21)\\
&\alpha\,(R_{\text{flow},1,b_2} && - \beta \,\text{div}\, R_{\text{flow},2,\boldsymbol{\nabla} b_2}) && = 0\,, && && (3.22)\\
&D_{dc_j} && - \lambda \,\text{div}\, R_{\text{illum},\boldsymbol{\nabla} dc_j} && = 0\,, && i = 1, \ldots, n_b\,, && (3.23)
\end{aligned}$$

where the scale indices ℓ have been omitted for the sake of readability. Regarding the data term, we can standardise also here its structure in terms of a generalised motion tensor. Since the inner argument is linear in the unknowns, we can write:

$$
\begin{aligned}
D^{\ell}_{\text{bright}} = \Psi\Bigg(\theta \cdot \bigg(\Big(f_{2,x}^{\ell,\boldsymbol{u}^{\ell}}, f_{2,y}^{\ell,\boldsymbol{u}^{\ell}}, -\boldsymbol{\phi}(f_1^{\ell})^{\top}, f_2^{\ell,\boldsymbol{u}^{\ell}} - \overbrace{\Big(\bar{\phi}(f_1^{\ell}) + \boldsymbol{c}^{\ell\top}\boldsymbol{\phi}(f_1^{\ell})\Big)}^{=\Phi(\boldsymbol{c}^{\ell}, f_1^{\ell})}\Big) \\
\cdot \Big(du^{\ell}, dv^{\ell}, \boldsymbol{dc}^{\ell\top}, 1\Big)^{\top}\bigg)^2\Bigg), \qquad (3.24) \\
= \Psi\Big(\theta \cdot (\boldsymbol{t}^{\top} \cdot \tilde{\boldsymbol{dw}})^2\Big), \\
= \Psi\Big(\theta \cdot \tilde{\boldsymbol{dw}}^{\top} \cdot \tilde{\boldsymbol{J}}^{\text{bright}} \cdot \tilde{\boldsymbol{dw}}\Big).
\end{aligned}
$$

Please note the similarities to the standard motion tensor [Bruhn, 2006]: The last component of the data vector $\boldsymbol{t}$ resembles a temporal derivative (with compensated first frame), and the vector $\tilde{\boldsymbol{dw}}$ captures all unknowns. However, the resulting extended motion tensor $\tilde{\boldsymbol{J}}^{\text{bright}} = \boldsymbol{t} \cdot \boldsymbol{t}^{\top}$ has size $(3 + n_b) \times (3 + n_b)$. Also the gradient constancy can be written in motion tensor notation, but since we removed the illumination coefficient increments from it, this motion tensor $\boldsymbol{J}^{\text{grad}}$ falls back to standard size 3×3.

As before, we abbreviate the appearing non-linear terms as follows:

$$
\begin{aligned}
\Psi'_{\text{bright}} &:= \nu\Psi'\Big(\theta \cdot \tilde{\boldsymbol{dw}}^{\top} \cdot \tilde{\boldsymbol{J}}^{\text{bright}} \cdot \tilde{\boldsymbol{dw}}\Big), && (3.25) \\
\Psi'_{\text{grad}} &:= (1-\nu)\Psi'\Big(\theta \cdot \boldsymbol{dw}^{\top} \cdot \boldsymbol{J}^{\text{grad}} \cdot \boldsymbol{dw}\Big), && (3.26) \\
\Psi'_{R_1} &:= \Psi'\Big(|\boldsymbol{\nabla} u + \boldsymbol{\nabla} du - \boldsymbol{a}|^2 + |\boldsymbol{\nabla} v + \boldsymbol{\nabla} dv - \boldsymbol{b}|^2\Big), && (3.27) \\
\Psi'_{R_2} &:= \Psi'\Big(|\boldsymbol{\mathcal{J}} \boldsymbol{a}|^2 + |\boldsymbol{\mathcal{J}} \boldsymbol{b}|^2\Big), && (3.28) \\
\Psi'^{i}_{\text{illum}} &:= \Psi'\Big(\sum_{j=1}^{n_b} \gamma_j \,(\boldsymbol{r}_i^{\top} \boldsymbol{\nabla}(c_j^{\ell} + dc_j^{\ell}))^2\Big), && (3.29) \\
\boldsymbol{D}_{\text{illum}} &:= [\boldsymbol{r}_1 \; \boldsymbol{r}_2]^{\top} \begin{pmatrix} \Psi'^{1}_{\text{illum}} & \\ & \Psi'^{2}_{\text{illum}} \end{pmatrix} [\boldsymbol{r}_1 \; \boldsymbol{r}_2], && (3.30)
\end{aligned}
$$

and state the remaining components of the Euler-Lagrange equations:

$$
\begin{aligned}
D_{du} = \Psi'_{\text{bright}} \cdot (\tilde{J}_{11}^{\text{bright}} du + \tilde{J}_{12}^{\text{bright}} dv + \tilde{J}_{13}^{\text{bright}} dc_1 + \ldots + \tilde{J}_{1(2+n_b)}^{\text{bright}} dc_{n_b} + \tilde{J}_{1(3+n_b)}^{\text{bright}}) \\
+ \Psi'_{\text{grad}} \cdot (J_{11}^{\text{grad}} du + J_{12}^{\text{grad}} dv + J_{13}^{\text{grad}}), \\
(3.31)
\end{aligned}
$$

and the data term contribution of the j-th coefficient equation is:

$$\begin{aligned} D_{dc_j} = \Psi'_{\text{bright}} \cdot (\tilde{J}^{\text{bright}}_{(2+j)1} du + \tilde{J}^{\text{bright}}_{(2+j)2} dv \\ + \tilde{J}^{\text{bright}}_{(2+j)3} dc_1 + ... + \tilde{J}^{\text{bright}}_{(2+j)(2+n_b)} dc_{n_b} + \tilde{J}^{\text{bright}}_{(2+j)(3+n_b)}) \,. \end{aligned} \tag{3.32}$$

Finally, the contribution of the illumination coefficient smoothness term involves the diffusion tensor $\boldsymbol{D}_{\text{illum}}$, c.f. (3.30). It reads:

$$R_{\text{illum},\boldsymbol{\nabla} dc_j} = \boldsymbol{D}_{\text{illum}} \cdot (\boldsymbol{\nabla} c_j + \boldsymbol{\nabla} dc_j) \,. \tag{3.33}$$

As before, the boundary conditions for the two flow Equations (3.17) and (3.18) read:

$$\boldsymbol{n}^\top (\boldsymbol{\nabla} u + \boldsymbol{\nabla} du - \boldsymbol{a}) = 0 \,, \text{ and } \quad \boldsymbol{n}^\top (\boldsymbol{\nabla} v + \boldsymbol{\nabla} dv - \boldsymbol{b}) = 0 \,, \tag{3.34}$$

and the Equations (3.19) – (3.22) are equipped with homogeneous Neumann boundary conditions, c.f. Equation (2.112). The boundary conditions for the last minimality condition (3.23) differ slightly from the previous ones:

$$\boldsymbol{n}^\top \boldsymbol{D}_{\text{illum}} \cdot (\boldsymbol{\nabla} c_j + \boldsymbol{\nabla} dc_j) = 0 \,, \quad j = 1, ..., n_b \,. \tag{3.35}$$

The discussed Euler-Lagrange equations constitute a system of $6 + n_b$ nonlinear PDEs. In order to resolve the last nonlinear terms c.f. (3.25) to (3.30), we again apply the lagged nonlinearity algorithm exactly as described in Section 2.4.4.

3.3 BTF Basis Learning

In the previous section, we have considered a variational energy functional that is to be minimised for the optic flow field and the illumination changes jointly. In this functional, we quantify the illumination changes with a brightness transfer function (BTF). This function is parametrised in terms of a linear combination of basis functions. Thus, the unknowns to be estimated jointly with the flow are the coefficients of this linear combination. This means we estimate new coefficients together with each new flow field, but the basis functions ϕ_j are learned offline from training data and remain the same afterwards.

These basis functions will be the topic of this section: We will discuss our estimation strategy for the mean BTF $\bar{\phi}$, the n_b basis functions ϕ_j, as well as the associated weights γ_j for the regulariser of the coefficient fields. Our basic

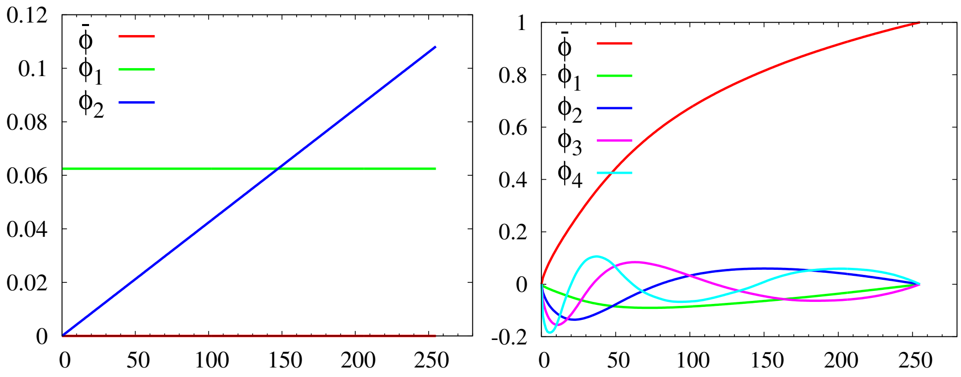


Figure 3.2: Bases from the literature. **Left:** Normalised affine basis. The additive basis only consists of the ϕ_1. **Right:** EMoR functions Grossberg and Nayar [2002].

strategy is inspired by the *Empirical Model of Response (EMoR)* of Grossberg and Nayar [2004], where the camera response function of imaging systems is also parametrised with a set of basis functions. However, our model acts on intensities instead of irradiances.

As already mentioned in Section 2, input intensities f are mapped to output intensities via the BTF:

$$\Phi(f) = \bar{\phi}(f) + \sum_{j=1}^{n_b} c_j \cdot \phi_j(f)\,. \tag{3.36}$$

Note that many kinds of polynomial and exponential illumination models can be represented using the appropriate basis functions. For instance, the standard model without any compensation can be obtained by the choice

$$\bar{\phi}(f) = f\,, \quad n_b = 0\,. \tag{3.37}$$

Similarly, the affine model of Negahdaripour and Yu [1993] fits into this framework by choosing

$$\bar{\phi}(f) = 0\,, \qquad \phi_1(f) = 1\,, \qquad \phi_2(f) = f\,. \tag{3.38}$$

The same holds also for the purely additive models in Cornelius and Kanade [1984] and Mukawa [1990], i.e. if instead $\phi_2(f) = 0$. In Figure 3.2 these bases are depicted.

The recent KITTI Vision Benchmark Suite Geiger et al. [2012] and also the new Middlebury stereo benchmark of Scharstein et al. [2014] offer huge sets of real-world image sequences together with ground truth optical flow fields. This gives us access to samples of input and output intensity levels of realistic scenarios. In particular, the availability of optical flow fields allows us to register consecutive frames and to analyse the behavior of the true BTF on a per-pixel basis.

3.3.1 General Strategy

Our general strategy to learn a basis from this massive amount of training data consists of three steps: First, we segment and cluster each training image spatially according to illumination changes along the ground truth flow. The segmentation is important, since we cannot expect that different image pairs provide fundamentally different *global* BTFs. Instead, we have to estimate multiple BTFs per image pair, since typical illumination changes such as drop shadows or specular reflections are *local* phenomena. In a second step, we use the segmented input images and compute for each region of each input image a separate BTF. This enables us to determine the sought local BTFs. Third, all these BTFs are used to perform a principal component analysis (PCA) in order to identify the most representative basis functions for the observed illumination changes. It is worth noting that these steps can be applied iteratively, i.e. the estimated basis functions can be used again to segment the input images and thus to obtain improved BTFs. Our basic basis estimation strategy is related to the work of Tieu and Miller [2002]. There, the change of colour between the first and second image is seen as a 3-D displacement in RGB-space. The displacement vectors of all colours of an image together are considered as a *colour flow*. Then, such colour flows are extracted from many input images and a suitable basis is estimated in terms of so-called *eigenflows*.

Let us now detail on the steps of our strategy.

3.3.2 Segmenting Illumination Changes

Let us assume that we are given the training image sequences with corresponding ground truth flows. This means for each pixel we are given the intensity of the first frame and the intensity of the corresponding pixel in the second frame. Since our goal is to discriminate image regions with distinct lighting situations, i.e. with different brightness transfer functions, we first have to determine the pointwise BTF for every pixel of each image pair. However, although this pointwise BTF can be arbitrarily complex, the given images provide *only one* constraint per pixel: The unknown BTF must map the intensity of this pixel in the first frame to the intensity of the corresponding pixel in the second frame.

Coefficient Estimation. To relax this extremely under-determined problem, let us now assume we are already given an estimate of the basis functions. Then, the sought pointwise BTF can be approximated using the given basis, and our task comes down to computing the n_b entries of the optimal coefficient vector $\boldsymbol{c}$ in each pixel. This lowers the degrees of freedom drastically in each pixel. Consequently, this problem fits perfectly into our variational model from Section 3.2, with the difference that we *only* have to solve for the coefficients $\boldsymbol{c}$. The ground truth optic flow is given and does not need to be estimated. However, as the ground truth might not be provided at every pixel (i.e. due to occlusions or due to sparse laser scans), we have to disable the data term at those positions where a flow vector is missing. Basically, this procedure leads to a variational inpainting method [Weickert and Welk, 2006], because if the data term is disabled, only the coefficient regularisation term contributes to the energy. Note that unlike in traditional inpainting scenarios, we are not interested in the coefficient values at positions of missing data. We only want to enforce global communication via the regularisation term in order to avoid isolated estimates and to remove the discussed ambiguities.

Clustering. Once the coefficients are found, we perform a K-Means clustering [Steinhaus, 1957] (usually $K = 4$) on the coefficients. This takes place exclusively in the n_b-dimensional coefficient space; spatial coordinates are intentionally ignored here in order to allow spatially disjoint regions belonging to the same segment. All pixels whose coefficients have been clustered together share a similar brightness transfer function and thus exhibit a similar lighting situation. Figure 3.3 shows an example where such a segmentation allows to distinguish regions in the image that undergo different brightening effects.

3.3.3 Estimating Brightness Transfer Functions

Given the previously computed segmentation, the next task is to estimate one brightness transfer function $g : \mathbb{R} \to \mathbb{R}$ per segment. To this end, we adopt the *global* idea of Grossberg and Nayar [2002] *locally*: For each segment we construct the intensity histogram h_1 of the pixels in the first frame as well as the histogram h_2 of the corresponding intensities in the second frame. In this context, we only consider pixels with valid optical

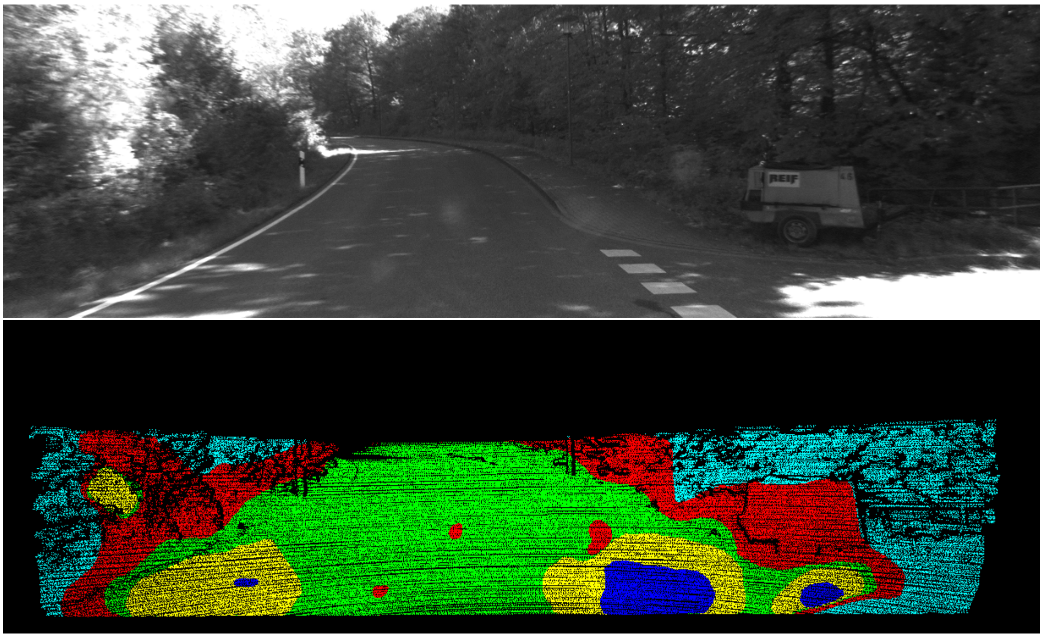


Figure 3.3: **Top:** Frame 1 of KITTI training sequence #114. **Top:** Corresponding K-Means segmentation. Each colour indicates a separate cluster, black pixels denote locations where ground truth is missing. The separation between the stronger brightening effect on the street and the weaker brightening effect in the environment becomes visible. Moreover, the interreflections at the windshield show off in terms of the three red spots on the street.

flow, i.e. a ground truth vector must be given and must not point out of the image domain. As a consequence, neither occlusions, disocclusions or out-of-image motion can spoil our result. Once both histograms have been created, we compute the BTF that transforms h_1 into h_2 by means of a histogram specification. A pseudo-code algorithm for this step is given in Algorithm 1. In this context, fully saturated segments or too small clusters may lead to wrong and unrealistic brightness transfer functions. To avoid this, we reject any segments in which more than 80% of all pixels have the same intensity, as well as segments in which more than one third of all possible intensities do not occur. Please note that the resulting function of the histogram specification is discrete and given by a vector $\boldsymbol{g} \in \mathbb{R}^{256}$ that is not parametrised in terms of basis functions and coefficients. In Figure 3.4, we depict the estimated basis functions for two different benchmarks. Although the Middlebury stereo benchmark contains much less training scenes, one has to keep in mind that each scene is available in three lighting situations. Moreover, colour images

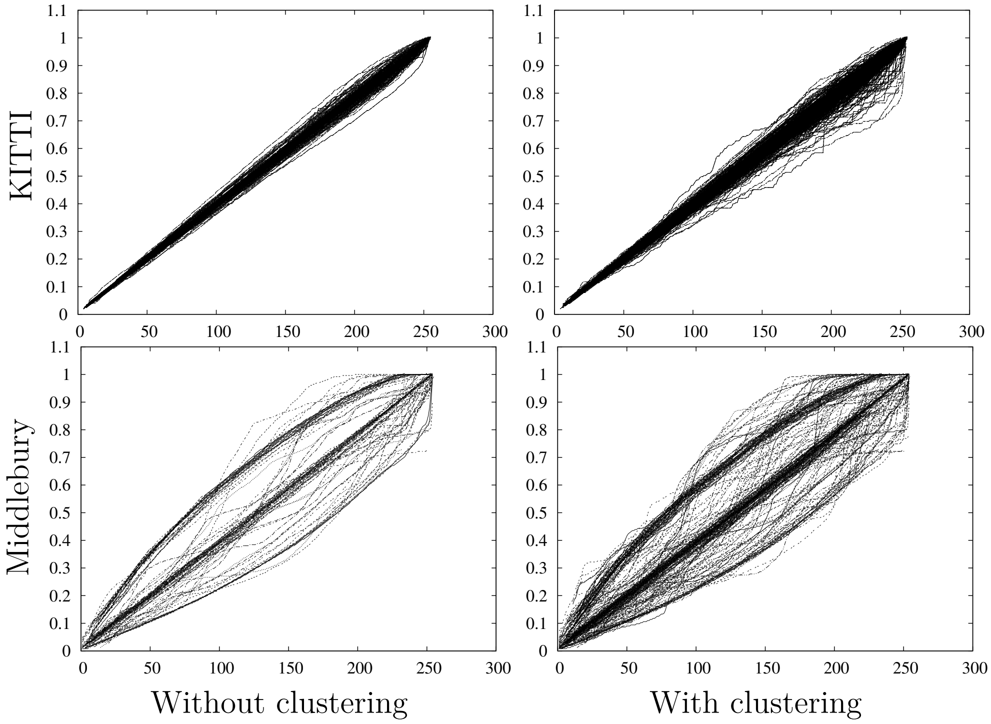


Figure 3.4: Result of our BTF extraction algorithm. **Left column:** Without clustering. **Right column:** After iterating with clustering. **Top row:** Result of using all 194 training images of the KITTI benchmark. As one can see, a clearly larger variety of BTFs can be found. **Bottom row:** Using the Middlebury stereo benchmark, already the initial extraction gives a large variety of BTFs. This is due to the various lighting situations that are provided.

Algorithm 1 BTF extraction algorithm.

```
$hist_1 \leftarrow$ histogram of clustered pixels in first frame
$hist_2 \leftarrow$ histogram of clustered pixels in second frame
$cummulative_1, cummulative_2, index_2 \leftarrow 0$
**for** $index_1 \leftarrow 0, 255$ **do**
    $cummulative_1 \leftarrow cummulative_1 + hist_1(index_1)$
    **while** $cummulative_2 < cummulative_1$ **do**
        $index_2 \leftarrow index_2 + 1$
        $cummulative_2 \leftarrow cummulative_2 + hist_2(index_2)$
    **end while**
    $BTF(index_1) = index_2$
**end for**
```

are available, allowing us to extract three BTFs per image. Along with the BTFs from other segments it serves as input for the following PCA.

3.3.4 Learning the Basis

After having performed the previous clustering and estimation steps on each of the p training image pairs we obtain $m \leq K \cdot p$ brightness transfer functions, so-called *observations*. In order to find one common set of basis functions for all of them, we perform a principal component analysis (PCA). After concatenating all observations $\boldsymbol{g}_i$ $(i = 1, \ldots, m)$ into a so-called *observation matrix*

$$\boldsymbol{G} = (\boldsymbol{g}_1 | \ldots | \boldsymbol{g}_m) \in \mathbb{R}^{256 \times m}, \tag{3.39}$$

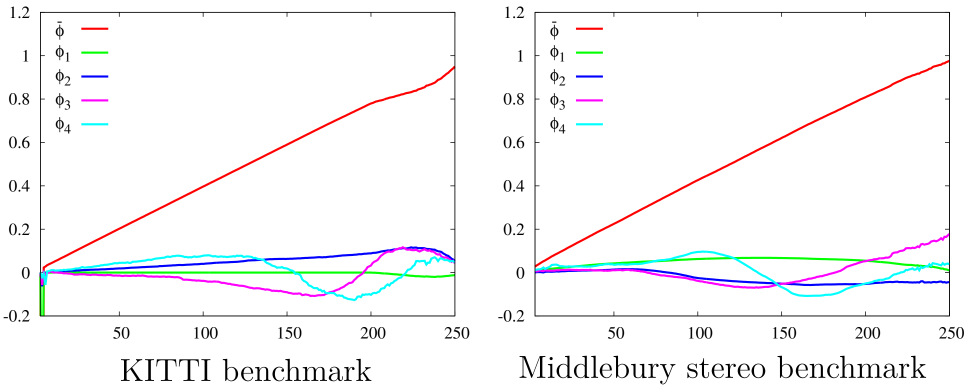
we compute the row-wise mean (i.e. the sample mean over all observations) $\bar{\boldsymbol{g}}$ of $\boldsymbol{G}$. Then we obtain the positive semi-definite covariance matrix $\boldsymbol{C}$ as:

$$\boldsymbol{C} = \boldsymbol{U}^\top \boldsymbol{\Sigma} \boldsymbol{U} = \frac{1}{m-1} \sum_{i=1}^{m} (\boldsymbol{g}_i - \bar{\boldsymbol{g}})(\boldsymbol{g}_i - \bar{\boldsymbol{g}})^\top. \tag{3.40}$$

A deep discussion of principal component analysis is out of the scope of this thesis; for details we refer to the book of Jolliffe [2002]. From this principal component decomposition, the sought basis functions ϕ_j $(j = 1, ..., n_b)$ can be found as the eigenvectors of the covariance matrix (the columns of $\boldsymbol{U}$). Moreover, the row-wise mean $\bar{\boldsymbol{g}}$ coincides with the 0-th basis function which is the mean brightness transfer function $\bar{\phi}$. Furthermore, the diagonal matrix $\boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{\sigma})$ contains the (non-negative) eigenvalues which represent the variance of the given data along the principal components. This is a well-suited estimate for the relative magnitude of the coefficients. Hence, we choose the weights γ_j in the anisotropic coefficient regularisation term (3.10) proportional to the inverse of the eigenvalues. More exactly, as another normalisation step, we divide the final weight by the sum of all weights:

$$\gamma_j = \frac{\sigma_j^{-1}}{\sum_{i=1}^{n} \sigma_i^{-1}} \tag{3.41}$$

Figure 3.5 shows the estimated bases for the KITTI Vision Benchmark Suite and compares it to an affine basis and the *EMoR* basis provided by Grossberg and Nayar [2004]. We can see that compared to the EMoR basis our basis functions for the KITTI benchmark rather model illumination changes in the upper part of the dynamic range. Moreover, the mean brightness transfer function is roughly linear, since we do not estimate a camera response function as in Grossberg and Nayar [2004], but a mapping between intensities (where identity is expected as average).



KITTI benchmark

Middlebury stereo benchmark

Figure 3.5: Our estimated bases for the KITTI and Middlebury stereo benchmark.

3.3.5 Iterating the Estimation

The strategy we have described so far assumes a basis to be given for the clustering step. Initially, however, only the training images and ground truth flows are given. Thus, in our first iteration loop we omit the clustering step, treat the whole images as one segment, and estimate one global brightness transfer function per image pair. This leads to a first estimate for the basis which allows us then to perform the clustering as described. The impact of iterating the estimation is demonstrated with two figures. First, Figure 3.4 makes the gain of variability in the extracted raw BTFs obvious. Especially for the KITTI benchmark this gain is important since the initial BTFs show only very little differences. Next, considering the resulting basis functions and their shapes, Figure 3.6 shows the impact of iterating. Here we depict the obtained initial and iterated bases for the KITTI [Geiger et al., 2012] as well as the recent Middlebury stereo benchmark [Scharstein et al., 2014]. The results of the two benchmarks show a slightly different behaviour: for the KITTI benchmark, the initial basis vectors are close to zero in the lower half of the dynamic range. After iteration, all basis vectors except ϕ_1 are clearly different from zero. This is not the case for the bases computed from the Middlebury training images. Here the initial estimate already shows clear deviations from zero in the lower dynamic range. Furthermore, mainly the higher basis vectors ϕ_3 and ϕ_4 change during the iteration. The other basis components only change very little. This can be explained from the fact that already the initially extracted global BTFs capture enough illumination changes. As a consequence, the iteration scheme does not discover completely new BTFs, as is the case for the KITTI benchmark.

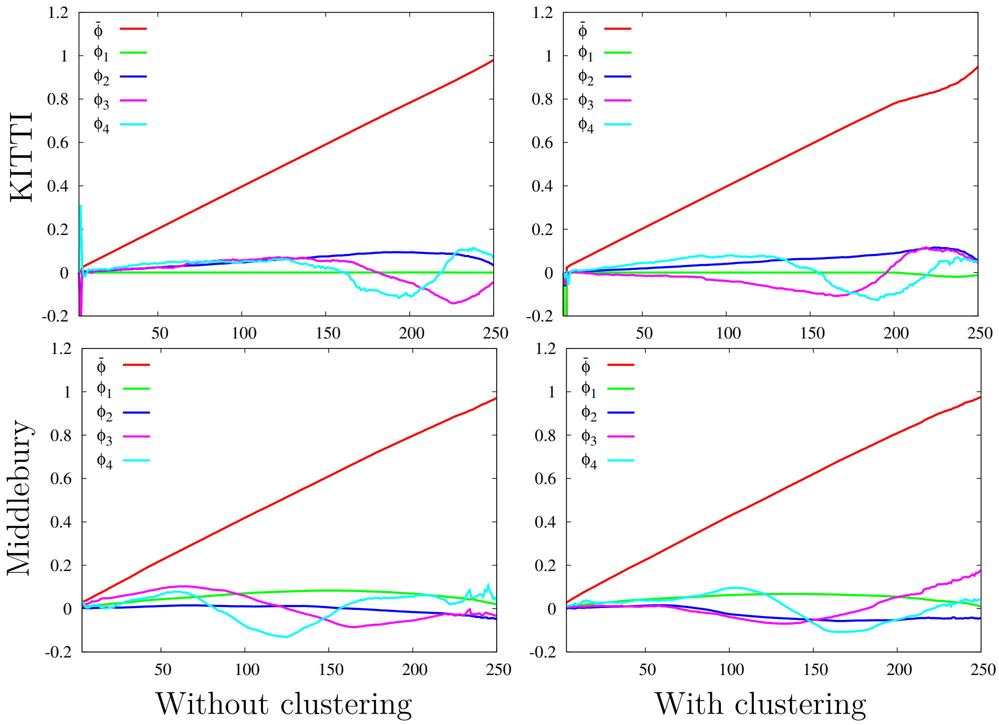


Figure 3.6: Impact of iterating the estimation for the KITTI (top) and Middlebury benchmark (bottom). The left plot shows the initially extracted basis (without clustering step), the right plot shows the iterated basis.

3.4 Experiments

Let us now come to the evaluation of our framework. To this end we will start by detailing on our experimental setup. After that we will compare different available bases, then we will evaluate some components of our framework. Finally, we will compare with the state-of-the-art in terms of public benchmark systems.

3.4.1 Experimental Setup

Choice of Parameters

Although our model contains a considerable number of parameters, effectively we only adjust the three main model parameters α, β, and λ. As in the previous chapter, the contrast parameters for the sub-quadratic functions have been chosen fixed for all experiments. Concerning the K-Means clustering step we kept $K = 3$ fixed as well. Generally, we stick to the described parameter optimisation strategy c.f. Section 2.5.1 to guarantee an absolutely

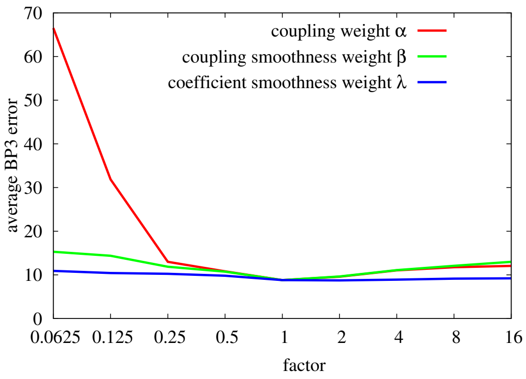


Figure 3.7: Effect of varying the three main model parameters. Plot shows the error when varying the optimal parameter by a factor. The coupling weight is the most important parameter.

fair comparison throughout all following evaluations.

To illustrate the behavior of this model under variations of its main parameters, we perform the following experiment. First, we find optimal parameters for the KITTI image sequence #15 ($\alpha = 0.174$, $\beta = 28$, and $\lambda = 0.14$). Next, we vary each of the three parameters by a factor of 2^k where $k \in \{-4, ..., 4\}$. The other two parameters are kept fix. The resulting average error measurements are plotted in Figure 3.7. As one can see, the crucial parameter is – as before – the weight of the coupling term α, the two other parameters have much less influence on the resulting error.

Implementation Details

Our implementation is based on the one from the previous chapter. The number of unknowns per pixel however increases by $n_c \cdot n_b$ because of the illumination coefficients. The complementary regularisation term on these coefficients leads to an anisotropic diffusion term that we discretise according to Weickert [1998].

Runtimes. The number of coefficients n_b has the largest impact on the runtime of our single core implementation. Because of this, the usage of colour image material leads to a significant increase of runtime (three times as many coefficient fields). Table 3.1 depicts the runtime of our implementation on an a Macbook Air with Intel Core i5 processor (1.3 GHz). As image sequence we chose the *Rubberwhale* sequence of Baker et al. [2011] (584 × 388 pixels, RGB). As before, we set the number of outer and inner solver iterations to 5 and 10, respectively.

Table 3.1: Runtimes of our method. Each coefficient field is subject to regularisation, thus the runtime is approximately proportional to the number of coefficients.

Number of coefficients n	1	2	3	4	5
Runtime [sec]	102.3	133.9	181.5	245.3	330.2

Table 3.2: Influence of different basis function on the resulting optic flow accuracy.

Configuration	*subset3newmiddle*	*subset3kitti*
Baseline (no comp.)	5.679 px	10.296 %
Affine basis	3.989 px	9.514 %
Initial basis	4.146 px	8.309 %
Iterated basis	**3.856 px**	**8.276 %**

3.4.2 Comparison of Basis Functions

In our first experiment, we investigate the usefulness of illumination estimation in general and analyse the impact of choosing different sets of basis functions on the quality of the flow estimation. To this end, we consider the subset of the KITTI benchmark [Geiger et al., 2012] as well as the selection of training sequences from the recent Middlebury stereo benchmark [Scharstein et al., 2014]. As we are using the parameter optimisation strategy described in Section 2.5.1 here for three parameters, we limit the subset to three images for each benchmark (*subset3kitti* and *subset3newmiddle*).

As first step, we compute a basis for both image sets using all available image sequences, c.f. Figure 3.5. We perform three iterations of the described basis estimation scheme and store the initial as well the intermediate results. In the next step, we compute the average error of our method – with different bases as well as without any illumination compensation – for each of the image sets. For each configuration, we have optimised the parameters of our model w.r.t. the adequate error measure using the ground truth. The results of this experiment are presented in Table 3.2. As one can see, all compensation schemes allow to decrease the error compared to the baseline method, i.e. with no illumination compensation and coefficient estimation. Moreover, our initial basis is comparable to the affine parametrisation of Gennert and Negahdaripour [1987]. However, we can observe a clear improvement with the result of our iteration scheme.

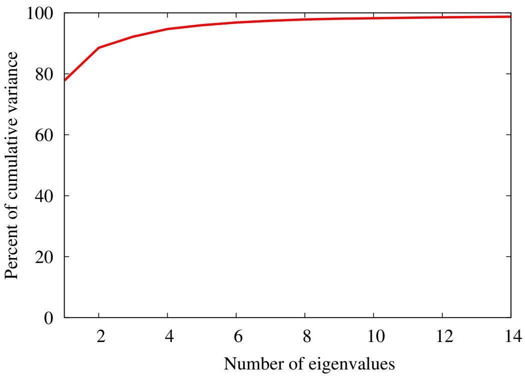


Figure 3.8: Cumulative energy content for each eigenvector.

Table 3.3: Number of bases n_b and resulting accuracy on two different image sets.

n_b	*subset3kitti*	*subset3newmiddle*
1	9.21 %	5.42 px
2	8.48 %	4.27 px
3	8.46 %	3.83 px
4	8.28 %	3.86 px
5	8.18 %	**3.70 px**
6	8.21 %	3.76 px
7	**8.13 %**	4.42 px
8	8.15 %	4.04 px

3.4.3 Component Evaluation

Number of Basis Vectors

Our parametrisation of the BTF comprises another parameter n_b which represents the number of basis vectors to be used. It clearly is a parameter for which a suitable value has to be chosen. In Figure 3.8, we plot the cumulative variance resulting from the PCA for the last iteration of our KITTI basis. As one can see from this plot, already the first principal component captures 77.7% of the variance of the input data, and the first 3 components cover more than 92%.

Apart form this theoretical consideration, we also evaluate the accuracy of our variational framework for several choices of n_b. The result is depicted in Table 3.3, where one can see that more components lead to higher accuracy. If not otherwise stated, we use $n_b = 4$ basis vectors as a compromise between accuracy and computational effort for the rest of our experiments.

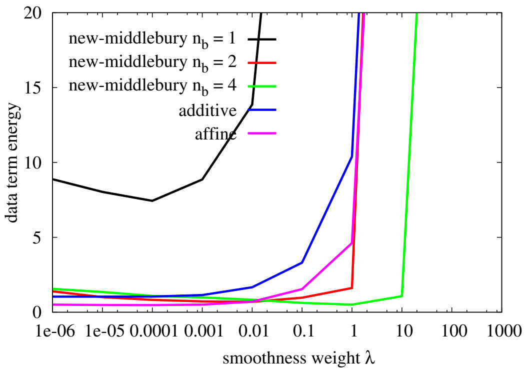


Figure 3.9: Data term contribution for varying smoothness parameters.

Additionally, we perform the following synthetical experiment to illustrate the benefit of our BTF paramterisation in a learned basis. We consider the Adirondack sequence [Scharstein et al., 2014], and solve our functional only for the illumination coefficients (ground truth flow is given). We do this for varying values of the illumination coefficient smoothness parameter λ, and evaluate the integral value of our data term. In Figure 3.9, the results of this experiment are depicted for different bases. Here one can see that for small smoothness weights, all parametrisations can represent the image data well, the data term integral is small. Differences become obvious for larger smoothness weights. Here, the coefficients are forced to be smooth while explaining the image data. As one can see, if two or more coefficients are used with a leared basis, the resulting data energy is lower than for the canonic bases (additive and affine). For $n_b = 4$, even for really large smoothness weights, the data energy stays low. This means that our learned model can represent the given image data well while at the same time being regularised strongly.

Coefficient Regularisation

Concerning the regularisation strategy for the illumination coefficients, we have tested an nonlinear isotropic alternative to the anisotropic term from Equation (3.10). This alternative term reads

$$R_{\text{illum}}^{\text{iso-nonlin}}(\boldsymbol{c}) = \Psi\left(\sum_{j=1}^{n_b} \gamma_j |\boldsymbol{\nabla} c_j|^2\right) . \tag{3.42}$$

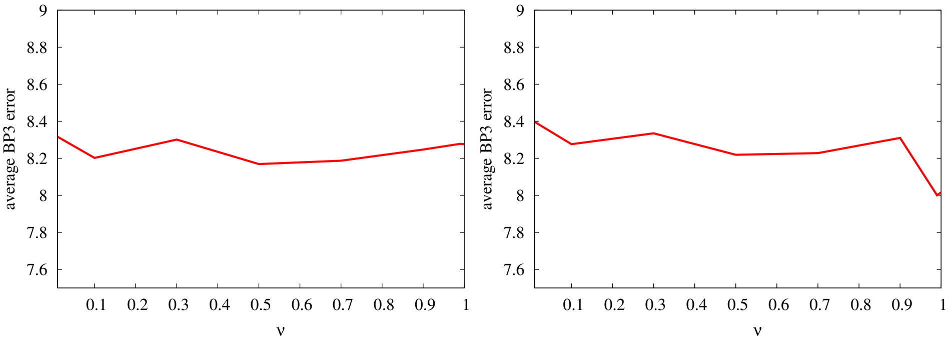
We run our optimisation strategy for both models on the known set of 10 KITTI sequences. As it turns out, the anisotropic regularisation does perform clearly better than the isotropic term: The average BP3 error increases from 12.0% to 12.6 % when switching from anisotropic to isotropic regularisation.

Gradient Constancy Assumption

We have also evaluated the necessity of the gradient constancy assumption in our model. To this end, we compute the best accuracy for different values of $\nu \in [0, 1]$, the parameter steering the influence of the gradient constancy assumption. In this context, the choice $\nu = 1$ disables the gradient constancy assumption completely, and $\nu = 0$ switches to pure gradient constancy. Note that the latter extreme case is not solvable in with our incremental energy formulation since we distributed the coefficient increments only to the intensity constancy assumption, c.f. Equation 3.15. The results of this experiment are presented in Figure 3.10. Regarding the left plot there, one can see that the gradient constancy assumption does not influence the accuracy too much. The minimal error value was obtained with $\nu = 0.5$. However, during our experiments we found that bright saturated areas lead to artifacts in the flow field, especially if the gradient constancy assumption is switched off. By introducing a space-variant weight into the data term that disables any constancy assumption whenever the intensity of any of the frames is outside the interval $[0, 250]$, we were able to diminish this effect. This small additional weight changes the behavior under variations of the gradient parameter, as can be seen in the right plot of Figure 3.10. Now, the best accuracy is in fact obtained if the gradient constancy assumption is switched off completely.

Behaviour under Synthetic Rescalings

Also the question how our estimation scheme behaves under synthetic rescalings is interesting. We adopt the experimental setup from the previous chapter (c.f. Section 2.5.3) and evaluate how the error behaves when varying the value of γ in the interval $[\frac{1}{3}, 3]$. The result of this experiment is depicted in Figure 3.11. It turns out that for values of $\gamma > 1$, the estimation performs quite robustly. However, for $\gamma < 1$, the error increases. This is plausible, since this case corresponds to a brightening of the images, leading to more and more saturated image regions. Traditional data terms cannot represent such situations, usually the robust function weights the constancy assumption down in this case. However, our estimation scheme can compensate the first frame to the maximal intensity. In such saturated regions however, this does rather harm the result since the image information is destroyed in saturated regions. This effect is reflected in Figure 3.11 by large errors for small γ-values.



With saturated regions

Without them

Figure 3.10: Accuracy under variations of the gradient constancy parameter ν. If $\nu = 0$, only the gradient constancy assumption is enabled. If $\nu = 1$, only the compensated grey value constancy assumption counts. **Left:** If saturated image regions are treated as non-saturated, the gradient constancy is important; best result for $\nu = 0.5$. **Right:** If the data term is weighted down in saturated regions, the gradient constancy assumption looses its importance; highest accuracy for $\nu = 0.99$.

3.4.4 Analysis of Transfer Functions and Coefficient Fields

Let us now shed light on the coefficients that are estimated jointly with the optical flow. To this end, we have picked one of the training sequences with moderate illumination changes, see Figure 3.12, and another sequence with severe illumination changes, see Figure 3.13. The two figures show the first frame of the respective sequence together with the estimated flow as well as the four computed coefficient fields. Furthermore, we have highlighted interesting locations in the images using coloured squares. The brightness transfer functions at those locations – that can be computed as linear combinations of the learned basis functions weighted by the estimated local coefficients – are jointly depicted in a graph using the corresponding colours.

For our first challenging example (Figure 3.12), the flow field appears reasonably accurate, which is confirmed by a BP3 error of only 8.81% – the highest accuracy we have been able to achieve on this image sequence with any method discussed in this thesis.

Since the lighting changes in that image sequence are rather global, the extracted brightness transfer functions are similar in shape. In fact, they only differ in the upper end of the dynamic range. As can be seen from the BTFs, the image becomes darker. This is mainly reflected by the strongly negative values in the coefficient field c_1 (that belongs to a positive basis

Figure 3.11: Behaviour of the estimation approach under synthetic γ-rescalings. The setup of this experiment is completely analogous to the γ-experiment from the previous chapter. One can see that except for extremely small values of γ, the method is robust against such rescalings.

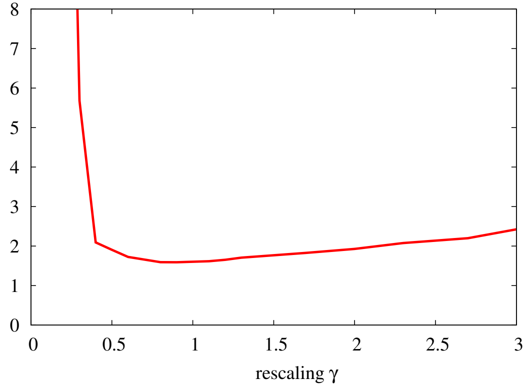


Table 3.4: Error statistics of our method for the bad pixel measure with varying thresholds (BP2 - BP5), averaged over all sequences of the KITTI evaluation benchmark.

Error	Out-Noc	Out-All	Avg-Noc	Avg-All
2 pixels	10.09 %	15.63 %	1.8 px	3.6 px
3 pixels	7.63 %	12.46 %	1.8 px	3.6 px
4 pixels	6.32 %	10.63 %	1.8 px	3.6 px
5 pixels	5.45 %	9.38 %	1.8 px	3.6 px

function). Moreover, slight local variations of the BTFs can be observed in the coefficient plots, in particular in the plots of the coefficient fields c_2, c_3, c_4. An example, where our model has actually estimated significantly differing BTFs for different parts of the image is presented in Figure 3.13. Particularly challenging in this sequence are the inter-reflections in the windshield in front of the camera. However, the flow field is still of reasonable quality (BP3 of 5.6%). Apart of the spatially varying BTFs, one can also observe that the inter-reflections are reproduced by the corresponding coefficient fields.

3.4.5 Comparison to the Literature

Let us now compare our method to other approaches from the literature. To this end, we evaluated our method on the KITTI test sequences using the optimised parameters (α, β, λ)=(0.26, 30, 1). The corresponding results are shown in Tables 3.4 and 2.16. Table 3.4 gives detailed information on the

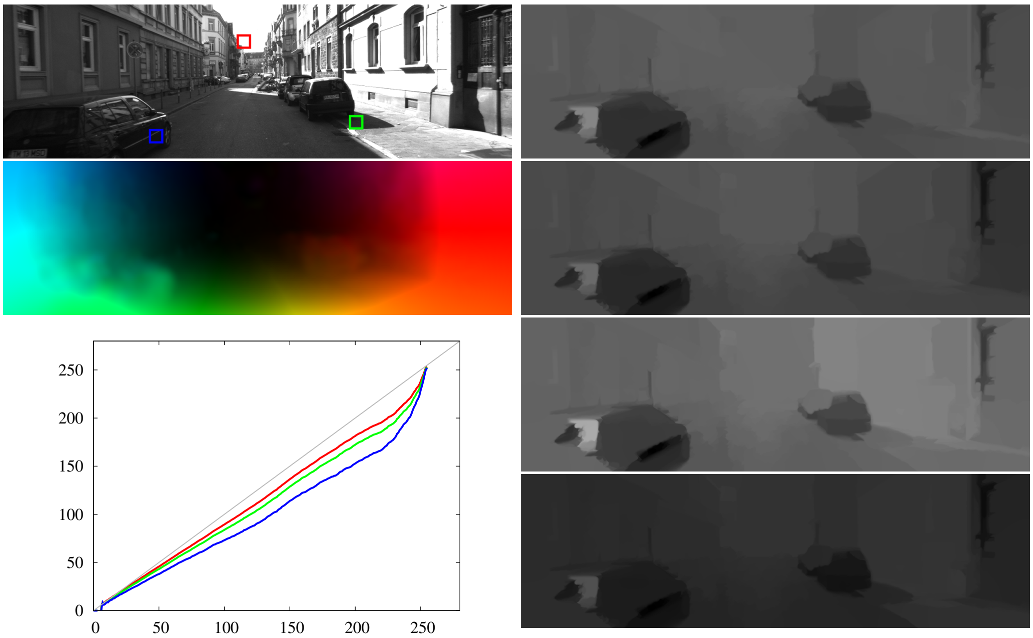


Figure 3.12: Estimated coefficients and BTFs. **Left column, from top to bottom:** First frame of KITTI training sequence #15 with three highlighted positions, estimated optical flow field (avg. BP3 error of 8.81 %), and plot of the three corresponding brightness transfer functions. Plot colours coincide with the marker colours. **Right column, from top to bottom:** Estimated coefficient fields c_1 to c_4. Coefficients have been shifted such that a grey value of 127 denotes a coefficient value of 0. Brighter values denote positive coefficients, darker values negative coefficients.

performance of our algorithm in *non-occluded* and *all* regions for different thresholds of the bad pixel error measure (BP2 - BP5). The main benchmark table has already been given in the previous chapter (Table 2.16) and is repeated here for convenience (Table 3.5) with better highlighting. It shows the performance of our algorithm compared to other *pure two-frame* optical flow methods *without stereo constraints* (such constraints are likely to fail in realistic scenarios with independently moving objects). As one can see, our conference submission with direct second-order regularisation ([Demetz et al., 2014] $\boldsymbol{\mathcal{H}}$) is among the leading optical flow approaches in this benchmark. In particular, when considering all pixels (i.e. also occluded regions), this method ranked first at the time of submission and is significantly more accurate than previous approaches. This clearly demonstrates that performing a joint estimation of illumination changes and motion can outperform methods discarding illumination information by using invariants. Additionally,

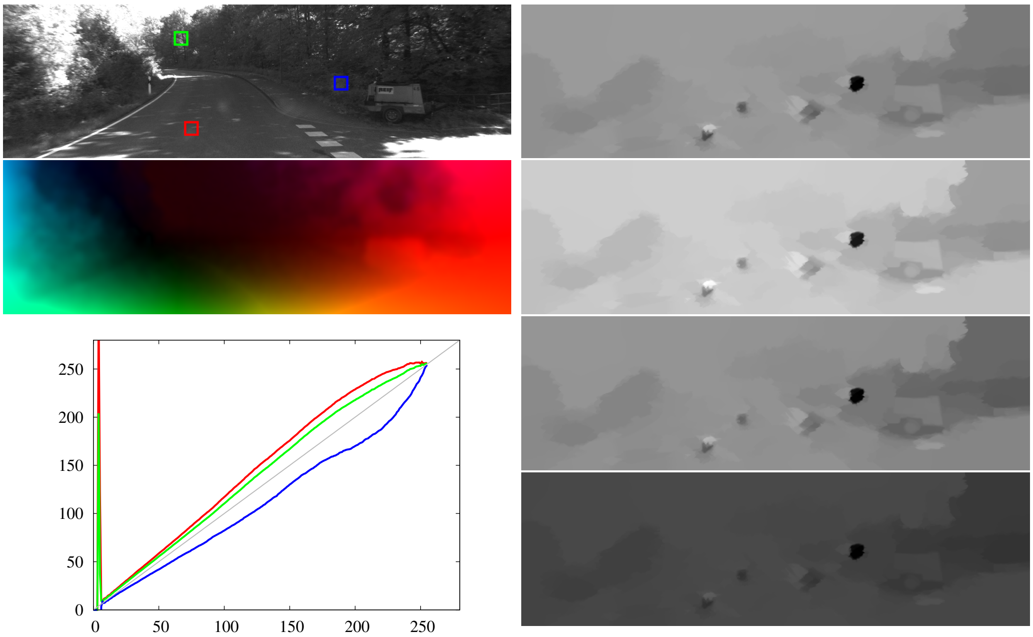


Figure 3.13: Estimated coefficients and BTFs. **Left column, from top to bottom:** First frame of KITTI training sequence #114 with three highlighted positions, estimated optical flow field, plot of the three corresponding brightness transfer functions. Plot colours coincide with the marker colours. **Right column, from top to bottom:** Estimated coefficient fields c_1 to c_4. Coefficients have been shifted such that a grey value of 127 denotes a coefficient of 0. Brighter values denote positive coefficients, darker values negative coefficients.

the table entry *Estimation w. TGV* represents the estimation scheme with coupled second-order regulariser exactly as presented in this chapter. Those results are perfectly comparable to our results from the previous chapter (green entry, [Demetz et al., 2015] CRT w. TGV).

3.5 Summary

In this chapter we extended our variational framework for optic flow computation. Our extension allowed to refrain from invariances for matching structures. Instead, we included the problematic appearance changes explicitly into the data term of our framework and modelled the spatial behaviour of these changes with a complementary anisotropic regulariser. This lead to a variational energy functional that we minimised for the optic flow and the illumination changes *jointly*.

Table 3.5: Top KITTI benchmark results as of March 31st, 2015. Table repeated for convenience. Only pure two-frame dense optic flow methods are shown. All methods of this chapter are highlighted in red. Our methods that have been discussed in the previous chapter are highlighted in green.

Method	**BP3** [%]				**AEE** [px]			
	noc		**occ**		**noc**		**occ**	
[Ranftl et al., 2014]	**5.93**	1	11.96	2	1.6	4	3.8	3
[Wei et al., 2014]	6.03	2	13.08	5	1.6	4	4.2	5
[Braux-Zin et al., 2013]	6.20	3	15.15	7	**1.5**	1	4.5	6
[Demetz et al., 2014] $\boldsymbol{\mathcal{H}}$	6.52	4	**11.03**	1	**1.5**	1	**2.8**	1
[Demetz et al., 2015] (CRT w. TGV)	6.71	5	12.09	3	2.0	8	3.9	4
[Vogel et al., 2013]	7.11	6	14.57	6	1.9	7	5.5	8
[Weinzaepfel et al., 2013]	7.22	7	17.79	8	**1.5**	1	5.8	9
Estimation w. TGV	7.63	8	12.46	4	1.8	6	3.6	2
[Rashwan et al., 2013]	7.91	9	18.90	13	2.0	8	6.1	10
[Mohamed et al., 2014]	8.67	10	18.78	12	2.4	11	6.7	13
[Timofte and Gool, 2015]	9.09	11	19.32	14	2.6	12	7.6	16
[Demetz et al., 2013] (CRT w. TV)	9.43	12	18.72	11	2.7	14	6.5	11
[Sun et al., 2014]	10.04	13	20.26	15	2.6	12	7.1	14
[Kennedy and Taylor, 2015]	10.22	14	18.46	10	2.0	8	5.0	7
[Sun et al., 2014]	10.49	15	20.64	16	2.8	15	7.2	15
[Hermann and Klette, 2013]	10.74	16	22.66	17	3.2	17	12.2	17
[Ranftl et al., 2012]	11.03	17	18.37	9	2.9	16	6.6	12

The first part of this chapter was devoted to the variational model behind this idea. We discussed the brightness transfer function that plays the central role of our extension. It is parametrised with a linear combination of basis functions, whose coefficients we regularised anisotropically. Finally, we showed how to minimise our model in terms of its associated Euler-Lagrange equations.

Besides this variational framework, another important component of our model was the estimation of a suitable basis. We discussed how to extract the brightness transfer function from an image sequence and how to estimate a basis form many samples of such functions.

In our experiments, we analysed the behaviour and performance of the developed method. We saw that proper regularisation of the illumination coefficients is very important, as otherwise any motion in the scene could be attributed to illumination changes, and vice-versa. We were also able to resign from the gradient constancy assumption that contradicted our general idea of compensating illumination changes such that no invariant features are necessary anymore.

The big advantage of our compensation ansatz for illumination changes is that our model can potentially cope with any (smooth) appearance changes. This is not the case for invariance-based models, where the choice of one particular invariance decides which appearance changes can be tackled, and which not. In that sense, our estimation framework is very flexible and can adapt to any situation, whereas invariances lead to a much more rigid constancy assumption with fewer degrees of freedom.

This additional degree of freedom also represents the main disadvantage of this framework, as it leads to an additional parameter that has to be chosen adequately.

Chapter 4

Optic Flow Scale Space

From a practical point of view, the most important parameters of the methods we have discussed so far are the various regularisation weights. Among those, the weight of the flow regularisation term has a crucial influence on the smoothness of the resulting optic flow field. The larger its value, the smoother will be the resulting flow. Such a behaviour is also well known in the context of image regularisation, and there exists a close relationship to parabolic image evolution equations that define so-called *image scale spaces* [Iijima, 1963; Iijima et al., 1973]. For the case of image scale spaces, the relations between the regularisation weight and the evolution time are very well understood [Scherzer and Weickert, 2000]. In the context of variational optic flow, however, the corresponding scale space evolution equations are not known yet. The purpose of this chapter is thus to derive these missing evolution equations for the optic flow scale space.

First, we give a reinterpretation of the classical variational methods of Horn and Schunck [1981] and of Nagel and Enkelmann [1986] as Whittaker-Tikhonov regularisations of the normal flow. We show that this requires to replace the Euclidean norm by a space-variant matrix-induced norm that respects the data constraints.

Then, we generalise this framework to a broader class of methods that also allows to come up with new models that have not been considered before. We show that they can offer better performance than the classical variational methods.

After that, we will make the transition from the regularisation framework to a scale-space representation. This leads to the novel concept of *optic flow scale-spaces*. They are parabolic evolutions of vector-valued data with the regularisation parameter as scale and the normal flow as initial state. However, we will see that there are important differences to many image scale-spaces: The optic flow scale spaces are not of divergence type and hence do not preserve the average value of the initial data. Moreover, due to the matrix-weighted norm, they turn out to be highly anisotropic.

Finally, we exploit the optic flow scale space evolution to automatically select the best scale that gives the most accurate optic flow field. As a parameter-free scale selection principle we employ the *Optimal Prediction Principle*, which is specifically tailored to the needs of optic flow estimation [Zimmer et al., 2011b].

This chapter bases on the publication [Demetz et al., 2011] published at the conference on Scale Space and Variational Methods in Computer Vision.

Related Work

The earliest occurrences of Gaussian scale-space theory go back to Iijima's pioneering work and its use in optical character recognition many decades ago [Iijima, 1963; Iijima et al., 1973]. Since then, scale-spaces have become versatile tools for analysing and understanding the multiscale structure of images; see e.g. the monographs [Florack, 1997; Lindeberg, 1994; Sporring et al., 1997; Weickert, 1998] and the references therein. While partial differential equations (PDEs) of evolution type provide a natural framework for most scale-space concepts [Alvarez et al., 1993], it has also been shown that variational regularisation methods create scale-spaces where the regularisation parameter acts as scale [Scherzer and Weickert, 2000].

The transformation we apply is related to a proposal by Schnörr [1993], who did not pursue this concept further. With respect to the interpretation of variational methods in terms of specific norms, vector spaces, and higher order manifolds, there is a huge amount of literature available; see e.g. Sochen et al. [2001] and the references therein. Particularly interesting in this context is the work of Ben-Ari and Sochen [2009] who derive a class of smoothness terms based on spatially varying norms induced by suitable embeddings of the flow field into higher dimensional vector spaces. Scale selection is a classical issue in Gaussian scale-space theory [Lindeberg, 1994]. More specifically, choosing optimal smoothness parameters is an enduring problem for almost all classes of scale-space and variational methods. In our context, the works by Krajsek and Mester [2006], Mrázek and Navara [2003] and the recent ideas by Zimmer et al. [2011b] are most relevant. While there has been some research on scale spaces for image sequences [Fagerström, 2007; Guichard, 1998; Lindeberg, 2013; Laptev et al., 2007], to our knowledge the concept of optic flow scale space has not been considered before.

4.1 Variational Optic Flow as Whittaker-Tikhonov Regularisation

As starting point of our derivations of an optic flow scale space we consider the classic method of Horn and Schunck [1981]. This variational method estimates the optic flow field $\boldsymbol{u} := (u, v)^\top = (u(x, y, z), v(x, y, z))^\top$ as the minimiser of the energy functional

$$E(\boldsymbol{u}) = \int_\Omega \Big((f_x u + f_y v + f_z)^2 + \alpha \left(\|\boldsymbol{\nabla} u\|^2 + \|\boldsymbol{\nabla} v\|^2 \right) \Big) \,\mathrm{d}\boldsymbol{x} \;, \tag{4.1}$$

where, as before, $\Omega \subset \mathbb{R}^2$ represents the spatial image domain, $f : \Omega \times [0, \infty) \to \mathbb{R}$ the grey value image sequence, $\| \cdot \|$ stands for the Euclidean norm, subscripts denote partial derivatives, and $\boldsymbol{\nabla} = (\partial_x, \partial_y)^\top$ is the spatial gradient operator.

The data term of the latter functional models the already linearised assumption that corresponding pixels in subsequent frames have similar grey value. Since it depends on the two unknown functions u and v, its solution is under-determined. Evidently, in order to find a unique solution, additional assumptions on u and v are needed. This is realised by the second term – the so-called smoothness term. It penalises variations of the solution and is weighted by the positive regularisation parameter α.

4.1.1 Regularisation in a Spatially Varying Norm

In the following, we consider a slightly modified version of the energy (4.1) with the additional terms $\epsilon^2(u^2 + v^2) + c$, where ϵ is a small positive constant and $c(x, y, z) = -\epsilon f_z^2 / (|\boldsymbol{\nabla} f|^2 + \epsilon^2)$ is a function which does not depend on the unknown and thus does not play a role in the actual minimisation. Later on, these terms will be useful for theoretical reasons. The modified energy then reads as

$$E(\boldsymbol{w}) = \int_\Omega \Big((f_x u + f_y v + f_z)^2 + \epsilon^2 (u^2 + v^2) + c + \alpha \left(\|\boldsymbol{\nabla} u\|^2 + \|\boldsymbol{\nabla} v\|^2 \right) \Big) \,\mathrm{d}\boldsymbol{x} \;. \tag{4.2}$$

Let us now reformulate the latter energy in an image regularisation framework. This will allow us to obtain a different much more intuitive understanding of the underlying variational model. Since a smoothness term is already present, the main task is now to derive a suitable similarity term. To this end, we make use of the following result (see also Abhau et al. [2009]): Let $\boldsymbol{u}_{n,\epsilon}$ be the regularised normal flow, given by

$$\boldsymbol{u}_{n,\epsilon} = \frac{-f_z \boldsymbol{\nabla} f}{|\boldsymbol{\nabla} f|^2 + \epsilon^2} \;, \tag{4.3}$$

and let $\boldsymbol{A} : \Omega \to \mathbb{R}^{2\times 2}$ be a symmetric positive definite matrix in every point of the image domain defined as

$$\boldsymbol{A}^2 = \boldsymbol{\nabla} f \boldsymbol{\nabla} f^\top + \epsilon^2 \boldsymbol{I} \ , \tag{4.4}$$

where $\boldsymbol{I}$ denotes the unit matrix. Then the following equivalence holds:

$$(f_x u + f_y v + f_z)^2 + \epsilon^2 (u^2 + v^2) + c = (\boldsymbol{u} - \boldsymbol{u}_n)^\top \boldsymbol{A}^2 \, (\boldsymbol{u} - \boldsymbol{u}_n) \ . \tag{4.5}$$

This can easily be verified by straightforward calculations, see Appendix B. Essentially this means that the data term of the modified functional (4.2) can be rewritten to a quadratic form. The concept of matrix-weighted norms allows us then to make the similarity the functional to an image regularisation-like energy [Bertero et al., 1988] obvious:

$$E(\boldsymbol{u}) = \int_\Omega \Big(\, \|\boldsymbol{u} - \boldsymbol{u}_n\|^2_{\boldsymbol{A}^2} + \alpha \left(\|\boldsymbol{\nabla} u\|^2 + \|\boldsymbol{\nabla} v\|^2 \right) \Big) \ \mathrm{d}\boldsymbol{x} \ . \tag{4.6}$$

In this context, for a symmetric positive definite matrix $\boldsymbol{M}$ the corresponding matrix-weighted norm is given by $\|\boldsymbol{x}\|^2_{\boldsymbol{M}} := \langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\boldsymbol{M}} = \boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{x}$. Note that due to the additional term in (4.2), the space-variant matrix $\boldsymbol{A}$ fulfils these requirements by construction everywhere.

Having performed the previous rewritings, the following insight becomes explicit: Essentially, the seminal variational optic flow method of Horn and Schunck can be interpreted as Whittaker-Tikhonov regularisation of the normal flow in a matrix-weighted spatially varying norm.

4.1.2 Analysis of the Matrix-Weighted Norm

The rewritten data term favours solutions that are similar to the regularised normal flow. However, the actual deviation is evaluated in rotated and rescaled coordinate system that is determined by the constraint matrix $\boldsymbol{A}$. Since the eigensystem of the latter matrix is obvious, we can easily analyse and understand the effect of the introduced matrix-weighted norm. The eigendecomposition of $\boldsymbol{A}^2$ given by

$$\boldsymbol{A}^2 = \left(|\boldsymbol{\nabla} f|^2 + \epsilon^2 \right) \frac{\boldsymbol{\nabla} f}{|\boldsymbol{\nabla} f|} \frac{\boldsymbol{\nabla} f^\top}{|\boldsymbol{\nabla} f|} \ + \ \epsilon^2 \, \frac{\boldsymbol{\nabla} f^\perp}{|\boldsymbol{\nabla} f|} \frac{\boldsymbol{\nabla} f^{\perp^\top}}{|\boldsymbol{\nabla} f|} \quad . \tag{4.7}$$

Thus, we can express the complete data term as

$$\|\boldsymbol{u} - \boldsymbol{u}_n\|^2_{\boldsymbol{A}^2} = \left(|\boldsymbol{\nabla} f|^2 + \epsilon^2 \right) \left\langle \boldsymbol{u} - \boldsymbol{u}_n, \frac{\boldsymbol{\nabla} f}{|\boldsymbol{\nabla} f|} \right\rangle^2 + \epsilon^2 \left\langle \boldsymbol{u} - \boldsymbol{u}_n, \frac{\boldsymbol{\nabla} f^{\ \perp}}{|\boldsymbol{\nabla} f|} \right\rangle^2 \ . \tag{4.8}$$

This shows that the central quantity being under consideration is the normal flow $\boldsymbol{u}_n$, or rather the difference between the actual solution and the normal flow $\boldsymbol{u}-\boldsymbol{u}_n$. This difference vector is then projected into the local eigensystem of $\boldsymbol{A}^2$. There, its component perpendicular to the image gradient is basically negligible (since ϵ^2 is small), while its component along the image gradient is the part that actually contributes.

This also confirms the classic explanation of the linearised grey value constancy assumption as a constraint line: The expression $f_x u + f_y v + f_t = 0$ defines a line perpendicular to the image gradient with distance $f_t / |\boldsymbol{\nabla} f|$ from the origin Horn and Schunck [1981]. Also the findings in Zimmer et al. [2011b] are in accordance with this interpretation: There, the authors rewrite the assumption into a projection of the difference vector $\boldsymbol{w} - \boldsymbol{w}_n$ onto the image gradient, which exactly comes down to our locally adapted norm for $\epsilon = 0$.

4.2 Generalisations

So far, our reformulation of the model of Horn and Schunck [1981] has only identified the matrix-induced norm $\|\cdot\|_{\boldsymbol{A}^2}$ to play a central role. Consequently, we now propose to generalise this idea to a class of norms with varying anisotropy. To this end, we alter the exponent of the constraint matrix inducing the norm. This yields a data term of the general form

$$\|\boldsymbol{u} - \boldsymbol{u}_n\|^2_{\boldsymbol{A}^{2-\beta}} \ , \tag{4.9}$$

with $0 \leq \beta \leq 2$. While theoretically possible, values of β beyond 2 do not make sense, since then the anisotropy of the norm would be inverted, i.e. the penalisation directions would be swapped. Our parametrisation of the norm has been chosen such that for $\beta = 0$ the original model of Horn and Schunck, and for $\beta = 2$ a pure decoupled vector-valued regularisation of the normal flow in the Euclidean norm is obtained (since the constraint matrix collapses to the identity matrix, i.e. $\boldsymbol{A}^{2-2} = \boldsymbol{I}$).

To establish a consistently extended model, we also equip the smoothness term with the same spatially varying norm. This leads to the regulariser

$$\|\boldsymbol{\nabla} u\|^2_{\boldsymbol{A}^{-\gamma}} + \|\boldsymbol{\nabla} v\|^2_{\boldsymbol{A}^{-\gamma}} \, , \tag{4.10}$$

where $\gamma \geq 0$. This generally anisotropic image-driven regulariser allows variations of the flow field across image edges but not along them. In the special case of $\gamma = \beta = 0$ the model corresponds to Whittaker-Tikhonov regularisation [Whittaker, 1923; Tikhonov, 1963] as used by Horn and Schunck [1981].

Nagel and Enkelmann. Another special case of our generalised model is obtained for $\gamma = 2$ and $\beta = 0$: Then, our method is very closely related to the method of Nagel and Enkelmann [1986], as the regularisation term of both methods can be written as

$$\boldsymbol{\nabla} u^\top \boldsymbol{D}\, \boldsymbol{\nabla} u + \boldsymbol{\nabla} v^\top \boldsymbol{D}\, \boldsymbol{\nabla} v\,, \tag{4.11}$$

where the eigenvectors of the matrix $\boldsymbol{D}$ coincide for both methods, and only the corresponding eigenvalues differ slightly:

$$\boldsymbol{D} = \left(\frac{\boldsymbol{\nabla} f}{|\boldsymbol{\nabla} f|}\frac{\boldsymbol{\nabla} f^\perp}{|\boldsymbol{\nabla} f^\perp|}\right)^\top \begin{pmatrix}\mu_1 & 0\\ 0 & \mu_2\end{pmatrix} \left(\frac{\boldsymbol{\nabla} f}{|\boldsymbol{\nabla} f|}\frac{\boldsymbol{\nabla} f^\perp}{|\boldsymbol{\nabla} f^\perp|}\right)\,. \tag{4.12}$$

For the original method of Nagel and Enkelmann [1986] we have the following eigenvalues:

$$\mu_1 = \frac{\epsilon^2}{|\boldsymbol{\nabla} f|^2 + 2\epsilon^2} \quad, \text{ and } \quad \mu_2 = \frac{|\boldsymbol{\nabla} f|^2 + \epsilon^2}{|\boldsymbol{\nabla} f|^2 + 2\epsilon^2}\,, \tag{4.13}$$

and for our method, the following eigenvalues result:

$$\mu_1 = \frac{1}{|\boldsymbol{\nabla} f|^2 + \epsilon^2} \quad, \text{ and } \quad \mu_2 = \frac{1}{\epsilon^2}\,. \tag{4.14}$$

By rescaling the smoothness weight of our method by the factor ϵ^2, we effectively rescale the eigenvalues. Then, the close similarity becomes obvious.

Recall that the proposed spatially varying norm naturally arises from the linearised constancy assumption in the data term. In this way our approach differs significantly from the work in Ben-Ari and Sochen [2009], which derives such a norm by embedding the flow in a higher dimensional vector space and thus disregards the data term throughout the derivation. Incorporating both generalisations, we finally consider the energy functional

$$E(\boldsymbol{u}) = \int_\Omega \Big(\, \|\, \boldsymbol{u} - \boldsymbol{u}_n \,\|^2_{\boldsymbol{A}^{2-\beta}} + \alpha \left(\|\boldsymbol{\nabla} u\|^2_{\boldsymbol{A}^{-\gamma}} + \|\boldsymbol{\nabla} v\|^2_{\boldsymbol{A}^{-\gamma}} \right) \Big)\, \mathrm{d}x\, \mathrm{d}y\ , \tag{4.15}$$

with α, β and γ as defined before. This energy functional forms the basis for our optic flow scale space introduced in the next section.

4.3 Optic Flow Scale Space

Let us now derive the actual evolution equations. The Euler-Lagrange equations associated with our functional (4.15) read

$$\boldsymbol{A}^{2-\beta} \cdot (\boldsymbol{u} - \boldsymbol{u}_n) - \alpha \cdot \begin{pmatrix} \operatorname{div}\left(\boldsymbol{A}^{-\gamma}\ \boldsymbol{\nabla} u\right) \\ \operatorname{div}\left(\boldsymbol{A}^{-\gamma}\ \boldsymbol{\nabla} v\right) \end{pmatrix} = 0\ . \tag{4.16}$$

with reflecting Neumann boundary conditions

$$\boldsymbol{n}^\top \boldsymbol{A}^{-\gamma}\, \boldsymbol{\nabla} u = 0 \quad \text{and} \quad \boldsymbol{n}^\top \boldsymbol{A}^{-\gamma}\, \boldsymbol{\nabla} v = 0 \tag{4.17}$$

Note that Equation (4.16) is vector-valued. In the next step, subtract the divergence terms and *divide* the factor $\boldsymbol{A}^{2-\beta}$ to the other side. This is valid since $\boldsymbol{A}$ is invertible by construction. We obtain:

$$\frac{\boldsymbol{u} - \boldsymbol{u}_n}{\alpha} \;=\; \boldsymbol{A}^{\beta-2} \begin{pmatrix} \operatorname{div}\left(\boldsymbol{A}^{-\gamma}\, \boldsymbol{\nabla} u\right) \\ \operatorname{div}\left(\boldsymbol{A}^{-\gamma}\, \boldsymbol{\nabla} v\right) \end{pmatrix}, \tag{4.18}$$

In this form, the relations to an implicit time discretisation of the following filter are obtious:

$$\boxed{\partial_t\, \boldsymbol{u} \;=\; \boldsymbol{A}^{\beta-2} \begin{pmatrix} \operatorname{div}(\boldsymbol{A}^{-\gamma}\, \boldsymbol{\nabla} u) \\ \operatorname{div}(\boldsymbol{A}^{-\gamma}\, \boldsymbol{\nabla} v) \end{pmatrix},} \tag{4.19}$$

with a single time step of size α, and the normal flow $\boldsymbol{u}_n$ as initial state at time $t = 0$:

$$\boldsymbol{u}(\,\cdot\,, 0) = \boldsymbol{u}_n\,. \tag{4.20}$$

Obviously this temporal evolution constitutes a scale space, whose evolution time t coincides with the regularisation parameter α of the associated energy. Interestingly, the initial state of our *optic flow scale space* is the regularised normal flow, which is the only component of the flow field that can be directly extracted from the image data.

Note that we have transformed the *regularisation-like* energy functional (4.15) into a *diffusion-like* coupled system of parabolic PDEs (4.19). In the context of image filtering, relations between such methods have been investigated by Scherzer and Weickert [2000].

4.4 Numerical Realisation

For solving the parabolic problem in (4.19) we use an explicit scheme. To this end we discretise the two flow components u and v by sampling them on a regular grid and stacking all rows in single vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N$, where N denotes the number of pixels. Using this single-index notation, we discretise the matrix $\boldsymbol{A}^{\beta-2}$ in pixel i by

$$\boldsymbol{A}_i^{\beta-2} = \begin{pmatrix} a_i & b_i \\ b_i & c_i \end{pmatrix}, \quad i = 1, \ldots, N\,. \tag{4.21}$$

Accordingly, we discretise the diffusive terms using finite differences and obtain a nonadiagonal diffusion matrix $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ (similar to Weickert [1998]). This leads to the following explicit scheme:

$$\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{pmatrix}^{k+1} = \left(\boldsymbol{I}+\tau\left(\begin{array}{ccc|ccc} a_1 & & & b_1 & & \\ & \ddots & & & \ddots & \\ & & a_N & & & b_N \\ \hline b_1 & & & c_1 & & \\ & \ddots & & & \ddots & \\ & & b_N & & & c_N \end{array}\right)\left(\begin{array}{c|c} \boldsymbol{D} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{D}\end{array}\right)\right)\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{pmatrix}^{k}, \tag{4.22}$$

where every iteration advances the evolution by the time step size $\tau > 0$. Thus, after k iterations $(\boldsymbol{u}^k, \boldsymbol{v}^k)^\top$ contains the flow field at scale $\alpha = k\cdot\tau$. As a consequence, this explicit scheme inherently samples the whole scale space up to the stopping time α in intervals of size τ. Thereby, the whole iteration matrix remains constant for all iterations, since all terms are exclusively image-driven.

In order to accelerate this explicit scheme, we made use of the *fast explicit diffusion (FED)* strategy Grewenig et al. [2010], which performs cycles of explicit iterations with varying time step sizes. In particular, up to 50% of the step sizes can exceed the stability limit significantly, while the overall process remains provably stable. By that, the order of the smoothing time reached in n steps can be increased from $O(n)$ to $O(n^2)$. In our case, this is very beneficial, since the maximal stable time step size of the explicit scheme can decrease drastically with increasing anisotropy or small choices of the parameter β.

4.5 Optimal Scale Selection

In the previous sections, we have set up a novel class of optic flow scale spaces which all evolve in the regularisation parameter α.

Evidently, in the optic flow setting there exists one distinct scale within each scale space that provides the flow with the highest accuracy. Since the optic flow on all scales is available by construction, we have access to the deep structure of this scale space and can exploit this information to perform an automatic scale selection. To this end, we adapt the *Optimal Prediction Principle* (OPP) of Zimmer et al. [2011b]. In short, this principle suggests to rate the quality of an optic flow field between the first and second frame according to its extrapolation quality from the first to the third frame. The underlying assumption is that the velocity of objects (or the camera) remains constant over time. Zimmer et al. [2011b] show that this simple assumption works very well for the automatic estimation of the smoothness weight α.

In our case, for a given flow field $\boldsymbol{u}=(u,v)^\top$ between the frames at time z and $z+1$, we assess the extrapolation quality by evaluating the *Average Data Constancy Error* (*ADCE*), which is based on the grey value constancy assumption without linearisation:

$$ADCE_{1,3}(\boldsymbol{u}) = \frac{1}{|\Omega|} \int_\Omega \Big(f(x+2u, y+2v, z+2) - f(x,y,z) \Big)^2 \,\mathrm{d}\boldsymbol{x} \,. \quad (4.23)$$

It is obvious that if the model assumptions hold, a *good* flow field will lead to *small* values of this error measure. Note that in contrast to the optimisation strategy of Zimmer et al. [2011b], we can exploit the following advantageous property of our numerical scheme: It explicitly evolves in the parameter α, hence after each iteration the flow field at cumulated time α is available, and the *ADCE* can be evaluated. This on-the-fly computation of the quality estimate is not possible for most other optic flow methods, because they typically require to solve a new system of equations for each value of α.

Besides the OPP, we also tried other schemes for automatic scale estimation. In particular, we investigated the performance of the decorrelation method of Mrázek and Navara [2003]. However, experiments indicated that the underlying assumptions do not hold for our optic flow scale space.

4.6 Experiments

In order to investigate the behaviour of our optic flow scale space we perform experiments on several image sequences that are publicly available and for which the ground truth flow field is known. In particular, we use the *Yosemite* sequence with and without clouds [Barron et al., 1994], the *New Marble*[1] sequence as well as the *Rubberwhale* sequence [Baker et al., 2011].

In our first experiment we compute samples of the scale space for the New Marble sequence at different evolution times and for several choices of β and γ. Figure 4.1 shows the corresponding flow fields, where colour encodes the direction and brightness indicates the magnitude of the displacements. Here, one can clearly see the scale space behaviour of the proposed diffusion-like optic flow process: Independently of β and γ, the initial state of all these scale spaces ($\alpha = 0$) is given by the noisy normal flow, while for larger values of α the flow fields become successively smoother. In this context, we make two observations: On the one hand, for $\gamma > 0$ discontinuities are preserved

[1] available from `http://i21www.ira.uka.de/image_sequences`

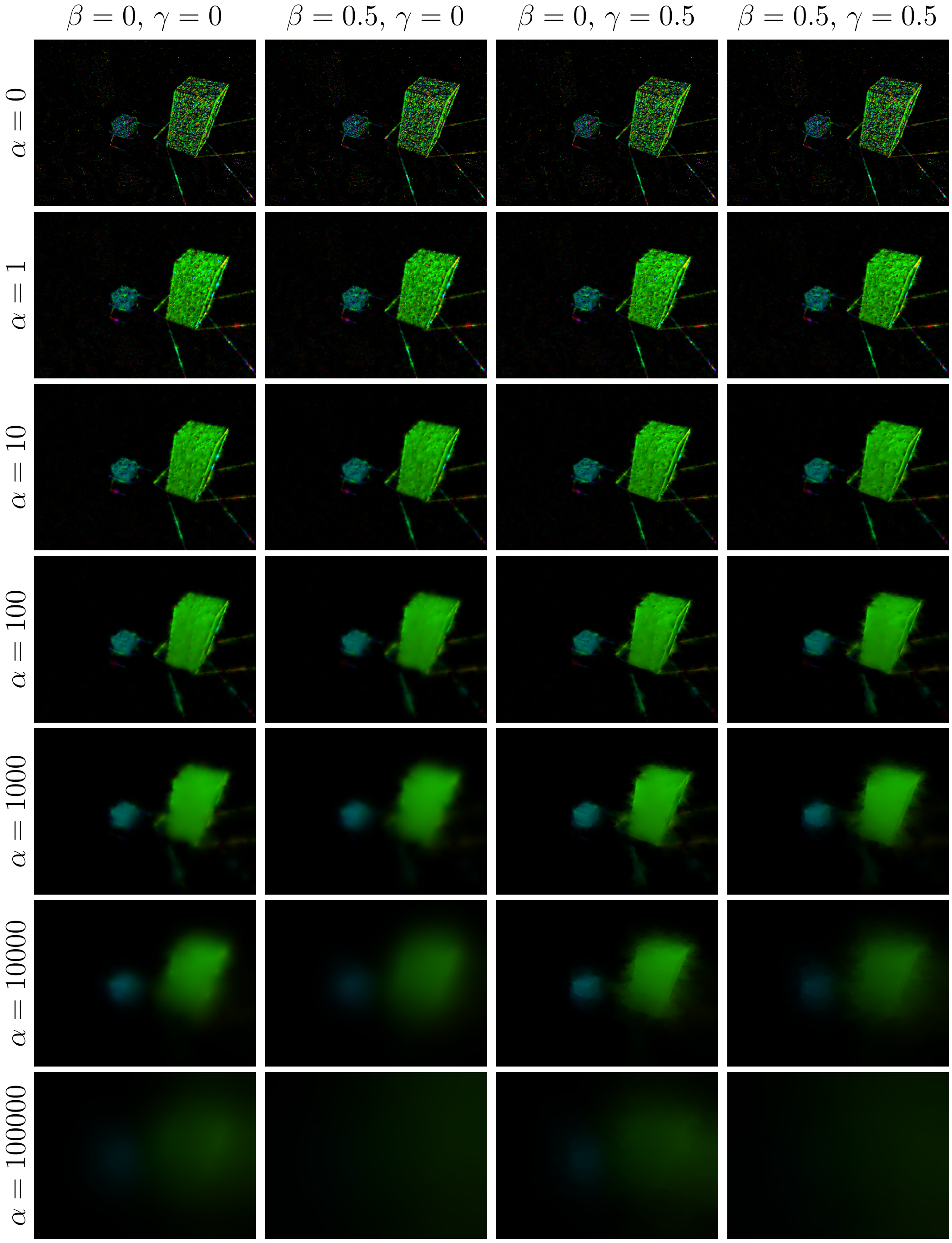


Figure 4.1: Scale space at different stopping times for varying β and γ.

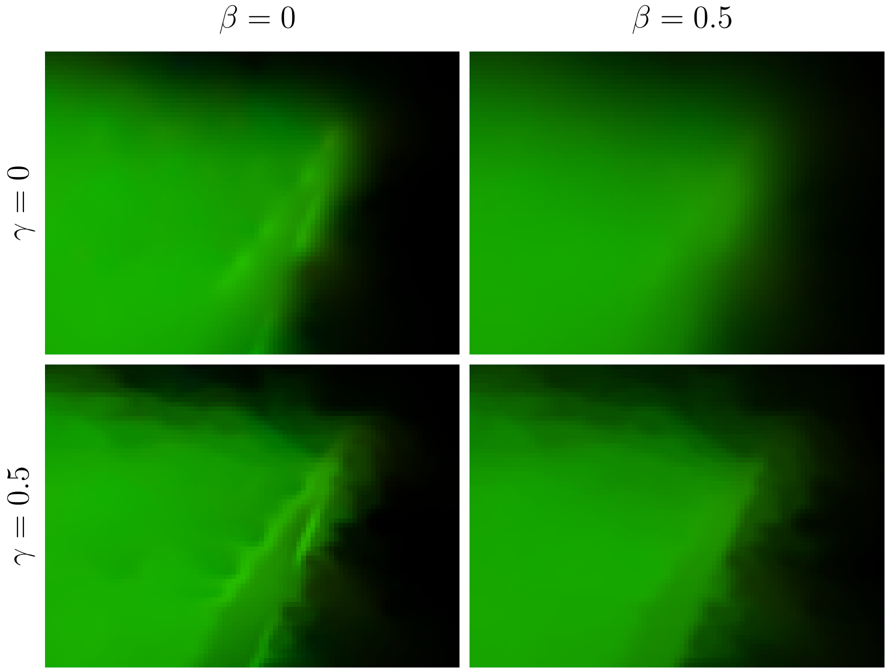


Figure 4.2: Zoom into the optic flow fields at scale $\alpha = 1000$ from Figure 4.1.

Table 4.1: Quantitative error measurements in terms of the *AAE* on different image sequences. For our method, β and γ have been optimised. The actual choices are given in brackets.

Image sequence	Horn / Schunck	Nagel / Enkelmann	Our method
New Marble	2.65	2.77	2.53 ($\beta=1.0$, $\gamma=0.4$)
Rubberwhale	10.58	9.27	9.04 ($\beta=0.5$, $\gamma=1.0$)
Yosemite	7.51	6.86	6.41 ($\beta=0.4$, $\gamma=0.9$)
Yosemite no clouds	2.82	3.63	2.76 ($\beta=0.2$, $\gamma=0.1$)

for a longer time, since the regulariser is then of image-driven anisotropic nature. On the other hand, results for $\beta > 0$ are slightly less noisy, since the larger eigenvalue of $\boldsymbol{A}^{2-\beta}$ – the one that depends on the magnitude of the image gradient – is now subject to a smaller exponent, cf. Equation (4.8). This becomes particularly visible in the magnifications shown in Figure 4.2.

In a second experiment, we compare the accuracy of the proposed scheme against the two special cases in our framework: Horn and Schunck ($\beta=\gamma=0$) and Nagel and Enkelmann ($\beta=0$, $\gamma=2$). This is done for the aforementioned image sequences by means of the *Average Angular Error* (*AAE*) Barron et al. [1994]. It should be noted that we keep the presmoothing scale fixed at $\sigma=1$ throughout all experiments, since its impact is not in the focus of our contribution. Table 4.1 demonstrates that our method consistently leads to improved results. In particular, it shows that the additional degrees of freedom β and γ provided by our general class of scale spaces can be beneficial.

In our third experiment, we investigate the automatic selection of the scale parameter α of our model using the OPP. To this end, we first juxtapose the graph of the estimated quality in terms of the *ADCE* with the graph of the measured accuracy given by the *AAE* in Figure 4.3 (a). This is done for the Yosemite sequence with clouds with $\beta=\gamma=0.5$. One can see that both graphs have a similar and well aligned shape. In particular, the minima of both curves are attained at almost the same position. Secondly, we compare the estimated values for the regularisation parameter α against those that are optimal with respect to the *AAE.* This is done for all four image sequences with $\beta=\gamma=0.5$ fixed. As one can see from Figure 4.3 (b), the OPP works very well in practice: In all cases, the *AAE* at the estimated scale is close to the one of the optimal scale.

In our final experiment we analyse how the two generalisation parameters

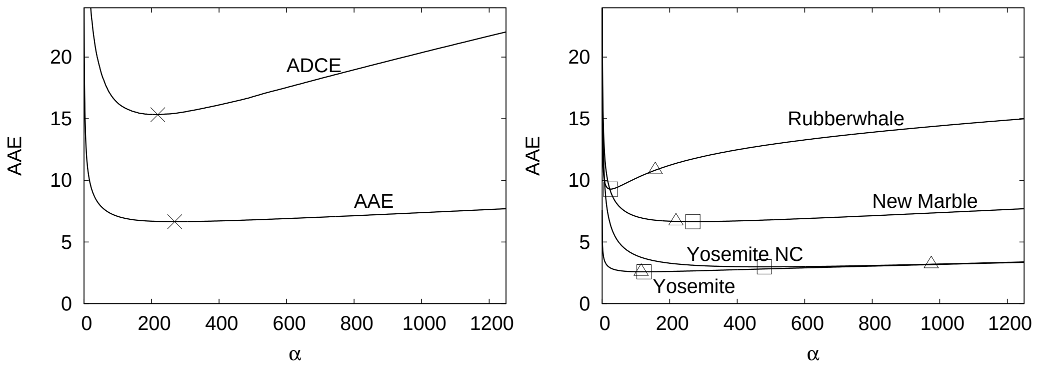


Figure 4.3: Automatic scale selection using the Optimal Prediction Principle. **(a) Left:** Graphs of the *ADCE* and *AAE* side-by side. Crosses denote the minimal value of the graph. **(b) Right:** Estimation results of the selected scale α for different image sequences. Triangles indicate the estimated value and squares denote the optimal choice.

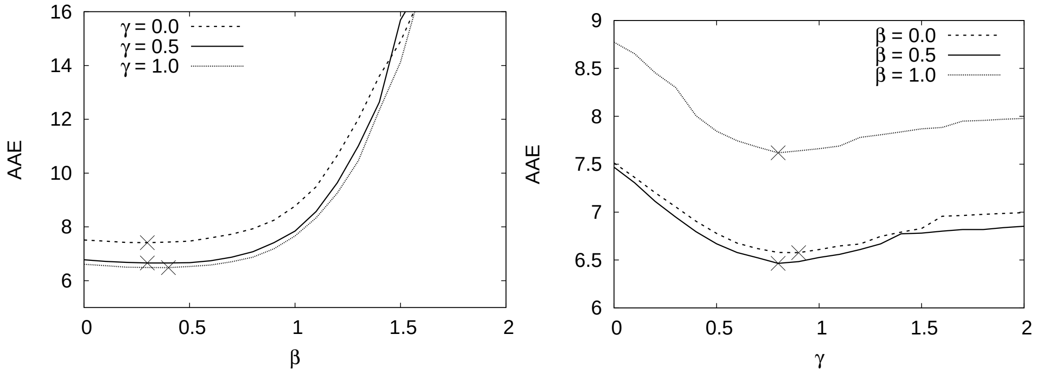


Figure 4.4: Influence of the parameters β and γ on the accuracy. **(a) Left:** Behaviour under variations of β for fixed γ. **(b) Right:** Ditto for varying γ and fixed β. In both cases, crosses indicate the minimum of the graphs.

β and γ influence the accuracy of the estimation. To this end, we have computed the *AAE* for $\beta, \gamma \in [0, 2]$ using the Yosemite sequence with clouds. As in the previous experiments, the selection of the stopping time α within each scale space has been performed automatically using the OPP. The resulting graphs in Figure 4.4 (a) and (b) show that for both parameters values larger than zero consistently improve the accuracy.

4.7 Summary

In this chapter we have developed and analysed the *optic flow scale space* (OFSS). As for traditional scale spaces, the OFSS originates from the regularisation parameter that virtually all variational optic flow methods have in common – larger values lead to smoother solutions. In order to derive the central evolution equations of the OFSS, we modified the original method of Horn and Schunk slightly such that we could transform the classical linearised grey value constancy assumption into a similarity term in a problem-specific space-variant norm. This allowed us to deduce the evolution equations of the optic flow scale space. Interestingly, the initial state of our evolution is the normal flow – an analogy to image scale spaces where the initial state is the input image. After that, we established connections to other optic flow methods by transferring the matrix weighted norm also to the smoothness term. Moreover, by varying the exponents of the constraint matrix, we were even able to derive novel models and corresponding scale spaces. In every iteration of our numerical scheme the iterate is a valid flow field which corresponds to the solution of the basic energy functional with the current process time as regularisation parameter. This fact is exploited by the optimal prediction principle that allows to determine an optimal scale heuristically. With our experiments we showed that the different generalisations in fact lead to different evolutions and that certain novel configurations allow to decrease the error.

Chapter 5

Conclusion and Outlook

5.1 Conclusion

The central topic of this thesis was optic flow, the 2-D displacement field that describes the motion of each position between consecutive frames of an image sequence. In particular, we concentrated on one of the major challenges of accurate and robust optic flow estimation: illumination and appearance changes. Such changes are a big problem since they contradict the traditional colour constancy assumption that many methods rely on. There are two generally different strategies to approach this problem, which we discussed in the second and third chapter of this work.

The prevailing way to handle illumination changes is via invariances. There exists a great variety of invariant features in the literature, on which we gave broad and structured overview in the first part of the second chapter. This discussion grouped all invariant features into classes of increasing invariance. The class with the highest degree of invariance we considered comprises transformations that are not affected by any monotonically increasing rescaling – so-called morphological invariance. For this class we proposed two novel features the complete rank transform and the complete census transform. These two transforms carry as much image information as possible in this class on invariance. Next, we presented a variational framework that allows to incorporate any of the discussed signatures and allows a fair comparison. Our extensive experiments evaluated the performance of the features among each other as well as among the state-of-the-art. We could show that our novel complete signatures are favorable against all other features. Moreover, the results on the public KITTI vision benchmark document that our resulting variational optic flow method competes with the state-of-the-art.

In general, the term *invariance* is considered a desirable property of computer vision algorithms. However, the closely related term *ignorance* has a much worse reputation: If a feature is invariant against a certain class of changes, it discards any information of that class – no matter if it is a nuisance due to lighting changes, or a part of valuable image information that should not be ignored. This change of perspective lead us to the idea that instead of ignoring some parts of the given information, we could try to estimate the appearance changes. To this end we developed a variational model in the third chapter of this dissertation that estimates optic flow and a so-called *brightness transfer function* (BTF) jointly. This model allowed us to express our assumptions about the flow as well as the changes in appearance of the scene mathematically: After compensating the first frame for the illumination changes with the BTF, the traditional colour constancy assumption should hold again everywhere. Moreover, both the optic flow as well as the BTF should exhibit some sort of spatial smoothness. We extended our model from the second chapter by these assumptions and showed experimentally that such a strategy is able to produce flow fields of highest accuracy.

Now that we have presented two alternatives to handle illumination changes, we are left with the question which of the two approaches is better. First of all, from a theoretical point of view, we think that the estimation scheme should be favorable, as there is no inherent limitation in the type of appearance change that could theoretically not be handled. The capability to potentially explain any plausible illumination change however comes at a price: there is an additional parameter that has to be chosen carefully. Moreover, the estimation scheme always runs the risk of attributing a change to the wrong cause – motion or illumination. Thus, the interplay between the data and regularisation terms is very complex. The invariance-based scheme does not have this degree of freedom, and only works with the part of the image information that the feature provides. While this limits the capabilities on the one hand, it also makes the parameter calibration much easier. One of our experiments illustrates this trade-off very well: For the frequently considered sequence #15 of the KITTI benchmark, the lowest error of 8.81% was achieved with our estimation framework. However, considering all 194 image sequences, the invariance-based framework with our CRT-based data term lead to the globally best result in this thesis. This illustrates that the estimation model is in fact capable to estimate highest accuracy – if its parameters are chosen optimally. Our CRT-based method however is more robust in this respect and reaches high accuracy without the need to fine-tune parameters for each new problem.

In the fourth chapter of this dissertation, we addressed another important component that practically all variational regularisation methods have in common: the smoothness parameter that weights the data against the smoothness assumption. In the context of image regularisation, close connections to scale spaces are well understood. However, for optic flow methods such an analysis had not been performed before. The key to our elegant formulation of the optic flow scale space was a small modification of the method of Horn and Schunck. This allowed us to reinterpret the data term as a similarity assumption in a space-variant norm. From this, we could derive parabolic evolution equations whose initial state is the normal flow. Analogously to image regularisation, the process time of our optic flow scale space evolution equations corresponds to the regularisation parameter of the optic flow functional.

5.2 Outlook

We have seen that the parameter choice is a major issue for the estimation scheme. One ad-hoc possibility to alleviate it would be to supplement a CRT-based data term. However, this can only lighten the problem, but cannot remove the degree of freedom completely. If time is not an issue, also the optimal prediction principle that was discussed in the last chapter might be helpful, however modifications w.r.t. illumination changes would be necessary, eventually via invariances.

Another interesting topic could be to alter our realisation of an illumination compensation via a local BTF into a global model that might be learned online.

The estimation scheme can be seen as first step towards finding an explanation of all changes in an image sequence. We did this in terms of a spatial and a tonal displacement that we estimated in each pixel. As a long term goal, one could think of going further and estimating richer models that include for instance geometry and reflectance. The hope would be that the parameters of such representations might be easier to calibrate. However, such schemes of course would loose general applicability and one has to be careful not to end up in a standard 3D reconstruction setting.

In the context of the optic flow scale space and the optimal prediciton principle, extending the applicability of our theoretical findings to advanced optic flow methods might be an interesting topic. Especially the application of the optimal prediction principle in each scale of the warping strategy might offer room for accuracy improvement.

Appendix A

Discretisation Details

The goal of this appendix is to sketch our discretisation of the coupled second order regularisation term that we use in our models from Chapters 2 and 3.

For the sake of simplicity, we consider the following regularisation energy functional, whose data term is much simpler than our terms for optic flow:

$$\begin{aligned} E(u,a,b) = &\underbrace{\int_\Omega \Psi_d\Big((u-f)^2\Big)\,\mathrm{d}\boldsymbol{x}}_{E_{\text{data}}} \\ &+\alpha\Bigg(\underbrace{\int_\Omega \Psi_c\Big(\Big\|\boldsymbol{\nabla} u - \tbinom{a}{b}\Big\|_2^2\Big)\,\mathrm{d}\boldsymbol{x}}_{E_{\text{coupl}}} + \beta \underbrace{\int_\Omega \Psi_s\Big(\|\boldsymbol{\nabla} a\|^2 + \|\boldsymbol{\nabla} b\|^2\Big)\,\mathrm{d}\boldsymbol{x}}_{E_{\text{smooth}}}\Bigg)\,, \end{aligned} \tag{A.1}$$

and where only one pair of auxiliary coupling variables $a, b : \Omega \rightarrow \mathbb{R}$ has to be considered. Our discretisation has one unknown for and u, a and b per pixel. The discretisation of the first term is trivial, we have:

$$E_{\text{data}}(\boldsymbol{u}) = \sum_{i=1}^{n}\sum_{j=1}^{m} \Psi_d\left(((u_{i,j} - f_{i,j})^2\right)\,, \tag{A.2}$$

where $\boldsymbol{u} \in \mathbb{R}^{n\times m}$ is the discetised version of u. To develop a discretised version of the second term, partial derivatives of u have to be approximated, which we do by averaging forward and backward finite differences, c.f. Weickert [1998]; Brox [2005]. We obtain:

$$
\begin{aligned}
E_{\text{coupl}}(\boldsymbol{u},\boldsymbol{a},\boldsymbol{b}) = \sum_{i=1}^{n}\sum_{j=1}^{m} \Psi_c \Bigg[& \frac{1}{2}\,\chi_{i,j}^{[1,n-1]\times[1,m]} \left(\frac{u_{i+1,j}-u_{i,j}}{h_x} - a_{i,j}\right)^2 \\
+ & \frac{1}{2}\quad \chi_{i,j}^{[2,n]\times[1,m]} \left(\frac{u_{i,j}-u_{i-1,j}}{h_x} - a_{i,j}\right)^2 \\
+ & \frac{1}{2}\,\chi_{i,j}^{[1,n]\times[1,m-1]} \left(\frac{u_{i,j+1}-u_{i,j}}{h_y} - b_{i,j}\right)^2 \\
+ & \frac{1}{2}\quad \chi_{i,j}^{[1,n]\times[2,m]} \left(\frac{u_{i,j}-u_{i,j-1}}{h_y} - b_{i,j}\right)^2 \Bigg],
\end{aligned}
\tag{A.3}
$$

where the indicator functions

$$
\chi_{i,j}^{a\times b} = \begin{cases} 1 & \text{if } i \in a \text{ and } j \in b \\ 0 & \text{else} \end{cases},
\tag{A.4}
$$

make sure we do not access the signal outside its domain. For the smoothness term on the coupling variables, we obtain in an analogous way for $\boldsymbol{a},\boldsymbol{b} \in \mathbb{R}^{n\times m}$:

$$
\begin{aligned}
E_{\text{smooth}}(\boldsymbol{u},\boldsymbol{a},\boldsymbol{b}) = \sum_{i=1}^{n}\sum_{j=1}^{m} & \\
\Psi_s \Bigg[& \frac{1}{2}\,\chi_{i,j}^{[1,n-1]\times[1,m]} \left(\frac{a_{i+1,j}-a_{i,j}}{h_x}\right)^2 + \frac{1}{2}\,\chi_{i,j}^{[2,n]\times[1,m]} \left(\frac{a_{i,j}-a_{i-1,j}}{h_x}\right)^2 \\
+ & \frac{1}{2}\,\chi_{i,j}^{[1,n]\times[1,m-1]} \left(\frac{a_{i,j+1}-a_{i,j}}{h_x}\right)^2 + \frac{1}{2}\,\chi_{i,j}^{[1,n]\times[2,m]} \left(\frac{a_{i,j}-a_{i,j-1}}{h_x}\right)^2 \\
+ & \frac{1}{2}\,\chi_{i,j}^{[1,n-1]\times[1,m]} \left(\frac{b_{i+1,j}-b_{i,j}}{h_x}\right)^2 + \frac{1}{2}\,\chi_{i,j}^{[2,n]\times[1,m]} \left(\frac{b_{i,j}-b_{i-1,j}}{h_x}\right)^2 \\
+ & \frac{1}{2}\,\chi_{i,j}^{[1,n]\times[1,m-1]} \left(\frac{b_{i,j+1}-b_{i,j}}{h_x}\right)^2 + \frac{1}{2}\,\chi_{i,j}^{[1,n]\times[2,m]} \left(\frac{b_{i,j}-b_{i,j-1}}{h_x}\right)^2 \Bigg].
\end{aligned}
\tag{A.5}
$$

The minimiser of the energy function is found by setting the derivative of E w.r.t. each $u_{i,j}$,$a_{i,j}$ and $b_{i,j}$ to zero and solving the resulting non-linear system of equations, c.f. Section 2.4.4. The correct boundary equations result naturally from the discretised energy.

Appendix B

Derivation of the Matrix-Weighted Norm

It remains to show that the data term of the modified functional from Equation (4.2) and the similarity term from Equation (4.6) are in fact equivalent. To this end, we transform both data terms to arrive at the same expression. We start with the modified Horn and Schunck data term:

$$\begin{aligned}
&(f_x u + f_y v + f_z)^2 + \epsilon^2(u^2+v^2) - \frac{\epsilon f_z^2}{|\boldsymbol{\nabla} f|^2+\epsilon^2} \\
&= (f_x^2+\epsilon^2)u^2 + 2 f_x f_y u\, v \qquad +2 f_x f_z u - \frac{\epsilon f_z^2}{|\boldsymbol{\nabla} f|^2+\epsilon^2} \\
&\qquad\qquad + (f_y^2+\epsilon^2)v^2 \qquad + f_y f_z v \\
&\qquad\qquad\qquad\qquad\qquad + f_z^2 \\
&= \boldsymbol{u}^\top \begin{pmatrix} f_x^2+\epsilon^2 & f_x f_y \\ f_x f_y & f_y^2+\epsilon^2 \end{pmatrix} \boldsymbol{u} + 2\boldsymbol{\nabla} f^\top \boldsymbol{u} f_z + f_z^2 (1 - \frac{\epsilon}{|\boldsymbol{\nabla} f|^2+\epsilon^2}) \\
&= \boldsymbol{u}^\top \boldsymbol{A}^2 \boldsymbol{u} + 2\boldsymbol{\nabla} f^\top \boldsymbol{u} f_z + f_z^2 \frac{|\boldsymbol{\nabla} f|^2}{|\boldsymbol{\nabla} f|^2+\epsilon^2}\,.
\end{aligned} \tag{B.1}$$

From the other end, using the abbreviation

$$\boldsymbol{u}_n = \frac{-f_z \boldsymbol{\nabla} f}{|\boldsymbol{\nabla} f|^2+\epsilon^2}\,, \tag{B.2}$$

for the regularised normal flow, we transform the similarity term as follows:

$$\begin{aligned}
&(\boldsymbol{u}-\boldsymbol{u}_n)^\top \boldsymbol{A}^2\,(\boldsymbol{u}-\boldsymbol{u}_n)\\
&=\boldsymbol{u}^\top\boldsymbol{A}^2\boldsymbol{u}-2\boldsymbol{u}^\top\boldsymbol{A}^2\boldsymbol{u}_n+\boldsymbol{u}_n^\top\boldsymbol{A}^2\boldsymbol{u}_n\\
&=\boldsymbol{u}^\top\boldsymbol{A}^2\boldsymbol{u}-2\boldsymbol{u}^\top(\boldsymbol{\nabla} f\boldsymbol{\nabla} f^\top+\epsilon^2\boldsymbol{I})\boldsymbol{u}_n+\boldsymbol{u}_n^\top\boldsymbol{\nabla} f\boldsymbol{\nabla} f^\top\boldsymbol{u}_n+\boldsymbol{u}_n^\top\epsilon^2\boldsymbol{u}_n\\
&=\boldsymbol{u}^\top\boldsymbol{A}^2\boldsymbol{u}-2\boldsymbol{u}^\top(\frac{-\boldsymbol{\nabla} f\boldsymbol{\nabla} f^\top\boldsymbol{\nabla} f f_z}{|\boldsymbol{\nabla} f|^2+\epsilon^2}+\frac{-\epsilon^2\boldsymbol{\nabla} f f_z}{|\boldsymbol{\nabla} f|^2+\epsilon^2})\\
&\qquad\qquad+\frac{-\boldsymbol{\nabla} f^\top f_z\boldsymbol{\nabla} f}{|\boldsymbol{\nabla} f|^2+\epsilon^2}\cdot\frac{-\boldsymbol{\nabla} f^\top\boldsymbol{\nabla} f f_z}{|\boldsymbol{\nabla} f|^2+\epsilon^2}+\epsilon^2\frac{-\boldsymbol{\nabla} f^\top\boldsymbol{\nabla} f f_z^2}{|\boldsymbol{\nabla} f|^2+\epsilon^2}\\
&=\boldsymbol{u}^\top\boldsymbol{A}^2\boldsymbol{u}-2\boldsymbol{u}^\top\frac{-f_z}{|\boldsymbol{\nabla} f|^2+\epsilon^2}(\boldsymbol{\nabla} f|\boldsymbol{\nabla} f|^2+\boldsymbol{\nabla} f\epsilon^2)\\
&\qquad\qquad+\frac{|\boldsymbol{\nabla} f|^4 f_z^2}{(|\boldsymbol{\nabla} f|^2+\epsilon^2)^2}+\epsilon^2\frac{|\boldsymbol{\nabla} f|^2 f_z^2}{(|\boldsymbol{\nabla} f|^2+\epsilon^2)^2}\\
&=\boldsymbol{u}^\top\boldsymbol{A}^2\boldsymbol{u}+2\boldsymbol{u}^\top f_z\boldsymbol{\nabla} f\frac{|\boldsymbol{\nabla} f|^2+\epsilon^2}{|\boldsymbol{\nabla} f|^2+\epsilon^2}+\frac{|\boldsymbol{\nabla} f|^2+\epsilon^2}{(|\boldsymbol{\nabla} f|^2+\epsilon^2)^2}(|\boldsymbol{\nabla} f|^2 f_z^2)\\
&=\boldsymbol{u}^\top\boldsymbol{A}^2\boldsymbol{u}+2\boldsymbol{u}^\top f_z\boldsymbol{\nabla} f+\frac{|\boldsymbol{\nabla} f|^2 f_z^2}{|\boldsymbol{\nabla} f|^2+\epsilon^2}\,.
\end{aligned}\tag{B.3}$$

As one can see, both computations result in the same expression, thus the equivalence holds. □

Own Publications

Demetz, O., Hafner, D., and Weickert, J. (2013). The complete rank transform: A tool for accurate and morphologically invariant matching of structures. In *Proceedings of the British Machine Vision Conference*. BMVA Press. **Awarded the Maria Petrou Prize for Invariance in Computer Vision.**

Demetz, O., Hafner, D., and Weickert, J. (2015). Morphologically invariant matching of structures with the complete rank transform. *International Journal of Computer Vision*, 113(3):220–232. Invited paper.

Demetz, O., Stoll, M., Volz, S., Weickert, J., and Bruhn, A. (2014). Learning brightness transfer functions for the joint recovery of illumination changes and optical flow. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision - ECCV 2014*, volume 8689, pages 455–471. Springer International Publishing.

Demetz, O., Weickert, J., Bruhn, A., and Welk, M. (2007). Beauty with variational methods: An optic flow approach to hairstyle simulation. In Sgallari, F., Murli, F., and Paragios, N., editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 825–836. Springer, Berlin.

Demetz, O., Weickert, J., Bruhn, A., and Zimmer, H. (2011). Optic flow scale space. In Bruckstein, A. M., ter Haar Romeny, B., Bronstein, A. M., and Bronstein, M. M., editors, *Scale-Space and Variational Methods in Computer Vision*, volume 6667 of *Lecture Notes in Computer Science*, pages 713–724. Springer.

Hafner, D., Demetz, O., and Weickert, J. (2013). Why is the census transform good for robust optic flow computation? In Kuijper, A., Pock, T., Bredies, K., and Bischof, H., editors, *Scale-Space and Variational Methods in Computer Vision*, volume 7893 of *Lecture Notes in Computer Science*, pages 210–221. Springer.

Hafner, D., Demetz, O., and Weickert, J. (2014). Simultaneous HDR and optic flow computation. In *Proc. IEEE International Conference on Pattern Recognition (ICPR)*, pages 2065–2070.

Hafner, D., Demetz, O., Weickert, J., and Reißel, M. (2015). Mathematical foundations and generalisations of the census transform for robust optic flow computation. *Journal of Mathematical Imaging and Vision*, 52(1):71–86.

Kramarev, V., Demetz, O., Schroers, C., and Weickert, J. (2013). Cross anisotropic cost volume filtering for segmentation. In Lee, K. M., Matsushita, Y., Rehg, J. M., and Hu, Z., editors, *Computer Vision - ACCV 2012*, volume 7724 of *Lecture Notes in Computer Science*, pages 803–814, Berlin. Springer.

Schroers, C., Zimmer, H., Valgaerts, L., Bruhn, A., Demetz, O., and Weickert, J. (2012). Anisotropic range image integration. In Pinz, A., Pock, T., Bischof, H., and Leberl, F., editors, *Pattern Recognition*, volume 7476 of *Lecture Notes in Computer Science*, pages 73–82, Berlin. Springer. **Awarded a DAGM-ÖAGM 2012 Paper Prize**.

Bibliography

Abhau, J., Belhachmi, Z., and Scherzer, O. (2009). On a decomposition model for optical flow. In Cremers, D., Boykov, Y., Blake, A., and Schmidt, F., editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 5681 of *Lecture Notes in Computer Science*, pages 126–139. Springer, Berlin. (Cited on page 125)

Acar, R. and Vogel, C. R. (1994). Analysis of bounded variation penalty methods for ill–posed problems. *Inverse Problems*, 10:1217–1229. (Cited on page 56)

Alvarez, L., Deriche, R., Papadopoulo, T., and Sánchez, J. (2007). Symmetrical dense optical flow estimation with occlusions detection. *International Journal of Computer Vision*, 75(3):371–385. (Cited on page 17)

Alvarez, L., Deriche, R., Weickert, J., and Sanchez, J. (2002). Dense disparity map estimation respecting image discontinuities: A PDE and scale-space based approach. *Journal of Visual Communication and Image Representation, Special Issue on Partial Differential Equations in Image Processing, Computer Vision and Computer Graphics*, 13(1/2):3–21. (Cited on page 60)

Alvarez, L., Esclarín, J., Lefébure, M., and Sánchez, J. (1999a). 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. (Cited on page 58)

Alvarez, L., Guichard, F., Lions, P.-L., and Morel, J.-M. (1993). Axioms and fundamental equations in image processing. *Archive for Rational Mechanics and Analysis*, 123:199–257. (Cited on pages 35 and 124)

Alvarez, L., Weickert, J., and Sánchez, J. (1999b). A scale-space approach to nonlocal optical flow calculations. In Nielsen, M., Johansen, P., Olsen,

O. F., and Weickert, J., editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 235–246. Springer, Berlin. (Cited on page 18)

Anandan, P. (1989). A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310. (Cited on pages 18 and 59)

Aujol, J.-F., Gilboa, G., Chan, T., and Osher, S. (2006). Structure-texture image decomposition–modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136. (Cited on page 30)

Ayvaci, A., Raptis, M., and Soatto, S. (2012). Sparse occlusion detection with optical flow. *International Journal of Computer Vision*, 97(3):322–338. (Cited on page 92)

Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J., and Szeliski, R. (2011). A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31. (Cited on pages 15, 70, 72, 74, 85, 111, and 131)

Barnes, C., Shechtman, E., Goldman, D. B., and Finkelstein, A. (2010). The generalized PatchMatch correspondence algorithm. In Daniilidis, K., Maragos, P., and Paragios, N., editors, *Computer Vision – ECCV 2010*, volume 6312 of *Lecture Notes in Computer Science*, pages 29–43. Springer. (Cited on page 93)

Barron, J. L., Fleet, D. J., and Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77. (Cited on pages 22, 131, and 134)

Barrow, H. G. and Tenenbaum, J. M. (1978). Recovering intrinsic scene characteristics from images. *Computer Vision Systems*, pages 3–26. (Cited on page 93)

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer Vision And Image Understanding*, 110(3):346–359. (Cited on page 34)

Belhumeur, P. N. and Kriegman, D. (1998). What is the set of images of an object under all possible lighting conditions? *International Journal of Computer Vision*, 28(3):245–260. (Cited on page 94)

Ben-Ari, R. and Sochen, N. (2009). A geometric framework and a new criterion in optical flow modeling. *Journal of Mathematical Imaging and Vision*, 33:178–194. (Cited on pages 124 and 128)

Bertero, M., Poggio, T. A., and Torre, V. (1988). Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889. (Cited on page 126)

Bhat, D. N. and Nayar, S. K. (1998). Ordinal measures for image correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):415–423. (Cited on page 40)

Black, M. J. and Anandan, P. (1991). Robust dynamic motion estimation over time. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 292–302. (Cited on pages 18, 53, and 56)

Black, M. J., Fleet, D. J., and Yacoob, Y. (2000). Robustly estimating changes in image appearance. *Computer Vision and Image Understanding*, 78(1):8–31. (Cited on page 94)

Blake, A. and Zisserman, A. (1987). *Visual Reconstruction.* MIT Press, Cambridge, MA. (Cited on page 56)

Blomgren, P. and Chan, T. (1998). Color tv: total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309. (Cited on page 30)

Braux-Zin, J., Dupont, R., and Bartoli, A. (2013). A general dense image matching framework combining direct and feature-based costs. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 185–192. IEEE Press. (Cited on pages 35, 37, 58, 63, 89, and 120)

Bredies, K., Kunisch, K., and Pock, T. (2010). Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526. (Cited on pages 57 and 97)

Brox, T. (2005). *From Pixels to Regions: Partial Differential Equations in Image Analysis.* PhD thesis, Faculty of Mathematics and Computer Science, Saarland University, Germany. (Cited on page 142)

Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In Pajdla, T. and Matas, J., editors, *Computer Vision – ECCV 2004*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer. (Cited on pages 18, 49, 56, 59, 61, 62, 87, and 96)

Brox, T. and Malik, J. (2011). Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):500–513. (Cited on pages 35 and 63)

Bruhn, A. (2006). *Variational Optic Flow Computation: Accurate Modelling and Efficient Numerics*. PhD thesis, Dept. of Computer Science, Saarland University, Saarbrücken, Germany. (Cited on pages 50, 69, and 101)

Bruhn, A. and Weickert, J. (2005). Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 749–755. IEEE Computer Society Press. (Cited on pages 54, 83, and 96)

Bruhn, A., Weickert, J., and Schnörr, C. (2005). Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231. (Cited on pages 17 and 81)

Butler, D. J., Wulff, J., Stanley, G. B., and Black, M. J. (2012). A naturalistic open source movie for optical flow evaluation. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *Computer Vision – ECCV 2012*, volume 7577 of *Lecture Notes in Computer Science*, pages 611–625. Springer. (Cited on pages 15, 72, 74, and 79)

Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., and Fua, P. (2012). BRIEF: Computing a Local Binary Descriptor Very Fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1281–1298. (Cited on page 39)

Chambolle, A. (1994). Partial differential equations and image processing. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 16–20. IEEE Computer Society Press. (Cited on pages 30 and 35)

Chambolle, A. and Lions, P.-L. (1997). Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188. (Cited on page 57)

Chan, C.-H., Goswami, B., Kittler, J., and Christmas, W. (2012). Local ordinal contrast pattern histograms for spatiotemporal, lip-based speaker authentication. *IEEE Transactions on Information Forensics and Security*, 7(2):602–612. (Cited on page 40)

Charbonnier, P., Blanc-Féraud, L., Aubert, G., and Barlaud, M. (1994). Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proc. IEEE International Conference on Image Processing*, volume 2, pages 168–172. (Cited on page 54)

Chen, Q. and Koltun, V. (2013). A simple model for intrinsic image decomposition with depth cues. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 241–248. (Cited on page 93)

Cohen, I. (1993). Nonlinear variational method for optical flow computation. In *Proc. Eighth Scandinavian Conference on Image Analysis*, volume 1, pages 523–530, Tromsø, Norway. (Cited on pages 18 and 56)

Cornelius, N. and Kanade, T. (1984). Adapting optical-flow to measure object motion in reflectance and X-ray image sequences. *Computer Graphics*, 18(1):24–25. (Cited on pages 93, 94, and 103)

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In Schmid, C., Soatto, S., and Tomasi, C., editors, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 886–893. (Cited on page 35)

Debevec, P. E. and Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. In *Proc. SIGGRAPH 97*, Annual Conference Series, pages 369–378. ACM Press. (Cited on page 93)

Dederscheck, D., Müller, T., and Mester, R. (2012). Illumination invariance for driving scene optical flow using comparagram preselection. In *Proc. IEEE Intelligent Vehicles Symposium (IV)*, pages 742–747. (Cited on page 92)

Didas, S. (2008). *Denoising and Enhancement of Digital Images – Variational Methods, Integrodifferential Equations, and Wavelets.* PhD thesis, Dept. of Mathematics, Saarland University. (Cited on page 57)

Drulea, M. and Nedevschi, S. (2013). Motion estimation using the correlation transform. *IEEE Transactions on Image Processing*, 22(8):3260–3270. (Cited on page 34)

Enkelmann, W. (1988). Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences. *Computer Vision, Graphics, and Image Processing*, 43(2):150 – 177. (Cited on page 59)

Fagerström, D. (2007). Spatio-temporal scale-spaces. In Sgallari, F., Murli, F., and Paragios, N., editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 326–337. Springer, Berlin. (Cited on page 124)

Florack, L. (1997). *Image Structure*, volume 10 of *Computational Imaging and Vision*. Kluwer, Dordrecht. (Cited on page 124)

Fouad, M. M., Dansereau, R. M., and Whitehead, A. D. (2009). Geometric registration of images with arbitrarily-shaped local intensity variations from shadows. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 201–204. (Cited on page 93)

Fröba, B. and Ernst, A. (2004). Face detection with the modified census transform. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 91–96. (Cited on page 38)

Fučik, S., Kratochvil, A., and Nečas, J. (1973). Kačanov–Galerkin method. *Commentationes Mathematicae Universitatis Carolinae*, 14(4):651–659. (Cited on page 65)

Garg, R., Roussos, A., and Agapito, L. (2013). A variational approach to video registration with subspace constraints. *International Journal of Computer Vision*, 104(3):286–314. (Cited on page 94)

Garrido, P., Valgaerts, L., Wu, C., and Theobalt, C. (2013). Reconstructing detailed dynamic face geometry from monocular video. *ACM Transactions on Graphics*, 32(6):158:1–158:10. (Cited on page 16)

Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361. (Cited on pages 15, 22, 23, 50, 52, 72, 73, 87, 103, 109, and 112)

Gelfand, I. M. and Fomin, S. V. (2000). *Calculus of Variations*. Dover, New York. (Cited on page 63)

Gennert, M. A. and Negahdaripour, S. (1987). Relaxing the brightness constancy assumption in computing optical flow. Technical Report 975, Artificial Intelligence Laboratory, Massachusetts Instituite of Technology. (Cited on pages 92, 93, and 112)

Gerig, G., Kübler, O., Kikinis, R., and Jolesz, F. A. (1992). Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11:221–232. (Cited on page 30)

Geusebroek, J. M., van den Boomgaard, R., Smeulders, A. W. M., and Geerts, H. (2001). Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12):1338–1350. (Cited on page 33)

Golland, P. and Bruckstein, A. M. (1997). Motion from color. *Computer Vision and Image Understanding*, 68(3):346–362. (Cited on page 32)

Grewenig, S., Weickert, J., and Bruhn, A. (2010). From box filtering to fast explicit diffusion. In Goesele, M., Roth, S., Kuijper, A., Schiele, B., and Schindler, K., editors, *Pattern Recognition*, volume 6376 of *Lecture Notes in Computer Science*, pages 543–552, Berlin. Springer. (Cited on page 130)

Grewenig, S., Weickert, J., Schroers, C., and Bruhn, A. (2013). Cyclic schemes for PDE-based image analysis. Technical Report 327, Department of Mathematics and Computer Science, Saarland University, Saarbrücken. (Cited on page 67)

Grossberg, M. D. and Nayar, S. K. (2002). What can be known about the radiometric response from images? In Heyden, A., Sparr, G., Nielsen, M., and Johansen, P., editors, *Computer Vision – ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 189–205. Springer. (Cited on pages 93, 103, and 105)

Grossberg, M. D. and Nayar, S. K. (2004). Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282. (Cited on pages 93, 95, 103, and 108)

Guichard, F. (1998). A morphological, affine, and Galilean invariant scale-space for movies. *IEEE Transacions on Image Processing*, 7(3):444–456. (Cited on page 124)

Haber, E. and Modersitzki, J. (2007). Intensity gradient based registration and fusion of multi-modal images. *Methods of Information in Medicine*, 46(3):292–299. (Cited on page 28)

HaCohen, Y., Shechtman, E., Goldman, D. B., and Lischinski, D. (2011). Non-rigid dense correspondence with applications for image enhancement. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2011)*, 30(4):70:1–70:9. (Cited on page 93)

Hager, G. D. and Belhumeur, P. N. (1998). Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1025–1039. (Cited on page 94)

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition. (Cited on page 17)

Haussecker, H. and Fleet, D. (2001). Estimating optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):661–673. (Cited on page 93)

Hermann, S. and Klette, R. (2013). Hierarchical scan-line dynamic programming for optical flow using semi-global matching. In Park, J.-I. and Kim, J., editors, *Computer Vision - ACCV 2012 Workshops*, volume 7729 of *Lecture Notes in Computer Vision*, pages 556–567. Springer, Berlin. (Cited on pages 89 and 120)

Hermosillo, G., Chefd'Hotel, C., and Faugeras, O. (2002). Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50(3):329–343. (Cited on page 34)

Hewer, A., Weickert, J., Scheffer, T., Seibert, H., and Diebels, S. (2013). Lagrangian strain tensor computation with higher order variational models. In Burghardt, T., Damen, D., Mayol-Cuevas, W., and Mirmehdi, M., editors, *Proc. British Machine Vision Conference*, Bristol, UK. BMVA Press. (Cited on page 58)

Horn, B. and Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, 17:185–203. (Cited on pages 15, 17, 18, 20, 47, 48, 49, 56, 123, 125, and 127)

Horn, B. K. P. (1974). Determining lightness from an image. *Computer Graphics and Image Processing*, 3:277–299. (Cited on page 93)

Horn, B. K. P. and Brooks, M. J., editors (1989). *Shape from Shading*. MIT Press, Cambridge, MA, USA. (Cited on page 93)

Huber, P. J. (1981). *Robust Statistics*. Wiley, New York. (Cited on page 53)

Iijima, T. (1962). Basic theory on normalization of pattern (in case of typical one-dimensional pattern). *Bulletin of the Electrotechnical Laboratory*, 26:368–388. In Japanese. (Cited on page 59)

Iijima, T. (1963). Theory of pattern recognition. *Electronics and Communications in Japan*, 1:123–134. In English. (Cited on pages 123 and 124)

Iijima, T., Genchi, H., and Mori, K. (1973). A theory of character recognition by pattern matching method. In *Proc. First International Joint Conference* *on Pattern Recognition*, pages 50–56, Washington, DC. In English. (Cited on pages 123 and 124)

Jolliffe, I. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer, New York. (Cited on page 108)

Kennedy, R. and Taylor, C.-J. (2015). Optical flow with geometric occlusion estimation and fusion of multiple frames. In Tai, X.-C., Bae, E., Chan, T.-F., and Lysaker, M., editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 8932 of *Lecture Notes in Computer Science*, pages 364–377. Springer. (Cited on pages 89 and 120)

Keys, R. (1981). Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(6):1153–1160. (Cited on pages 67 and 72)

Kim, T. H., Lee, H. S., and Lee, K. M. (2013). Optical flow via locally adaptive fusion of complementary data costs. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 3344–3351. (Cited on page 55)

Krajsek, K. and Mester, R. (2006). A maximum likelihood estimator for choosing the regularization parameters in global optical flow methods. In *Proc. IEEE International Conference on Image Processing (ICIOP)*, pages 1081–1084. (Cited on page 124)

Lai, S. and Vemuri, B. C. (1998). Reliable and efficient computation of optical flow. *International Journal of Computer Vision*, 29(2):87–105. (Cited on pages 50 and 51)

Land, E. H. and McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11. (Cited on page 93)

Laptev, I., Caputo, B., SchÃijldt, C., and Lindeberg, T. (2007). Local velocity-adapted motion events for spatio-temporal recognition. *Computer Vision and Image Understanding*, 108(3):207–229. (Cited on page 124)

Lindeberg, T. (1994). *Scale-Space Theory in Computer Vision*. Kluwer, Boston. (Cited on page 124)

Lindeberg, T. (2013). Scale selection properties of generalized scale-space interest point detectors. *Journal of Mathematical Imaging and Vision*, 46(2):177–210. (Cited on page 124)

Liu, C., Yuen, J., and Torralba, A. (2011). SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):978–994. (Cited on page 35)

Lowe, D. L. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110. (Cited on page 34)

Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, volume 2, pages 674–679, San Francisco. Morgan Kaufmann Publishers Inc. (Cited on pages 17 and 49)

Lysaker, M., Lundervold, A., and Tai, X. C. (2003). Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590. (Cited on page 57)

Mann, S., Manders, C., and Fung, J. (2003). The lightspace change constraint equation (LCCE) with practical application to estimation of the projectivity+gain transformation between multiple pictures of the same subject matter. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages III–481–4. (Cited on page 92)

Mei, X., Sun, X., Zhou, M., Jiao, S., Wang, H., and Zhang, X. (2011). On building an accurate stereo matching system on graphics hardware. In *Proc. IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 467–474. (Cited on page 37)

Meister, A. (2008). *Numerik linearer Gleichungssysteme*. Vieweg, Braunschweig, 3rd edition. (Cited on page 69)

Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630. (Cited on page 35)

Mileva, Y., Bruhn, A., and Weickert, J. (2007). Illumination-robust variational optical flow with photometric invariants. In Hamprecht, F. A., Schnörr, C., and Jähne, B., editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 152–162. Springer. (Cited on pages 31, 33, and 34)

Mittal, A. and Ramesh, V. (2006). An intensity-augmented ordinal measure for visual correspondence. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 849–856. (Cited on page 40)

Modersitzki, J. (2009). *FAIR: Flexible Algorithms for Image Registration.* Fundamentals of Algorithms. SIAM, Philadelphia. (Cited on pages 28 and 42)

Mohamed, M., Rashwan, H., Mertsching, B., Garcia, M., and Puig, D. (2014). Illumination-robust optical flow using a local directional pattern. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(9):1499–1508. (Cited on pages 89 and 120)

Mohamed, M. A. and Mertsching, B. (2012). TV-L1 optical flow estimation with image details recovering based on modified census transform. In Bebis, G., Boyle, R., Parvin, B., Koracin, D., Fowlkes, C., Wang, S., Choi, M.-H., Mantler, S., Schulze, J., Acevedo, D., Mueller, K., and Papka, M., editors, *Advances in Visual Computing*, volume 7431 of *Lecture Notes in Computer Science*, pages 482–491. Springer. (Cited on page 38)

Molnár, J., Chetverikov, D., and Fazekas, S. (2010). Illumination-robust variational optical flow using cross-correlation. *Computer Vision and Image Understanding*, 114(10):1104–1114. (Cited on page 34)

Mrázek, P. and Navara, M. (2003). Selection of optimal stopping time for nonlinear diffusion filtering. *International Journal of Computer Vision*, 52(2-3):189–203. (Cited on pages 124 and 131)

Mukawa, N. (1990). Estimation of shape, reflection coefficients and illuminant direction from image sequences. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 507–512. (Cited on pages 92, 93, and 103)

Müller, T., Rabe, C., Rannacher, J., Franke, U., and Mester, R. (2011). Illumination robust dense optical flow using census signatures. In Mester, R. and Felsberg, M., editors, *Pattern Recognition*, volume 6835 of *Lecture Notes in Computer Science*, pages 236–245. Springer. (Cited on page 37)

Nagel, H. H. (1983a). Constraints for the estimation of displacement vector fields from image sequences. In *Proc. International Joint Conference on Artificial Intelligence*, pages 945–951. (Cited on page 59)

Nagel, H.-H. (1983b). Displacement vectors derived from second-order intensity variations in image sequences. *Computer Vision, Graphics, and Image Processing*, 21(1):85–117. (Cited on page 27)

Nagel, H.-H. and Enkelmann, W. (1986). 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. (Cited on pages 18, 58, 98, 123, and 128)

Negahdaripour, S. and Yu, C.-H. (1993). A generalized brightness change model for computing optical flow. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 2–11. (Cited on pages 93 and 103)

Nir, T., Bruckstein, A. M., and Kimmel, R. (2008). Over-parameterized variational optical flow. *International Journal of Computer Vision*, 76(2):205–216. (Cited on page 94)

Nir, T., Kimmel, R., and Bruckstein, A. (2005). Variational approach for joint optic-flow computation and video restoration. Technical report, Technion Israel Institute of Technology. (Cited on page 17)

Nocedal, J. and Wright, S. (2006). *Numerical Optimization.* Springer, New York. (Cited on pages 58, 62, and 98)

Ochs, P., Malik, J., and Brox, T. (2013). Segmentation of moving objects by long term video analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1187–1200. (Cited on page 16)

Onkarappa, N. and Sappa, A. (2014). Speed and texture: an empirical study on optical-flow accuracy in ADAS scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 15(1):136–147. (Cited on page 16)

Papenberg, N., Bruhn, A., Brox, T., Didas, S., and Weickert, J. (2006). Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158. (Cited on page 27)

Perona, P. and Malik, J. (1990). Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639. (Cited on page 98)

Pharr, M. and Humphreys, G. (2010). *Physically Based Rendering, Second Edition: From Theory To Implementation.* Morgan Kaufmann, San Francisco, 2nd edition. (Cited on page 18)

Pietikäinen, M., Hadid, A., Zhao, G., and Ahonen, T. (2011). *Computer Vision Using Local Binary Patterns.* Springer, London. (Cited on page 39)

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, 3rd edition. (Cited on pages 40 and 71)

Puxbaum, P. and Ambrosch, K. (2010). Gradient-based modified census transform for optical flow. In Bebis, G., Boyle, R. D., Parvin, B., Koracin, D., Chung, R., Hammoud, R. I., Hussain, M., Tan, K.-H., Crawfis, R., Thalmann, D., Kao, D., and Avila, L., editors, *Proc. International Symposium on Advances in Visual Computing (ISVC)*, volume 6453 of *Lecture Notes in Computer Science*, pages 437–448. Springer. (Cited on page 44)

Radke, R. J. (2012). *Computer Vision for Visual Effects.* Cambridge University Press. (Cited on page 16)

Rakêt, L. L., Roholm, L., Bruhn, A., and Weickert, J. (2012). Motion compensated frame interpolation with a symmetrical optical flow constraint. In Bebis, G., Boyle, R., Parvin, B., Koracin, D., Fowlkes, C., Wang, S., Choi, M.-H., Mantler, S., Schulze, J., Acevedo, D., Mueller, K., and Papka, M., editors, *Advances in Visual Computing*, volume 7431 of *Lecture Notes in Computer Science*, pages 447–457. Springer. (Cited on page 16)

Ranftl, R., Bredies, K., and Pock, T. (2014). Non-local total generalized variation for optical flow estimation. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, volume 8689 of *Lecture Notes in Computer Science*, pages 439–454. Springer. (Cited on pages 39, 89, and 120)

Ranftl, R., Gehrig, S., Pock, T., and Bischof, H. (2012). Pushing the limits of stereo using variational stereo estimation. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 401–407. (Cited on pages 37, 58, 88, 89, 97, and 120)

Rashwan, H., Mohamed, M., Garcia, M., Mertsching, B., and Puig, D. (2013). Illumination robust optical flow model based on histogram of oriented gradients. In Weickert, J., Hein, M., and Schiele, B., editors, *Pattern Recognition*, volume 8142 of *Lecture Notes in Computer Science*, pages 354–363, Berlin. Springer. (Cited on pages 35, 89, and 120)

Rother, C., Kiefel, M., Zhang, L., Schölkopf, B., and Gehler, P. V. (2011). Recovering intrinsic images with a global sparsity prior on reflectance. In

Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K., editors, *Advances in Neural Information Processing Systems (NIPS)*, pages 765–773. Neural Information Processing Systems Foundation. (Cited on page 93)

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268. (Cited on pages 29 and 56)

Sadek, R., Facciolo, G., Arias, P., and Caselles, V. (2013). A variational model for gradient-based video editing. *International Journal of Computer Vision*, 103(1):127–162. (Cited on page 16)

Santos-Victor, J., Sandini, G., CurOtto, F., and Garibaldi, S. (1993). Divergent stereo for robot navigation: learning from bees. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 434–439. (Cited on page 16)

Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., and Westling, P. (2014). High-resolution stereo datasets with subpixel-accurate ground truth. In Jiang, X., Hornegger, J., and Koch, R., editors, *Pattern Recognition*, volume 8753 of *Lecture Notes in Computer Science*, pages 31–42. Springer International Publishing. (Cited on pages 73, 74, 92, 103, 109, 112, and 114)

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42. (Cited on page 23)

Scherzer, O. and Weickert, J. (2000). Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12(1):43–63. (Cited on pages 123, 124, and 129)

Schneevoigt, T., Schroers, C., and Weickert, J. (2014). A dense pipeline for 3D reconstruction from image sequences. In Jiang, X., Hornegger, J., and Koch, R., editors, *Pattern Recognition*, volume 8753, pages 629–640. Springer International Publishing. (Cited on page 16)

Schnörr, C. (1993). On functionals with greyvalue-controlled smoothness terms for determining optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1074–1079. (Cited on pages 28 and 124)

Schnörr, C. (1994). 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. IEEE Computer Society Press. (Cited on page 56)

Schönemann, T. and Cremers, D. (2006). Near real-time motion segmentation using graph cuts. In Franke, K., Müller, K. R., Nickolay, B., and Schäfer, R., editors, *Pattern Recognition*, volume 4174 of *Lecture Notes in Computer Science*, pages 455–464. Springer Berlin Heidelberg. (Cited on page 50)

Setzer, S., Steidl, G., and Teuber, T. (2011). Infimal convolution regularizations with discrete L1-type functionals. *Communications in Mathematical Sciences*, 9(3):797–827. (Cited on page 57)

Sevilla-Lara, L., Sun, D., Learned-Miller, E.-G., and Black, M.-J. (2014). Optical flow estimation with channel constancy. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, volume 8689 of *Lecture Notes in Computer Science*, pages 423–438. Springer International Publishing. (Cited on page 63)

Shi, J. and Tomasi, C. (1994). Good features to track. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 593–600. (Cited on page 16)

Shulman, D. and Herve, J.-Y. (1989). Regularization of discontinuous flow fields. In *Proc. Workshop on Visual Motion*, pages 81–86. IEEE Computer Society Press. (Cited on pages 18 and 56)

Sloane, N. J. A. and Plouffe, S. (1995). *The Encyclopedia of Integer Sequences*. Academic Press, San Diego. (Cited on page 41)

Soatto, S. (2009). Actionable information in vision. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 2138–2145. (Cited on page 45)

Sochen, N., Kimmel, R., and Bruckstein, F. (2001). Diffusions and confusions in signal and image processing. *Journal of Mathematical Imaging and Vision*, 14(3):195–210. (Cited on page 124)

Sporring, J., Nielsen, M., Florack, L., and Johansen, P., editors (1997). *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*. Kluwer, Dordrecht. (Cited on page 124)

Stein, F. (2004). Efficient Computation of Optical Flow Using the Census Transform. In Rasmussen, C. E., Bülthoff, H. H., Schölkopf, B., and

Giese, M. A., editors, *Pattern Recognition*, volume 3175 of *Lecture Notes in Computer Science*, pages 79–86. Springer. (Cited on pages 16 and 38)

Steinbrücker, F., Sturm, J., and Cremers, D. (2011). Real-time visual odometry from dense RGB-D images. In *Workshop on Live Dense Reconstruction with Moving Cameras at the IEEE International Conference on Computer Vision (ICCV).* (Cited on page 16)

Steinhaus, H. (1957). Sur la division des corps matériels en parties. *Bulletin de l'Académie Polonaise des Sciences, Classe 3*, 4:801–804. (Cited on page 105)

Stoll, M., Volz, S., and Bruhn, A. (2013). Adaptive integration of feature matches into variational optical flow methods. In Lee, K. M., Rehg, J., Matsushita, Y., and Hu, Z., editors, *Computer Vision - ACCV 2012*, volume 7726 of *Lecture Notes in Computer Science*, pages 1–14, Berlin. Springer. (Cited on pages 35 and 63)

Sullivan, G., Ohm, J., Han, W.-J., and Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1649–1668. (Cited on page 16)

Sun, D., Roth., S., and Black, M. J. (2010). Secrets of optical flow estimation and their principles. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2432–2439. (Cited on pages 58 and 62)

Sun, D., Roth, S., and Black, M. J. (2014). A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision*, 106(2):115–137. (Cited on pages 89 and 120)

Tang, F., Lim, S. H., Chang, N. L., and Tao, H. (2009). A novel feature descriptor invariant to complex brightness changes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2631–2638. (Cited on page 40)

Tieu, K. and Miller, E. G. (2002). Unsupervised color constancy. In Thrun, S. and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*, pages 1303–1310. MIT Press, Cambridge, MA. (Cited on pages 94, 95, and 104)

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady*, 4:1035–1038. (Cited on pages 18, 56, and 127)

Timofte, R. and Gool, L. V. (2015). Sparseflow: Sparse matching for small to large displacement optical flow. In *Proc. IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1100–1106. IEEE. (Cited on pages 89 and 120)

Tretiak, O. and Pastor, L. (1984). Velocity estimation from image sequences with second order differential operators. In *Proc. IEEE International Conference on Pattern Recognition (ICPR)*, pages 16–19. (Cited on page 27)

Trobin, W., Pock, T., Cremers, D., and Bischof, H. (2008). An unbiased second-order prior for high-accuracy motion estimation. In Rigoll, G., editor, *Pattern Recognition*, volume 5096 of *Lecture Notes in Computer Science*, pages 396–405, Berlin. Springer. (Cited on page 57)

Uras, S., Girosi, F., Verri, A., and Torre, V. (1988). A computational approach to motion perception. *Biological Cybernetics*, 60(2):79–87. (Cited on page 27)

Valgaerts, L., Bruhn, A., Zimmer, H., Weickert, J., Stoll, C., and Theobalt, C. (2010). Joint estimation of motion, structure and geometry from stereo sequences. In Daniilidis, K., Maragos, P., and Paragios, N., editors, *Computer Vision – ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, pages 568–581. Springer, Berlin. (Cited on pages 51 and 100)

van de Weijer, J. and Gevers, T. (2004). Robust optical flow from photometric invariants. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 1835–1838. (Cited on page 33)

Vogel, C., Roth, S., and Schindler, K. (2013). An evaluation of data costs for optical flow. In Weickert, J., Hein, M., and Schiele, B., editors, *Pattern Recognition*, volume 8142 of *Lecture Notes in Computer Science*, pages 343–353, Berlin. Springer. (Cited on pages 30, 45, 58, 85, 89, and 120)

Vogel, C. R. and Oman, M. E. (1996). Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238. (Cited on page 65)

Volz, S., Bruhn, A., Valgaerts, L., and Zimmer, H. (2011). Modeling temporal coherence for optical flow. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1123. (Cited on page 98)

Wang, O., Schroers, C., Zimmer, H., Gross, M., and Sorkine-Hornung, A. (2014). Videosnapping: Interactive synchronization of multiple videos. *ACM Transactions on Graphics*, 33(4):77:1–77:10. (Cited on page 16)

Wang, Z., Fan, B., and Wu, F. (2011). Local intensity order pattern for feature description. In *IEEE International Conference on Computer Vision (ICCV)*, pages 603–610. (Cited on page 40)

Wedel, A., Pock, T., Zach, C., Cremers, D., and Bischof, H. (2008). An improved algorithm for TV-L1 optical flow. In Cremers, D., Rosenhahn, B., Yuille, A. L., and Schmidt, F. R., editors, *Statistical and Geometrical Approaches to Visual Motion Analysis*, volume 5604 of *Lecture Notes in Computer Science*, pages 23–45. Springer. (Cited on pages 29 and 30)

Wei, D., Liu, C., and Freeman, W. T. (2014). A data-driven regularization model for stereo and flow. In *Proc. IEEE International Conference on 3D Vision (3DV)*, pages 277–284. (Cited on pages 89 and 120)

Weickert, J. (1998). *Anisotropic Diffusion in Image Processing.* Teubner, Stuttgart. (Cited on pages 111, 124, 130, and 141)

Weickert, J. and Schnörr, C. (2001). A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal on Computer Vision*, 45(3):245–264. (Cited on page 58)

Weickert, J. and Welk, M. (2006). Tensor field interpolation with PDEs. In Weickert, J. and Hagen, H., editors, *Visualization and Processing of Tensor Fields*, pages 315–325. Springer. (Cited on page 105)

Weinzaepfel, P., Revaud, J., Harchaoui, Z., and Schmid, C. (2013). Deepflow: Large displacement optical flow with deep matching. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1385–1392. (Cited on pages 89 and 120)

Weiss, Y. (2001). Deriving intrinsic images from image sequences. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 68–75. (Cited on page 93)

Werlberger, M., Pock, T., and Bischof, H. (2010). Motion estimation with non-local total variation regularization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2464–2471. (Cited on pages 34 and 58)

Whittaker, E. T. (1923). A new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:65–75. (Cited on page 127)

Witkin, A., Terzopoulos, D., and Kass, M. (1987). Signal matching through scale space. *International Journal of Computer Vision*, 1(2):133–144. (Cited on pages 18 and 59)

Witkin, A. P. (1983). Scale-space filtering. In *Proc. Eighth International Joint Conference on Artificial Intelligence*, volume 2, pages 945–951, Karlsruhe, West Germany. (Cited on page 59)

Xu, L., Jia, J., and Matsushita, Y. (2010). Motion detail preserving optical flow estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1293–1300. (Cited on page 55)

Yuan, J., Schnörr, C., and Steidl, G. (2007). Simultaneous optical flow estimation and decomposition. *SIAM Journal on Scientific Computing*, 29(6):2283–2304. (Cited on page 57)

Zabih, R. and Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. In Eklundh, J.-O., editor, *Computer Vision – ECCV 1994*, volume 800 of *Lecture Notes in Computer Science*, pages 151–158. Springer, Berlin. (Cited on pages 16, 36, 37, and 39)

Zach, C., Pock, T., and Bischof, H. (2007). A duality based approach for realtime TV-L1 optical flow. In Hamprecht, F., Schnörr, C., and Jähne, B., editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 214–223. Springer, Berlin. (Cited on page 38)

Zeidler, E. (1990). *Nonlinear Functional Analysis and its Applications II/B: Nonlinear Monotone Operators.* Springer, Berlin. (Cited on page 65)

Zimmer, H., Bruhn, A., and Weickert, J. (2011a). Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. *Computer Graphics Forum (Proceedings of Eurographics)*, 30(2):405–414. (Cited on page 16)

Zimmer, H., Bruhn, A., and Weickert, J. (2011b). Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388. (Cited on pages 19, 34, 50, 51, 58, 76, 81, 83, 98, 124, 127, 130, and 131)

Zimmer, H., Bruhn, A., Weickert, J., Valgaerts, L., Salgado, A., Rosenhahn, B., and Seidel, H.-P. (2009). Complementary optic flow. In Cremers, D., Boykov, Y., Blake, A., and Schmidt, F., editors, *Energy Minimization*

Methods in Computer Vision and Pattern Recognition, volume 5681 of *Lecture Notes in Computer Science*, pages 207–220. Springer, Berlin. (Cited on page 87)