Variationelle 3D-Rekonstruktion aus Stereobildpaaren und Stereobildfolgen

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

vorgelegt von

Levi Valgaerts

Saarbrücken, 2011

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

Vision and Image Processing Group, Cluster of Excellence MMCI,
Universität des Saarlandes, 66041 Saarbrücken

Tag des Kolloquiums

29.09.2011

Dekan

Prof. Dr. Holger Hermanns

Prüfungsausschuss

Prof. Dr. Bernt Schiele (Vorsitz)
Universität des Saarlandes

Prof. Dr. Joachim Weickert (1. Gutachter)
Universität des Saarlandes

Prof. Dr. Thomas Brox (2. Gutachter)
Universität Freiburg

Dr. Andrés Bruhn
Universität des Saarlandes

Kurzzusammenfassung – Short Abstract

Kurzzusammenfassung – Deutsch

Diese Arbeit befasst sich mit der 3D Rekonstruktion und der 3D Bewegungsschätzung aus Stereodaten unter Verwendung von Variationsansätzen, die auf dichten Verfahren zur Berechnung des optischen Flusses beruhen. Im ersten Teil der Arbeit untersuchen wir ein neues Anwendungsgebiet von dichtem optischen Fluss, nämlich die Bestimmung der Fundamentalmatrix aus Stereobildpaaren. Indem wir die Abhängigkeit zwischen der geschätzten Stereogeometrie in Form der Fundamentalmatrix und den berechneten Bildkorrespondenzen geeignet ausnutzen, sind wir in der Lage, im zweiten Teil der Arbeit eine gekoppelte Bestimmung der Fundamentalmatrix und des optischen Flusses vorzuschlagen, die zur einer Erhöhung der Genauigkeit beider Schätzungen führt. Im Gegensatz zu vielen existierenden Verfahren berechnet unser gekoppelter Ansatz dabei die Lage der Kameras und die 3D Szenenstruktur nicht einzeln, sonderen bestimmt sie in einem einzigen gemeinsamen Optimierungsschritt. Dem Prinzip der gemeinsamen Schätzung weiter folgend koppeln wir im letzten Teil der Arbeit die dichte 3D Rekonstruktion der Szene zusätzlich mit der Bestimmung der zugehörigen 3D Bewegung. Dies wird durch die Intergation von räumlicher und zeitlicher Information aus mehreren Stereobildpaaren in ein neues Modell zur Szenenflussschätzung realisiert.

Short Abstract – English

This work deals with 3D reconstruction and 3D motion estimation from stereo images using variational methods that are based on dense optical flow. In the first part of the thesis, we will investigate a novel application for dense optical flow, namely the estimation of the fundamental matrix of a stereo image pair. By exploiting the high interdependency between the recovered stereo geometry and the established image correspondences, we propose a coupled refinement of the fundamental matrix and the optical flow as a second contribution, thereby improving the accuracy of both. As opposed to many existing techniques, our joint method does not solve for the camera pose and scene structure separately, but recovers them in a single optimisation step. True to our principle of joint optimisation, we further couple the dense 3D reconstruction of the scene to the estimation of its 3D motion in the final part of this thesis. This is achieved by integrating spatial and temporal information from multiple stereo pairs in a novel model for scene flow computation.

Eidesstattliche Versicherung

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

Saarbrücken, den 16. Oktober, 2011

Levi Valgaerts

Zusammenfassung

Variationsansätze zur Bestimmung des optischen Flusses haben sich als die genauesten und am besten verstandenen Verfahren für die Bewegunsschätzung aus Bildfolgen etabliert. Sie beruhen auf der Minimierung eines einzigen globalen Energiefunktionals, das die zeitliche Konstanz gewisser Bildeigenschaften modelliert (wie z.B. des Grauwerts), während es gleichzeitig sicherstellt, dass das gesuchte Verschiebungsvektorfeld eine Form von Regularität aufweist. Im Laufe der Zeit wurden solche Verfahren so weiterentwickelt, dass sie in der Lage sind, große Verschiebungen zu schätzen, robust unter Rauschen und fehlerhafte Daten zu sein, scharfe Bewegungsgrenzen zu liefern – und dies alles in Echtzeit für praktische Anwendungen. Trotz dieser Fortschritte hat relativ wenig Forschung auf dem eng verwandten Gebiet der Stereo-Rekonstruktion stattgefunden, einem Bereich, der typischerweise von anderen Ansätzen zur Korrespondenzfindung dominiert wird.

In dieser Arbeit zeigen wir anhand von drei Beiträgen, dass diese Unterteilung in verschiedene Anwendungsgebiete hinfällig ist. Zum einen untersuchen wir ein neues Anwendungsszenario für dichten optischen Fluss: die Berechnung der Fundamentalmatrix. Diese Matrix beschreibt die projektive Geometrie eines Stereo-Bildpaars und bildet die Grundlage für viele Algorithmen zur 3D-Rekonstruktion. Zum anderen schlagen wir einen neuen Variationsansatz vor, der die Fundamentalmatrix und den optischen Fluss gleichzeitig als Minimierer eines gemeinsamen Energiefunktionals berechnet. Konzeptionell bestimmt dieses Verfahren die relative Lage und Orientierung der beiden Kameras sowie die Struktur der Szene in einem gemeinsamen Optimierungsrahmen, wodurch eine dichte Rekonstruktion der abgebildeten Szene möglich wird. Als dritten Beitrag erweitern wir die gemeinsame Schätzung von Korrespondenzen und Stereo-Geometrie auf den Fall mehrerer Bilderpaare. In diesem Zusammenhang entwicklen wir ein Variationsansatz, der für zwei aufeinander folgende Bildpaare einer Stereosequenz die Fundmentalmatrix und die Verschiebungsvektorfelder zwischen allen Bildern bestimmt. Neben einer dichter Rekonstruktion der abgebildeten Szene ermöglicht dieser Ansatz die Berechnung der tatsächlichen 3D-Bewegung von Objekten in der Szene, den so genannten Szenenfluss.

Der Schwerpunkt dieser Arbeit liegt auf einer konsequenten Modellierung als gemeinsames Optimierungsproblem und auf der gleichzeitige Schätzung von Geometrie, Struktur und Bewegung der zugrunde liegenden 3D-Szene. Wir gehen diese Aufgabe in einer sehr systematischen Weise an, indem wir zunächst untersuchen, ob die Bestimmung der Stereogeometrie aus dichtem optischen Fluss genau genug ist, um als Grundlage für weitere integrierte Ansätze zu dienen. In einem ausführlichen Vergleich mit zwölf merkmalsbasierten Verfahren zeigen wir, dass dichte Flussfelder eine konkurrenzsfähige Schätzung der Geometrie mit Sub-Pixel Genauigkeit erlauben. Anschließend demonstrieren wir, dass eine gemeinsame Berechnung des optischen Flusses und der Fundamentalmatrix nicht nur eine stabilisierende Wirkung auf das Ergebnis hat, sondern auch zu einer Steigerung der Genauigkeit gegenüber einer sequentiellen Schätzung führt. In einem letzten Schritt koppeln wir schließlich die Rekonstruktion der Szene und die Bestimmung der 3D-Bewegung. Neben diesem allgemeinen Modell für den Szenenfluss stellen wir noch einige weitere Verbesserungen vor. Im Rahmen der hierarchischen Optimierung, die typischerweise zur Schätzung grosser Verschiebungen verwendet wird, führen wir eine kompakte Tensornotation ein, die eine Normalisierung der Modellannahmen ermöglicht. Auf diese Weise

machen wir nicht nur die Abweichungen von Modellannahmen als Entfernungen in der Bildebene interpretierbar, sondern können auch ihre Äquivalenz zu weit verbreiteten geometrischen Fehlermaßen im Bereich des Maschinensehens zeigen. Darüber hinaus bestimmen wir explizit Verdeckungen, die aufgrund der Bewegung der Szene oder der Kameras entstanden sind, und schliessen die zugehörigen Korrespondenzen von der Schätzung aus. Alle diese Modellierungsschritte führen schließlich zu einem Ansatz, der aktuelle Verfahren in der Literatur zur Berechnung des Szenenflusses hinsichtlich der Genauigkeit der Schätzung übertrifft – obwohl diese explizit von der Kenntnis der zugrunde liegenden Stereo-Geometrie Gebrauch machen.

Variational 3D Reconstruction from Stereo Image Pairs and Stereo Sequences

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

by

Levi Valgaerts

Saarbrücken, 2011

Thesis Supervisor: Prof. Dr. Joachim Weickert

Thesis Advisor: Dr. Andrés Bruhn

Referees: Prof. Dr. Joachim Weickert, Prof. Dr. Thomas Brox

Mathematical Image Analysis Group, Faculty of Mathematics and Informatics,
Saarland University, 66041 Saarbrücken

Vision and Image Processing Group, Cluster of Excellence MMCI,
Universität des Saarlandes, 66041 Saarbrücken

Abstract

Variational optical flow methods have established themselves as one of the most accurate and well understood techniques for motion estimation from image sequences. They minimise a single global energy functional that models the temporal constancy of certain image properties, e.g. the image grey value, while imposing some form of regularity on the displacement field. Over the course of time, optical flow algorithms have been specialised to cope with large displacements, to be robust under noise and corrupted data, to provide sharp motion boundaries and to perform in real-time for practical applications. Despite this progress, much less research has been done in the context of stereo reconstruction, a closely related field that has been typically dominated by other correspondence methods.

In this thesis we bridge this separation into different application domains by three contributions. First we investigate a new application for dense optical flow, namely the computation of the fundamental matrix. The fundamental matrix is the intrinsic representation of the projective geometry of a stereo image pair and forms the basis for many reconstruction algorithms. Secondly, we propose a new variational model that recovers the fundamental matrix and the optical flow simultaneously as the minimisers of a single energy functional. Conceptually, this method recovers the relative position of the two cameras and the structure of the scene in a single optimisation step, thereby providing a dense reconstruction of the depicted scene. As a third contribution, we extend our framework for the joint estimation of correspondences and stereo geometry to multiple images. For two stereo image pairs taken at two different time steps, we develop a variational method that simultaneously recovers the fundamental matrix and all optical flows between the images. In addition to providing a dense reconstruction of the depicted scene, this approach allows us to recover the actual 3D motion of the objects in the scene, the so-called scene flow.

The focus of this thesis lies on a consistent modelling in a variational framework and on the simultaneous estimation of stereo geometry, 3D scene structure and 3D scene motion. We will approach this task in a very systematic way by first asking ourselves if the recovery of the stereo geometry from dense optical flow is accurate enough to serve as a basis for more integrated methods. In an extensive comparison with twelve widely-used sparse feature based estimation techniques we demonstrate that we can estimate the stereo geometry with competitive sub-pixel quality. In a next step, we show that a coupled solution of the optical flow and the fundamental matrix will not only have a stabilising effect on the outcome, but will at the same time push the accuracy bounds further down. In a last step, we finally couple the reconstruction of the scene to the retrieval of its 3D motion field. Aside from this novel generalised model for scene flow, we introduce several further refinements. Within the multi-resolution framework required to handle large displacements, we introduce a compact notation that makes a normalisation of the linearised model assumptions possible. This way, we make the deviations from model assumptions interpretable as distances in the image plane and can show their equivalence to widely-known geometrical error measures from computer vision. In addition we explicitly detect occlusions due to scene motion and changes in camera viewpoint and exclude them from the estimation process. All these design steps ultimately lead to an approach that even outperforms recent scene flow techniques that make explicit use of a given stereo geometry.

To My Father

Let us finish it
We're almost there

Acknowledgements

My most sincere thanks go out to my thesis advisor Dr. Andrés Bruhn, whose profound insights and constant support helped guide me throughout my Ph.D. period. A lot of the ideas presented in this thesis have originated from our many fruitful discussions and I owe much of what I learned as a researcher to him. I am equally grateful that I was offered a position at his newly founded Vision and Image Processing Group and I wish him all the best throughout his future scientific career.

I wish to thank Prof. Joachim Weickert from the Mathematical Image Analysis Group at Saarland University for giving me the opportunity to work in one of the finest teams for image processing and computer vision. During my work there I was able to combine some of my biggest interests and develop a deeper understanding of the field of computer vision. I also wish to thank Prof. Christian Theobalt from the Graphics, Vision and Video Group at the Max Planck Institute for Informatics for showing interest in my work through our past and current collaborations and for sharing many of his creative visions of the future of machine vision. Furthermore, I thank Prof. Thomas Brox from the University of Freiburg for accepting the task of reviewing this thesis and Prof. Nassir Navab from the Technical University of Munich for introducing me to the field of computer vision.

I would like to thank all my colleagues at the Mathematical Image Analysis Group for the great time we had. More specifically I would like to thank Dr. Bernhard Burgeth, Dr. Martin Welk, Dr. Michael Breuß, Dr. Stephan Didas, Dr. Luis Pizarro, Dr. Simon Setzer, Natalia Slesareva, Irena Galić, Oliver Demetz, Sven Grewenig, Kai Hagenburg, Christian Schmaltz, Dr. Oliver Vogel, Gabriela Ghimpeteanu and Marcus Hargarter. In particular I would like to express gratitude to the colleagues with whom I shared the office for most of the time: to Henning Zimmer and Markus Mainberger for all the work related and non-work related discussion that we had and for becoming good friends over the years. Last but not least, I would like to thank two irreplaceable forces at the group: Pascal Gwosdek for helping out with about every computer related problem imaginable and Ellen Wintringer for taking care of even the smallest organisational detail.

I thank my colleagues Adrian Alexa, Cristián Madrigal, and Carsten Stoll from the Max Institute of Informatics for many shared moments in leisure and in work.

I also thank Yong Chul Yu and Sebastian Volz for working as student programmers on the visualisation tool. I wish them both all the best during their Ph.D. studies.

I thank all my former colleagues of the Computational Science and Engineering program at the Technical University Munich for becoming good friends, in particular Veselin Dikov, Gordana Stojceska and Niko Manopulo.

At the home front, I would like to thank my former study colleagues and friends Philippe Schram and Niko Renard, and especially Bart Grauwels for his efforts of staying in touch.

I would like to thank all the people that were present at the day of the defense for their interest and support.

Last but not least I would like to express my most profound gratitude to my parents and to Jasmina Bogojeska for their unlimited support and patience. This work would not have been possible without them.

Saarbrücken, October 16, 2011 Levi Valgaerts

Contents

1 Introduction

1.1 Motivation

Visual perception is one of our most important senses and is unlike any other responsible for the way we experience the world in which we live. Making an automated system, like a computer, derive the same information from images as we do, has been a goal for many years and has given rise to the fields of digital image processing and computer vision.

An important step in building a visual system is granting it the capability of inferring a three-dimensional model of the world from two-dimensional image information. Since depth information can not be recovered from a single image alone, at least two cameras are necessary to produce a *3D reconstruction* of the observed scene. To reconstruct an object that is seen in the left image of a stereo image pair, it needs to be identified in the right image as well. The reconstruction problem is therefore intrinsically related to the *image correspondence problem*. In this thesis we will solve the correspondence problem by estimating the two-dimensional displacement field that maps pixels in the left image to their new location in the right image, the so-called *optical flow*. In contrast to traditional optical flow, however, the displacement field between two stereo images can not be arbitrary and for its computation specific geometric constraints have to be taken into account.

Aside from recovering the static structure of the observed scene, we will also be interested in determining its dynamic behaviour. This property is related to how the objects in the scene move within their environment and how they change their shape. The actual motion of the scene can be described by a three-dimensional displacement field that is referred to as the *scene flow*. The estimation of scene flow includes a time component and its retrieval is not possible from two stereo images alone. Instead, its computation requires a temporal series of stereo pairs in the form of a stereo image sequence. This gives the problem an extra dimension, as objects need not only be identified in the left and the right image, but need to be tracked additionally throughout the image sequences. Scene flow estimation thus not only relies on the computation of a constrained optical flow for the 3D reconstruction of the scene, but has furthermore a strong connection to classical unrestricted optical flow estimation for the recovery of the motion of the scene.

Fields of Application. The 3D reconstruction from two stereo images has a direct influence on algorithms that seek for a more complete model of the scene from multiple views [SCD+06, BBH08]. With the increasing importance of online image data bases and visual search on the Internet, these algorithms have found new applications. One example is photo tourism [SSS06, ASS+09], where large unstructured collections of photographs are used to reconstruct cities and viewpoints that enable exploration of the data base in 3D. Closely related is the generation of intermediate viewpoints and the seamless navigation

through multiple video streams by exploiting the common 3D geometry of the underlying scene [BBPP10]. In application areas that are traditionally dominated by 2D techniques, three-dimensional cues are steadily integrated because they often lead to a more discriminative outcome. This is for instance the case for face recognition [BBK05, KPT+07].

Since optical flow can be regarded as the projection of the scene flow on the image plane, it is not surprising that there exists an overlap of application domains. Classical examples of motion detection are vehicle navigation and the segmentation of moving objects, which are an integral part of robotics and drivers assistance systems [WMR+09, UWSI10]. Another exciting field of research of scene flow is that of human face capture [MCT09, CGZZ10, BHPS10], where movements and expressions are detected that can tell us something about emotions or even about speech [LW08]. Strongly related are techniques for markerless motion capture and pose tracking [MHK06, dAST+08, SRBW11] that have received interest from the medical field, as well as from the entertainment industry.

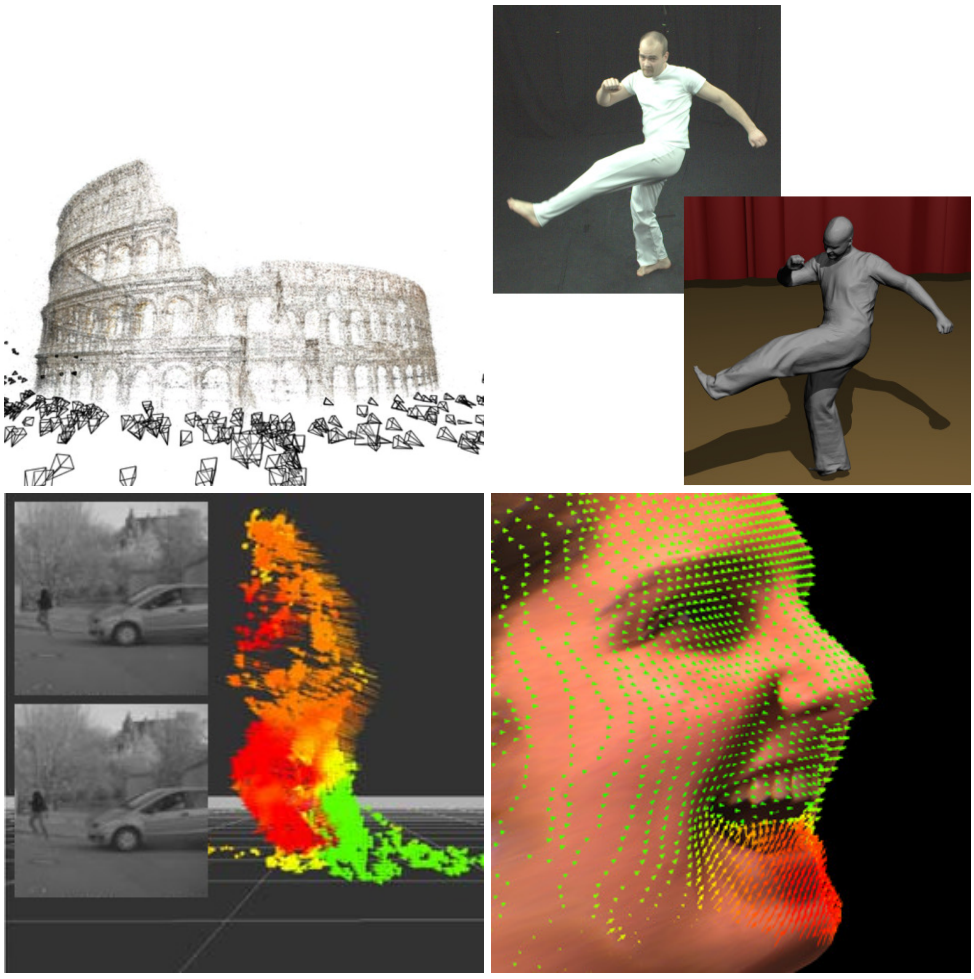

Fig. 1.1: Application areas of stereo reconstruction and scene flow computation. **Top Row: (a)** Reconstruction of cities and monuments from large online image data bases (from [ASS+09]). **(b)** Markerless motion capture and pose tracking (from [dAST+08]). **Bottom Row: (c)** Vehicle navigation and pedestrian avoidance (from [WMR+09]). **(d)** Capturing of facial expressions and motion (from [VBZ+10]).

1.2 Overview

The goal of this thesis is the development of variational methods for the task of stereo reconstruction and scene flow computation from stereo image pairs and stereo image sequences. The stereo reconstruction problem is composed of two highly interconnected sub-problems, namely the recovery of the camera motion and the estimation of the scene structure. The camera motion refers to the change in view point from one stereo image to the other and can be associated with one global entity, the *fundamental matrix*. The scene structure, on the other hand, is defined by the arrangement of the individual three-dimensional points in space and how they project on corresponding two-dimensional points in different images. It is therefore inherently linked to the local change in position of each pixel and thus to the concept of *dense optical flow*. Scene flow is the dense motion of the three-dimensional points in the scene and it can be thought of as a correspondence between two reconstructions of the scene at consecutive time instances.

Our general contribution exists in exploiting the high interdependency between fundamental matrix computation and optical flow estimation in a stereo setting to improve the accuracy of both. On a higher level, the emphasis of this thesis ultimately lies on the development of high performance methods for the *joint computation of camera motion, scene structure and scene flow* in a variational framework. We will approach this task in a very systematic way. First we ask ourselves if the derivation of the stereo geometry from a given dense optical flow field makes sense and if it is accurate enough to provide the basis for more integrated methods. In doing so we introduce a novel application area of dense optical flow, namely the estimation of the fundamental matrix from a stereo image pair. In a next step we will show that a coupled solution of the optical flow and the fundamental matrix will not only have a stabilising effect on the outcome, but will at the same time push the accuracy bounds further down. As opposed to many existing techniques, our joint methodology does not treat the sub-problems of camera motion and scene structure estimation separately, but fuses them into a single optimisation problem. True to our principle of joint optimisation, we couple the reconstruction of the scene to the retrieval of its three-dimensional motion field in a last step. This is achieved by integrating spatial and temporal information from multiple stereo pairs in a novel scene flow model.

In the following we will give a more detailed overview of the important concepts in this thesis and of the three contributions discussed above.

Variational Optical Flow. Variational methods for optical flow computation determine the displacement field between two images as the minimiser of a suitable energy functional. Ever since the seminal approach of Horn and Schunck [HS81], this energy functional mostly consists of two terms: a *data term* that imposes constancy on characteristic image features and a *smoothness term* that regularises the possibly non-unique solution of the data term by locally filling in information from neighbouring regions. A general formulation of this energy is given by

$$\mathcal{E}(\boldsymbol{w}) = \int_{\Omega} \big(\underbrace{\mathcal{E}_{\mathrm{D}}(g_1, g_2, \boldsymbol{w})}_{\text{data term}} + \alpha \underbrace{\mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w})}_{\text{smoothness term}} \big) \, \mathrm{d}\boldsymbol{x} \ , \tag{1.1}$$

where $\boldsymbol{x} = (x, y)^\top$ is a point in the rectangular image domain Ω and $g_1(x, y)$ and $g_2(x, y)$ are the intensity values of the two images between which the optical flow $\boldsymbol{w} = (u, v)^\top$ is estimated. Further, $\nabla = (\partial_x, \partial_y)^\top$ stands for the spatial gradient operator and the regularisation weight α determines the degree of smoothness of the solution.

Apart from their high accuracy [BBPW04, ZPB07a, NBK08, BBM09, ZBW+09], variational methods for optical flow estimation offer two main advantages over other approaches for estimating image correspondences:

- *No Hidden Model Assumptions.* All assumptions on the data and the type of solution are integrated in one energy functional. This automatically leads to a transparent and consistent modelling and a minimisation that does not rely on hidden intermediate processing steps. In this thesis we only design methods that minimise a single energy containing all model assumptions and all problem unknowns.

- *Dense Flow Fields without Gross Outliers.* Due to the *filling-in* of flow information from other image regions by the smoothness term, variational methods provide us with a dense flow field. By construction, they are thus global methods and result in an image correspondence for each pixel without requiring any final interpolation steps. Moreover, no gross outliers in the correspondences are created because of the combination of robust data constraints and global smoothness assumptions.

Estimation of the Fundamental Matrix. The optical flow between a pair of images is normally unrestricted and can be the result of motion in the scene as well as of the movement of the camera. Now we will make the assumption that the scene is either static or, equivalently, that the images are taken from different view points at the same time instance. In this case, both images form a stereo pair and the optical flow will be restricted by the underlying stereo geometry. A complete description of the geometry of two stereo images is provided by the fundamental matrix F [HZ00, FLP01], a 3×3 matrix which associates with each point in the image a line in the other image, on which its corresponding point has to lie. If a point $(x, y)^\top$ in the first image lies on the epipolar line of its corresponding point $(x + u, y + v)^\top$ in the second image and vice versa, both points are said to satisfy the *epipolar constraint*

$$\begin{pmatrix} x+u \\ y+v \\ 1 \end{pmatrix}^\top F \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0 \ . \tag{1.2}$$

This constraint prescribes that $(x + u, y + v)^\top$ has to lie on the *epipolar line* $F(x, y, 1)^\top$, or equivalently that $(x, y)^\top$ has to lie on the epipolar line $F^\top(x + u, y + v)^\top$.

The epipolar constraint has to be satisfied *in each point* of the image. Provided that the dense optical flow $(u, v)^\top$ is known, the entries of the fundamental matrix can thus be inferred by imposing the constraint (1.2) in all pixels. In fact, this makes the problem highly overdetermined since far less image correspondences are theoretically necessary to determine the nine entries of F. In practice, however, the computed correspondences are not exact and including a large amount of them will improve the stability and the robustness of the estimation process. The quality of the fundamental matrix will additionally benefit

from the previously mentioned advantages of dense variational optical flow, namely their high matching accuracy and their inherent robustness against large outliers.

Our proposed method of estimating the fundamental matrix from all correspondences provided by dense optical flow contrasts with well-established techniques that compute the stereo geometry by matching a small set of sophistically optimised interest points (e.g. [HS88, Low04, WAH93, LF96, TM97, HZ00, FLP01]). These *sparse feature based* methods spend most of their effort on singling out only the *very best* correspondences and many elaborate statistical approaches [FB81, RL87, RFP08] have been developed to achieve this. Despite the fact that ideas from dense optical flow computation have been widely adopted in many matching tasks, including stereo matching, the estimation of the fundamental matrix has generally been the exclusive domain of sparse techniques.

★ ***Contribution 1.*** Our first main goal is to introduce a novel application area for optical flow by demonstrating that dense methods can be very beneficial for estimating the fundamental matrix. Important with this respect, is a juxtaposition with the currently leading class of techniques that are based on matching sparse image correspondences. By comparing our results with those obtained by feature based techniques, we identify cases in which dense methods have advantages over sparse approaches.

Joint Estimation of the Optical Flow and the Fundamental Matrix. The matching process of variational optical flow methods is guided by the assumption that certain image features remain constant along the motion trajectory. One of the most common constancy assumptions in this context is the assumption that the image brightness does not change under the apparent motion. This can be expressed as

$$g_2(\boldsymbol{x} + \boldsymbol{w}) - g_1(\boldsymbol{x}) = 0 \ . \tag{1.3}$$

In practice, the brightness constancy constraint is approximated by a first order Taylor expansion, either early on in the model or later during the minimisation. It then becomes

$$g_{2x}u + g_{2y}v + g_2(\boldsymbol{x}) - g_1(\boldsymbol{x}) = 0 \ , \tag{1.4}$$

where g_{2x} and g_{2y} stand for the spatial derivatives of $g_2(\boldsymbol{x})$ in x and y direction. The above equation is clearly insufficient to determine the two optical flow components u and v uniquely, a problem referred to as the *aperture problem*. With his respect, the linearised brightness constancy constraint is a line constraint that tells us on which line the matching point has to lie, but not exactly where. As illustrated in Fig. 1.2 (b), only the component of the optical flow orthogonal to this line, the so-called *normal flow*, can be determined. The aperture problem may be overcome by assuming constancy on other image features, such as the image gradient, but these should not depend linearly on the brightness. In the extreme case, all neighbouring points share the same brightness value (homogeneous region) and the matching process is undefined. This is shown in Fig. 1.2 (a).

The epipolar constraint (1.2) can be considered as an additional line constraint on the optical flow. The requirement that a corresponding point has to lie on its epipolar line can thus be used as an extra cue for disambiguating the possibly multiple solutions of the data term. This is shown in Fig. 1.2 (c). If the fundamental matrix is known, the epipolar lines are fixed and the epipolar constraint can be imposed as a *hard constraint*.

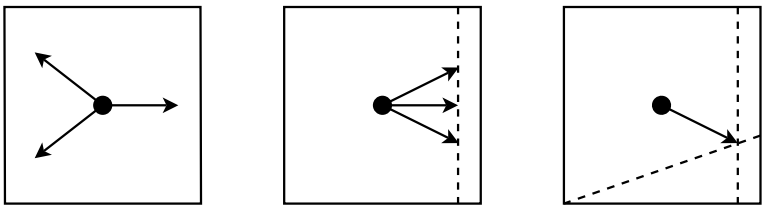

Fig. 1.2: The aperture problem (partially adapted from [Bru06]). **(a)** No image information, the correspondence can lie anywhere. **(b)** Linearised brightness constancy assumption in a textured region, the correspondence lies on a line. **(c)** Additional epipolar constraint, the correspondence can be determined uniquely.

This reduces the originally two-dimensional matching process to a one-dimensional search along the epipolar lines [ADSW02, SBW05, BAS07]. A disadvantage of this approach is that the quality of the optical flow depends on how well the fundamental matrix has been estimated. Conversely, the same is also true for the quality of the fundamental matrix when we estimate it from a given fixed optical flow field. In this light it thus makes sense to estimate both quantities simultaneously, such that the accuracy of one is not limited by the method used to compute the other. This can be achieved by adding an *epipolar term* to the optical flow functional that imposes the epipolar constraint as a *soft constraint*

$$\mathcal{E}(\boldsymbol{w}, F) = \int_{\Omega} \big(\mathcal{E}_{\mathrm{D}}(g_1, g_2, \boldsymbol{w}) \; + \; \alpha \, \mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w}) \; + \; \beta \underbrace{\mathcal{E}_{\mathrm{E}}(\boldsymbol{w}, F)}_{\text{epipolar term}} \big) \, \mathrm{d}\boldsymbol{x} \; . \tag{1.5}$$

By minimising $\mathcal{E}(\boldsymbol{w}, F)$ for $\boldsymbol{w}$ and F, we are introducing a feedback between both unknowns, such that the optical flow can benefit from an improved stereo geometry and vice versa. Moreover, estimating all quantities within a single optimisation framework puts the overall approach on a more solid theoretical basis and guarantees an equal (symmetric) treatment of geometry (epipolar lines) and flow (correspondences).

★ ***Contribution 2.*** Motivated by the promising results of our first contribution, we propose a new variational model that recovers the fundamental matrix and the optical flow simultaneously as the minimisers of a single energy functional. We demonstrate that this joint approach further improves the quality of the fundamental matrix estimation and yields better results than an optical flow approach without epipolar geometry estimation.

Stereo Reconstruction. The joint estimation of the fundamental matrix and the optical flow solves the dense reconstruction problem in a single optimisation step. With the help of the recovered fundamental matrix, we can now project each corresponding point pair back into space and find the associated three-dimensional point. This type of reconstruction, however, is mostly not metrically correct. A more truthful reconstruction can be obtained under the assumption that certain information about the cameras is known, such as the focal length. In this case, the fundamental matrix can be expressed in coordinates that have been normalised by the camera parameters and will take on a special form known as the *essential matrix* [Lon81]. In this thesis we will compute the essential matrix from

the fundamental matrix as an intermediate step in obtaining a realistic reconstruction of the scene. The question that immediately arises is why we do not directly estimate the essential matrix? First of all, we have experienced that there is no relevant gain over the already high accuracy of the geometry estimated via the fundamental matrix. Secondly, estimating the fundamental matrix makes our approach more generally applicable, even if no information at all is available about the camera system. Finally, and more importantly, there is a modelling reason for not doing so: The essential matrix does not live in the pixel space that is commonly used for optical flow computation and its retrieval would therefore cut apart the symmetric treatment of flow and geometry in a joint model.

Joint Stereo Reconstruction and Scene Flow Estimation. Since depth information is required to determine three-dimensional motion, scene flow can not be computed without estimating the scene structure as well. In contrast to *structure from motion* [IA99, TZ99], scene flow does not relate to a static world. Instead, objects in the scene are allowed to move freely and in a non-rigid fashion. Thus, for estimating scene flow, stereo sequences are required that provide at least two views per time instance. The minimal set-up for scene flow computation is depicted in Fig. 1.3, where (g_1, g_2) is a stereo pair at time t and (g_3, g_4) represents a stereo pair at the next time instance $t + 1$.

By adding the time dimension to our problem, we obtain a mixture of pure optical flow estimation (scene motion) and the estimation of stereo geometry and stereo correspondences (scene reconstruction). Existing algorithms often treat stereo and motion independently and rely on a sequential computation of the scene flow and structure [PAT96, ZK01, VBR$^+$05, PKF07, WRV$^+$08]. However, to improve the quality of the estimation it is important that motion and shape estimation are coupled. To estimate all the optical flows between the images and all the fundamental matrices between the stereo pairs in a joint variational framework, we will minimise an extended energy of the form

$$\begin{aligned}\mathcal{E}(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k, F_1, \dots, F_l) = \int_\Omega \Bigg(& \sum_{i=1}^{k} \mathcal{E}_{\mathrm{D}i}(g_1, \dots, g_n, \boldsymbol{w}_1, \dots, \boldsymbol{w}_k) \\ & + \sum_{i=1}^{k} \alpha_i \, \mathcal{E}_{\mathrm{S}i}(\nabla \boldsymbol{w}_i) \\ & + \sum_{i=1}^{l} \beta_i \, \mathcal{E}_{\mathrm{E}i}(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k, F_i) \Bigg) \, \mathrm{d}\boldsymbol{x} \ . \end{aligned} \tag{1.6}$$

In the above general energy, the enumeration index n stands for the number of images in the model, k for the number of optical flows between them and l for the number of image pairs between which a fundamental matrix can be computed. For our four-frame model of Fig. 1.3, $n = 4$, $k = 6$ and $l = 2$. We will see in this thesis that the six pairs of optical flow variables $\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_6$ are redundant and that their number can be reduced by taking into account the connections that exist between all image correspondences. It is additionally reasonable to assume that the fundamental matrix remains constant between the two time instances t and $t + 1$ if the stereo system does not change too much under operation. This will eventually lead to a minimal parameterisation for the joint estimation of camera motion, scene structure and scene flow from stereo image sequences.

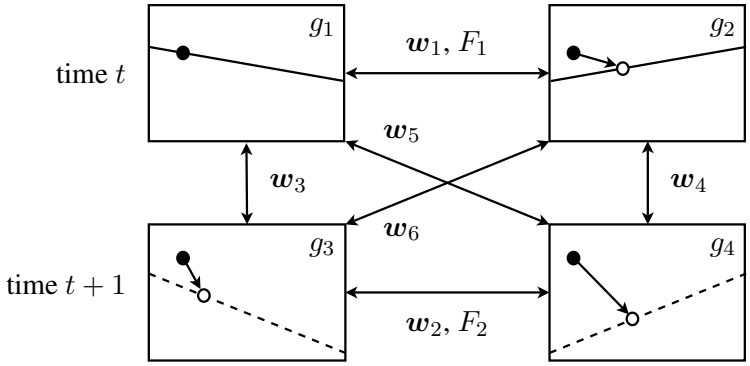

Fig. 1.3: The four-frame set-up used for computing scene flow. The original location of a point in image g_1 is indicated by a black dot, the new location in the other images by a white dot. The stereo correspondences have to lie on corresponding epipolar lines.

★ ***Contribution 3.*** As a third main contribution, we extend our framework for the joint estimation of correspondences and stereo geometry to multiple images. For the four-frame case, consisting of two stereo image pairs with a temporal offset, we develop a variational method that simultaneously reconstructs the scene at two different time instances. Because we recover the fundamental matrix in the process, our method is able to compute the three-dimensional scene flow for sequences with an unknown and arbitrary stereo geometry.

1.3 Related Work

After giving an overview of the thesis and the most important contributions therein, we turn our attention to related work.

Estimation of the Fundamental Matrix. We start by sketching existing approaches for estimating the epipolar geometry. The epipolar geometry is the relation that underlies two stereo views and can be described by a single entity, the fundamental matrix [HZ00, FLP01]. The fundamental matrix is a 3×3 matrix of rank 2 that is defined up to a scale factor. For *uncalibrated* images – i.e. images taken by cameras for which no information about orientation, pose or internal settings is available – it is possible to estimate the fundamental matrix solely from image correspondences by means of the epipolar constraint [HZ00, FLP01]. Most techniques establish these correspondences by matching a sparse set of characteristic image features. Such feature based techniques proceed in different stages:

- *Feature Extraction.* Classical examples of image features include edges [MH80, Can86], Förstner-Harris features [FG87, HS88] and KLT features [LK81, TK91, ST94]. Matching can for instance be achieved by thresholding the correlation score of the surrounding image patch. The set of correspondences obtained this way is often ambiguous since several points might be paired to the same one [ZDFL95]. More recently, scale invariant SIFT [Low04] and SURF [BETV08] features have been proposed that identify locations of interest in scale-space [Wit83, Lin94] and

associate with each of them a highly distinctive descriptor vector. These type of features can be matched correctly with a higher probability. Comparative studies by Mikolajczyk and Schmid [MS05] have repeatedly put forward SIFT as one of the most accurate local matching algorithms to date.

- *Fundamental Matrix Estimation.* Feature based approaches for estimating the fundamental matrix go back to Longuet-Higgins [Lon81], who introduced the 8-point algorithm to compute the essential matrix, the equivalent of the fundamental matrix for internally calibrated cameras. The 8-point algorithm excels in simplicity because of its linear nature but it has the disadvantage that the quantity being minimised has no geometrical interpretation. To overcome this shortcoming, Weng *et al.* [WHA89] and Luong and Faugeras [LF96] proposed nonlinear techniques involving geometrically meaningful measures such as the distance of a point to its epipolar line. Hartley and Zisserman [HZ00] recommend a Maximum Likelihood (ML) estimation that minimises the distance of a point to the manifold determined by the parameterisation of the fundamental matrix. The ML estimate is optimal from a statistical viewpoint if a Gaussian error model is assumed [WAH93] and is sometimes called the *Gold Standard algorithm*. A non-linear technique that makes use of the minimal set of point correspondences that is theoretically needed to compute the fundamental matrix is the 7 point algorithm [HZ00]. As opposed to the 8-point algorithm, the 7-point algorithm directly imposes the rank 2 constraint but can lead to multiple solutions and can not be extended to an arbitrary number of features.

- *Outlier Removal.* In practice, a feature based estimation method has to be able to deal with false correspondences arising from the lack of geometrical constraints in the matching process. This has led to a multitude of robust extensions that can handle a relatively large amount of outliers: M-estimators [Hub81], Least Median of Squares (LMedS) [RL87] and the Random Sample Consensus (RANSAC) [FB81] number among such robust techniques. In recent years numerous RANSAC variants have been developed to further increase robustness against localisation noise (LO-RANSAC [CMO04]) and degenerate configurations (DEGENSAC [CWM05] and QDEGSAC [FP06]). An overview of robust estimators in the context of fundamental matrix computation is given in [TM97, Zha98, Ste99, RFP08].

  Clearly, the quality of feature based methods relies on the quality of the random sampling approach. However, one should not forget that features may also suffer from well-known localisation errors due to their computation in scale-space; see e.g. [WB94, ZGS$^+$09].

In addition to feature based methods there also exists a limited number of approaches that estimate the fundamental matrix directly from image information [SHS07].

Estimation of Dense Image Correspondences. Now that we have discussed sparse feature based methods for estimating the fundamental matrix, we review dense techniques for establishing correspondences within a global energy minimisation framework. Typical representatives are variational methods for computing the optical flow and methods for estimating the correspondences from stereo pairs.

- *Optical Flow.* Variational optical flow methods minimise an energy functional that models temporal constancy of image properties via a data term and regularity of the flow field via a smoothness term. The quadratic data term of the seminal model of Horn and Schunck [HS81] has been gradually extended with robust variants that combine several constancy assumptions in order to cope with noise, occlusions and illumination changes [BBPW04, KMK05, MBW07, SPC09a, WPB10]. To respect discontinuities in the optical flow, smoothness terms have been proposed that take into account image edges [NE86, AELS99], edges in the evolving flow field [Sch94, WS01a] or both [ZBW$^+$09, ZBW11]. Similar extensions have been introduced in a probabilistic setting using discrete models [BA91, MP98a, SRLB08] or by coupled systems of partial differential equations [PVPO94]. More recently, variational optical flow estimation has been coupled to feature based descriptor matching in order to combine their respective advantages [Val07] and to allow for the recovery of the large motion of small objects [BBM09, BM10]. The minimisation of variational optical flow models mostly proceeds by means of a gradient descent algorithm or via a discretisation of the Euler-Lagrange equations. In recent years, fast numerical schemes have played an increasingly important role in the real-time application of optical flow [BWKS06, ZPB07a, GZG$^+$10].

- *Stereo Correspondences.* The basic idea of using dense, energy-based methods is not restricted to optical flow computation, it can also be applied to depth estimation from stereo pairs. If the epipolar geometry of a stereo image pair is known, the correspondence problem even reduces to a one dimensional search – the search along epipolar lines. Most successful stereo correspondence methods are either discrete or continuous optimisation techniques [SS02]. Discrete methods model the images and displacement field as Markov random fields and try to find the most probable displacement for the given data. Minimisation is usually done by means of graph cuts [KZ02], belief propagation [KSK06, FH06] or dynamic programming [LSY06]. While discrete approaches allow a better optimisation of the energy by constraining the depth values to a discrete subset, continuous variational methods can be advantageous when smooth transitions are favoured. Variational methods decompose the optical flow along the epipolar lines [ADSW02, SBW05], restrict the estimation process in horizontal direction [BAS07] after rectifying the images, or solve for the unknown depth in every pixel [RD96, STV03]. Since the latter approaches will generally lead to a highly non-linear energy, not many convincing minimisation strategies exist that solve directly for the unknown depth [SGC10]. Aside from methods that include the epipolar constraint as a hard constraint, there is also a limited number of approaches that feed a precomputed epipolar geometry into an optical flow method by means of a soft constraint [WPB$^+$08].

Estimation of the Epipolar Geometry from Optical Flow. Related to our first contribution are methods that recover stereo information from dense image correspondences. Early ideas that couple optical flow with the estimation of the epipolar geometry go back to the differential form of the epipolar constraint by Viéville and Faugeras [VF95] and Brooks *et al.* [BCB97]. Based on these works, Ohta and Kanatani [OK95] and Kanatani *et al.* [KSO$^+$00] have presented statistical evaluations of the estimation of the stereo geo-

metry and the depth from monocular image sequences. These studies suffered, however, from inaccurate optical flow methods and a lack of absolute performance measures. Moreover, in such a differential setting optical flow is regarded as the infinitesimal velocity field of an image sequences , thereby strictly separating *structure from motion* from *wide baseline stereo*. In this thesis we do not make such a distinction, since recent optical flow methods are able to cope with both small and large displacements. In another early work, Irani and Anandan [IA99] already advocate the use of dense correspondences for the estimation of motion models in a structure from motion setting. Only the more recent work of Strecha *et al.* [SFV04] is known to us in which dense optical flow correspondences are used for the calibration of a stereo rig, but no quantitative results are reported.

Coupled Estimation of the Epipolar Geometry and Correspondences. Our second contribution is a coupled model for optical flow and epipolar geometry estimation, where a rigid motion model enters the functional in the form of a soft constraint with unknown fundamental matrix entries. In the optical flow model of Nir *et al.* [NBK08] this happens via an explicit parameterisation of the displacements by unknown motion coefficients. Also close in spirit to our ideas are feature based ML methods that simultaneously estimate the fundamental matrix while correcting an initial set of point correspondences [WAH93, HZ00]. These methods are, however, sparse in nature and differ by the distance measure that is being minimised. Another more recent sparse method that pairs the epipolar constraint with the brightness constancy assumption is that of Saragih and Goecke [SG07]. In an early work, Hanna [Han91] iteratively estimates camera motion and dense structure parameters in a joint refinement process by using the optical flow as an intermediate representation. Contrary to our approach, the problem is formulated in a differential setting and the optimisation performed locally. A simultaneous estimation of the fundamental matrix and the three-dimensional surface is proposed in an uncalibrated discrete setting by Schlesinger *et al.* [SFS04]. Despite a joint model formulation, proper initialisation is required to bootstrap the method. More recently it has been argued that it can be beneficial to couple the computation of epipolar geometry and optical flow fields within a joint energy functional [VBW08, WCPB09].

Estimation of Scene Flow. We conclude this related work section by providing a short survey of methods that retrieve the three-dimensional motion of the scene, i.e. the scene flow. As mentioned before, scene flow can not be computed without estimating the scene structure as well. Existing scene flow algorithms often treat this motion and stereo component independently. In fact, most of them rely on a sequential computation of the scene flow and structure [PAT96, ZK01, VBR$^+$05, PKF07, WRV$^+$08]. To improve the quality of the estimation, however, some methods exploit the spatial and temporal dependencies in the image sequence [ZK01, CK02, MS06, HD07, CPMK09]. Among those methods that solve for the scene flow and structure simultaneously, variational approaches play a major role. A small number of these techniques parameterise the problem directly in 3D space [CPMK09, BMK10]. Others are based on optical flow computation [ZK01, MS06, HD07, WVM$^+$08] and have consistently improved their accuracy over the years in the wake of increasing optical flow accuracy.

Closest related to the scene flow algorithm that we propose as the third main contribution in this thesis are the methods of [ZK01, MS06, HD07], which jointly compute spatial and

temporal motion fields by minimising a single energy. In particular the method of Huguet and Devernay [HD07] uses similar data constraints as our approach. As opposed to our energy, it applies a joint smoothness term to all appearing displacement fields. A more adequate separate treatment of the smoothness term is proposed by Wedel *et al.* [WVM+08] who decouple the estimation of structure and motion to achieve real-time performance. In their case, however, the separate smoothness term does not yield more accurate results than their preceding work with joint regularisation [WRV+08]. Apart from these methods that parameterise the displacements in terms of image coordinates, there are also techniques that work directly in 3D space. Such techniques include methods based on reprojection errors [CPMK09], space carving and nonlinear optimisation [VBSK00, CK02], deformable meshes [FP08], Markov Random Fields [IM06] and direct depth estimation [BMK10]. In contrast to our approach, all these methods rely on a previous calibration step, since they are either based on rectified sequences or require explicit camera information.

1.4 Outline

Different parts of the work presented in this thesis have been published at conferences [VBW08, VBZ+10] and in journals [VBMW11]. In the following we give an outline of the thesis and shortly discuss the content of the different chapters. We make use of this summary to highlight additional contributions that have not received consideration in the overview section 1.2.

A Brief History of Stereo Geometry. Chapter 2 lays the basis for our work by discussing the image formation process and introducing the mathematical tools that are necessary to establish relations between the two- and three-dimensional world. We give an introduction to projective geometry and provide an elementary camera model that can be abstracted by a single matrix operation. We will see that one image alone does not allow us to perform a reconstruction of the depicted scene and that at least a second camera has to be added to overcome the depth ambiguity. The chapter then introduces the most important description of the geometry of an uncalibrated stereo pair, the fundamental matrix, and its counterpart for known internal camera parameters, the essential matrix. The estimation of the fundamental matrix and the subsequent extraction of the essential matrix will allow us to reconstruct a scene up to scale in the camera coordinate system.

Variational Optical Flow. Chapter 3 is devoted to the modelling and numerical solution of variational optical flow methods. With the estimation of the fundamental matrix as a later application in mind, we will consider here classical unrestricted optical flow between two stereo images. Typical for a stereo setting is that our optical flow model has to be able to cope with the large displacements that can arise from a substantial change in viewpoint. To this end, we focus on choices for the data term that can handle large displacements and are at the same time robust under noise and illumination changes. With respect to the smoothness term, we discuss the classification of discontinuity preserving regularisation strategies according to [WS01a], as well as more recent ideas that extend this taxonomy. A suitable combination of data and smoothness constraints will then give rise to the model prototype of Brox *et al.* [BBPW04] that will serve as a baseline method for most of the

techniques proposed in this thesis. We finally move on to the minimisation and discretisation of our prototypical model, where the coarse-to-fine warping strategy will play a central role in avoiding local minima. An incremental computation based on the motion tensor will further assure convexity of the final energy.

Fundamental Matrix from Dense Optical Flow. In Chapter 4 we collect our knowledge from the two previous chapters and propose a two-step method for estimating the epipolar geometry of a stereo pair. Our technique first establishes a dense set of image correspondences by means of optical flow and then estimates the fundamental matrix by imposing the epipolar constraint in each pixel. The estimation is based on a modified version of the 8-point algorithm [Lon81] that exploits the large amount of optical flow correspondences and their inherent robustness against large outliers in a reweighted least-squares framework. Via a systematic juxtaposition of our dense method with existing sparse feature based methods, we demonstrate that modern optical flow methods can serve as novel approaches for estimating the epipolar geometry with competitive quality and identify scenarios where dense techniques should be preferred over sparse ones.

While these key results are of a more general nature, we have to restrict our methodology to prototypical representatives. As a baseline method for accurate dense optical flow, we choose the approach of Brox *et al.* [BBPW04]. For feature based approaches we consider two feature matching algorithms (KLT [ST94] and SIFT [Low04]), three random sampling algorithms (LMedS [RL87], LORANSAC [CMO04] and DEGENSAC [CWM05]), and two different distance measures (epipolar distance [FLP01] and reprojection error [HZ00]). This comes down to twelve different variants of feature based methods.

Optical Flow for Uncalibrated Stereo Images. In Chapter 5 we demonstrate that not only the estimation of the epipolar geometry from dense optical flow is promising, but that there exists also a benefit in the opposite direction: Knowing the epipolar geometry can have a stabilising effect on the estimation of the optical flow. To this end we introduce a novel technique that estimates the optical flow and the fundamental matrix simultaneously in a single variational framework. In contrast to existing stereo methods that make explicit use of a given precomputed fundamental matrix, we do not require the epipolar geometry to be known beforehand. This makes our approach more flexible and applicable in *uncalibrated* stereo settings, where no knowledge about the stereo geometry is available.

In practice, we achieve this goal by incorporating the epipolar constraint as a *soft constraint* into the optical flow functional. This allows the optical flow and the fundamental matrix to correct each other during a joint optimisation process, which ensures a set of image correspondences that is most consistent with the estimated stereo geometry. Besides yielding better optical flow results than approaches that do not estimate the epipolar geometry in the process, our experiments will show that our joint estimation method further improves the fundamental matrix estimation of the two-step method of Chapter 4. Since the modelling in a variational framework is very general, our strategy can profit directly from any progress in future optical flow research. This will be demonstrated by replacing the baseline optical flow approach of Brox *et al.* [BBPW04] by the more recent method of Zimmer *et al.* [ZBW11]. We conclude the chapter with a discussion of the advantages and shortcomings of our joint variational method.

Scene Flow for Uncalibrated Stereo Sequences. In Chapter 6 we propose a variational scene flow method for uncalibrated stereo sequences. We do this by integrating the spatial and temporal information from two stereo pairs in a global energy functional while simultaneously estimating the unknown stereo geometry on consecutive time steps. Assuming that the internal camera parameters are known, our method allows to recover the dense scene structure and the dense scene flow up to a scale factor.

Apart from this novel generalised model, we make four additional contributions: (i) Within the multi-resolution framework required to handle large displacements, we extend the motion tensor notation of optical flow [BW05, Bru06] to multi-dimensional linearised data constraints and to the soft epipolar constraint. This notation allows us to normalise these constraints such that deviations from model assumptions can be interpreted as geometrical distances. (ii) Secondly, we show the equivalence of the normalised epipolar constraint to two widely-used distance measures that have so far only been encountered in the context of feature based methods in Chapter 4: the epipolar distance [FLP01] and the Sampson error [WHA89]. (iii) Thirdly, we propose a regularisation strategy that penalises discontinuities in the different displacement fields separately. This makes sense, since motion and depth continuities do not necessarily coincide. (iv) As a final contribution, we explicitly detect occlusions due to scene motion and changes in camera viewpoint and exclude them from the estimation process. Experiments on synthetic calibrated and uncalibrated data and on real-world sequences demonstrate the benefits of our contributions. We even outperform recent techniques that make explicit use of a given stereo geometry.

2 A Brief History of Stereo Geometry

The work presented in this thesis deals with methods that aim to infer properties of the three-dimensional world from two-dimensional images. To be able to do this we must understand how the real world relates to its projected image and vice versa. In other words, we must investigate how images are formed and what the connections are between images of the same scene. Only by this understanding we will be able to perform the inverse task of recovering the position of a point in the 3D world from its 2D projections.

This chapter lays the basis for our work by introducing the mathematical tools that are necessary to establish relations between the 2D and 3D world. In Sec. 2.1 we will start with a short overview of projective geometry which is the representation of choice for the actions of a projective camera. Sec. 2.2 is devoted to single view geometry and provides an elementary camera model that can be abstracted by the operation of a single *projection matrix*. We will see that one image alone does not allow us to perform a 3D reconstruction of the depicted scene and that at least a second camera has to be added to overcome the depth ambiguity. Two view geometry will be the topic of Sec. 2.3, which introduces the most important matrix entity of this thesis: the *fundamental matrix*. While the fundamental matrix is a complete description of the geometric relation between two stereo images, it does not suffice for a truthful reconstruction of the scene. This is where a third matrix, the *essential matrix*, comes into play. Trough its recovery in Sec. 2.4 we will close in on our final goal: a reconstruction of the scene up to scale in the camera coordinate system.

2.1 Projective Geometry in a Nutshell

Central to the mechanism of image formation is the concept of projection, which associates a point in the image plane with points in the real world. For describing projections and the relations between different images of the same scene, traditional Euclidean geometry has proven inadequate. For instance, parallel lines in $\mathbb{R}^3$ will generally not be parallel anymore after they have been projected onto the image plane in $\mathbb{R}^2$. In fact they will meet in a *vanishing point*, which is the projection of a world point that lies infinitely far away from the observer. Parallelism is therefore clearly not preserved by projection, and neither are angles and distances. This is where *projective geometry* enters the game. Projective geometry describes a larger class of transformations than the usual translation and rotation of Euclidean geometry, including the perspective projection of a camera. In particular, projective geometry makes it possible to reason about infinity, as illustrated before by the projected intersection of parallel lines: Points at infinity, which are not defined in Euclidean space, are not treated any differently in projective space and can even be mapped to finite points. This makes projective geometry an important tool in the study of the geometry of multiple images and their relation to the three-dimensional world.

The two-dimensional projective space $\mathbb{P}^2$ will be used to describe entities in the image plane. It is a simple extension of the Euclidean space $\mathbb{R}^2$, obtained by adding a 1 to the coordinates of a point $\boldsymbol{x} = (x, y)^\top \in \mathbb{R}^2$. This results in the coordinate triple $(x, y, 1)^\top$. To ensure a one-to-one correspondence between the Euclidean and the projective coordinates, we additionally demand that $(x, y, 1)^\top$ and $(kx, ky, k)^\top$, with $k \in \mathbb{R}_0$, represent the same projective point. The points of $\mathbb{P}^2$ are thus represented by equivalence classes of coordinate triples and are equal when they differ by a common scale factor. Because of the scale factor, the projective coordinates of a point $\boldsymbol{x}$ are also called *homogeneous coordinates* and will be denoted further by $\boldsymbol{x}_\mathrm{h}$. Conversely, the two Euclidean coordinates can be obtained by dividing the homogeneous coordinates by their last entry. Points with the homogeneous coordinate $(x, y, 0)^\top$ have no Euclidean equivalent but can be interpreted as the limit of a point in $\mathbb{R}^2$ in the direction $(x, y)^\top$: They are points at infinity.

Just as points, lines in $\mathbb{P}^2$ are represented by a homogeneous three-coordinate vector. This shared representation allows for the elegant formulation of several properties in which the roles of points and lines can be interchanged. The incidence of a point $\boldsymbol{x}_\mathrm{h} = (x, y, 1)^\top$ and a line $\boldsymbol{l} = (a, b, c)^\top$, for instance, can be expressed by the equation

$$\boldsymbol{l}^\top \boldsymbol{x}_\mathrm{h} = \boldsymbol{x}_\mathrm{h}^\top \boldsymbol{l} = 0 \ , \tag{2.1}$$

which is nothing else than the standard equation of a line $ax + by + c = 0$. Moreover, the properties of intersection of two lines and join of two points can be expressed in an equivalent way, namely by the cross product of their respective coordinates. As such, the line joining two points $\boldsymbol{x}$ and $\boldsymbol{x}'$ can be written as

$$\boldsymbol{x}_\mathrm{h} \times \boldsymbol{x}'_\mathrm{h} = [\boldsymbol{x}_\mathrm{h}]_\times \boldsymbol{x}'_\mathrm{h} \ , \tag{2.2}$$

where $[\boldsymbol{x}_\mathrm{h}]_\times$ is the skew-symmetric matrix (up to scale)

$$\begin{pmatrix} 0 & -1 & y \\ 1 & 0 & -x \\ -y & x & 0 \end{pmatrix} \ , \tag{2.3}$$

whose left and right null-space are $\boldsymbol{x}_\mathrm{h}$ [HZ00, FLP01]. The symmetric role that points and lines play in $\mathbb{P}^2$ is commonly known as the *duality principle* and greatly facilitates the understanding of many projective relations, including the all-important epipolar constraint.

The extension of the three-dimensional Euclidean space $\mathbb{R}^3$ to the corresponding projective space $\mathbb{P}^3$ is carried out in the same way as for $\mathbb{R}^2$. This time the equivalence classes are the three-dimensional points and planes that are both represented by a vector of four coordinates defined up to a scale factor. The homogeneous coordinates of the three-dimensional point $\boldsymbol{X} = (X, Y, Z)^\top \in \mathbb{R}^3$ are $(X, Y, Z, 1)^\top$ and will be denoted by $\boldsymbol{X}_\mathrm{h}$. Join and incidence relations for points and planes in $\mathbb{P}^3$ are derived analogously as for points and lines in $\mathbb{P}^2$ and there exists a similar duality for both. The representation of lines in $\mathbb{P}^3$ is based on so-called Plücker coordinates that satisfy a quadratic constraint known as the Plücker relation [HZ00, FLP01]. The treatment and duality of lines in $\mathbb{P}^3$ is generally more complex than in $\mathbb{P}^2$ and falls outside the scope of this thesis.

2.2 One View

A camera performs a transformation between the 3D world and the 2D image plane. One of the simplest mathematical models for this mapping is provided by the *pinhole camera* model depicted in Fig. 2.1 (a). It works by means of a central projection of points in space onto a plane. Using projective geometry, this mapping can be represented by a matrix operation. In the following we will derive a general form of the projection matrix of a camera and analyse how it encodes the specific camera properties.

2.2.1 A Simple Camera Model

At first we assume that the centre of projection $\boldsymbol{C}$ coincides with the origin of the Euclidean world coordinate system. As shown in Fig. 2.1 (a) the Z-axis of this system is chosen to be perpendicular to the image plane and is called the *optical axis*. It intersects the image plane in positive direction in the *principal point* $\boldsymbol{p}$, which in turn serves as the origin of the 2D image coordinate system. The distance from the camera centre $\boldsymbol{C}$ to the image plane is known as the focal length and will be denoted by f.

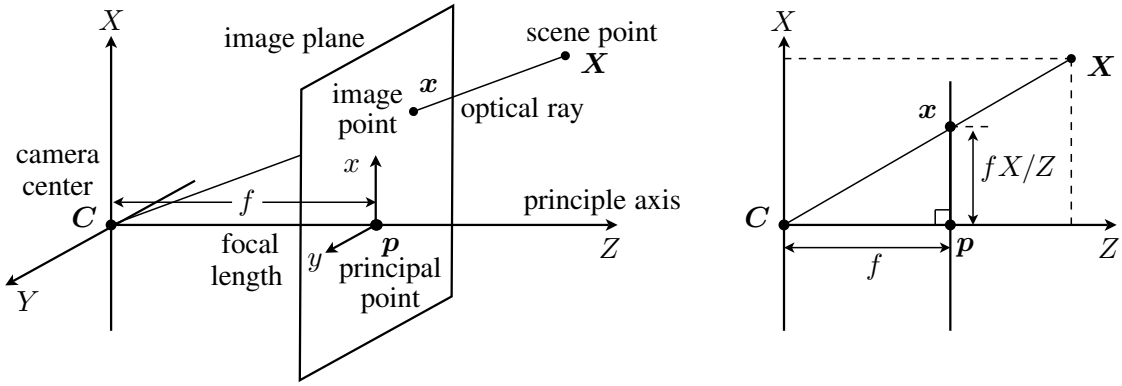

Fig. 2.1: The pinhole camera geometry. **Left: (a)** Overview. **Right: (b)** X-Z cross section.

A 3D world point $\boldsymbol{X} = (X, Y, Z)^\top$ is projected onto the point $\boldsymbol{x} = (fX/Z, fY/Z)^\top$ where the *optical ray*, the line joining $\boldsymbol{X}$ and $\boldsymbol{C}$, meets the image plane. This relationship can be easily derived from Fig. 2.1 (b) by means of the theorem of intersecting lines. If the world point and the image point are expressed in homogeneous coordinates, the central projection is simply a linear mapping and can be written as

$$\boldsymbol{x}_\mathrm{h} = \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{pmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = P\boldsymbol{X}_\mathrm{h} \ , \tag{2.4}$$

where an empty entry denotes 0. The 3×4 matrix P is called the camera *projection matrix* and has the form

$$P = \mathrm{diag}\,(f, f, 1)\,(I^3, \mathbf{0}^3) \ , \tag{2.5}$$

where I^3 is the 3×3 identity matrix, $\mathbf{0}^3$ the 3×1 zero vector and $\mathrm{diag}\,(a, b, c)$ denotes the 3×3 diagonal matrix with diagonal entries a, b and c.

Definition 2.1. *The 3×4 matrix P which represents a linear projective mapping from points in $\mathbb{P}^3$ to points in $\mathbb{P}^2$ is called the camera projection matrix, or short, camera matrix.*

2.2.2 A General Camera Model

The camera matrix provides a complete description of the projective action of a pinhole camera, but its formulation in Eq. (2.5) is due to our particular choice of coordinate system. In general, the world and image coordinate axes are not aligned with the camera, as illustrated in Fig. 2.2. In the following we will investigate a more general form of the camera matrix that maps points in a world coordinate frame to pixel coordinates.

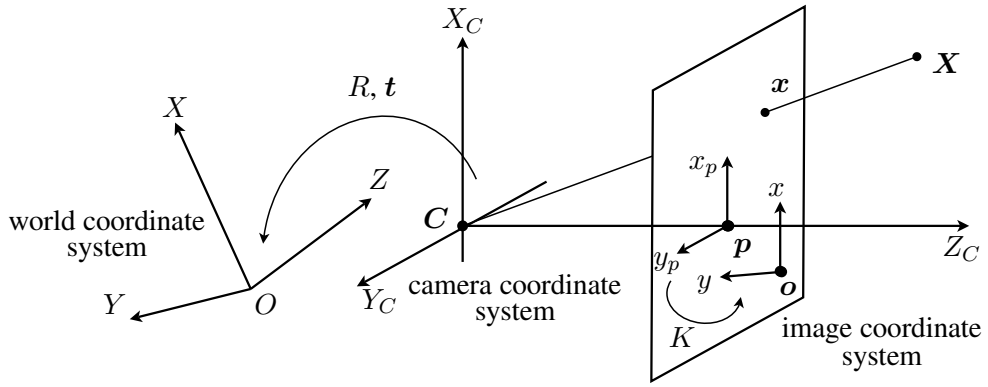

Fig. 2.2: The general pinhole camera geometry.

2.2.2.1 Intrinsic Camera Parameters

In most image processing and computer vision applications, the origin of the pixel coordinate system in the image plane is chosen to be in the upper left corner of the image, and not in the principle point $\boldsymbol{p}$. If we take into account the offset of $\boldsymbol{p} = (x_0, y_0)^\top$ from the origin, the image point $\boldsymbol{x}$ corresponding to $\boldsymbol{X} = (X, Y, Z)^\top$ has the coordinate $(fX/Z + x_0, fY/Z + y_0)^\top$. As a consequence, the camera matrix becomes

$$P = K(I^3, \mathbf{0}^3) \quad \text{with} \quad K = \begin{pmatrix} f & & x_0 \\ & f & y_0 \\ & & 1 \end{pmatrix} \quad , \tag{2.6}$$

the 3×3 camera *calibration matrix*. In the above expression, empty matrix entries indicate a zero. The calibration matrix describes a change of coordinate system within the image plane and can be associated with the internal characteristics of the actual camera device. For a general pinhole camera, this is not only restricted to a change in origin, such as described by calibration matrix K, but can also involve a scaling in both axial directions and a possible skew due to the non-orthogonality of the coordinate system. A general calibration matrix is therefore an upper triangular matrix with 5 degrees of freedom describing the camera *intrinsics or internal parameters*. For our purposes, however, the focal length f and the principle point $\boldsymbol{p}$ are the most important intrinsic parameters and definition (2.6) of K with its 3 degrees of freedom is sufficient. Both f and $\boldsymbol{p}$ are expressed in pixel units and when their values are known the camera is said to be *internally calibrated*.

2.2.2.2 Extrinsic Camera Parameters

If we consider only one camera, we are free to choose the world coordinate system equal to the camera coordinate system such that the camera matrix can be written as $P = K(I^3, \mathbf{0}^3)$. When two cameras are involved, however, we have to take into account the *rotation* and *translation* of at least one of them with respect to the world system. In this case the camera matrix will take on a more general form. If we denote by $\boldsymbol{X}_C$ the coordinates of a world point $\boldsymbol{X}$ with respect to the camera coordinate system, then the mapping between the world system and the camera system can be expressed as

$$\boldsymbol{X}_C = R(\boldsymbol{X} - \boldsymbol{C}) = R\boldsymbol{X} + \boldsymbol{t} \ . \tag{2.7}$$

Here R is the 3×3 rotation matrix that aligns the two systems and $\boldsymbol{C}$ the world coordinate of the camera centre. We have further introduced the 3×1 vector $\boldsymbol{t} = -R\boldsymbol{C}$ to make the translation part less explicit on R and $\boldsymbol{C}$. In homogeneous coordinates we can write

$$\begin{aligned} \boldsymbol{X}_{C\,\mathrm{h}} &= \begin{pmatrix} R & \mathbf{0}^3 \\ \mathbf{0}^{3\top} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{X} \\ 1 \end{pmatrix} + \begin{pmatrix} \boldsymbol{t} \\ 0 \end{pmatrix} & (2.8) \\ &= \begin{pmatrix} R & \boldsymbol{t} \\ \mathbf{0}^{3\top} & 1 \end{pmatrix} \boldsymbol{X}_{\mathrm{h}} \ . & (2.9) \end{aligned}$$

With the help of expression (2.6) we can now write

$$\boldsymbol{x}_{\mathrm{h}} = K(I^3, \mathbf{0}^3)\ \boldsymbol{X}_{C\,\mathrm{h}} = K(I^3, \mathbf{0}^3) \begin{pmatrix} R & \boldsymbol{t} \\ \mathbf{0}^{3\top} & 1 \end{pmatrix} \boldsymbol{X}_{\mathrm{h}} = K(R, \boldsymbol{t})\ \boldsymbol{X}_{\mathrm{h}} \ . \tag{2.10}$$

The camera matrix P of a general pinhole camera is

$$P = K(R, \boldsymbol{t}) \ , \tag{2.11}$$

where R and $\boldsymbol{t}$ are the *extrinsics or external camera parameters* associated with the pose and orientation of the camera with respect to the world. When the internal and external parameters of a camera are known, the camera is said to be *fully calibrated*.

2.2.2.3 Properties of the General Camera Matrix

In the following we summarise the most important properties of the camera projection matrix $P = K(R, \boldsymbol{t}) = (KR, K\boldsymbol{t})$:

- The matrix P is a homogeneous quantity that is defined up to a scale factor and has in general 11 degrees of freedom: 5 from the calibration matrix K, 3 from the rotation matrix R and 3 from the translation vector $\boldsymbol{t}$. The camera matrix has rank 3, because the rank of the submatrix KR is 3.

- The camera rotation R and the internal parameters K can be recovered by decomposing the 3×3 submatrix KR into the product of an upper-triangular matrix K and an orthogonal matrix R via the QR-decomposition [GV89]. The calibration matrix K must have a positive diagonal.

- The homogeneous point $\boldsymbol{C}_{\mathrm{h}}$ is the right null-space of P, which can be easily shown by considering $P\boldsymbol{C}_{\mathrm{h}} = K(R, -R\boldsymbol{C})(\boldsymbol{C}^{\top}, 1)^{\top} = \boldsymbol{0}^3$. The camera centre $\boldsymbol{C}$ is thus the only point in space for which the image is undefined. Denoting the last column of P by $\boldsymbol{p}_4$, the camera centre can be computed as [HZ00]

$$\boldsymbol{C} = -(KR)^{-1}\boldsymbol{p}_4 \ . \tag{2.12}$$

- The *depth* of a 3D point $\boldsymbol{X}$ with respect to the camera P is the distance to the plane through the camera centre parallel to the image plane. It can be computed as [HZ00]

$$\frac{\mathrm{sgn}\big(\det\,(KR)\,\big)w}{||\boldsymbol{m}_3||} \ , \tag{2.13}$$

  where the vector $\boldsymbol{m}_3$ is the third row of the submatrix KR and w the homogeneous scale factor stemming from the projection $P\boldsymbol{X}_{\mathrm{h}} = w(x, y, 1)^{\top}$. Formula (2.13) is independent of the scaling of $\boldsymbol{X}_{\mathrm{h}}$ and P and it can be used to determine if $\boldsymbol{X}$ lies in front of or behind the camera.

Remark on the General Camera Matrix. In our derivation of the camera matrix we have only considered Euclidean world transformations, such as rotations and translations. These transformations are in accordance with our concept of the real world and are sufficient for most practical purposes. The camera matrix is, however, in essence a projective device between two projective spaces. If we want to take into account the larger class of projective transformations of $\mathbb{P}^3$, we should represent relation (2.7) between the camera and world coordinates by an invertible 4×4 homogeneous matrix H^4 as $\boldsymbol{X}_{C\,\mathrm{h}} = H^4\boldsymbol{X}_{\mathrm{h}}$. Similarly, a projective transformation represented by a 3×3 matrix H^3 could be applied to the image coordinates in $\mathbb{P}^2$. In its most general form the camera matrix is therefore composed of a projective transformation of 3D space, a projection of 3D space on a 2D image and a projective transformation of the 2D image. It can then be written as

$$P = H^3(I^3, \boldsymbol{0}^3)H^4 = (M, \boldsymbol{p}_4) \ . \tag{2.14}$$

The rank of P has to be 3, because for a smaller rank the range of the camera mapping would be a line or a point, and not the whole image plane. The formulation (2.14) also includes parallel projections with a camera centre at infinity and a singular submatrix M. In this work we do not consider such special cameras and restrict ourselves to perspective cameras with a finite camera centre for which $M = KR$ is unique.

2.3 Two Views

The projective mapping of a camera constitutes a reduction of our world representation by one dimension. Inverting this projection, i.e. recovering a 3D point from its image, is a difficult task since we try to go from a poor 2D representation to a richer 3D representation. If we take a look back at the basic camera model in Fig. 2.1 (a), we see that the projection of a 3D point on the image plane represents a whole incoming ray of light. Therefore, the image point $\boldsymbol{x}$ is not only the projection of its corresponding 3D point $\boldsymbol{X} = (X, Y, Z)^{\top}$, but also the projection of all points $(kX, kY, kZ)^{\top}$, $k \in \mathbb{R}_0$, on the optical ray through $\boldsymbol{x}$. As a consequence there is a *depth ambiguity* which makes it impossible to infer the position of a 3D scene point from a single image alone.

2.3.1 Reconstruction from Two Views

If we have two cameras, we can consider the projections of the same 3D point in the two images and find the intersection of their optical rays. Such a *stereo* set-up is shown in Fig. 2.3, where a 3D point $\boldsymbol{X}$ is observed by two cameras, one with camera centre in $\boldsymbol{C}$ and one with camera centre in $\boldsymbol{C}'$. To be able to infer the position of the two optical rays through $\boldsymbol{X}$, we now need to know two things: a *full description of the geometry of both cameras*, i.e. the camera matrices P and P', and the pair of *corresponding image points* $\boldsymbol{x}$ and $\boldsymbol{x}'$ on which $\boldsymbol{X}$ is projected by P and P'. For known camera matrices and image correspondences, we can compute $\boldsymbol{X}$ by solving the system of equations

$$\begin{cases} P\boldsymbol{X}_{\mathrm{h}} &= \boldsymbol{x}_{\mathrm{h}} \\ P'\boldsymbol{X}_{\mathrm{h}} &= \boldsymbol{x}'_{\mathrm{h}} \end{cases} . \tag{2.15}$$

The camera matrices P and P' encode the relative position of the right camera with respect to the left one. This is often referred to as the *camera motion*. In order for the system (2.15) to have a solution, there has to be at least a non-zero translation between the two camera centres, otherwise the two optical rays through the corresponding points coincide and the depth ambiguity remains. If the two images are taken from different view points, there is a change in position between the corresponding image points $\boldsymbol{x}$ and $\boldsymbol{x}'$. As we will show for a special type of camera geometry, this change in position is determined by the position of the 3D point $\boldsymbol{X}$ in space. For a given camera geometry, the displacement field induced in the image plane is therefore directly related to the *scene structure*.

Determining the relative camera pose and the scene structure are highly interconnected sub-problems and together they are referred to as the *reconstruction problem*. Before moving on to the design of methods for estimating structure and motion from image correspondences, we first have to analyse the basic relations that underly these correspondences. For a solution of (2.15) to exist, the coordinates of $\boldsymbol{x}$ and $\boldsymbol{x}'$ can not be arbitrary but have to satisfy a certain constraint. This brings us to the epipolar constraint and the representation of the projective geometry of two stereo views by the fundamental matrix.

2.3.2 The Fundamental Matrix

When a point $\boldsymbol{x}$ in the left image is given, the position of the corresponding point $\boldsymbol{x}'$ in the right image can not be arbitrary. This is because the 3D point $\boldsymbol{X}$ has to lie on the optical ray through $\boldsymbol{x}$. As a consequence, $\boldsymbol{x}'$ is constrained to lie on the projection of the optical ray in the right image. This constraint is known as the *epipolar constraint* and, by denoting the image of the optical ray by its projective coordinate $\boldsymbol{l}'$, we can write it as

$$\boldsymbol{x}_{\mathrm{h}}'^{\top}\boldsymbol{l}' = 0 \ . \tag{2.16}$$

Analogously, $\boldsymbol{x}$ has to lie on $\boldsymbol{l}$, the projection of the optical ray through $\boldsymbol{x}'$ in the left image. The lines $\boldsymbol{l}'$ and $\boldsymbol{l}$ are called the *epipolar lines* of $\boldsymbol{x}$ and $\boldsymbol{x}'$. As illustrated in Fig. 2.3, the epipolar lines are the intersection of the image planes and the *epipolar plane* through $\boldsymbol{C}$, $\boldsymbol{C}'$ and $\boldsymbol{X}$. The line that connects the two camera centres $\boldsymbol{C}$ and $\boldsymbol{C}'$ is called the *baseline* and it intersects the image planes in the points $\boldsymbol{e}$ and $\boldsymbol{e}'$. These points are called the epipoles. The left epipole $\boldsymbol{e}$ is the projection of $\boldsymbol{C}'$ on the left image plane and

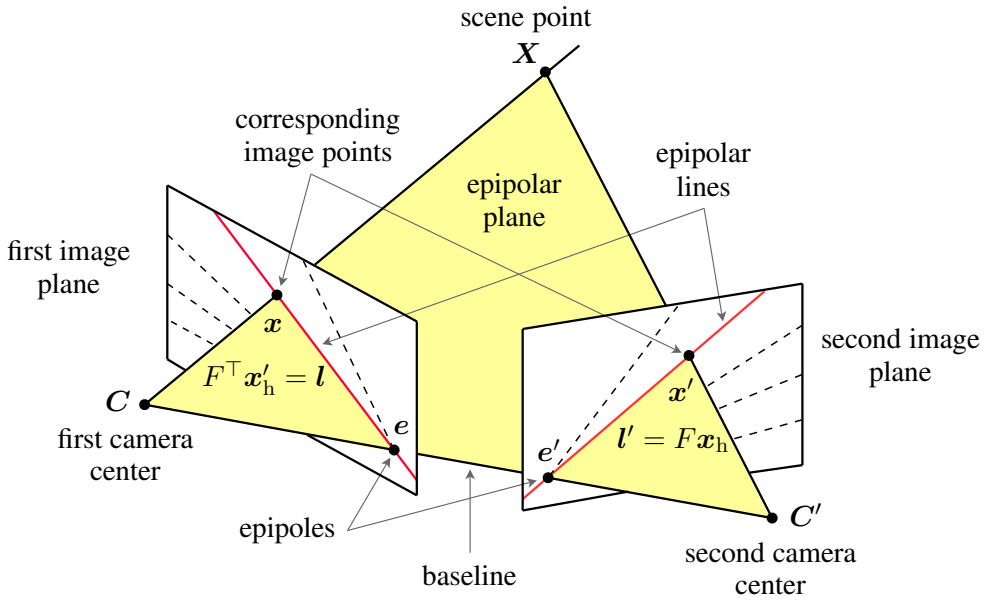

Fig. 2.3: The epipolar geometry of a stereo image pair.

the right epipole $\boldsymbol{e}'$ is the projection of $\boldsymbol{C}$ on the right image plane. Since the baseline is contained in the epipolar plane of every space point, it is clear that the epipolar lines of all corresponding points intersect in the epipoles, thereby forming a pencil of lines.

The epipolar geometry describes a mapping between points and lines, more precisely between a point in one image and its corresponding epipolar line in the other image. In the following we will formalise this mapping by assuming that the two camera projection matrices P and P' are known. We consider the following parameterization of the optical ray through $\boldsymbol{x}$ in the left image [XZ96, HZ00, FLP01]:

$$\boldsymbol{X}_{\mathrm{h}}(\lambda) = P^{+}\boldsymbol{x}_{\mathrm{h}} + \lambda \boldsymbol{C}_{\mathrm{h}} \ , \tag{2.17}$$

where P^{+} is the pseudo-inverse of P [Pen56, PTVF92, HJ94]. To verify that $\boldsymbol{X}_{\mathrm{h}}(\lambda)$ indeed lies on the optical ray through $\boldsymbol{x}$, we consider the projection:

$$P\boldsymbol{X}_{\mathrm{h}}(\lambda) = PP^{+}\boldsymbol{x}_{\mathrm{h}} + \lambda P\boldsymbol{C}_{\mathrm{h}} = \boldsymbol{x}_{\mathrm{h}} \ , \tag{2.18}$$

where we have made use of the properties $PP^{+} = I^3$ and $P\boldsymbol{C}_{\mathrm{h}} = \mathbf{0}^3$. The projections of the space points $P^{+}\boldsymbol{x}_{\mathrm{h}}$ and $\boldsymbol{C}_{\mathrm{h}}$ on the right image plane are $P'P^{+}\boldsymbol{x}_{\mathrm{h}}$ and the right epipole $P'\boldsymbol{C}_{\mathrm{h}} = \boldsymbol{e}'_{\mathrm{h}}$. The epipolar line $\boldsymbol{l}'$ is the join of these two points and can be written as the cross product of their projective coordinates

$$\boldsymbol{l}' = [\boldsymbol{e}'_{\mathrm{h}}]_{\times} P'P^{+}\boldsymbol{x}_{\mathrm{h}} = F\boldsymbol{x}_{\mathrm{h}} \ . \tag{2.19}$$

Here, the matrix

$$F = [\boldsymbol{e}'_{\mathrm{h}}]_{\times} P'P^{+} \ , \tag{2.20}$$

that describes the correspondence of $\boldsymbol{x}_{\mathrm{h}}$ and $\boldsymbol{l}$ is defined as the *fundamental matrix*.

Definition 2.2. *The 3×3 matrix F, which represents a projective mapping from points in the left image plane to the pencil of epipolar lines through $\boldsymbol{e}'$ in the right image plane, is called the fundamental matrix of the two views.*

If we now insert result (2.19) for the epipolar line $\boldsymbol{l}'$ into equation (2.16), we obtain the following expression for the epipolar constraint

$$\boldsymbol{x}_{\mathrm{h}}'^{\top} F \, \boldsymbol{x}_{\mathrm{h}} = 0 \ . \tag{2.21}$$

Note that the epipolar constraint in this form does not depend on any specific knowledge about the camera matrices P and P'. Instead, equation (2.21) describes the relation between corresponding points in the image plane and the fundamental matrix of the two views. For a complete description of the two-dimensional projective geometry of two stereo images, it is therefore not necessary to have full calibration details of the cameras but it is sufficient to know the fundamental matrix between the images. A pair of views for which the fundamental matrix is known, is said to be *weakly calibrated*.

2.3.2.1 Properties of the Fundamental Matrix

Here we summarise the most important properties of the fundamental matrix F:

- The matrix F is a homogeneous quantity that is defined up to a scale factor and has 7 degrees of freedom: the 8 degrees of freedom of a homogeneous 3×3 matrix minus the constraint that F is of *rank* 2. The singularity of F is apparent from definition (2.20) where $[\boldsymbol{e}_{\mathrm{h}}']_{\times}$ has rank 2 and $P'P^{+}$ rank 3. The rank deficiency can also be understood from the fact that all points of an epipolar line $\boldsymbol{l}$ will be mapped by F to the same entity, namely the corresponding epipolar line $\boldsymbol{l}'$.

- The fundamental matrix is independent of the scene structure. It only depends on the internal and external parameters of the two cameras and can be computed from the camera matrices via formula (2.20). Without any knowledge about the camera matrices, however, F can be computed solely from a set of corresponding image points using the epipolar constraint (2.21). Techniques that estimate the fundamental matrix from image correspondences will be the topic of Chapter 4.

- If F is the fundamental matrix between the cameras P and P', then $F^{\top}$ is the fundamental matrix between the cameras P' and P. $F^{\top}$ represents a projective mapping from the right image plane to the pencil of epipolar lines through $\boldsymbol{e}$. As a consequence, an epipolar line in the left image can be written as

$$\boldsymbol{l} = F^{\top} \boldsymbol{x}_{\mathrm{h}}' \ . \tag{2.22}$$

- All epipolar lines $\boldsymbol{l}' = F\boldsymbol{x}_{\mathrm{h}}$ in the right image contain the epipole $\boldsymbol{e}'$, such that $\boldsymbol{e}_{\mathrm{h}}'^{\top} F \boldsymbol{x}_{\mathrm{h}} = 0, \forall \boldsymbol{x}$. This means that $\boldsymbol{e}_{\mathrm{h}}'^{\top} F = \mathbf{0}^3$ and thus that the right epipole $\boldsymbol{e}_{\mathrm{h}}'$ is the left null-space of F. Equivalently, the left epipole $\boldsymbol{e}_{\mathrm{h}}'$ is the right null-space of F. Algorithmically, the epipoles are found as the singular vector belonging to the smallest singular value of F or $F^{\top}$ via singular value decomposition (SVD) [TH84].

- A pair of camera matrices (P, P') uniquely determines a fundamental matrix F via Eq. (2.20). On the other hand, F only determines a pair of camera matrices up to a multiplication by a projective transformation. To see this, we consider the 4×4 matrix H representing a projective transformation of the world coordinates. Then

$$P\boldsymbol{X} = PHH^{-1}\boldsymbol{X} \quad \text{and} \quad P'\boldsymbol{X} = P'HH^{-1}\boldsymbol{X} \ , \tag{2.23}$$

hold for a 3D point $\boldsymbol{X}$. The corresponding points of $\boldsymbol{X}$ under (P, P') are thus the same as the corresponding points of $H^{-1}\boldsymbol{X}$ under $(PH, P'H)$. This means that the fundamental matrices for the camera pairs (P, P') and $(PH, P'H)$ are the same and that F is invariant under a projective transformation of 3D space.

2.3.3 Special Motion Types

The fundamental matrix for a general camera motion can be expressed as a function of the intrinsic and extrinsic parameters of the camera pair. As an example we will compute the fundamental matrix for a motion arising from a pure translation *without* rotation. We will further derive properties of the epipolar geometry for the special case where the translation is parallel to the image plane, the so-called *ortho-parallel* camera set-up. This configuration plays an important role in classical stereo vision because the corresponding image points lie on horizontal lines, thus greatly simplifying the correspondence problem.

2.3.3.1 Translating and Rotating Cameras

Since F is invariant under projective transformations of 3D space, we can assume without loss of generality that the left camera coincides with the world coordinate system. If we denote by R and $\boldsymbol{t}$ the rotation and translation between the two cameras, we can write the camera matrices as $P = K(I^3, \mathbf{0}^3)$ and $P' = K'(R, \boldsymbol{t})$ with camera centres $\boldsymbol{C} = \mathbf{0}^3$ and $\boldsymbol{C}' = -(KR)^{-1}K\boldsymbol{t}$. The epipoles are then

$$\boldsymbol{e}_{\mathrm{h}} = P\boldsymbol{C}'_{\mathrm{h}} = KR^{\top}\boldsymbol{t} \quad \text{and} \quad \boldsymbol{e}'_{\mathrm{h}} = P'\boldsymbol{C}_{\mathrm{h}} = K'\boldsymbol{t} \ . \tag{2.24}$$

Substituting the camera matrices into the explicit formula (2.20) for the fundamental matrix leads to an expression of F in function of K, K', R and $\boldsymbol{t}$:

$$\begin{aligned} F &= [\boldsymbol{e}'_{\mathrm{h}}]_{\times} P' P^{+}, && (2.25)\\ &= [\,K'\boldsymbol{t}\,]_{\times} K'(R, \boldsymbol{t}) \begin{pmatrix} K^{-1} \\ \mathbf{0}^{3\top} \end{pmatrix}, && (2.26)\\ &= [\,K'\boldsymbol{t}\,]_{\times} K' R\, K^{-1}, && (2.27)\\ &= K'^{-\top}[\boldsymbol{t}]_{\times} R\, K^{-1} \ . && (2.28) \end{aligned}$$

In the last step we made use of the rule for commuting a skew-symmetric matrix $[\boldsymbol{t}]_{\times}$ with a non-singular matrix K' that says $[\,K'\boldsymbol{t}\,]_{\times} = K'^{-\top}[\boldsymbol{t}]_{\times}K'^{-1}$ [HZ00, FLP01].

2.3.3.2 Translating Cameras

We now assume that the camera centre translates without a rotation of the camera, i.e. $R = I^3$. If we further assume for simplicity that both cameras have the same calibration matrix $K = K'$, we can write the fundamental matrix according to formula (2.28) as

$$F = K^{-\top}[\boldsymbol{t}]_{\times} K^{-1} = [\,K\boldsymbol{t}\,]_{\times} \ . \tag{2.29}$$

Moreover, it is easy to see that the epipoles in the left and the right image are equal:

$$\boldsymbol{e}_{\mathrm{h}} = P\boldsymbol{C}'_{\mathrm{h}} = K\boldsymbol{t} = P'\boldsymbol{C}_{\mathrm{h}} = \boldsymbol{e}'_{\mathrm{h}} \ , \tag{2.30}$$

where the equality signs stand for an equality up to a scale factor. Since the fundamental matrix (2.29) is skew-symmetric [1], it further holds for any point $\boldsymbol{x}$ that $\boldsymbol{x}_{\mathrm{h}}^{\top} F \boldsymbol{x}_{\mathrm{h}} = \boldsymbol{x}_{\mathrm{h}}^{\top} \boldsymbol{l} = 0$. Each point $\boldsymbol{x}$ in the left image lies thus on its corresponding epipolar line $\boldsymbol{l}$. Since the corresponding point $\boldsymbol{x}'$ also lies on $\boldsymbol{l}$, it follows that the epipolar line is the line of apparent motion of $\boldsymbol{x}$ and the epipole $\boldsymbol{e} = \boldsymbol{e}'$ the vanishing point of the motion for all image points.

2.3.3.3 Ortho-parallel Cameras

We can simplify things even further by assuming that the baseline vector $\boldsymbol{t}$ lies within the image plane. Without loss of generality we can choose $\boldsymbol{t}$ to lie on the X-axis, i.e. $\boldsymbol{t} = (-b, 0, 0)^{\top}$ for a camera centre $\boldsymbol{C}$ that moves to $(b, 0, 0)^{\top}$. According to Eq. (2.30), the epipole now takes on the form

$$\boldsymbol{e}'_{\mathrm{h}} = K\boldsymbol{t} = (1, 0, 0)^{\top} \tag{2.31}$$

and represents the direction of the x-axis in the image plane. As a result, all epipolar lines for the ortho-parallel scenario are horizontal and all corresponding points have the same y-coordinate. The fundamental matrix thus becomes

$$F = [\boldsymbol{e}'_{\mathrm{h}}]_{\times} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad . \tag{2.32}$$

Disparity of a Point. The horizontal displacement between $\boldsymbol{x}'$ and $\boldsymbol{x}$ is often referred to as the *disparity* d. In the following we derive an expression of the disparity of $\boldsymbol{x}$ in terms of the depth of its corresponding 3D point $\boldsymbol{X} = (X, Y, Z)^{\top}$. If we assume that $(x, y)^{\top}$ are the inhomogeneous coordinates of $\boldsymbol{x}$, then we can write its homogeneous coordinates as $P\boldsymbol{X}_{\mathrm{h}} = K\boldsymbol{X} = (Zx, Zy, Z)^{\top}$. The homogeneous coordinates of the corresponding image point $\boldsymbol{x}'$ are found by projecting $\boldsymbol{X} = (X, Y, Z)^{\top}$ on the right image plane

$$\begin{aligned} P'\boldsymbol{X}_{\mathrm{h}} &= K(I^3, \boldsymbol{t})\boldsymbol{X}_{\mathrm{h}} && (2.33) \\ &= K\boldsymbol{X} + K\boldsymbol{t} && (2.34) \\ &= (Zx - f\,b, Zy, Z)^{\top} \ , && (2.35) \end{aligned}$$

with $Kt = (-f\,b, 0, 0)^{\top}$ for a calibration matrix of the form (2.6). The disparity is now the change in x-coordinate between the points $(x - f\,b/Z, y)^{\top}$ and $(x, y)^{\top}$

$$d = -\frac{f\,b}{Z} \ , \tag{2.36}$$

where f is the focal length, b the signed baseline distance between the camera centres and Z the depth of $\boldsymbol{X}$. When certain internal and external camera properties are given, it is thus possible to recover the scene depth by estimating the displacement of the points in the image. We further see that the disparity is inversely proportional to the scene depth. This means that the image of an object close to the camera will undergo a larger apparent motion than that of an object far away from the camera. A disparity of zero assumes that the object is infinitely far away from the cameras.

1. for a skew-symmetric matrix $A \in \mathbb{R}^{n \times n}$, it holds that $\boldsymbol{a}^{\top} A\, \boldsymbol{a} = 0$ for all $\boldsymbol{a} \in \mathbb{R}^n$

Reconstruction from Ortho-parallel Cameras. For the special case of ortho-parallel cameras there exists a simple relation between the coordinates of a 3D space point $\boldsymbol{X} = (X, Y, Z)^\top$, its projection $(x, y)^\top$ in the left image and the corresponding disparity d. This relation can be used to recover $\boldsymbol{X}$ and is given by

$$\begin{pmatrix} x \\ y \\ d \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} X \\ Y \\ -b \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ 0 \end{pmatrix} \quad , \tag{2.37}$$

where $(x_0, y_0)^\top$ are the coordinates of the principal point. This equality results directly from Eq. (2.36) and from the projection of $\boldsymbol{X}$ by the camera matrix $P = K(I^3, \mathbf{0}^3)$.

Application: Image Rectification. If the fundamental matrix F is known for an arbitrary stereo pair, both images can be transformed to make the epipolar geometry simpler. More precisely we can compute two 3×3 projective mappings H and H' such that

$$F = {H'}^\top [(1, 0, 0)^\top]_\times H \quad , \tag{2.38}$$

where $[(1, 0, 0)^\top]_\times$ is the fundamental matrix (2.32) for the ortho-parallel camera set-up. The transformed points $H\boldsymbol{x}_h$ and $H'\boldsymbol{x}'_h$ will thus satisfy the ortho-parallel stereo geometry and will have the same y-coordinate. This process is known as *image rectification* and most proposed algorithms [HG93, Fau93, Har99, LZ99, FLP01] compute the rectifying transformations H and H' by mapping the epipoles of both images to infinity. An example of image rectification obtained with the method outlined in [HZ00] is given in Fig. 2.4. It can be observed that the epipolar lines of the rectified images coincide and that corresponding points lie on the same horizontal line. If one of the epipoles lies inside the image, care has to be taken that the rectified images do not become infinitely large and appropriate algorithms have to be used [PKV99].

2.4 Two Calibrated Views

We have previously seen that a 3D point $\boldsymbol{X}$ can be reconstructed from its image correspondences $\boldsymbol{x}$ and $\boldsymbol{x}'$ by solving the system of equations (2.15), a process that is known as *triangulation*. To be able to determine $\boldsymbol{X}$, however, the camera projection matrices need to be determined. In this section we discuss how a pair of camera matrices can be obtained if the fundamental matrix has been determined from a set of corresponding image points between two stereo views. We will show that for uncalibrated cameras a pair of camera matrices can be extracted from the fundamental matrix up to a projective transformation of space. This projective ambiguity is the best that we can achieve for a reconstruction from image correspondences alone. It can only be resolved by taking into account additional information about the scene or the cameras. One such type of information that is often readily available for a camera, are the intrinsic parameters. We will see that for internally calibrated cameras the fundamental matrix takes on a more specialised form, the essential matrix, from which the camera matrices can be retrieved up to a scale factor.

Fig. 2.4: An example of image rectification. **Top Row: (a) + (b)** The two original images with the corresponding epipolar lines superimposed as white lines. The epipolar geometry has been estimated with the method proposed in Chapter 4. **Bottom Row: (c) + (d)** Rectified images obtained by mapping the epipoles to infinity using the method of [HZ00].

2.4.1 The Projective Ambiguity

Assume that the fundamental matrix F has been determined from a set of image correspondences, then we can extract a pair of *canonical camera matrices* as [LV96, HZ00, FLP01]

$$P = (I^3, \mathbf{0}^3) \quad \text{and} \quad P' = ([\boldsymbol{e}'_{\mathrm{h}}]_\times F, \boldsymbol{e}'_{\mathrm{h}}) \ . \tag{2.39}$$

The projective invariance of the fundamental matrix has the important implication that this is not the only pair of camera matrices that corresponds to F. According to result (2.23), any two camera pairs (P_1, P'_1) and (P_2, P'_2) corresponding to the same fundamental matrix F, differ by a projective transformation H and their reconstructions are related by $\boldsymbol{X}_{2h} = H^{-1}\boldsymbol{X}_{1h}$. Because a projective transformation of the cameras and the structure does not change the projected points, a reconstruction from image correspondences alone is at best up to a projective ambiguity of three-dimensional space [Fau92, HGC92]. That it is in

general impossible to recover all intrinsic and extrinsic parameters from the fundamental matrix alone is also obvious from equality (2.28): Even if we take both calibration matrices to be equal and assume that $\boldsymbol{t}$ is only known up to a scale factor, we still have a total number of 10 camera parameters, whereas F has only 7 degrees of freedom.

2.4.2 The Essential Matrix

The *essential matrix* is the specialisation of the fundamental matrix for known internal parameters of the cameras and was introduced by Longuet-Higgins in [Lon81]. It describes a mapping between points and epipolar lines in coordinates that have been *normalised* by the camera calibration matrices. If K and K' are the calibration matrices of P and P', then the normalised image points are obtained by $\hat{\boldsymbol{x}}_{\mathrm{h}} = K^{-1}\boldsymbol{x}_{\mathrm{h}}$ and $\hat{\boldsymbol{x}}'_{\mathrm{h}} = K'^{-1}\boldsymbol{x}'_{\mathrm{h}}$. Using formulation (2.28) of the fundamental matrix as a function of the camera parameters we can write the epipolar constraint in terms of the normalised image coordinates

$$\boldsymbol{x}'^{\top}_{\mathrm{h}} K'^{-\top}[\boldsymbol{t}]_{\times} R\, K^{-1}\boldsymbol{x}_{\mathrm{h}} = 0 \quad \Longleftrightarrow \tag{2.40}$$

$$\hat{\boldsymbol{x}}'^{\top}_{\mathrm{h}}[\boldsymbol{t}]_{\times} R\, \hat{\boldsymbol{x}}_{\mathrm{h}} = 0 \quad \Longleftrightarrow \tag{2.41}$$

$$\hat{\boldsymbol{x}}'^{\top}_{\mathrm{h}} E\, \hat{\boldsymbol{x}}_{\mathrm{h}} = 0 \ . \tag{2.42}$$

Definition 2.3. *The essential matrix of a pair of cameras* $P = K(I^3, \mathbf{0}^3)$ *and* $P' = K'(R, \boldsymbol{t})$ *is the* 3×3 *matrix*

$$E = [\boldsymbol{t}]_{\times} R \ . \tag{2.43}$$

Just as the fundamental matrix $F = [\boldsymbol{e}'_{\mathrm{h}}]_{\times} P' P^{+}$, the essential matrix E can be factored as the product of a skew-symmetric matrix and a second non-singular matrix. Whereas, the second matrix is not uniquely defined in the case of the fundamental matrix, it is given by the rotation matrix between the cameras for the essential matrix. Moreover, the skew-symmetric part of the essential matrix is related to the direction of the translation vector $\boldsymbol{t}$, making E only dependent on the camera pose. The essential matrix is therefore a Euclidean description of the camera geometry, as opposed to the fundamental matrix, which is a projective description.

2.4.2.1 Properties of the Essential Matrix

In the following we summarise the most important properties of the essential matrix E:

- The essential matrix is a homogeneous quantity with five degrees of freedom: three from $\boldsymbol{t}$ and three from R, minus the scale ambiguity.
- The relation between the fundamental matrix and the essential matrix is given by

$$E = K'^{\top} F K \ . \tag{2.44}$$

  From this relation it is clear that E inherits the rank two constraint from F.

- A camera pair (P, P') that has been extracted from the essential matrix is correct up to a displacement and a scaling of three-dimensional space. Let such a *similarity transformation* be given by the 4×4 matrix

$$H = \lambda \begin{pmatrix} R^3 & \boldsymbol{t}^3 \\ \mathbf{0}^{3\top} & 1 \end{pmatrix} \ , \tag{2.45}$$

  where R^3 and t^3 are a 3D rotation and translation and λ an overall scaling. Then it is easily verified that (P, P') and $(PH, P'H)$ have the same calibration matrices. An ambiguity up to a similarity transformation thus exists for calibrated cameras.

- An important property of the essential matrix is that it has two equal singular values while the third one is zero. This means that the SVD of E can be written as

$$E = U \operatorname{diag}(\sigma, \sigma, 0)\, V^\top \ . \tag{2.46}$$

  The zero singular value simply follows from the rank two constraint. To verify further that an SVD with two equal non-zero singular values indeed represents an essential matrix we consider the following factorisation of the diagonal matrix

$$\operatorname{diag}(\sigma, \sigma, 0) = \sigma\, [\boldsymbol{t}_0]_\times R_0 \ , \tag{2.47}$$

  where we have defined the generic skew-symmetric matrix $[\boldsymbol{t}_0]_\times$ and the rotation matrix R_0 as

$$[\boldsymbol{t}_0]_\times = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad R_0 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \ . \tag{2.48}$$

  By substituting this factorisation into the SVD (2.46) and by taking into account that U and V are orthogonal matrices, we can now write E as follows

$$E = \sigma\ \underbrace{\left(U\, [\boldsymbol{t}_0]_\times U^\top\right)}_{[\boldsymbol{t}]_\times}\ \underbrace{\left(U R_0\, V^\top\right)}_{R} \ . \tag{2.49}$$

  Since $U R_0\, V^\top$ is still a rotation matrix and since it can be verified that $U\, [\boldsymbol{t}_0]_\times U^\top$ is still a skew-symmetric matrix [GV89], the matrix E can be decomposed as the product of a skew-symmetric matrix $[\boldsymbol{t}]_\times$ and a rotation matrix R. Per definition, E is thus an essential matrix. At first glance this decomposition has 6 degrees of freedom, 3 from U and V each. Due to the fact that the two non-zero singular values are equal, however, the SVD is not unique and one degree of freedom is lost.

2.4.3 Reconstruction up to Scale

The essential matrix can be estimated from image correspondences for a pair of camera matrices with known internal parameters. This can be achieved either directly from the normalised epipolar constraint (2.42) or via the fundamental matrix with the help of relation (2.44). As opposed to the fundamental matrix, the camera pose can be extracted from the factorisation of the essential matrix as the product of a skew-symmetric matrix and a rotation matrix. We will see that this leads to four possible choices for the two camera matrices, but that only one choice corresponds to a physical camera pair. Because it is not possible to determine the baseline distance from the essential matrix, the computed 3D reconstruction will differ from the true reconstruction by an *overall scale factor*.

2.4.3.1 Recovery of the Camera Translation and Rotation

If we assume that the left camera is given by $P = K(I^3, \mathbf{0}^3)$ and that the SVD of the essential matrix is given by $E = U\mathrm{diag}\,(\sigma, \sigma, 0)\, V^\top$, we can recover the translational and rotational parts of $P' = K'(R, \boldsymbol{t})$ from the factorisation (2.49). We start by noting that $\boldsymbol{t}$ is the right null-space of the matrix $[\boldsymbol{t}_\times]$ and that therefore $U\,[\boldsymbol{t}_0]_\times U^\top \boldsymbol{t} = 0$ must hold. It can be verified that this requirement is satisfied for the choice

$$\boldsymbol{t} = U\,(0, 0, 1)^\top = \boldsymbol{u}_3 \ , \tag{2.50}$$

where $\boldsymbol{u}_3$ is the last column of U. By setting $\boldsymbol{t}$ to the last left singular vector of E we have implicitly fixed the size of the baseline to $|\boldsymbol{t}| = 1$. This is a valid choice since the true magnitude of the translation can not be recovered due to the scale ambiguity of the essential matrix. This ambiguity also applies to the sign of $\boldsymbol{t}$ and therefore

$$\boldsymbol{t} = \boldsymbol{u}_3 \quad \text{or} \quad \boldsymbol{t} = -\boldsymbol{u}_3 \ . \tag{2.51}$$

Before moving on to the computation of R, we point out that the same factorisation as (2.49) also holds for the transposed of the matrices $[\boldsymbol{t}_0]_\times$ and R_0. This is clear from inspection and it has been shown [Lon81, Har92, HZ00] that both factorisations of E in terms of a skew-symmetric matrix and a rotation matrix are unique. As a consequence, there are two possible choices for R

$$R = UR_0\,V^\top \quad \text{or} \quad R = UR_0^\top\,V^\top \ . \tag{2.52}$$

In total there are thus *four* rotation-translation pairs $(R, \boldsymbol{t})$ that can be associated with the same essential matrix, a result that was already mentioned by Longuet-Higgins [Lon81]. Hartley [Har92] further developed this and similar methods for retrieving the camera translation and rotation have been reported by Faugeras [Fau92, Fau93].

2.4.3.2 Selection of the Camera Matrices

From the previous results we obtain four possibilities for the camera matrix P':

$$P' = K'(UR_0V^\top, \quad \mathbf{u}_3) \quad \text{or} \tag{2.53}$$

$$P' = K'(UR_0V^\top, -\mathbf{u}_3) \quad \text{or} \tag{2.54}$$

$$P' = K'(UR_0^\top V^\top, \quad \mathbf{u}_3) \quad \text{or} \tag{2.55}$$

$$P' = K'(UR_0^\top V^\top, -\mathbf{u}_3) \ . \tag{2.56}$$

The first two camera matrices only differ by the direction of the baseline. It can be shown that the third camera differs from the first one by a rotation of 180° about the baseline. The same is true for the fourth and the second camera. All four camera pairs (P, P') are illustrated in Fig. 2.5, where the rows show the pairs with reversed baseline and the columns the pairs with reversed orientation. However, for only one of these configurations a reconstructed point $\mathbf{X}$ will lie in front of both cameras. It is therefore sufficient to determine the reconstruction of a single image correspondence and select the camera pair for which its depth (2.13) with respect to both cameras is positive. This would correspond to the camera pair (a) in the figure. It has to be noted that although the camera rotation has been retrieved unambiguously from the essential matrix, the baseline has been normalised for convenience. Moving the two camera centres along the baseline does indeed not change the essential matrix but only leads to a scaling of the reconstructed scene.

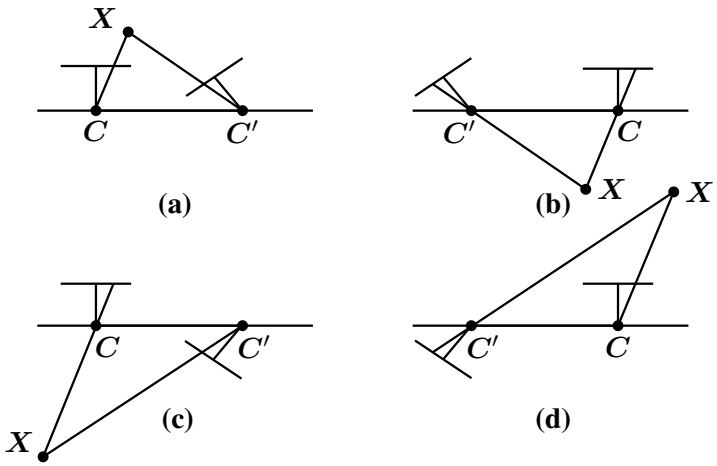

Fig. 2.5: The four possible camera pairs from the essential matrix (adapted from [HZ00]). The camera pairs **(a) + (b)** and **(c) + (d)** correspond to a reversed baseline, while the pairs **(a) + (c)** and **(b) + (d)** correspond to a reversed orientation.

Remark on the SVD of E. If we determine E from an *estimate* of F via Eq. (2.44), the SVD of E will in practice be given by $E = UDV^\top$, with $D = \mathrm{diag}(\sigma_1, \sigma_2, 0)$ and $\sigma_1 \geq \sigma_2 \geq 0$. The closest true essential matrix in Frobenius norm E' is then obtained by averaging σ_1 and σ_2 as $E' = U\ \mathrm{diag}((\sigma_1+\sigma_2)/2, (\sigma_1+\sigma_2)/2, 0)V^\top$ [HZ00]. To compute the two camera matrices from the essential matrix, an expression for E' is, however, not needed because only the matrices U and $V^\top$ are used in the computation of P and P'. Since the singular vectors of E and E' are the same, the camera matrices can be retrieved from the SVD of E without explicitly computing E'. It is thus possible that E is not an essential matrix, but this is not a requirement for the reconstruction method described here.

2.4.3.3 Reconstruction by Triangulation

For calibrated ortho-parallel cameras, the reconstruction of a 3D point from its two projections can be found via the explicit relation (2.37). For general (converging) cameras with known camera matrices, on the other hand, we have no choice but to solve the system (2.15). This system is an equality between homogeneous vectors and by keeping in mind that $\boldsymbol{x}_\mathrm{h}$ and $P\boldsymbol{X}_\mathrm{h}$ have to be parallel, the unknown scale factor can be eliminated by rewriting the equations with the help of the cross product [HZ00]

$$\begin{cases} \boldsymbol{x}_\mathrm{h} \times P\boldsymbol{X}_\mathrm{h} &= \mathbf{0}^3 \\ \boldsymbol{x}'_\mathrm{h} \times P'\boldsymbol{X}_\mathrm{h} &= \mathbf{0}^3 \end{cases} . \tag{2.57}$$

It now turns out that only four of the six equations of this new system are linearly independent, but this still suffices to determine the four coordinates of $\boldsymbol{X}_\mathrm{h}$ uniquely. In practice, there will be errors in the measured point correspondences and in the estimated camera matrices, such that a solution in a least squares sense is preferred.

Besides this simple linear algorithm, there are two triangulation methods that find the unknown point $\boldsymbol{X}$ in a geometrically interpretable way. As can be seen in Fig. 2.6 (a), triangulation comes down to finding the intersection of the optical rays through $\boldsymbol{x}$ and

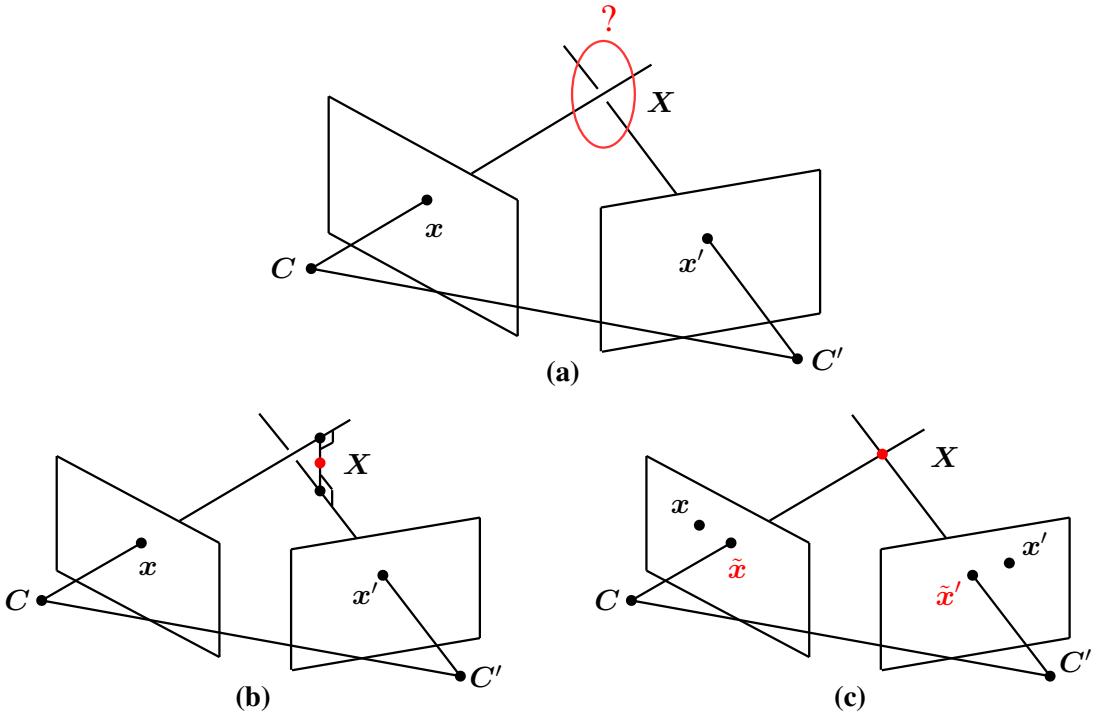

Fig. 2.6: **(a)** The geometrical interpretation of the triangulation process. **(b)** Finding the closest point to the optical rays. **(c)** Minimising the reprojection error.

$\boldsymbol{x}'$ (thereby closing the *triangle* $\boldsymbol{CXC}'$). Due to the earlier mentioned estimation errors, these two rays generally do not intersect in space and an approximate solution for $\boldsymbol{X}$ is needed. A first possible solution is depicted in Fig. 2.6 (b) and is given by the midpoint of the line segment that intersects both optical rays orthogonally. This midpoint corresponds to the 3D point that lies closest to both optical rays and it can be found in a linear way [TV98]. A second solution exist in finding an image point pair $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}')$ that lies close to the measured point pair $(\boldsymbol{x}, \boldsymbol{x}')$ and satisfies the epipolar constraint exactly. The optical rays through $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{x}}'$ will intersect in a unique point $\boldsymbol{X}$ that can be found by solving the equation system (2.57) in $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{x}}'$. This idea is shown in Fig. 2.6 (c). The latter method minimises the so-called *reprojection error* between $(\boldsymbol{x}, \boldsymbol{x}')$ and $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}')$ [HZ00] and is closely related to the non-linear technique described in Sec. 4.2.3.2.

2.5 Summary

In this chapter we have given an overview of the most important concepts of single view geometry and uncalibrated and calibrated two view stereo geometry. We have introduced three important projective entities that take on the form of three homogeneous matrices and have explored their specific properties. We presented the triangulation equations that allow to reconstruct a 3D point from its 2D projections and a pair of camera matrices and concluded the chapter with one of the most significant results of calibrated stereo, namely the recovery of the camera matrices up to scale via the essential matrix.

3 Variational Optical Flow

This chapter forms the second pillar of this thesis. It is devoted to the modelling and numerical solution of variational optical flow methods. It will pave the road for the development of novel energy-based correspondence methods and is crucial for the understanding of the variational techniques presented in Chapters 5 and 6.

The chapter naturally break down into two parts. The first part, Sec. 3.1, addresses modelling aspects and covers different choices of *data* and *smoothness terms*. In case of the data term, the focus lies on the constancy of appropriate image features and its robustness under noise and illumination changes. For the smoothness term, the preservation of semantically important discontinuities in the flow field is of interest. Sec. 3.2 comprises the second part of this chapter and deals with minimisation and discretisation aspects.

An important property of the correspondence methods examined in this thesis, is their ability to cope with the large displacements that can arise from changes in viewpoint. In this chapter, most important concepts of variational optical flow will be introduced by example of a model prototype which will serve as a baseline method for the variational techniques introduced later in this thesis. Contrary to small displacement approaches, such as the classical optical flow method of Horn and Schunck [HS81], constancy assumptions will appear in our prototypical functional in their original implicit form, and not as linearised expressions. This will have its effect on the choice of the minimisation strategy, as the resulting energy is generally not convex. An import aspect of the energy optimisation therefore lies in breaking down the initial non-convex problem into convex sub-problems, such that a minimiser can be found by means of established global solvers. Here, a coarse-to-fine warping strategy plays a central role in avoiding meaningless local minima, while an incremental computation assures convexity of the energy.

3.1 Modelling of Large Displacement Optical Flow

To formulate an energy for the variational computation of optical flow between two stereo images, we start from a scalar-valued image sequence $g(x, y, t) : \Omega \times [0, \infty) \to \mathbb{R}$, defined over a rectangular image domain $\Omega \subset \mathbb{R}^2$. In general, g has been obtained by convolving the original image sequence g_0 with a Gaussian kernel K_σ of standard deviation σ

$$g = K_\sigma * g_0 \ . \tag{3.1}$$

This *presmoothing* step removes noisy high frequency components in the input data and guarantees that $g \in \mathcal{C}^\infty$, which is a prerequisite for the well-posedness of any optical flow method. Furthermore, the Taylor approximation of the constancy assumptions, that is performed either early on in the model or later during minimisation, only holds for large displacements if g has a certain degree of smoothness.

Without loss of generality we now assume that the left and right image of an uncalibrated stereo pair are embedded as two consecutive frames in the image sequence:

$$g_{\mathrm{l}}(x, y) = g(x, y, t) \quad \text{and} \quad g_{\mathrm{r}}(x, y) = g(x, y, t + 1) \ . \tag{3.2}$$

By defining the left and right image this way, we have removed the explicit dependency on the time variable t. This makes sense in a stereo setting, because the optical flow field is not regarded anymore as the instantaneous velocity of an image point, but as the displacement between the projections of a static scene point onto the two stereo images.

The optical flow will be denoted by the two-component vector $\boldsymbol{w} = (u, v)^\top$ and is found by minimising an energy functional of the general form

$$\mathcal{E}(\boldsymbol{w}) = \int_\Omega \left(\mathcal{E}_{\mathrm{D}}(\boldsymbol{w}) + \alpha\, \mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w})\right) \mathrm{d}\boldsymbol{x} \ , \tag{3.3}$$

where $\boldsymbol{x} = (x, y)^\top$ is a point in the image domain Ω and $\nabla \boldsymbol{w} := (\nabla u, \nabla v)^\top$. Energy (3.3) consists of two terms: the *data term* $\mathcal{E}_{\mathrm{D}}(\boldsymbol{w})$, which relates the image data to the unknown optical flow, and the *smoothness term* $\mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w})$, which regularises the possibly non-unique solution of the data term. While the data term formalises the assumption that characteristic image features stay constant along the motion trajectory, thereby matching corresponding objects in both stereo frames, the smoothness term imposes the assumption that neighbouring points are likely to belong to the same object and thus undergo a similar motion. The *smoothness weight* $\alpha > 0$ serves as a regularisation parameter, the larger its value, the stronger the penalisation of flow gradients and the smoother the flow field.

3.1.1 The Data Term

To retrieve corresponding structures in two images, we must impose constancy assumptions on suitable image features. These have to be chosen in accordance with the illumination conditions during acquisition and the prevailing type of motion. In the following, we will design a data term that is based on a combination of two such image features and which has proven to be successful in the context of motion estimation: the image brightness and the image gradient. Additionally, we require our method to be able to handle large displacements. Therefore we refrain from linearising the constancy assumptions, as opposed to what is commonly done for small-displacement optical flow methods.

3.1.1.1 Constancy of the Image Brightness

The oldest and most frequently used constancy assumption in literature is the assumption that the grey value of a pixel does not change along its path of motion. This assumption has been used in the classical approaches of Horn and Schunck [HS81] and Nagel and Enkelmann [NE86] and is only valid if the lighting conditions for the left and the right frame are the same. The constancy of image brightness can be formulated as

$$g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) - g_{\mathrm{l}}(\boldsymbol{x}) = 0 \ . \tag{3.4}$$

To impose this constraint in an energy minimisation framework, we penalise deviations of the grey value difference from zero by means of the quadratic data term [HS81, LK81]

$$\mathcal{E}_{\mathrm{D}}(\boldsymbol{w}) = |g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) - g_{\mathrm{l}}(\boldsymbol{x})|^2 \ . \tag{3.5}$$

3.1.1.2 Constancy of the Image Gradient

The assumption that the image brightness is invariant under the occurring motion is often violated. Particularly in real-world situations, constancy should be imposed on image features that are less sensitive to illumination changes. An efficient strategy in this context is the consideration of derivatives. To account for additive illumination changes in the scene, we can for instance assume that the spatial brightness gradient does not change under its displacement [TP84, Sch93, PBB$^+$06]. Since the spatial gradient is a two-component vector, we obtain the following two constraints

$$\partial_x g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}) - \partial_x g_{\mathrm{l}}(\boldsymbol{x}) = 0, \tag{3.6}$$
$$\partial_y g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}) - \partial_y g_{\mathrm{l}}(\boldsymbol{x}) = 0 \ . \tag{3.7}$$

Because specific knowledge on the illumination conditions is often not available or incomplete, several constancy assumptions can be combined by means of a linear combination. A combination of the gradient and the grey value constancy assumption is obtained by squaring the constraints (3.6) and (3.7) and adding them to the data term (3.5). This gives rise to the following data term [BBPW04]

$$\mathcal{E}_{\mathrm{D}}(\boldsymbol{w}) = |g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}) - g_{\mathrm{l}}(\boldsymbol{x})|^2 + \gamma\, |\nabla g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}) - \nabla g_{\mathrm{l}}(\boldsymbol{x})|^2 \ , \tag{3.8}$$

where γ a positive weight that steers the influence of the gradient constancy assumption. Aside from an improved performance under varying illumination, the combination of several image features has the advantage of including more information than a data term that is based on a single constraint. We will see in Sec. 3.2.1.1 that equation (3.4) alone is not sufficient to determine the two unknowns u and v uniquely, but that it is possible to obtain a local solution of the data term by combining several independent data constraints.

3.1.1.3 Constancy of Other Image Features

To cope with additive illumination changes, higher order derivatives – such as the spatial Hessian [PBB$^+$06] – can be considered as well. Both the gradient and the Hessian contain directional information and any constancy of these image features implies a constancy on their orientation. As a consequence, these constraints are useful for the estimation of translational or slow rotational motion, but might lead to slightly poorer results if fast rotations are dominating [Bru06]. As a remedy, rotationally invariant features have been proposed like the gradient magnitude and the Laplacian [RD96, PBB$^+$06]. Despite the theoretical shortcomings of the gradient constancy assumption, it has nevertheless proven very successful in combination with the brightness constancy constraint. In this work we will therefore opt for this combination and do not consider additional data constraints.

3.1.1.4 Non-quadratic Penalisation

The data term that we have considered so far penalises deviations from constancy assumptions quadratically. As a consequence, outliers – such as those arising from noise and from violations of model assumptions in occluded areas – will have a large influence on the estimation process. To provide the data term with robustness against these outliers it is desirable to penalise them less severely [Hub81, HRRS86, RL87]. This can be achieved

by using a non-quadratic penalisation function $\Psi(s^2)$, where s denotes the data constraint [BA91, MP98a, BBPW04]. In the light of the Taylor linearisation of the data constraints that will be carried out during the minimisation in Sec. 3.2.1.1, $\Psi(s^2)$ is generally chosen to be convex in s. This is necessary to guarantee the convexity and well-posedness of the resulting energy and to allow the use of globally convergent numerical solvers. An example of a function that penalises its argument sub-quadratically is the regularised L_1-norm

$$\Psi(s^2) = \sqrt{s^2 + \epsilon^2} \ , \tag{3.9}$$

where ϵ is a small positive constant. Replacing the quadratic penaliser in $\mathcal{E}_{\mathrm{D}}(\boldsymbol{w})$ by the proposed L_1-function, results in the data term

$$\mathcal{E}_{\mathrm{D}}(\boldsymbol{w}) = \Psi \left(|g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) - g_{\mathrm{l}}(\boldsymbol{x})|^2 + \gamma \, |\nabla g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) - \nabla g_{\mathrm{l}}(\boldsymbol{x})|^2 \right) \ . \tag{3.10}$$

3.1.1.5 Colour Information

In most cases, the use of colour images leads to slightly better results than for grey value versions of the same images, because extra information is available for the matching process. To integrate colour information in our methods, we consider a multi-channel variant of the data term in which the three colour channels are coupled as follows:

$$\mathcal{E}_{\mathrm{D}}(\boldsymbol{w}) = \Psi \left(\sum_{i=1}^{3} |g_{\mathrm{r}\,i}(\boldsymbol{x}+\boldsymbol{w}) - g_{\mathrm{l}\,i}(\boldsymbol{x})|^2 + \gamma \sum_{i=1}^{3} |\nabla g_{\mathrm{r}\,i}(\boldsymbol{x} + \boldsymbol{w}) - \nabla g_{\mathrm{l}\,i}(\boldsymbol{x})|^2 \right) \ , \tag{3.11}$$

where g_1, g_2 and g_3 represent the three RGB colour channels of an image g. Please note that we seek a common displacement field for all three channels.

3.1.2 The Smoothness Term

We have seen that the data term might be insufficient to determine a locally unique solution for the optical flow. To overcome this problem, variational methods make use of an additional term that regularises the solution in these regions: the smoothness term or regulariser. While the data term includes prior knowledge of the scene, the smoothness term makes assumptions on the solution itself. The simplest optical flow regulariser, for instance, – the quadratic regulariser of Horn and Schunck [HS81]

$$\mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w}) = |\nabla u|^2 + |\nabla v|^2 \tag{3.12}$$

– penalises deviations from smoothness of the flow field. As shown by Weickert and Schnörr [WS01a], the design of the smoothness term is closely related to the modelling of diffusion processes for vector-valued images. In fact, the smoothness term will give rise to a diffusion process that is responsible for the propagation of neighbourhood information to locations where no unique solution for the data term exists. This *filling-in* effect is also the reason why variational optical flow methods give dense flow fields as opposed to many local methods that produce incomplete or sparse displacement fields. Because we consider a two-image stereo setup in this thesis, we will only focus on spatial regularisers and not on spatial-temporal regularisers that assume smoothness over multiple frames.

3.1.2.1 The Euler-Lagrange Equations

According to the calculus of variations [Els61, GF00], a minimiser $(u, v)^\top$ of the energy (3.3) necessarily has to fulfil the *Euler-Lagrange equations*. They are given by the system of partial differential equations (PDEs)

$$0 = \partial_u \mathcal{E}_{\mathrm{D}} - \alpha \left(\partial_x \left(\partial_{u_x} \mathcal{E}_{\mathrm{S}} \right) + \partial_y \left(\partial_{u_y} \mathcal{E}_{\mathrm{S}} \right) \right), \tag{3.13}$$

$$0 = \underbrace{\partial_v \mathcal{E}_{\mathrm{D}}}_{\substack{\text{data} \\ \text{(reaction)}}} - \alpha \underbrace{\left(\partial_x \left(\partial_{v_x} \mathcal{E}_{\mathrm{S}} \right) + \partial_y \left(\partial_{v_y} \mathcal{E}_{\mathrm{S}} \right) \right)}_{\substack{\text{smoothness} \\ \text{(diffusion)}}} , \tag{3.14}$$

with homogeneous Neumann (reflecting) boundary conditions. An alternative way of looking at these equations is by interpreting them as the steady state of a diffusion-reaction system of a vector-valued image $(u, v)^\top$, where the diffusion part results from the smoothness term $\mathcal{E}_{\mathrm{S}}$ and the reaction part from the data term $\mathcal{E}_{\mathrm{D}}$. In [WS01a] this relation is used to classify the different regularisers depending on the type of diffusion process that they induce. In the following we will take a closer look at this classification.

3.1.2.2 A Classification of Spatial Regularisers for Optical Flow

The simple Horn and Schunck regulariser can be related to *homogeneous* diffusion and comes down to a Gaussian blurring of the optical flow components. As a consequence, the motion field estimated with this regulariser will be smoothed equally in all directions, resulting in less sharp and dislocated flow edges. For good results, it is of course necessary that the semantically important edges are preserved during estimation and that smoothing across flow discontinuities is inhibited as much as possible. Table 3.1 gives an overview of discontinuity preserving regularisers and their corresponding diffusion processes based on [WS01a] and [ZBW11]. In the following we will describe them in more detail.

Flow-driven Isotropic Regularisers. *Flow-driven* regularisers take into account the discontinuities of the unknown displacement field and introduce a feedback of the evolving flow into the smoothing process. One way of achieving this effect is by replacing the quadratic penaliser of the homogeneous smoothness term (3.12) by a non-quadratic function $\Psi(s^2)$ [SH89, Sch94]. Thus, we obtain the *isotropic flow-driven* regulariser

$$\mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w}) = \Psi \left(|\nabla u|^2 + |\nabla v|^2 \right) \ . \tag{3.15}$$

As can be seen in Table 3.1, this regulariser is associated to non-linear diffusion with a scalar-valued *diffusivity function* $\Psi'(s^2)$. The diffusivity is positive and generally decreasing in s in order to reduce the amount of smoothing at high gradient locations. The corresponding diffusion part of the Euler-Lagrange equation for the u-component is

$$\operatorname{div} \Big(\Psi' \left(|\nabla u|^2 + |\nabla v|^2 \right) \nabla u \Big) \ . \tag{3.16}$$

Flow-driven Anisotropic Regularisers. The scalar valued diffusivity $\Psi'(s^2)$ does not contain any directional information and does not distinguish between smoothing across or along flow edges. While diffusion across discontinuities is often not desired, smoothing along them can be advantageous, especially with respect to noisy boundaries. Such *anisotropic* behaviour can be obtained by means of a *diffusion tensor* D which steers the direction and intensity of the local diffusion process. Following the ideas of Weickert and Schnörr [WS01a], a suitable diffusion tensor is for instance given by

$$D(\nabla u, \nabla v) = \Psi'(\nabla u \nabla u^\top + \nabla v \nabla v^\top) := \sum_{i=1}^{2} \Psi'(\mu_i)\, \boldsymbol{v}_i\, \boldsymbol{v}_i^\top \ , \tag{3.17}$$

where μ_1 and μ_2 are the eigenvalues and $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ the eigenvectors of the 2×2 tensor $\nabla u \nabla u^\top + \nabla v \nabla v^\top$. While $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ represent the orthogonal directions of highest and lowest contrast in the flow field, μ_1 and μ_2 represent the corresponding contrast values. According to definition (3.17), the tensor-valued diffusivity function $\Psi'(\nabla u \nabla u^\top + \nabla v \nabla v^\top)$ applies the diffusivity $\Psi'(s^2)$ to the eigenvalues of its argument, while keeping its eigenvectors fixed. This way, a diffusion tensor is obtained that mainly inhibits smoothing across dominant flow features i.e. in the direction of highest contrast, but much less along edges. The resulting diffusion process can be obtained by replacing the scalar-valued diffusivity in the divergence expression (3.16) by the diffusion tensor D:

$$\operatorname{div}\left(D\ \nabla u\right) = \operatorname{div}\Big(\Psi'\left(\nabla u \nabla u^\top + \nabla v \nabla v^\top\right) \nabla u\Big) \ . \tag{3.18}$$

The corresponding flow-driven anisotropic regulariser $\mathcal{E}_{\mathrm{S}}$ is listed in Tab. 3.1.

Remark on the Diffusion Tensor Notation. By writing divergence expression (3.16) as

$$\operatorname{div}\Big(\Psi'\left(|\nabla u|^2 + |\nabla v|^2\right) I^2\ \nabla u\Big) \ , \tag{3.19}$$

with I^2 the 2×2 identity matrix, we can also write isotropic diffusion in a common divergence form $\operatorname{div}\left(D\ \nabla u\right)$. As a result, both isotropic and anisotropic regularisers can be classified by the diffusion tensor of their corresponding vector-valued diffusion process.

Image-driven Isotropic and Anisotropic Regularisers. *Image-driven* regularisers work along the same line as flow-driven regularisers, but adapt their smoothing behaviour to edges in the image data instead of edges in the evolving flow field. Isotropic [AELS99], as well as anisotropic [Nag83] techniques have been proposed, and both give rise to linear diffusion processes with a locally varying but constant diffusivity function or diffusion tensor. A notable example of an image-driven anisotropic technique is the method of Nagel and Enkelmann [Nag83, NE86], which uses the regularised projection matrix

$$D(\nabla g_{\mathrm{I}}) = \frac{1}{|\nabla g_{\mathrm{I}}|^2 + 2\epsilon^2} \left(\nabla g_{\mathrm{I}}^{\perp} \nabla g_{\mathrm{I}}^{\perp\top} + \epsilon^2 I^2\right) \ , \tag{3.20}$$

as diffusion tensor. Here, $\nabla g_{\mathrm{I}}^{\perp}$ is a vector perpendicular to ∇g_{I} and ϵ a small regularisation constant. While image-driven techniques can give very sharp and well localised flow boundaries, they have the disadvantage of oversegmentation artifacts. This is because not all image edges coincide with flow edges, such as in the presence of texture.

Tab. 3.1: A classification of optical flow regularisers based on their corresponding vector-valued diffusion process (adapted from [WS01a], [ZBV+08] and [ZBW11]).

	optical flow regulariser $\mathcal{E}_{\mathrm{S}}$	diffusion tensor D of the vector-valued diffusion process $\operatorname{div}(D\,\nabla u)$
flow-driven isotropic [Sch94]	$\Psi\left(\lvert\nabla u\rvert^2 + \lvert\nabla v\rvert^2\right)$	$\Psi'\left(\lvert\nabla u\rvert^2 + \lvert\nabla v\rvert^2\right) I^2$
flow-driven anisotropic [WS01a]	$\operatorname{tr}\Psi\left(\nabla u\nabla u^\top + \nabla v\nabla v^\top\right)$	$\Psi'\left(\nabla u\nabla u^\top + \nabla v\nabla v^\top\right)$
image-driven isotropic [AELS99]	$\Psi'\left(\lvert\nabla g_{\mathrm{I}}\rvert^2\right)\left(\lvert\nabla u\rvert^2 + \lvert\nabla v\rvert^2\right)$	$\Psi'\left(\lvert\nabla g_{\mathrm{I}}\rvert^2\right) I^2$
image-driven anisotropic [Nag83]	$\nabla u^\top D(\nabla g_{\mathrm{I}})\,\nabla u$ $+\nabla v^\top D(\nabla g_{\mathrm{I}})\,\nabla v$	$\frac{1}{\lvert\nabla g_{\mathrm{I}}\rvert^2+2\epsilon^2}\left(\nabla g_{\mathrm{I}}^\perp \nabla g_{\mathrm{I}}^{\perp\top} + \epsilon^2 I^2\right)$
flow- and image-driven anisotropic [ZBW+09]	$\Psi\left((\boldsymbol{r}_1^\top\nabla u)^2 + (\boldsymbol{r}_1^\top\nabla v)^2\right)$ $+(\boldsymbol{r}_2^\top\nabla u)^2 + (\boldsymbol{r}_2^\top\nabla v)^2$	$\Psi'\left((\boldsymbol{r}_1^\top\nabla u)^2 + (\boldsymbol{r}_1^\top\nabla v)^2\right)\boldsymbol{r}_1\boldsymbol{r}_1^\top + \boldsymbol{r}_2\boldsymbol{r}_2^\top$
PDE based flow-driven anisotropic [ZBV+08]	–	$\Psi'\Big(K_\rho * \left(\nabla u_\theta\nabla u_\theta^\top + \nabla v_\theta\nabla v_\theta^\top\right)\Big)$

Regularisers that Extend the Classification of Weickert and Schnörr. An anisotropic smoothness term that adapts its smoothing directions to the image structure, but steers its smoothing strength with respect to the flow contrast, was presented by Sun *et al.* [SRLB08] in a discrete setting. Such regularisation behaviour can be classified as *joint image- and flow-driven* and can produce sharp flow edges without the effect of oversegmentation. In contrast to the regulariser of Nagel and Enkelmann, which uses $\nabla g_{\mathrm{I}}^\perp$ for directional information, the smoothness term of Sun *et al.* obtains a more robust estimate of the local direction by analysing the eigenvectors of the *structure tensor* [FG87]

$$K_\rho * \left(\nabla g_{\mathrm{I}} \nabla g_{\mathrm{I}}^\top\right) \quad , \tag{3.21}$$

with an integration scale $\rho > 0$. A first remark with respect to this type of regularisation, is that the directional information from the structure tensor is not consistent with the imposed brightness and gradient constraints of the data term. It is therefore more natural to take into account the directional information provided by the data constraints and steer the regularisation process with respect to constraint edges instead of image edges. To this end, Zimmer *et al.* [ZBW+09, ZBW11] recently proposed to additionally include gradient channel information and analyse the eigenvectors $\boldsymbol{r}_i$, $i \in \{1, 2\}$ of

$$K_\rho * \Big(\big(\nabla g_{\mathrm{l}} \nabla g_{\mathrm{l}}^\top\big) + \gamma \ \big(\nabla g_{\mathrm{l}x} \nabla g_{\mathrm{l}x}^\top + \nabla g_{\mathrm{l}y} \nabla g_{\mathrm{l}y}^\top\big) \Big) \ . \tag{3.22}$$

This *regularisation tensor* can be regarded as a generalisation of the structure tensor which is *complementary* to the data term (3.8). The regulariser of Zimmer *et al.* then reads

$$\mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w}) = \Psi\left((\boldsymbol{r}_1^\top \nabla u)^2 + (\boldsymbol{r}_1^\top \nabla v)^2\right) + (\boldsymbol{r}_2^\top \nabla u)^2 + (\boldsymbol{r}_2^\top \nabla v)^2 \ , \tag{3.23}$$

where the regularisation direction is now adapted to the constraint directions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, whereas the magnitude of the regularisation depends on the flow contrast encoded in the directional derivatives $\boldsymbol{r}_i^\top \nabla u$ and $\boldsymbol{r}_i^\top \nabla v$. Since, the data term mainly constrains the flow in the direction of the largest eigenvalue of the regularisation tensor, a single robust penalisation in $\boldsymbol{r}_1$-direction is proposed. In the orthogonal $\boldsymbol{r}_2$-direction, a quadratic penalisation ensures a strong filling-in effect of missing information. The separate penalisation of both directions additionally ensures rotational invariance in the smoothness term. For a unified formulation of all previously mentioned smoothness terms using the regularisation tensor notation and for an extension to colour images we refer to [ZBW11].

Smoothing Strategies that Do not Fit the Classification of Weickert and Schnörr. For flow-driven anisotropic optical flow, the diffusion tensor (3.17) includes a coupling between the two flow components u and v of the displacement field. This way, the desired anisotropic behaviour is ensured because the eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ of the structure detector $\nabla u \nabla u^\top + \nabla v \nabla v^\top$ are in general not parallel to the gradients of u and v. Inspired by the diffusion filter of [Wei94b], one could improve the anisotropic behaviour of the tensor $\nabla u \nabla u^\top + \nabla v \nabla v^\top$ in two ways: (i) By regularising the components u and v by a Gaussian convolution of standard deviation θ, staircaising artifacts and problems with noise can be reduced [CLMC92]. (ii) By integrating neighbourhood information via a convolution of the tensor entries with a Gaussian kernel of standard deviation ρ, we can overcome undesired cancellation effects in the gradients ∇u_θ and ∇v_θ. Taking these extensions into account, the resulting diffusion tensor would take on the form

$$D(\nabla u_\theta, \nabla v_\theta) = \Psi'\Big(K_\rho * \big(\nabla u_\theta \nabla u_\theta^\top + \nabla v_\theta \nabla v_\theta^\top\big)\Big) \ . \tag{3.24}$$

Originally proposed in the setting of one dimensional stereo matching [ZBV+08], this type of diffusion process can only be formulated within the PDE framework of the Euler-Lagrange equations. Because of the Gaussian regularisation of the unknowns, a corresponding energy formulation is not known and consequently such *PDE based flow-driven anisotropic smoothing* falls outside the classification of Table 3.1. PDE based design can lead to more powerful smoothing strategies that have shown their benefit in the context of image denoising [Wei94b, Wei00] and PDE based inpainting [GWW+05].

3.1.2.3 The Diffusivity Function

An appropriate choice for $\Psi(s^2)$ is the previously introduced L_1-norm (3.9) [BBPW04]. It corresponds to the well known *total variation* (TV) diffusivity [ROF92] given by

$$\Psi'(s^2) = \frac{1}{2\sqrt{s^2 + \epsilon^2}} \ . \tag{3.25}$$

Other choices for the diffusivity can even lead to an enhancement of flow edges due to the backward diffusion that they might induce. An example of this type of function is the Perona-Malik diffusivity [PM87, PM90] with contrast parameter $c > 0$

$$\Psi'(s^2) = \frac{c^2}{s^2 + c^2} \ . \tag{3.26}$$

3.1.3 Our Model Prototype

The model that will serve as an example of variational optical flow in this chapter and that will form the basis for the design of novel correspondence methods in this thesis, is the method proposed by Brox *et al.* and Papenberg *et al.* [BBPW04, PBB$^+$06]. This model prototype integrates four successful concepts that add to its robustness and high accuracy: (i) Firstly, it combines the brightness and gradient constancy assumptions to obtain optimal performance in the presence of additive illumination changes. (ii) These constancy assumptions are expressed as an *implicit* dependency of the optical flow on the image data, which allows a correct treatment of large displacements. (iii) To render the model more robust with respect to noise and occlusions, a joint non-quadratic penalisation function based on the L_1-norm is applied to the combined data constraints. (iv) Finally, this method makes use of an isotropic flow-driven regularisation strategy based on the total variation, which allows for the preservation of meaningful discontinuities in the flow field.

We can combine these four concepts in a single variational framework by joining the data term (3.10) and the smoothness term (3.15) into one energy functional

$$\begin{aligned} \mathcal{E}(\boldsymbol{w}) = \int_\Omega \Big(& \Psi\left(|g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}) - g_{\mathrm{l}}(\boldsymbol{x})|^2 + \gamma\, |\nabla g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}) - \nabla g_{\mathrm{l}}(\boldsymbol{x})|^2 \right) \\ & + \alpha\, \Psi\left(|\nabla \boldsymbol{w}|^2 \right) \Big) \, \mathrm{d}\boldsymbol{x} \ , \end{aligned} \tag{3.27}$$

where the flow gradient $|\nabla \boldsymbol{w}|^2 := |\nabla u|^2 + |\nabla v|^2$. In both the data term and the smoothness term, Ψ is the regularised L_1 penaliser (3.9) with $\epsilon = 0.001$. It has to be noted that we use a variant of the method of Brox *et al.* with spatial regularisation, and not with spatio-temporal regularisation as proposed originally in [BBPW04]. In this thesis we will further exploit colour information if it is available. For experiments with colour sequences we will therefore replace the data term in the energy above by its multi-channel version (3.11). For modelling purposes and for describing the minimisation and discretisation of the energy, however, we will adhere to the grey-value formulation for simplicity. Extending the solution strategies to multiple image channels is in all cases straightforward.

While the method above will serve as our baseline algorithm for the development of novel correspondence methods, we will additionally integrate our ideas into a more sophisticated optical flow method for the experiments of Sec. 5.2. This method goes back to Zimmer *et al.* [ZBW11] and applies a separate robustification of the brightness and gradient constancy assumptions in the data term. In addition, it incorporates the flow- and image-driven anisotropic regularisation of (3.23) and includes an automatic estimation of the smoothness weight. The use of this recent technique will serve as an illustration that our modelling ideas are not limited to a specific variational method, but that they can be directly integrated into more advanced optical flow approaches as well.

3.2 Minimisation and Numerical Solution

The energy functional of our optical flow prototype (3.27) has one important property that has a major influence on the minimisation strategy that we will use: the desired flow field depends implicitly on the data term. This raises two problems that we have to overcome

- *Multiple Local Minima.* The proposed energy functional will in general be *non-convex*. As a result there will be multiple local minima, aside from the global minimum that we are looking for. Since the Euler-Lagrange equations only constitute a necessary condition for a minimiser, they will be satisfied by all minima and have thus *no unique solution*. As a consequence, gradient descent-based methods are no longer globally convergent and might find the minimum closest to the initialisation.

- *Implicit Terms.* If we derive the Euler-Lagrange equations of the energy functional, we obtain terms that depend implicitly on the optical flow. These terms can not be discretised by standard schemes and have to be made explicit in the unknowns.

While there is no general consensus on how to proceed in the case of multi-modal functionals, the most common approach is to combine an incremental computation of the optical flow with a coarse-to-fine multi-scale approach. An incremental computation assumes that an initial estimate of the flow is known, such that the original implicit constraints can be approximated by a series of constraints that are *linearised* with respect to an unknown flow increment. Embedding this approach in a coarse-to-fine framework then solves the linearised problem for *successively smoothed representations* of the input images. In this context, it is usually assumed that the linearised constancy assumptions are valid at the coarsest image scale, and that the solution can be refined henceforth by an incremental computation on the subsequently finer scales. More important, the hierarchical multi-scale procedure will reduce the chance of getting trapped in erroneous local minima since small local minima in the energy tend to vanish at coarse image scales. Two examples of coarse-to-fine minimisation are shown in Fig. 3.1. A non-hierarchical minimisation of the original energy will easily get trapped in a local minimum if the initialisation is bad. This is illustrated by the red arrow. With the help of an increasingly coarse scale representation of the original problem, the global minimum can be found or at least approximated sufficiently well. In practice, the solution on the current scale will then serve as an initialisation on the next finer scale, a process that is illustrated by the green arrows.

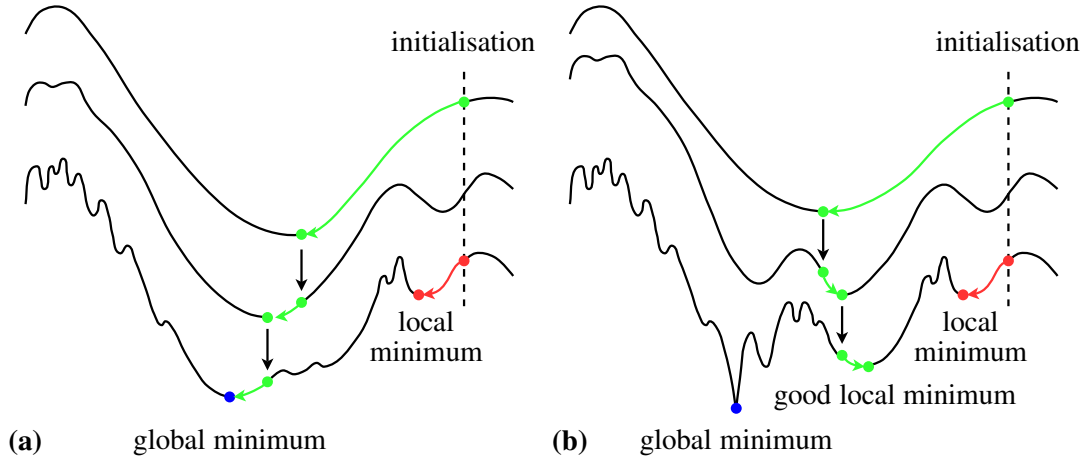

Fig. 3.1: Coarse-to-fine minimisation of a non-convex energy functional (adapted from [Bru06]). **Red:** Non-hierarchical minimisation of the original energy. **Green:** Coarse-to-fine minimisation. **(a)** Global minimum found. **(b)** Useful local minimum found.

3.2.1 Coarse-to-fine Warping Strategies

With respect to the incremental coarse-to-fine estimation of the optical flow, two strategies are often considered: (i) a fixed point iteration on the Euler-Lagrange equations [BBPW04] or (ii) the approximation of the original functional by a series of energies on every scale level [MP98a, MP02]. The first strategy starts from the Euler-Lagrange equations of the original energy and introduces a semi-implicit iteration scheme in these equations. The coarse-to-fine strategy is then embedded in a fixed point iteration on the Euler-Lagrange equations and all implicit expressions are linearised within this fixed point iteration. The second strategy formulates the energy functional directly in terms of the unknown flow increments. Linearisation is then performed within the functional itself and a minimiser is found for each scale level. Because the latter strategy requires a compensation of the images for the already computed motion field, a so-called image warping, such techniques are also known as *warping methods*. It has been shown in [BBPW04] that both incremental coarse-to-fine strategies are related and that the warping method can be derived as a hierarchical fixed point iteration on the Euler-Lagrange equations for our specific optical flow prototype (3.27). For some variational models in this thesis, however, – such as the models described in Chap. 6 – it is difficult to derive the Euler-Lagrange equations without first reformulating the energy in terms of the flow increments in a coarse-to-fine framework. To adopt a common strategy for the minimisation of all models in this thesis, we will therefore stick to the second warping strategy.

In the following we discuss how the energy will be formulated on each scale level and how the minimisation is embedded in the coarse-to-fine framework. Assuming that a solution for the optical flow $\boldsymbol{w} = (u, v)^\top$ is available from a coarser scale, we aim at expressing the total energy in terms of an unknown *small* flow increment $d\boldsymbol{w} = (du, dv)^\top$. To achieve this, we first resolve the data term, which depends implicitly on the optical flow, and introduce a tensor notation that makes the convexity of the resulting energy explicit.

3.2.1.1 Linearisation in the Data Term: The Motion Tensor

Let us first discuss the differential form of the data term. Using a first order Taylor expansion at the *warped image location* $\boldsymbol{x} + \boldsymbol{w}$, we can linearise the brightness difference with respect to the flow increment $d\boldsymbol{w}$ and obtain the approximation

$$g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}+d\boldsymbol{w})-g_{\mathrm{l}}(\boldsymbol{x}) \approx g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w})+\partial_x g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w})\, du+\partial_y g_{\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w})\, dv-g_{\mathrm{l}}(\boldsymbol{x}) \ . \quad (3.28)$$

If we introduce the following abbreviations

$$g_{\mathrm{r}x} = \partial_x g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) \ , \quad g_{\mathrm{r}y} = \partial_y g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) \ , \quad g_z = g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) - g_{\mathrm{l}}(\boldsymbol{x}) \ , \quad (3.29)$$

we can write the linearised term in (3.28) as an inner product

$$\boldsymbol{j}^\top \boldsymbol{d}_{\mathrm{h}} = g_{\mathrm{r}x} du + g_{\mathrm{r}y} dv + g_z \ , \quad (3.30)$$

where we have defined the homogeneous motion increment $\boldsymbol{d}_{\mathrm{h}} := (du, dv, 1)^\top$ and the extended gradient vector

$$\boldsymbol{j} := (g_{\mathrm{r}x}, g_{\mathrm{r}y}, g_z)^\top \ . \quad (3.31)$$

The linearised brightness constancy assumption can now be expressed by the equation

$$\boldsymbol{j}^\top \boldsymbol{d}_{\mathrm{h}} = 0 \ . \quad (3.32)$$

This equation can be seen as a large displacement variant of the classical *optical flow constraint* (OFC) of Horn and Schunck [HS81], as given by Eq. (1.4). Only this time, the temporal brightness difference is replaced by g_z, which can be understood as a brightness difference along the motion trajectory. It is clear that equation (3.32) expresses the constraint that the point $(du, dv)^\top$ has to lie on the line represented by $\boldsymbol{j}$. It does, however, not tell us where exactly on this line the point has to lie. Since every point on the line is a valid flow vector, the OFC alone is not sufficient to determine a unique local solution. This problem is known as the *aperture problem* [BPT88].

Introducing the additional abbreviations

$$g_{\mathrm{r}xx} = \partial_{xx} g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) \ , \quad g_{\mathrm{r}xy} = \partial_{xy} g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) \ , \quad g_{\mathrm{r}yy} = \partial_{yy} g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) \quad (3.33)$$

$$g_{xz} = \partial_x g_z \ , \qquad g_{yz} = \partial_y g_z \ , \quad (3.34)$$

we can also write the two gradient component differences as inner products

$$\partial_x g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w} + d\boldsymbol{w}) - \partial_x g_{\mathrm{l}}(\boldsymbol{x}) \ \approx \ \boldsymbol{j}_x^\top \boldsymbol{d}_{\mathrm{h}}, \quad (3.35)$$

$$\partial_y g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w} + d\boldsymbol{w}) - \partial_y g_{\mathrm{l}}(\boldsymbol{x}) \ \approx \ \boldsymbol{j}_y^\top \boldsymbol{d}_{\mathrm{h}} \ , \quad (3.36)$$

where we have defined the vectors

$$\boldsymbol{j}_x := \partial_x \boldsymbol{j} = (g_{\mathrm{r}xx}, g_{\mathrm{r}xy}, g_{xz})^\top \quad \text{and} \quad \boldsymbol{j}_y := \partial_y \boldsymbol{j} = (g_{\mathrm{r}xy}, g_{\mathrm{r}yy}, g_{yz})^\top \ . \quad (3.37)$$

The linearised gradient constancy assumption can thus be expressed by the two equations

$$\boldsymbol{j}_x^\top \boldsymbol{d}_{\mathrm{h}} \ = \ 0, \quad (3.38)$$

$$\boldsymbol{j}_y^\top \boldsymbol{d}_{\mathrm{h}} \ = \ 0 \ . \quad (3.39)$$

Now, it immediately becomes clear why it can be advantageous to include extra constancy assumptions in the data term: Each linearised assumption adds a new line constraint to the problem, which might help to disambiguate the solution.

If we now insert the three line constraints into the data term of energy (3.27), we can simplify the notation by writing the squared inner products as a quadratic form:

$$\begin{aligned} \mathcal{E}_{\mathrm{D}}(d\boldsymbol{w}) &= \Psi\left((\,\boldsymbol{j}^{\top}\boldsymbol{d}_{\mathrm{h}}\,)^2 + \gamma\,(\boldsymbol{j}_x^{\top}\boldsymbol{d}_{\mathrm{h}}\,)^2 + \gamma\,(\boldsymbol{j}_y^{\top}\boldsymbol{d}_{\mathrm{h}}\,)^2 \right), && (3.40)\\ &= \Psi\left(\boldsymbol{d}_{\mathrm{h}}^{\top} J\,\boldsymbol{d}_{\mathrm{h}}\right) \quad . && (3.41) \end{aligned}$$

Here, the symmetric 3×3 *motion tensor* [BW05]

$$J := \boldsymbol{j}\,\boldsymbol{j}^{\top} + \gamma\,\boldsymbol{j}_x\,\boldsymbol{j}_x^{\top} + \gamma\,\boldsymbol{j}_y\,\boldsymbol{j}_y^{\top} \quad , \tag{3.42}$$

provides coupling between the components of the flow increment. The motion tensor for the multi-channel data term (3.11) can be simply obtained by adding the motion tensors of each of the three image channels, as illustrated in [ZBW+09].

The Rank of the Motion Tensor. It has been shown in [Bru06] that the rank of the 2×2 submatrix of J containing the spatial derivatives determines the type of solution provided by the data term. Only for a full rank, the solution is unique, and this corresponds to at least two independently varying image features. For a rank smaller than two, we are confronted with the aperture problem: The solution has either an ambiguity along a line – as in the case of a single data constraint – or within the whole image plane – as in the case of homogeneous image regions (where the first two entries of $\boldsymbol{j}$, $\boldsymbol{j}_x$ and $\boldsymbol{j}_y$ are zero).

3.2.1.2 The Differential Form of the Energy

If we combine the above data term with linearised constraints and the smoothness term expressed in function of the flow increment, we obtain the following differential form of the energy that has to be minimised at each level of the coarse-to-fine approach

$$\mathcal{E}(d\boldsymbol{w}) = \int_{\Omega} \left(\Psi\left(\boldsymbol{d}_{\mathrm{h}}^{\top} J\,\boldsymbol{d}_{\mathrm{h}}\right) + \alpha\,\Psi\big(\,|\nabla(\boldsymbol{w}+d\boldsymbol{w})|^2\,\big)\right) \mathrm{d}\boldsymbol{x} \quad .$$

Deriving the Euler-Lagrange equations of $\mathcal{E}(d\boldsymbol{w})$ with respect to $d\boldsymbol{w}$ leads to the following system of coupled partial differential equations:

$$\begin{aligned} 0 \;=\;& \Psi'\left(\boldsymbol{d}_{\mathrm{h}}^{\top} J\,\boldsymbol{d}_{\mathrm{h}}\right)(J_{11}\,du + J_{12}\,dv + J_{13}) \\ & -\alpha\,\mathrm{div}\Big(\Psi'\big(\,|\nabla(\boldsymbol{w}+d\boldsymbol{w})|^2\,\big)\nabla(u+du)\Big), && (3.43)\\ 0 \;=\;& \Psi'\left(\boldsymbol{d}_{\mathrm{h}}^{\top} J\,\boldsymbol{d}_{\mathrm{h}}\right)(J_{12}\,du + J_{22}\,dv + J_{23}) \\ & -\alpha\,\mathrm{div}\Big(\Psi'\big(\,|\nabla(\boldsymbol{w}+d\boldsymbol{w})|^2\,\big)\nabla(v+dv)\Big) \quad . && (3.44) \end{aligned}$$

Here, J_{kl}, for $k, l \in \{1, 2, 3\}$, denotes the kl-th entry of J. It has to be noted that these equations are *non-linear* due to the function Ψ'. Because Ψ' is strictly convex, however, they have a *unique* solution which can be found by any globally convergent algorithm.

3.2.1.3 Hierarchical Minimisation

To embed the solution of the Euler-Lagrange equations in a coarse-to-fine framework, we use a representation of the images g_r and g_l that are smoothed at different scales. If we associate a variable k to the current image scale, we can start the coarse-to-fine refinement with a very smooth version of the images at the coarsest scale and use successively less smoothed image representations in the subsequent iterations. The smoothed images can for instance be obtained by convolving g_r and g_l with a Gaussian kernel of decreasing standard deviation or by using downsampled versions of the original images with a steadily increasing resolution. If the equation system (3.43) - (3.44) has been solved for the motion increments du^k and dv^k on the current image scale $(g_\mathrm{r}^k, g_\mathrm{l}^k)$, we can update the total optical flow by adding the motion increments to the already known flow

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k + d\boldsymbol{w}^k \quad . \tag{3.45}$$

After this update has been performed, we move on to the next image scale $(g_\mathrm{r}^{k+1}, g_\mathrm{l}^{k+1})$, where the motion tensor is recomputed by warping g_r^{k+1} with the new estimate $\boldsymbol{w}^{k+1}$. The Euler-Lagrange equations (3.43) - (3.44) are then solved again, but this time for the motion increments du^{k+1} and dv^{k+1}. Adding these increments to $\boldsymbol{w}^{k+1}$ gives us a new update for the total optical flow. This process of successive refinement through smoothed image representations is repeated until we reach the original image scale.

Algorithmically, we implement the coarse-to-fine framework by means of a *multi-resolution technique* [BAHH92, BA96, MP98b] that downsamples the image data to make use of the different resolution levels. Because of the smaller image sizes on the coarser levels, this approach is more efficient than the well-known *scale-space strategy* [AWS99], that considers Gaussian smoothed variants of the input data at a constant resolution. To resample the original images to a coarser resolution level and to interpolate the total optical flow $\boldsymbol{w} + d\boldsymbol{w}$ to the next finer level, we apply an area-based interpolation scheme as in [BWKS05]. The image warping by the known optical flow $\boldsymbol{w}$ is realised by a backward image registration approach that is based on bilinear interpolation [MP98a, BBPW04]. We further reduce the resolution from one level to the next according to a downsampling factor $\eta \in (0, 1)$. For higher values ($\eta \in [0.90, 0.95]$) we generally obtain more accurate results, while small values ($\eta \in [0.5, 0.7]$) allow for a faster computation. In this thesis the emphasis is on accuracy rather than on run time and we therefore set $\eta = 0.95$.

3.2.2 Discretisation

In the previous section we have seen how the non-convex energy of our optical flow prototype can be minimised in a coarse-to-fine framework. By expressing the energy in function of the unknown optical flow increments, we obtained a system of Euler-Lagrange equations for every level of the multi-resolution pyramid. In this section we present a discrete version of this coupled system of non-linear partial differential equations and discuss how its structure can help to find a solution by means of efficient numerical algorithms.

Tab. 3.2: Discretisations of averaging and differential operators (adapted from [Bru06]).

one-sided averaging	$\mathrm{M}_x^{\pm}\left([z]_{i,j}\right) := \frac{[z]_{i\pm1,j}+[z]_{i,j}}{2}$
	$\mathrm{M}_y^{\pm}\left([z]_{i,j}\right) := \frac{[z]_{i,j\pm1}+[z]_{i,j}}{2}$
one-sided differences	$\mathrm{D}_x^{\pm}\left([z]_{i,j}\right) := \pm\frac{[z]_{i\pm1,j}-[z]_{i,j}}{h}$
	$\mathrm{D}_y^{\pm}\left([z]_{i,j}\right) := \pm\frac{[z]_{i,j\pm1}-[z]_{i,j}}{h}$
central differences	$\mathrm{D}_x\left([z]_{i,j}\right) := \frac{[z]_{i+1,j}-[z]_{i-1,j}}{2h}$
	$\mathrm{D}_y\left([z]_{i,j}\right) := \frac{[z]_{i,j+1}-[z]_{i,j-1}}{2h}$
squared differences	$\mathrm{D}_x^2\left([z]_{i,j}\right) := \frac{1}{2}\left(\mathrm{D}_x^+\left([z]_{i,j}\right)\right)^2 + \frac{1}{2}\left(\mathrm{D}_x^-\left([z]_{i,j}\right)\right)^2$
	$\mathrm{D}_y^2\left([z]_{i,j}\right) := \frac{1}{2}\left(\mathrm{D}_y^+\left([z]_{i,j}\right)\right)^2 + \frac{1}{2}\left(\mathrm{D}_y^-\left([z]_{i,j}\right)\right)^2$
gradient magnitude	$\left\lvert\mathrm{D}^2\left([z]_{i,j}\right)\right\rvert := \sqrt{\mathrm{D}_x^2\left([z]_{i,j}\right) + \mathrm{D}_y^2\left([z]_{i,j}\right)}$

3.2.2.1 Discretisation Aspects

To solve the Euler-Lagrange equations numerically, we discretise them by means of finite differences [MG80, EBY99]. To this end we consider the unknown increments $du(x, y)$ and $dv(x, y)$ on a rectangular pixel grid with a grid spacing of h_x in x-direction and h_y in y-direction. In this thesis we assume square pixel sizes such that we can set $h_x = h_y = h$. We further denote by $[\,du\,]_{i,j}$ and $[\,dv\,]_{i,j}$ the approximations of du and dv at pixel location (i, j), with $i = 1, \ldots, n_x$ and $j = 1, \ldots, n_y$. The total number of pixels is thus given by $n = n_x \times n_y$. It has to be noted that in this section the image size n refers to the resampled image size on the current level of the multi-resolution pyramid. Due to the notational burden, however, we will refrain from indexing all variables by the current level.

Before we move on to a discretised version of the Euler-Lagrange equations, we first discuss the discretisation of the differential operator arising from the smoothness term. In the most general case, this differential operator can be written as a divergence expression involving a symmetric 2×2 diffusion tensor

$$\operatorname{div}(D\,\nabla z) = \partial_x\left(D_{11}\,\partial_x\,z\right) + \partial_x\left(D_{12}\,\partial_y\,z\right) + \partial_y\left(D_{12}\,\partial_x\,z\right) + \partial_y\left(D_{22}\,\partial_y\,z\right)\ . \quad (3.46)$$

The terms $\partial_x\left(D_{11}\,\partial_x\,z\right)$ and $\partial_y\left(D_{22}\,\partial_y\,z\right)$ are discretised by backward and forward differences, while $\partial_x\left(D_{12}\,\partial_y\,z\right)$ and $\partial_y\left(D_{12}\,\partial_x\,z\right)$ are discretised by central differences

$$\partial_x\left(D_{11}\,\partial_x z\right) \approx \mathrm{D}_x^-\big(\,\mathrm{M}_x^+(\,[D_{11}]_{i,j})\;\mathrm{D}_x^+(\,[z]_{i,j}\,)\big), \quad (3.47)$$

$$\partial_x\left(D_{12}\,\partial_y z\right) \approx \mathrm{D}_x\big(\qquad [D_{12}]_{i,j}\;\;\mathrm{D}_y\,(\,[z]_{i,j}\,)\big), \quad (3.48)$$

$$\partial_y\left(D_{12}\,\partial_x z\right) \approx \mathrm{D}_y\big(\qquad [D_{12}]_{i,j}\;\;\mathrm{D}_x\,(\,[z]_{i,j}\,)\big), \quad (3.49)$$

$$\partial_y\left(D_{22}\,\partial_y z\right) \approx \mathrm{D}_y^-\big(\,\mathrm{M}_y^+(\,[D_{22}]_{i,j})\;\mathrm{D}_y^+(\,[z]_{i,j}\,)\big)\ , \quad (3.50)$$

where $[D_{kl}]_{i,j}$ for $k, l \in \{1, 2\}$ denotes the discretised kl-th entry of the tensor D. The applied difference operators are clarified in Table 3.2. This finite difference approximation of the divergence expression can be obtained by discretising the original energy [Wei96]. For isotropic flow-driven regularisation, such as the one in our energy prototype, the diffusion tensor is a diagonal matrix, more specifically $D = \Psi'\big(\,|\nabla(\boldsymbol{w}+d\boldsymbol{w})|^2\,\big)I^2$. In this case, the discretisations (3.47) and (3.50) will result in a communication of the central pixel (i, j) with its four neighbours in axial directions. In the case of anisotropic regularisation, the mixed terms $\partial_x(D_{12}\,\partial_y z)$ and $\partial_y(D_{12}\,\partial_x z)$ have to be considered as well. These will result in an additional communication of the central pixel with its neighbours in diagonal direction. To ensure a stable discretisation, which bounds over- and undershoots in the solutions, the anisotropic stencil weights have to satisfy the non-negativity requirement. Since this property is generally not guaranteed by the discretisation (3.47) - (3.50) we use the anisotropic stencil proposed in [Wei96]. All spatial derivatives of the image data, used in the computation of the motion tensor entries, are approximated using fourth-order finite differences with stencil $(1, -8, 0, 8, -1)/12h$.

3.2.2.2 The Discrete Euler-Lagrange Equations

The discrete Euler-Lagrange equations, that correspond to our optical flow prototype, form a system of non-linear equations in the unknowns $[\,du\,]_{i,j}$ and $[\,dv\,]_{i,j}$. To analyse the structure of this system, we first write the Euler-Lagrange equations (3.43) - (3.44) in point-based notation for a pixel location (i, j). The equations for $[\,du\,]_{i,j}$ and $[\,dv\,]_{i,j}$ are

$$\begin{aligned}
0 \;=\;& [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{11}]_{i,j}\,[\,du\,]_{i,j} + [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{12}]_{i,j}\,[\,dv\,]_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]_{i,j}\big)}{2} \frac{\big([\,du\,]_{\tilde{i},\tilde{j}} - [\,du\,]_{i,j}\big)}{h^2} \\
& + [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{13}]_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]_{i,j}\big)}{2} \frac{\big([\,u\,]_{\tilde{i},\tilde{j}} - [\,u\,]_{i,j}\big)}{h^2} \qquad (3.51) \\
0 \;=\;& [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{12}]_{i,j}\,[\,du\,]_{i,j} + [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{22}]_{i,j}\,[\,dv\,]_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]_{i,j}\big)}{2} \frac{\big([\,dv\,]_{\tilde{i},\tilde{j}} - [\,dv\,]_{i,j}\big)}{h^2} \\
& + [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{23}]_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]_{i,j}\big)}{2} \frac{\big([\,v\,]_{\tilde{i},\tilde{j}} - [\,v\,]_{i,j}\big)}{h^2}\;, \qquad (3.52)
\end{aligned}$$

for $i = 1, \ldots, n_x$ and $j = 1, \ldots, n_y$. Here, $\mathcal{N}_l(i, j)$ denotes the set of the two neighbours of the pixel (i, j) in the axis direction l, with $l \in \{x, y\}$. Further, $[\Psi'_{\mathrm{D}}]_{i,j}$ and $[\Psi'_{\mathrm{S}}]_{i,j}$ denote the approximations of $\Psi'\left(\boldsymbol{d}_{\mathrm{h}}^{\top} J\, \boldsymbol{d}_{\mathrm{h}}\right)$ and $\Psi'\big(\,|\nabla(\boldsymbol{w}+d\boldsymbol{w})|^2\,\big)$ respectively in the

pixel (i, j). For the TV diffusivity (3.25), they can be computed as

$$[\Psi'_{\mathrm{D}}]_{i,j} = \frac{1}{2\sqrt{\big([\,du\,]_{i,j}, [\,dv\,]_{i,j}, 1\big)[J\,]_{i,j}\big([\,du\,]_{i,j}, [\,dv\,]_{i,j}, 1\big)^\top + \epsilon^2}}, \tag{3.53}$$

$$[\Psi'_{\mathrm{S}}]_{i,j} = \frac{1}{2\sqrt{\big|\mathrm{D}^2\big([\,u\,]_{i,j} + [\,du\,]_{i,j}\big)\big|^2 + \big|\mathrm{D}^2\big([\,v\,]_{i,j} + [\,dv\,]_{i,j}\big)\big|^2 + \epsilon^2}}\ , \tag{3.54}$$

where $[J\,]_{i,j}$ is the discrete approximation of the motion tensor and the discrete gradient magnitude operator $\big|\mathrm{D}^2\ ([z]_{i,j})\big|$ is defined in Table 3.2. It has to be noted that the non-linearity of the equation system (3.51)-(3.52) is due to the expressions $[\Psi'_{\mathrm{D}}]_{i,j}$ and $[\Psi'_{\mathrm{S}}]_{i,j}$ that depend on the unknowns $[\,du\,]_{i,j}$ and $[\,dv\,]_{i,j}$. The convexity of the corresponding energy functional nevertheless guarantees a unique solution.

3.2.2.3 Structure of the Non-linear System

In the point-based Euler-Lagrange equations we have already gathered those terms that directly depend on the unknown increments $[\,du\,]_{i,j}$ and $[\,dv\,]_{i,j}$ and those terms that only depend indirectly on the increments via the expressions $[\Psi'_{\mathrm{D}}]_{i,j}$ and $[\Psi'_{\mathrm{S}}]_{i,j}$. To see how these dependencies influence the structure of the overall equation system we first arrange the unknowns via a row-major ordering in two $n \times 1$ vectors

$$\mathbf{du} := \big([\,du\,]_{1,1}, \ldots, [\,du\,]_{n_x,n_y}\big)^\top, \tag{3.55}$$

$$\mathbf{dv} := \big([\,dv\,]_{1,1}, \ldots, [\,dv\,]_{n_x,n_y}\big)^\top\ . \tag{3.56}$$

To keep an overview, we additionally introduce the following compact notations

$$\mathbf{u} := \big([\,u\,]_{1,1}, \ldots, [\,u\,]_{n_x,n_y}\big)^\top, \tag{3.57}$$

$$\mathbf{v} := \big([\,v\,]_{1,1}, \ldots, [\,v\,]_{n_x,n_y}\big)^\top, \tag{3.58}$$

$$\mathbf{j}_{kl} := \big([J_{kl}]_{1,1}, \ldots, [\,J_{kl}\,]_{n_x,n_y}\big)^\top \quad \text{for } k, l \in \{1, 2, 3\}, \tag{3.59}$$

$$\mathbf{J}_{kl} := \mathrm{diag}\,\big([J_{kl}]_{1,1}, \ldots, [\,J_{kl}\,]_{n_x,n_y}\big) \quad \text{for } k, l \in \{1, 2, 3\}, \tag{3.60}$$

$$\boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv}) := \mathrm{diag}\,\big([\Psi'_{\mathrm{D}}]_{1,1}, \ldots, [\Psi'_{\mathrm{D}}]_{n_x,n_y}\big)\ . \tag{3.61}$$

The discrete Euler-Lagrange equation system can now be written as

$$\underbrace{\left(\begin{pmatrix} \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\,\mathbf{J}_{11} & \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\,\mathbf{J}_{12} \\ \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\,\mathbf{J}_{12} & \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\,\mathbf{J}_{22} \end{pmatrix} - \alpha \begin{pmatrix} \mathbf{L}(\mathbf{du}, \mathbf{dv}) & 0^n \\ 0^n & \mathbf{L}(\mathbf{du}, \mathbf{dv}) \end{pmatrix}\right)}_{A(\boldsymbol{p})} \underbrace{\begin{pmatrix} \mathbf{du} \\ \mathbf{dv} \end{pmatrix}}_{\boldsymbol{p}}$$

$$+ \underbrace{\begin{pmatrix} \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\,\mathbf{j}_{13} \\ \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\,\mathbf{j}_{23} \end{pmatrix} - \alpha \begin{pmatrix} \mathbf{L}(\mathbf{du}, \mathbf{dv}) & 0^n \\ 0^n & \mathbf{L}(\mathbf{du}, \mathbf{dv}) \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}}_{\boldsymbol{c}(\boldsymbol{p})} = \underbrace{\begin{pmatrix} \mathbf{0}^n \\ \mathbf{0}^n \end{pmatrix}}_{\boldsymbol{b}}\ , \tag{3.62}$$

where $\mathbf{L}(\mathbf{du}, \mathbf{dv})$ is the discrete version of the differential operator arising from the smoothness term. It is a *symmetric negative semidefinite* $n \times n$ matrix that has a *pentadiagonal structure* resulting from the five non-zero weights in the discretisation stencil of the divergence expression [Wei94a]. By construction, the matrices $\mathbf{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})$, $\mathbf{J}_{11}$ and $\mathbf{J}_{22}$ are symmetric positive semidefinite and, as a result, the overall $2n \times 2n$ system matrix A shares this property as well [Bru06, SBK10]. For non-constant images one can even show that A is positive definite, which means that it is invertible and that a unique solution exists [Bru06]. It has to be noted that A has a very sparse structure and that there is only a point-wise coupling (i.e. in each pixel) between $\mathbf{du}$ and $\mathbf{dv}$ via the off-diagonal block $\mathbf{J}_{12}$ and a neighbourhood coupling (i.e. between the five neighbouring pixels) via the five non-zero diagonals of $\mathbf{L}(\mathbf{du}, \mathbf{dv})$. The remaining terms in Eq. (3.62) have been gathered in the $2n \times 1$ vector $\boldsymbol{c}$, while the right hand side $\boldsymbol{b}$ is zero. If we represent the $2n$ unknowns by the parameter vector $\boldsymbol{p}$, we are thus solving a non-linear system of the form

$$f(\boldsymbol{p}) = \boldsymbol{b} \ , \tag{3.63}$$

where the non-linear operator f can be split into a matrix operation and a vector addition

$$f(\boldsymbol{p}) = A(\boldsymbol{p})\,\boldsymbol{p} + \boldsymbol{c}(\boldsymbol{p}) \ , \tag{3.64}$$

that are both depending on the variable $\boldsymbol{p}$. As we will see next, this particular structure will grant us the possibility to approach the problem as a series of linear problems that can be solved by means of established linear techniques.

3.2.3 Solvers

To solve the non-linear system of discrete Euler-Lagrange equations we extend the classical Gauß-Seidel method for solving linear systems of equations to the non-linear case and make use of advanced multigrid schemes to speed up computation.

3.2.3.1 The Lagged Diffusivity Method

The non-linear dependencies that arise in the discrete Euler-Lagrange equations (3.62) from the use of the robust function Ψ are collected in the matrix $A(\boldsymbol{p})$ and the vector $\boldsymbol{c}(\boldsymbol{p})$. If we now evaluate both operators $A(\boldsymbol{p})$ and $\boldsymbol{c}(\boldsymbol{p})$ for an already known solution $\boldsymbol{p}^k$, we can consider the following linear system

$$A(\boldsymbol{p}^k)\,\boldsymbol{p}^{k+1} = \left(\boldsymbol{b} - \boldsymbol{c}(\boldsymbol{p}^k)\right) \ . \tag{3.65}$$

Because the system matrix $A(\boldsymbol{p})$ is symmetric positive definite for all $\boldsymbol{p}$, it is invertible such that the update $\boldsymbol{p}^{k+1}$ can be found by

$$\boldsymbol{p}^{k+1} = A^{-1}(\boldsymbol{p}^k)\left(\boldsymbol{b} - \boldsymbol{c}(\boldsymbol{p}^k)\right) \ . \tag{3.66}$$

By doing so, we have actually introduced a fixed-point iteration with respect to the old or *lagged* solution that decomposes the original non-linear problem in a series of linear problems. This approach, which is often encountered in literature for solving systems of non-linear equations, is called the *lagged diffusivity method* [KNPS68, FKN73, CM99].

The linear system (3.65) that has to be solved at each fixed-point iteration step can be written in point-based notation as

$$
\begin{aligned}
0 \;=\;& [\Psi'_{\mathrm{D}}]^k_{i,j}\,[J_{11}]_{i,j}\,[\,du\,]^{k+1}_{i,j} + [\Psi'_{\mathrm{D}}]^k_{i,j}\,[J_{12}]_{i,j}\,[\,dv\,]^{k+1}_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} \left([\,du\,]^{k+1}_{\tilde{i},\tilde{j}} - [\,du\,]^{k+1}_{i,j}\right) \\
& + [\Psi'_{\mathrm{D}}]^k_{i,j}\,[J_{13}]_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} \left([\,u\,]_{\tilde{i},\tilde{j}} - [\,u\,]_{i,j}\right) \qquad (3.67) \\
0 \;=\;& [\Psi'_{\mathrm{D}}]^k_{i,j}\,[J_{12}]_{i,j}\,[\,du\,]^{k+1}_{i,j} + [\Psi'_{\mathrm{D}}]^k_{i,j}\,[J_{22}]_{i,j}\,[\,dv\,]^{k+1}_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} \left([\,dv\,]^{k+1}_{\tilde{i},\tilde{j}} - [\,dv\,]^{k+1}_{i,j}\right) \\
& + [\Psi'_{\mathrm{D}}]^k_{i,j}\,[J_{23}]_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} \left([\,v\,]_{\tilde{i},\tilde{j}} - [\,v\,]_{i,j}\right) \quad , \qquad (3.68)
\end{aligned}
$$

for $i = 1,\ldots,n_x$ and $j = 1,\ldots,n_y$. This system has to be solved for the unknowns $[\,du\,]^{k+1}_{i,j}$ and $[\,dv\,]^{k+1}_{i,j}$ while keeping the expressions $[\Psi'_{\mathrm{D}}]_{i,j}$ and $[\Psi'_{\mathrm{S}}]_{i,j}$ fixed in the solutions $[\,du\,]^{k}_{i,j}$ and $[\,dv\,]^{k}_{i,j}$ of the previous iteration step.

3.2.3.2 Basic Solver

A direct solution of the linear system (3.65) by inverting the system matrix A does not come into consideration as the size of the problem would have severe repercussions on memory requirement and computation time. Instead, various iterative methods exist that speed up the computation by embedding the solution in a fixed-point iteration while replacing the true inverse by a sufficiently good approximation that is cheap enough to determine. Often this approximation is a triangular matrix, such that the system can be solved in every iteration step by a simple point-wise forward substitution. The idea of iteratively solving simplified versions of the original linear system, forms the basis for some of the most frequently used relaxation techniques to date, including the *Gauß-Seidel* and *Successive Overrelaxation* (SOR) method [You71]. The convergence of these methods is guaranteed for any symmetric positive definite system matrix and the sparsity of our specific problem furthermore allows for an optimal linear time complexity of $O(n)$. Because of the better smoothing behaviour of the classical Gauß-Seidel method of high error frequencies, we will prefer this method over SOR in view of a multigrid extension afterwards.

The Gauß-Seidel Relaxation. To formalise the Gauß-Seidel method for our linear system (3.65), we denote by p_i, $1 \leq i \leq 2n$, the i-th entry of the parameter vector $\boldsymbol{p}$ and by $A_{ij}(\boldsymbol{p})$, $1 \leq j \leq 2n$ the ij-th entry of the system matrix $A(\boldsymbol{p})$. The update instruction for the i-th unknown then reads

$$p_i^{k+1,m+1} = \frac{1}{A_{ii}(\boldsymbol{p}^k)} \left(b_i - c_i(\boldsymbol{p}^k) - \sum_{j=1}^{i-1} A_{ij}(\boldsymbol{p}^k)\, p_j^{k+1,m+1} - \sum_{j=i+1}^{n} A_{ij}(\boldsymbol{p}^k)\, p_j^{k+1,m} \right) , \quad (3.69)$$

where m denotes the iteration index of the linear solver. Here, it is usually unnecessary and computationally overly expensive to solve the linear system (3.65) exactly, because the weights $[\Psi'_{\mathrm{D}}]$ and $[\Psi'_{\mathrm{S}}]$ have been computed for an old solution and we are thus only solving an approximation to the true Euler-Lagrange equations. It is in fact more advantageous with respect to computational efficiency to only perform a few Gauß-Seidel relaxation steps before moving on to the update of the lagged non-linear expressions [WHS+01, Vog02, FSHW04]. In the extreme case one may even think of updating the non-linear expressions after each iteration of the basic solver, in which case the solver index m collapses to the lagged diffusivity iteration index k. This technique, where the non-linear expressions are kept fixed for only one Gauß-Seidel iteration, is known as the lagged-diffusivity method with *frozen coefficients*.

The Coupled Point Gauß-Seidel Relaxation. In this thesis we are additionally interested in exploiting the coupling that exists between the unknowns in the equation system. With regard to this, we will perform in the forward substitution step a simultaneous update of both flow increments du and dv in the same pixel. This variant of the Gauß-Seidel technique that computes the unknowns in a block-wise fashion is known as the *coupled point relaxation* (CPR) method [Wes92]. To formalise the CPR method for our linear system, we denote by $\boldsymbol{p}_i$, $1 \leq i \leq n$, the 2×1 vector that contains the two discretised flow increments that belong to a certain pixel i. This time, $A_{ij}(\boldsymbol{p})$ denotes the 2×2 sub-matrix of $A(\boldsymbol{p})$ that only contains those entries that relate to the two pixels i and j. Then, the CPR iteration instruction with frozen coefficients for the i-th pixel can be formulated as

$$\boldsymbol{p}_i^{k+1} = A_{ii}^{-1}(\boldsymbol{p}^k) \left(\boldsymbol{b}_i - \boldsymbol{c}_i(\boldsymbol{p}^k) - \sum_{j=1}^{i-1} A_{ij}(\boldsymbol{p}^k)\, \boldsymbol{p}_j^{k+1} - \sum_{j=i+1}^{n} A_{ij}(\boldsymbol{p}^k)\, \boldsymbol{p}_j^{k} \right) . \quad (3.70)$$

In essence we are thus solving a 2×2 equation system in every pixel. If we denote again by (i, j) the location of a certain pixel in the image grid, we can write this system in point-based notation as

$$\begin{pmatrix} [\,du\,]_{i,j}^{k+1} \\ [\,dv\,]_{i,j}^{k+1} \end{pmatrix} = \begin{pmatrix} [M_{11}]_{i,j}^{k} & [M_{12}]_{i,j}^{k} \\ [M_{12}]_{i,j}^{k} & [M_{22}]_{i,j}^{k} \end{pmatrix}^{-1} \begin{pmatrix} [r_1]_{i,j}^{k} \\ [r_2]_{i,j}^{k} \end{pmatrix} , \quad (3.71)$$

for $i = 1, \ldots, n_x$ and $j = 1, \ldots, n_y$. The matrix entries are given by

$$[M_{11}]^k_{i,j} = [\Psi'_{\mathrm{D}}]^k_{i,j} [J_{11}]_{i,j} + \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2}, \tag{3.72}$$

$$[M_{22}]^k_{i,j} = [\Psi'_{\mathrm{D}}]^k_{i,j} [J_{22}]_{i,j} + \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2}, \tag{3.73}$$

$$[M_{12}]^k_{i,j} = [\Psi'_{\mathrm{D}}]^k_{i,j} [J_{12}]_{i,j} \ , \tag{3.74}$$

and the right hand side by

$$\begin{aligned}
[r_1]^k_{i,j} = & - [\Psi'_{\mathrm{D}}]^k_{i,j} [J_{13}]_{i,j} \\
& + \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}^-_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} [\, du \,]^{k+1}_{\tilde{i},\tilde{j}} \\
& + \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}^+_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} [\, du \,]^{k}_{\tilde{i},\tilde{j}} \\
& + \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} \left([\, u \,]_{\tilde{i},\tilde{j}} - [\, u \,]_{i,j}\right),
\end{aligned} \tag{3.75}$$

$$\begin{aligned}
[r_2]^k_{i,j} = & - [\Psi'_{\mathrm{D}}]^k_{i,j} [J_{23}]_{i,j} \\
& + \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}^-_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} [\, dv \,]^{k+1}_{\tilde{i},\tilde{j}} \\
& + \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}^+_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} [\, dv \,]^{k}_{\tilde{i},\tilde{j}} \\
& + \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]^k_{i,j}\right)}{2h^2} \left([\, v \,]_{\tilde{i},\tilde{j}} - [\, v \,]_{i,j}\right) \ ,
\end{aligned} \tag{3.76}$$

where $\mathcal{N}^-(i,j)$ is the set of neighbours of pixel (i,j) in the direction of the l-axis that have already been updated (i.e. for which $\tilde{i} < i$ and $\tilde{j} < j$) and $\mathcal{N}^+(i,j)$ the set of neighbours that still need to be updated (i.e. for which $\tilde{i} > i$ and $\tilde{j} > j$).

3.2.3.3 Multigrid Methods

The equation systems in this thesis have in common that there only exists a local coupling between the unknowns via the small neighbourhood used to discretise the differential operator (3.46). Due to the resulting slow exchange of information between consecutive iteration steps, classical relaxation techniques, such as the Gauß-Seidel method,

have difficulties suppressing low error frequencies in the solution for this type of coupling [Bra77, Hac85, BHM00]. As a consequence, these methods may exhibit poor convergence behaviour and may require many iterations before producing an estimate that is close enough to the true solution. As a remedy, hierarchical solvers have been developed that solve additional coarse grid representations of the original equation system to achieve better error reduction characteristics. Such *multigrid techniques* belong to the fastest numerical schemes to date for solving linear and nonlinear systems of equations.

The V-cycle. Crucial to the idea of multigrid methods, is that the low frequency components in the *error* – defined here as the difference between the current estimate and the true solution – will be suppressed slowly by a Gauß-Seidel-like method, while higher frequency components will be suppressed more efficiently. Since low frequency components will reappear as higher ones on a coarser grid representation, however, they can in turn be attenuated by applying the relaxation technique on multiple grid levels. It is important to note that this frequency analysis only applies to the error of the solution and not to the solution itself, as the desired optical flow is allowed to contain both low and high frequency components. To improve the convergence behaviour of non-hierarchical solvers, multigrid methods therefore introduce *a correction step that estimates the error of the current fine grid solution on a coarser grid level*, and then correct the fine grid solution by this coarse grid error. A two-grid correction cycle typically contains the following five steps:

1. *Presmoothing Relaxation:* High frequency components in the estimation error are removed by performing a fixed number of CPR Gauß-Seidel iterations on the original fine grid equation system.

2. *Coarse Grid Transfer:* Next, a coarse grid version of the equation system has to be created. Since we are not interested in computing the actual optical flow at a coarser level, but the error, we do not transfer the original equation system. Instead we build a coarse grid version of the so-called *residual equation system*. The residual equation system has as a solution the error of the current optical flow estimate and can be derived as a modification of the original equation system [Bru06].

3. *Coarse Grid Solution:* The low frequency components that remain in the error after the presmoothing relaxation step reappear on the coarser grid as higher frequencies. These can now be attenuated by applying the basic CPR Gauß-Seidel solver to the coarse version of the residual equation system.

4. *Fine Grid Transfer and Correction:* The estimated coarse grid error is transferred back to the fine grid level and is used to correct the optical flow from the presmoothing relaxation step. This is done by simply adding the error to the current estimate.

5. *Postsmoothing Relaxation:* Possible high frequency components that have been introduced by transferring the coarse grid error to the fine grid are eliminated by performing a final number of Gauß-Seidel iterations on the original equation system.

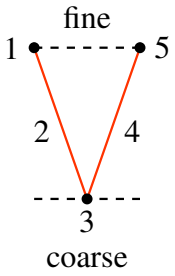

Fig. 3.2: A two-grid V-cycle.

A graphical representation of these five steps is shown in Fig. 3.2. The two grid levels are represented by dotted horizontal lines and the numbers correspond to the steps listed above. Each dot in the graph denotes the execution of a number of relaxation steps, while the red lines stand for the intergrid transfer operations. Because of the shape of the two-grid correction cycle, it is generally referred to as a *V-cycle*. It is of course possible to include more grid levels than the two depicted here. In such case, a third grid would provide a correction for the solution on the second grid and so on. In the case of multiple resolution levels we can even think of performing a second correction step on each grid level for improved accuracy. This will give rise to a recursive scheme that is known as a *W-cycle*.

The Full Multigrid Cycle. For the best performance, however, we will additionally apply a strategy that we have used before in the context of energy minimisation and embed the V- or W-cycles in a coarse-to-fine framework. This means that we do not directly start with our multigrid method on the finest grid level, but instead perform our correction cycles at coarser levels and use the result as initialisation for the next finer one. Aside from speeding up the computation by providing the V- and W-cycles with a good initialisation on each grid level, this strategy has the additional benefit of avoiding bad local minima, especially for nonlinear problems such as ours. This type of multigrid method is called *full multigrid* and has a proven linear time complexity of $O(n)$ for simple linear problems. Although a full multigrid approach requires the creation and the solution of a large number of equation systems, the overall speedup comes from the fact that less iterations of the basic solver are required and that the equation systems on the coarser grid levels can be solved much cheaper than the one on the finest grid. A graphical representation of a full multigrid cycle using four grid levels and 2 W-cycles per level is depicted in Fig. 3.3 (b).

Nested Hierarchical Ideas. The hierarchical full multigrid method is meant as a fast solver for the Euler-Lagrange equations. We keep in mind, however, that the Euler-Lagrange equations have to be solved on each level of a hierarchical coarse-to-fine approach themselves (see Section 3.2.1.3). The big picture thus involves two multi-resolution pyramids as depicted in Fig. 3.3 – (a) one for the coarse-to-fine warping strategy and (b) one for the full multigrid scheme. These two pyramids are generally not the same and differ in the following way: (i) While the coarse-to-fine warping pyramid often requires a downsampling factor of $\eta = 0.9$ or higher for good accuracy, a factor of $\eta = 0.5$ is already sufficient for the multigrid pyramid. (ii) The coarse-to-fine pyramid furthermore consist of resampled versions of the original images. The multigrid pyramid, on the other hand, is rather a hierarchy of equation systems, obtained by resampling the coefficients of the original equation system and recomputing the penaliser weights that depend non-linearly on the solution. To resample the coefficients and to interpolate the error to the next finer level we can, nevertheless, apply the same area-based scheme that we used for resampling the images in the coarse-to-fine warping approach [BWKS05]. The nested hierarchy discussed here, leads to a sophisticated solution strategy that guarantees high accuracy results, but can make the implementation at the same time quite challenging.

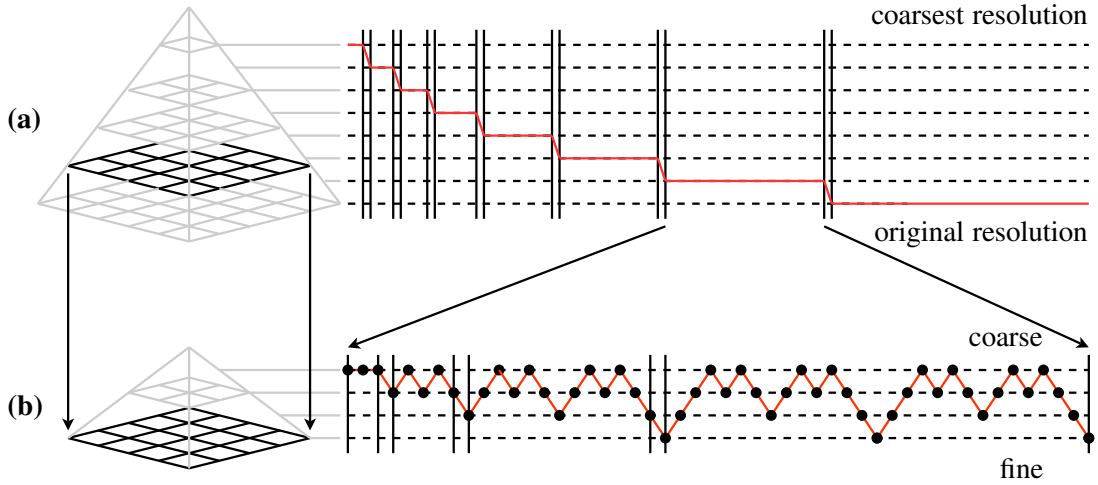

Fig. 3.3: The hierarchical ideas used in this thesis. **(a)** The coarse-to-fine warping strategy used to minimise the energy. Starting on the coarsest image scale, we successively refine the optical flow by using the estimate on the current scale as an initialisation for the solution on the next scale. The red line depicts the transition through the image pyramid and the approximate computation time. The main motivation for this approach is to reduce the probability of ending up in a bad local minimum. **(b)** The full multigrid strategy used to solve the Euler-Lagrange equations. This is a combination of the previous coarse-to-fine strategy and the multigrid correction idea. The red line depicts the transition through the multigrid pyramid and the black dots the execution of a certain number of iterations of the basic solver. Note that the different levels are visited multiple times to improve accuracy. The main motivation behind full multigrid is a reduction of computation time.

3.3 Summary

In this chapter we discussed the modelling and numerical solution of variational optical flow methods for large displacements. First we reviewed the design of the data term, where the emphasis was on the combination of constancy assumptions that are robust under illumination changes and perform well in the presence of noise, outliers and different motion types. We then moved on to choices for the smoothness term and identified different edge-preserving regularisation strategies based on the classification of the induced vector-valued diffusion process. A robust data term incorporating the brightness and gradient constancy and a smoothness term based on TV regularisation were finally combined in a model prototype that will serve as a baseline algorithm for novel correspondence methods in this thesis. In the second part of the chapter we investigated the minimisation of this model prototype via an incremental coarse-to-fine solution. A linearisation of the constancy assumptions with respect to a small flow increment lead to the concept of motion tensor, which makes the coupling between the flow components and the convexity of the resulting energy explicit. Provided with this simplified notation, we were able to clearly lay out the discretisation of the Euler-Lagrange equations and propose an efficient numerical scheme under the form of the lagged diffusivity method.

4 Fundamental Matrix from Dense Optical Flow

A three-dimensional point can be reconstructed from its projections if the camera matrices of both stereo views are known. The classical way of determining a pair of camera matrices is by relating a carefully chosen set of 3D control points to their projections on the image plane. Such a *full camera calibration* is mostly performed with the help of a calibration pattern of known dimensions [HZ00, FLP01]. For many image data, however, such as data from online repositories or hand-held cameras, a full calibration is not possible because the image source is unknown or because the stereo system changes during operation. In these cases it makes sense to ease the requirement from fully to weakly calibrated and compute the fundamental matrix as a first step towards a projective reconstruction. As mentioned earlier in Chap. 2, the estimation of the fundamental matrix does not require any prior knowledge about the scene and can be based on image correspondences alone.

In this chapter we introduce a new application of optical flow: *the dense estimation of the fundamental matrix of two stereo images*. To this end, we will combine the concepts from the two previous chapters in a two-step method that first establishes a set of dense image correspondences by computing a displacement vector in every pixel and then estimates the fundamental matrix by imposing the epipolar constraint on all matches. Apart from their high accuracy, variational optical flow methods offer two potential advantages for the computation of the fundamental matrix: (i) Due to the filling-in effect they provide a dense flow field, and thus a huge amount of correspondences, that can increase the robustness of the estimation process. (ii) They do not create gross outliers because of the combination of robust data constraints and a global smoothness assumptions.

Mismatches, on the other hand, are inherent to the widely-used class of *sparse feature based methods*. These techniques put most of their effort in selecting a small sophistically optimised set of feature correspondences that is most useful for the computation of the fundamental matrix. In this chapter we compare our dense results with those obtained by state-of-the-art feature based techniques and we will identify scenarios in which dense methods have a clear advantage over sparse approaches.

Our dense two-step method for estimating the fundamental matrix from optical flow is outlined in Sec. 4.1. In Sec. 4.2 we will give an overview of the different steps that make up sparse feature based methods and discuss twelve variants that incorporate the current state of the art. We will present a systematic juxtaposition of sparse and dense methods in Sec. 4.3 and demonstrate that modern optical flow based methods can serve as novel approaches for estimating the epipolar geometry with competitive quality. Throughout this chapter, we assume that the images have already been corrected for radial distortion. The linear methods discussed here, however, can be extended to simultaneously recover the fundamental matrix and the radial distortion [Fit01].

4.1 Fundamental Matrix Estimation from Optical Flow

In Chapter 3 we have described how dense optical flow can be computed between two stereo images by means of the variational method of Brox *et al.* [BBPW04]. In this section we will discuss how we can estimate the fundamental matrix from the set of correspondences that is provided by the optical flow. We will do this by studying a linear technique that computes the fundamental matrix directly from the epipolar constraint by means of a least squares fit on all the available correspondences. To minimise the least squares energy we consider two strategies: A total least squares solution in which we optimise all entries of the fundamental matrix simultaneously and an ordinary least squares solution in which we repeatedly fit a reduced number of parameters to the correspondence data. The main advantage of these linear methods is that they provide a good tradeoff between accuracy and ease of implementation. On the other hand, least squares fits are known to be relatively sensitive to noise in the correspondence data. To reduce the negative effect of outliers we will therefore introduce a robust weighing of the epipolar constraint, which will lead to a non-linear optimisation in a reweighted least squares framework.

4.1.1 The 8-point Algorithm of Longuet-Higgins

If the optical flow $\boldsymbol{w}$ has been estimated between two stereo images by minimising the energy (3.27), we can establish an image correspondence in every point of the left image. More precisely, the computed optical flow relates to every point $\boldsymbol{x} = (x, y)^\top$ in the left image a corresponding point in the right image that is given by

$$\boldsymbol{x}' = (x', y')^\top = \boldsymbol{x} + \boldsymbol{w} = (x + u, y + v)^\top \ . \tag{4.1}$$

In practice, we do this at the discrete pixel locations, resulting in a finite number of matches. We exclude points that are warped outside the image domain by the optical flow because the data term cannot be evaluated in these regions, leading to less reliable correspondences.

In order to develop a least squares based method for estimating the fundamental matrix F from the dense set of image correspondences provided by the optical flow, we first rewrite the epipolar constraint between $\boldsymbol{x}$ and $\boldsymbol{x}'$ as a product of two vectors [FLP01]

$$\boldsymbol{x}_{\mathrm{h}}'^{\top} F \, \boldsymbol{x}_{\mathrm{h}} = \boldsymbol{s}^\top \boldsymbol{f} = 0 \ . \tag{4.2}$$

Here, the 9×1 constraint vector $\boldsymbol{s}$ and parameter vector $\boldsymbol{f}$ are defined as

$$\boldsymbol{s} := (x\,x',\, y\,x',\, x',\, x\,y',\, y\,y',\, y',\, x,\, y,\, 1)^\top , \tag{4.3}$$

$$\boldsymbol{f} := (F_{11},\, F_{12},\, F_{13},\, F_{21},\, F_{22},\, F_{23},\, F_{31},\, F_{32},\, F_{33})^\top \ , \tag{4.4}$$

where F_{ij} for $i, j \in \{1, 2, 3\}$ denotes the ij-th entry of the fundamental matrix F. If we now consider equation (4.2) for n given point correspondences $(\boldsymbol{x}_i, \boldsymbol{x}_i')$, for $1 \leq i \leq n$, we can compose the following homogeneous system of which $\boldsymbol{f}$ has to be a solution

$$S \, \boldsymbol{f} = \boldsymbol{0}^9 \ , \tag{4.5}$$

where the $n \times 9$ system matrix S is formed by the n constraint vectors

$$S = \left(\boldsymbol{s}_1^\top, \ldots, \boldsymbol{s}_n^\top\right)^\top \ . \tag{4.6}$$

If we assume for a moment that the correspondences are exact (noise-free), the system matrix S can be at most of rank 8 for a solution to exist, because $\boldsymbol{f}$ is only defined up to a scale factor. Moreover, for $n \geq 8$ *exact* point correspondences, S will have exactly rank 8 and the fundamental matrix $\boldsymbol{f}$ will be given by the right null-space of S.

4.1.1.1 Total Least Squares Solution

Basic Idea. In practice, however, established point correspondences are *not exact* and it makes sense to include more than 8 point matches to achieve a stable solution. For the estimation of the fundamental matrix from optical flow, for instance, we wish to include all $n = n_x\, n_y$ available image correspondences, where $n_x \times n_y$ is the image resolution. In such case, the homogeneous system (4.5) becomes overdetermined and the rank of S generally becomes 9. To obtain a solution that is different from the trivial solution $\boldsymbol{f} = \mathbf{0}^9$, we can formulate the problem in a least squares sense by defining the following energy

$$\mathcal{E}(\boldsymbol{f}) = \sum_{i=1}^{n} (\boldsymbol{s}_i^\top \boldsymbol{f})^2 = \|S\,\boldsymbol{f}\|^2 \ , \tag{4.7}$$

while imposing the explicit constraint on the Frobenius norm of F

$$\|F\|_{\text{Frob}}^2 = \|\boldsymbol{f}\|^2 = \boldsymbol{f}^\top \boldsymbol{f} = 1 \ . \tag{4.8}$$

This approach is commonly known as a *total least squares (TLS)* solution [Lon81, vV91, TM97] and corresponds to an optimal Maximum Likelihood estimate of the fundamental matrix if the noise in the elements of $\boldsymbol{s}$ is assumed to be Gaussian with constant variance.

Solution as an Eigenvalue Problem. To find a minimiser of the energy (4.7) under the constraint (4.8), we make use of the method of the Lagrange multipliers and look for critical points of the Lagrangian

$$\mathcal{L}(\boldsymbol{f}, \lambda) = \sum_{i=1}^{n} (\boldsymbol{s}_i^\top \boldsymbol{f})^2 + \lambda(1 - \boldsymbol{f}^\top \boldsymbol{f}) \ . \tag{4.9}$$

Before differentiating the Lagrangian, we simplify things by rewriting $\mathcal{L}(\boldsymbol{f}, \lambda)$ as follows

$$\begin{aligned} \mathcal{L}(\boldsymbol{f}, \lambda) &= \boldsymbol{f}^\top \left(\sum_{i=1}^{n} \boldsymbol{s}_i \boldsymbol{s}_i^\top \right) \boldsymbol{f} + \lambda(1 - \boldsymbol{f}^\top \boldsymbol{f}) && (4.10) \\ &= \boldsymbol{f}^\top S^\top S\, \boldsymbol{f} + \lambda(1 - \boldsymbol{f}^\top \boldsymbol{f}) \ , && (4.11) \end{aligned}$$

where the 9×9 matrix $S^\top S$ is symmetric positive definite by construction. The critical points of $\mathcal{L}(\boldsymbol{f}, \lambda)$ are found by setting the derivatives with respect to $\boldsymbol{f}$ and λ to zero. If we denote by $\nabla_{\boldsymbol{f}}$ the gradient operator $(\partial_{f_1}, \cdots, \partial_{f_9})^\top$, it can be easily verified that a critical point of the Lagrangian has to fulfil the set of equations

$$\begin{aligned} \nabla_{\boldsymbol{f}}\, \mathcal{L}(\boldsymbol{f}, \lambda) &= \left(S^\top S - \lambda I\right) \boldsymbol{f} = \mathbf{0}^9, && (4.12) \\ \partial_\lambda\, \mathcal{L}(\boldsymbol{f}, \lambda) &= 1 - \|\boldsymbol{f}\|^2 = 0 \ , && (4.13) \end{aligned}$$

where we have dropped the factor 2 from the differentiation because of the equality to zero. Upon closer inspection, these equations constitute an *eigenvalue problem* where (4.12) is the eigenvalue equation of $S^\top S$ and (4.13) the usual requirement that each eigenvector has to be of unit norm. As a consequence, all eigenvector-eigenvalue pairs of $S^\top S$ are critical points of the Lagrangian. Because of its convex nature, however, only one eigenvector is a minimiser of the energy $\mathcal{E}(\boldsymbol{f})$. If we represent by $\boldsymbol{v}_i$ the normalised eigenvector of $S^\top S$ belonging to the i-th eigenvalue μ_i, then the energy in $\boldsymbol{v}_i$ evaluates to

$$\mathcal{E}(\boldsymbol{v}_i) = \|S\,\boldsymbol{v}_i\|^2 = \boldsymbol{v}_i^\top S^\top S\,\boldsymbol{v}_i = \boldsymbol{v}_i^\top \mu_i\,\boldsymbol{v}_i = \mu_i. \tag{4.14}$$

In other words, the energy of an eigenvector of $S^\top S$ is the corresponding eigenvalue. The unit norm vector $\boldsymbol{f}$ that minimises $\mathcal{E}(\boldsymbol{f})$ is thus the eigenvector belonging to the smallest eigenvalue of $S^\top S$. This simple algorithm for the estimation of the fundamental matrix from n image correspondences was already formulated by Longuet-Higgins in [Lon81]. In literature, it is generally known as the *8-point algorithm* and it can be implemented numerically with the help of the Jacobi method [Sch88, GV89, PTVF92].

4.1.1.2 Ordinary Least Squares Solution

Basic Idea. In the TLS approach, the trivial solution to the least squares problem is avoided by imposing a constraint on the Frobenius norm of the fundamental matrix. As an alternative, we can think of avoiding the zero solution by setting one of the entries of the fundamental matrix fixed to a value other than zero. This is equivalent to a parameterization of the fundamental matrix by the eight remaining entries of $\boldsymbol{f}$. For this type of solution, we implicitly assume that noise can only be present in the corresponding entry of the constraint vector $\boldsymbol{s}$. It is well known that this leads to an *ordinary least squares* (OLS) problem, which is solved via the corresponding normal equation [vV91, SB02]. The disadvantage of this approach with respect to the TLS technique, is that the entries of $\boldsymbol{f}$ do not play an equal role and that all nine possible normal equations have to be solved.

Solution via the Normal Equation. If we assume that fundamental matrix entry f_j, for any $1 \leq j \leq 9$, has been set to 1, we can denote by $\tilde{\boldsymbol{f}} := (f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_9)^\top$ the vector of the 8 remaining entries of $\boldsymbol{f}$. If we further denote by s_{ij} the j-th entry of the i-th constraint vector $\boldsymbol{s}_i$, for $1 \leq i \leq n$, we can express by $\tilde{\boldsymbol{s}}_i := (s_{i1}, \ldots, s_{i\,j-1}, s_{i\,j+1}, \ldots, s_{i9})^\top$ the constraint vector obtained by dropping the j-th entry of $\boldsymbol{s}_i$. Then the least squares energy (4.7) can be formulated in function of the parameter vector $\tilde{\boldsymbol{f}}$ as

$$\mathcal{E}(\tilde{\boldsymbol{f}}) = \sum_{i=1}^{n} (\tilde{\boldsymbol{s}}_i^\top \tilde{\boldsymbol{f}} + s_{ij})^2 = \|A\,\tilde{\boldsymbol{f}} - \boldsymbol{b}\|^2 \ . \tag{4.15}$$

Here, the $n \times 8$ matrix A equals the matrix S without the j-th column. The $n \times 1$ right hand side vector $\boldsymbol{b}$ collects the entries $-s_{ij}$, $1 \leq i \leq n$. The minimiser of this energy is now given by the solution of the classical normal equation

$$\tilde{\boldsymbol{f}} = (A^\top A)^{-1} A^\top \boldsymbol{b} \ . \tag{4.16}$$

Since all nine entries of the fundamental matrix can be chosen 1, a solution to all nine possible normal equations has to be found. This additionally prevents that a zero entry of F would be erroneously represented by a non-zero value. The OLS solution that produces a minimum value for the original energy (4.7) is finally retained.

4.1.2 Robust Estimation of the Fundamental Matrix

If we use image correspondences that have been established by an optical flow method, we expect that there will be moderate outliers that do not fully satisfy the epipolar constraint. Such outliers can be due to noise in the input data or due to effects that have not been modelled by the optical flow method, such as occlusions or multiplicative illumination changes. Since a Gaussian noise distribution with zero mean is assumed in a least squares framework, outliers can severely degrade the estimation process, which in turn may lead to a loss of accuracy. To reduce the effects of outlying data, we can apply the same strategy as for optical flow modelling and replace the squared penalisation by a sub-quadratic robust function. This will lead to a set of *non-linear* equations, the solution of which can be formulated as a reweighted least squares problem.

4.1.2.1 Reweighted Total Least Squares Solution

Basic Idea. To account for outliers in the correspondence data, we estimate the fundamental matrix with a robust version of the 8-point algorithm. This is achieved by replacing the quadratic penaliser in the TLS energy (4.7) by another function of the argument:

$$\mathcal{E}(\boldsymbol{f}) = \sum_{i=1}^{n} \Psi\left((\boldsymbol{s}_i^\top \boldsymbol{f})^2\right). \tag{4.17}$$

Here, $\Psi(s^2)$ is a positive, symmetric and in general convex function in s that grows sub-quadratically. In our applications we will set Ψ to the regularised L_1 norm (3.9). As a result, the least squares problem is transformed to the minimisation of the sum of absolute values. Our choice of robust function can be motivated by the fact that a minimiser of the sum of absolute differences can be associated to the median of the measurements [RL87], while a least squares solution is related to the mean value. As a statistical measure, it is well-known that the median is much less sensitive to outliers than the mean value.

Solution by Reweighted Least Squares. As in the previous section, we avoid the trivial solution via the explicit constraint $\|\boldsymbol{f}\|^2 = 1$ and minimise the energy (4.17) by looking for critical points of the Lagrangian

$$\mathcal{L}(\boldsymbol{f}, \lambda) = \sum_{i=1}^{n} \Psi\left((\boldsymbol{s}_i^\top \boldsymbol{f})^2\right) + \lambda(1 - \boldsymbol{f}^\top \boldsymbol{f}) \ . \tag{4.18}$$

The derivative of $\mathcal{L}(\boldsymbol{f}, \lambda)$ with respect to $\boldsymbol{f}$ can be written as

$$\nabla_{\boldsymbol{f}} \mathcal{L}(\boldsymbol{f}, \lambda) \ = \ \sum_{i=1}^{n} \Psi'\left((\boldsymbol{s}_i^\top \boldsymbol{f})^2\right) \nabla_{\boldsymbol{f}} (\boldsymbol{s}_i^\top \boldsymbol{f})^2 \ - \ 2\lambda \boldsymbol{f}, \tag{4.19}$$

$$= \ 2 \sum_{i=1}^{n} \Psi'\left((\boldsymbol{s}_i^\top \boldsymbol{f})^2\right) \begin{pmatrix} s_{i\,1}\ \boldsymbol{s}_i^\top \boldsymbol{f} \\ \vdots \\ s_{i\,9}\ \boldsymbol{s}_i^\top \boldsymbol{f} \end{pmatrix} - \ 2\lambda \boldsymbol{f}, \tag{4.20}$$

$$= \ 2 \Big(\sum_{i=1}^{n} \Psi'\left((\boldsymbol{s}_i^\top \boldsymbol{f})^2\right) \boldsymbol{s}_i \boldsymbol{s}_i^\top \boldsymbol{f} \qquad - \lambda \boldsymbol{f} \Big), \tag{4.21}$$

$$= 2\left(S^\top W(\boldsymbol{f})\, S - \lambda I\right)\boldsymbol{f} \ , \tag{4.22}$$

where s_{ij} is the j-th entry of $\boldsymbol{s}_i$, for $1 \leq j \leq 9$, and where we have defined the matrix $W(\boldsymbol{f})$ as the $n \times n$ diagonal matrix containing the positive non-linear expressions:

$$W(\boldsymbol{f}) = \mathrm{diag}\left(\Psi'\left((\boldsymbol{s}_1^\top \boldsymbol{f})^2\right), \ldots, \Psi'\left((\boldsymbol{s}_n^\top \boldsymbol{f})^2\right)\right) \ . \tag{4.23}$$

Critical points of the Lagrangian are thus solutions of the non-linear set of equations

$$\left(S^\top W(\boldsymbol{f})\, S - \lambda I\right)\boldsymbol{f} \ = \ \mathbf{0}^9, \tag{4.24}$$

$$1 - \|\boldsymbol{f}\|^2 \ = \ 0 \ . \tag{4.25}$$

This system of equations is very similar to the eigenvalue problem (4.12) - (4.13) of the 8-point algorithm discussed previously, only this time the system matrix $S^\top W(\boldsymbol{f})\, S$ itself is depending on the fundamental matrix $\boldsymbol{f}$. To find a solution, we propose an iterative scheme that is similar to the lagged-diffusivity method that is used in optical flow computation for handling the non-linearities in the Euler-Lagrange equations. To this end we decompose the non-linear system into a series of linear eigenvalue problems by means of a fixed-point iteration around the current solution. We can achieve this by fixing the system matrix for a current estimate of the fundamental matrix $\boldsymbol{f}^k$

$$\left(S^\top W(\boldsymbol{f}^k)\, S - \lambda I\right)\boldsymbol{f}^{k+1} \ = \ \mathbf{0}^9, \tag{4.26}$$

$$1 - \|\boldsymbol{f}^{k+1}\|^2 \ = \ 0 \ , \tag{4.27}$$

and seeking the update $\boldsymbol{f}^{k+1}$. Since $W(\boldsymbol{f}^k)$ is positive definite, the 9×9 system matrix $S^\top W(\boldsymbol{f}^k) S$ is symmetric positive definite with real non-negative eigenvalues. We can thus compute $\boldsymbol{f}^{k+1}$ via the 8-point algorithm, and use it in the next iteration step to refine the solution further. The weights in $W(\boldsymbol{f})$ are recomputed in every iteration step and therefore this approach comes down to a *reweighted total least squares* (RTLS) problem. Because the calculation of the weights requires an estimate of the fundamental matrix and vice versa, we use the standard 8-point algorithm to obtain an initial estimate.

4.1.2.2 Reweighted Ordinary Least Squares Solution

Basic Idea. In the same way as we did for the ordinary least squares approach in Sec. 4.1.1.2, we can parameterise the robust energy by eight of the nine entries of the fundamental matrix, while choosing the remaining entry fixed. To avoid representing a zero entry by a non-zero value, we have to evaluate the energy for all nine possible solutions.

Solution by Reweighted Normal Equations. Expressing the robust least squares energy (4.17) in function of the reduced parameter vector $\tilde{\boldsymbol{f}}$, we obtain

$$\mathcal{E}(\tilde{\boldsymbol{f}}) = \sum_{i=1}^{n} \Psi\left((\tilde{\boldsymbol{s}}_i^\top \tilde{\boldsymbol{f}} + s_{ij})^2\right). \tag{4.28}$$

To find a minimiser of $\mathcal{E}(\tilde{\boldsymbol{f}})$, we set its derivative with respect to $\tilde{\boldsymbol{f}}$ to zero. Keeping in mind that $\boldsymbol{f}$ only differs from $\tilde{\boldsymbol{f}}$ by a 1 in the j-th entry we can write

$$\Psi\left((\tilde{\boldsymbol{s}}_i^\top \tilde{\boldsymbol{f}} + s_{ij})^2\right) = \Psi\big((\boldsymbol{s}_i^\top \boldsymbol{f})^2\big) \quad \text{for } 1 \leq j \leq 9 \ , \tag{4.29}$$

such that the derivative becomes

$$\nabla_{\tilde{\boldsymbol{f}}}\,\mathcal{E}(\tilde{\boldsymbol{f}}) \;=\; \sum_{i=1}^{n}\Big(\Psi'\big((\boldsymbol{s}_i^\top \boldsymbol{f})^2\big)\,\tilde{\boldsymbol{s}}_i\tilde{\boldsymbol{s}}_i^\top\tilde{\boldsymbol{f}} + \Psi'\big((\boldsymbol{s}_i^\top \boldsymbol{f})^2\big)\,\tilde{\boldsymbol{s}}_i\; s_{ij}\Big), \tag{4.30}$$

$$\;=\; A^\top W(\boldsymbol{f})\,A\tilde{\boldsymbol{f}} - W(\boldsymbol{f})\,A^\top\boldsymbol{b}, \tag{4.31}$$

where the diagonal matrix $W(\boldsymbol{f})$ is the same as the one defined in equation (4.23). As in the reweighted total least squares case, we can fix the system matrix $A^\top W(\boldsymbol{f})\,A$ and the right hand side $W(\boldsymbol{f})\,A^\top\boldsymbol{b}$ for a current estimate of the fundamental matrix $\boldsymbol{f}^k$. We will designate this type of minimisation as a *reweighted ordinary least squares* (ROLS) solution. The ROLS update instruction for $\tilde{\boldsymbol{f}}^{\,k+1}$ is given by the normal equation solution

$$\tilde{\boldsymbol{f}}^{\,k+1} = (A^\top W(\boldsymbol{f}^k)\,A)^{-1}W(\boldsymbol{f}^k)\,A^\top\boldsymbol{b}\;. \tag{4.32}$$

4.1.3 Data Normalisation

The 8-point algorithm is not invariant to similarity transformations, such as translation, rotation and scaling of the image coordinates. At the same time, the eigenvalue problem of the TLS and the normal equation of the OLS are generally poorly conditioned because of the different orders of magnitude of the homogeneous point coordinates for typical image sizes. It is therefore advisable that all point correspondences are expressed in a fixed coordinate frame prior to the application of the linear methods discussed earlier. In [Har97], Hartley proposes a data *normalisation* procedure that transforms the corresponding points $\boldsymbol{x}_i$ and $\boldsymbol{x}'_i$, for $1 \le i \le n$, by two mappings T and T', such that the transformed points $T\boldsymbol{x}_{\mathrm{h}i}$ and $T'\boldsymbol{x}'_{\mathrm{h}i}$ have the projective coordinate $(1,1,1)^\top$ on average. In practice, T and T' are computed as a combination of a translation and a scaling, such that:

- The centroid of the transformed points comes to lie in the origin of the image.
- The average distance of the transformed points to the origin becomes $\sqrt{2}$.

Keeping this in mind, both transformations will then take on the form of a 3×3 matrix

$$T = \begin{pmatrix} s & 0 & s\,\bar{x} \\ 0 & s & s\,\bar{y} \\ 0 & 0 & 1 \end{pmatrix} \quad\text{and}\quad T' = \begin{pmatrix} s' & 0 & s'\,\bar{x}' \\ 0 & s' & s'\,\bar{y}' \\ 0 & 0 & 1 \end{pmatrix}\;. \tag{4.33}$$

Here $\bar{\boldsymbol{x}} = (\bar{x},\bar{y})^\top$ and $\bar{\boldsymbol{x}}' = (\bar{x}',\bar{y}')^\top$ are the centroids of the two point sets $\boldsymbol{x}_i$ and $\boldsymbol{x}'_i$, for $1 \le i \le n$, and s and s' are scale factors. While the coordinates of $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{x}}'$ denote the shifts in x- and y-direction that place the centroids in the origin, s and s' scale the images such that the distance of the transformed points to $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{x}}'$ becomes $\sqrt{2}$ on average. Such a normalisation step has been widely regarded as indispensable in combination with the linear estimation techniques presented in this section [HZ00, FLP01]. Because we experienced a dramatic gain in accuracy and stability for our robust extensions as well, we will from here on always perform a data normalisation of the optical flow correspondences prior to the fundamental matrix estimation via RTLS or ROLS.

To express the robust energy (4.17) in function of the normalised correspondence data, we rewrite the epipolar constraint as

$$\boldsymbol{x}'^{\top}_{\mathrm{h}} T'^{\top} \hat{F}\, T\, \boldsymbol{x}_{\mathrm{h}} = \hat{\boldsymbol{s}}^{\top} \hat{\boldsymbol{f}} = 0 \ , \tag{4.34}$$

where $\hat{\boldsymbol{f}}$ is the parameter vector for the fundamental matrix $\hat{F}$ of the transformed data and $\hat{\boldsymbol{s}}$ the constraint vector for the transformed data. Now $\hat{\boldsymbol{f}}$ can be found by minimising

$$\mathcal{E}(\hat{\boldsymbol{f}}) = \sum_{i=1}^{n} \Psi\left((\hat{\boldsymbol{s}}_i^{\top} \hat{\boldsymbol{f}})^2\right), \tag{4.35}$$

either in a RTLS or ROLS framework. A fundamental matrix for the original non-normalised data can be recovered as $F = T'^{\top}\, \hat{F}\ T$. Hartley noted, however, that if $\hat{F}$ is the minimiser of the *transformed energy* 4.35 in a least squares sense, $T'^{\top}\ \hat{F}\ T$ is not necessarily the minimiser of the original energy (4.17) in a least squares sense.

4.1.4 Enforcing the Rank 2 Constraint

A second issue posed by the methods described in this section concerns the rank of F. The solution of a least squares fit will in general not satisfy the singularity constraint, such that it is common to perform a *rank enforcement* step after the estimation. This can be done, for instance, by replacing the final solution with the closest rank 2 approximation in Frobenius norm [TH84]. It can be shown that this comes down to setting the smallest singular value of F to zero. Thus, if $U\,\mathrm{diag}\,(\sigma_1, \sigma_2, \sigma_3)\,V^{\top}$, with $\sigma_1 \geq \sigma_2, \geq \sigma_3 \geq 0$, is the SVD of a least squares solution, the closest rank two estimate is given by

$$F = U\,\mathrm{diag}\,(\sigma_1, \sigma_2, 0)\,V^{\top} \ . \tag{4.36}$$

If the estimation of the fundamental matrix is based on normalised correspondences data, the rank enforcement is performed before the denormalisation step.

4.2 Feature Based Methods for Comparison

We will compare the estimation of the fundamental matrix from optical flow with up to twelve variants of feature based techniques that are frequently encountered in literature. These feature based methods extract a sparse set of n distinctive point pairs and minimise an energy of the general form

$$\mathcal{E}(F) = \sum_{i=1}^{n} d_i^2 \ , \tag{4.37}$$

where the *distance measure* d can be interpreted as a geometrical distance in the image plane. Based on these distance measures, we divide the feature based methods in two different classes. The first class minimises the distance of a point to its corresponding epipolar line, while the second class minimises the so-called reprojection error in a Maximum Likelihood (ML) framework. We will refer to the first class of feature based technique as method class F1 and to the second class of feature based techniques as F2. In the following we will give an overview of the different steps that make up these two classes.

4.2.1 Feature Extraction and Matching

Under *feature or interest point extraction*, one traditionally understands the selection of a set of salient image locations with distinctive neighbourhood information. Classical examples include edges [MH80, Can86] and corners [HS88]. Once a certain number of features has been extracted in both images, correspondences must be established between them. This can for instance be achieved by thresholding the correlation score of the image patches surrounding the features. As opposed to dense optical flow, the set of correspondences found this way is *sparse* and often ambiguous since one point might be paired to several other points [ZDFL95]. Simple correlation based methods are sensitive to viewpoint changes and deformations and are characterised by a low matching repeatability. Therefore, recent techniques make use of a so called *descriptor*: a high-dimensional vector representation of the spatial neighborhood of the point of interest. Matching is then based on a comparison of the Euclidian or the Mahalanobis distance [Mah36] in descriptor space. To obtain a sparse set of feature correspondences for our comparison, we apply two widely used feature matching algorithms which will be discussed here.

4.2.1.1 The Scale Invariant Feature Transform

The first feature matching algorithm that we consider is the *Scale Invariant Feature Transform* (*SIFT*) [Low99, Low04]. SIFT features are characterised by a sub-pixel location and are detectable under changes in location, scale and orientation. To this end, SIFT identifies locations of interest in scale-space [Wit83, Lin94] and associates with each of them a high dimensional descriptor vector. This descriptor representation is based on normalized gradient information of a local image patch surrounding the feature and is designed to be invariant with respect to scale and rotation and partially invariant under affine distortions and illumination changes. This results in a set of distinctive image features that can be matches correctly between two or more images with high probability. Matching of SIFT features is mostly done by means of a nearest neighbour search in descriptor space. To avoid false matches, the distance to the closest neighbour is compared with the distance to the second closest neighbour. The rationale behind this is that a correct match will have its closest neighbour significantly closer than the closest incorrect match; a feature with no correct match, on the other hand, will have several neighbours that are comparably close in descriptor space. Comparative studies by Mikolajczyk and Schmid [MS05] have put forward SIFT as one of the most accurate local matching algorithms to date.

4.2.1.2 The Kanade-Lucas-Tomasi Tracker

While SIFT has become an accepted standard for stereo matching and object recognition, it may be outperformed in small-baseline scenarios by methods that are specifically tailored towards small displacements. To account for such cases, we consider as a second feature matching algorithm the *Kanade-Lucas-Tomasi Tracker (KLT)*, [LK81, TK91]. The KLT algorithm looks for local maxima of the eigenvalues of the structure tensor [FG87] and tries to detect the same features in the second image by minimising the intensity difference over a small local neighbourhood. While KLT feature extraction is closely related to the detection of classical points of interest, such as Harris corners [HS88], the tracking is essentially done by means of the sparse Lucas-Kanade optical flow method [LK81].

4.2.2 Inlier Selection and Initialisation

The number of feature correspondences that is returned by a matching algorithm is generally controlled by thresholding a quality measure, such as the distance ratio between the first and the second nearest neighbour for SIFT descriptors and the smallest eigenvalue of the structure tensor for KLT features. Choosing the threshold less strict often guarantees a larger number of tentative correspondences, but at the same time increases the portion of false matches that have an adverse effect on the estimation of the fundamental matrix. Two robust techniques that are frequently used in computer vision to reduce the influence of such gross outliers are the *Random Sampling Consensus* (*RANSAC*) [FB81] and the *Least Median of Squares* (*LMedS*) [RL87]. In contrast to other robust methods that include as many correspondences as possible, RANSAC and LMedS repeatedly estimate the fundamental matrix from randomly sampled minimal data sets in a so-called *hypothesize-and-verify* framework. Since the fundamental matrix has only 7 degrees of freedom, the size of the minimal correspondence set is 7. To generate a hypothesis, the 7-point algorithm [HZ00] is used which finds the rank 2 matrix that satisfies the cubic constraint [1]

$$\det\left(\alpha F_1 + (1-\alpha)F_2\right) = 0 \ . \tag{4.38}$$

Here, F_1 and F_2 are the fundamental matrices obtained from the two smallest right singular vectors $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ of the 7×9 system matrix (4.6).

4.2.2.1 Selection Criterion and Sample Count

In the case of RANSAC, the minimal correspondence set is enlarged by including correspondences that are consistent with the estimated fundamental matrix, the so called consensus set. This is done by computing the distances d for all n matches, and comparing it to an *inlier threshold*:

$$\begin{cases} \text{inlier} & \text{if } d_i^2 < t^2 \\ \text{outlier} & \text{if } d_i^2 \geq t^2 \end{cases} \quad \text{for } 1 \leq i \leq n \ . \tag{4.39}$$

The threshold t is chosen such that with a certain probability a data point is an inlier. If certain assumptions are made about the probability distribution of the distances, t can be expressed in terms of the standard deviation σ of the set of inliers. Since neither the true set of inliers, nor the correct fundamental matrix are known, an estimate of σ is often not available [2], such that t is often chosen empirically in practice. RANSAC ultimately returns as best solution the fundamental matrix that gives the largest consensus set.

LMedS on the other hand, returns the fundamental matrix for which the median of the squared distances is smallest over all random samples. To determine the set of inliers, criterion (4.39) is applied for $t = \sqrt{3.84}\,\sigma_{\mathrm{r}}$, where the robust standard deviation σ_{r} is estimated from the distance median as proposed by Rousseeuw and Leroy [RL87]

$$\sigma_{\mathrm{r}} = 1.4826\,(1 + 5/(n-8))\,\underset{i}{\mathrm{median}}|d_i| \ . \tag{4.40}$$

1. The 7-point algorithm thus generates three hypotheses per sample that have to be verified by RANSAC or LMedS.

2. The robust estimation of the standard deviation itself could be done by means of a random sampling technique, such as LMedS.

The number of random samples k that have to be drawn, is generally chosen such that with a probability of $p = 0.99$, at least one of the samples is free from outliers. If we denote by o the fraction of outliers in the data, the minimum number of samples of size 7 that has to be drawn can then be computed as [HZ00, FLP01]

$$k = \log(1-p)/\log\left(1-(1-o)^7\right) \quad . \tag{4.41}$$

Mostly, the true value of o is not known and k can not be computed beforehand. The true fraction of outliers can, however, be estimated adaptively during the sampling process as the ratio of the current number of outliers and the total number of correspondences. Every time a consensus set is found that is larger than the previously largest one, the value of o becomes smaller and k is reduced according to Eq. (4.41). The random sampling algorithm ends as soon as the number of samples taken is equal or greater than k.

4.2.2.2 RANSAC Extensions

In the last decade several extensions have been proposed to improve the performance of the random sampling algorithms. Hereby the focus has been primarily on RANSAC due to its capability of dealing with a large proportion of mismatches. To incorporate the current state of the art in the field, we consider in this thesis two such RANSAC extensions.

LORANSAC. The first extension has been proposed by Chum *et al.* [CMK03, CMO04] to reduce the influence of noise in the correspondence data and to simultaneously achieve a run-time that is closer to the theoretically predicted number of samples that has to be drawn (as given by Eq. (4.41)). It consists of performing a *local optimisation* (*LO*) step for each estimated fundamental matrix that has a larger support than all hypotheses generated so far. The LO-step comes down to applying an inner RANSAC loop that draws a fixed number of samples (20) from the current set of inliers. Since the samples are only drawn from correspondences that have been labeled as inliers, the chance is high that an improved hypothesis will be generated. The LO-step then returns this high quality fundamental matrix to the standard RANSAC algorithm, which will likely meet its termination criterion more rapidly. Since the LO-step is only invoked if a new maximum in the number of inliers is reached, it has hardly any impact on the overall run-time of RANSAC.

DEGENSAC. From the same authors comes a second extension to classical RANSAC that helps overcome the ambiguity that can occur in the estimation process when the majority of points lie on a dominant plane. In [CWM05] Chum *et al.* show that if five or more correspondences in a sample of size 7 are related by a planar homography (a projective mapping between two planes), the estimated epipolar geometry might be incorrect, yet consistent with a high number of correspondences. For such *degenerate configurations*, the *DEGENSAC* algorithm simultaneously estimates a fundamental matrix and a planar homography and uses model selection to choose a correct solution. More specifically, this model selection is performed by taking samples from the correspondences that are outliers to the estimated homography and computing the fundamental matrix from this outlier sample and the homography by means of the plane-and-parallax algorithm [HZ00].

Both LORANSAC and DEGENSAC extensions are implemented as nested RANSAC loops and can be combined in a standard RANSAC loop for improved robustness.

4.2.3 Minimisation of a Geometrical Distance Measure

For the estimation of the fundamental matrix from optical flow in Sec. 4.1, the distance measure in the general energy (4.37) took on the form of the *algebraic distance* [HZ00]

$$d = \boldsymbol{s}^\top \boldsymbol{f} \quad . \tag{4.42}$$

While the algebraic distance clearly has the advantage of being linear in the fundamental matrix entries, it does not represent a meaningful distance in the image plane. Moreover, it depends on the scaling of the fundamental matrix and its variance depends on the location of the feature correspondences in the image [TM97]. Because these properties are generally not regarded as beneficial for feature based estimation, alternative measures have been developed over time representing geometrically meaningful distances in the image. Two such distance measures and their minimisation will be discussed next in more detail. At the end of the section we conclude with a summary of all distance measures that are covered in this chapter. This overview can be found in Table 4.1 on page 72.

4.2.3.1 Method Class F1: Minimisation of the Epipolar Distance.

Basic Idea. As a geometrically meaningful distance measure that does not depend on the scale of F, Luong and Faugeras [LF96] and Faugeras *et al.* [FLP01] propose to minimise the squared *epipolar distance* over all inliers $(\boldsymbol{x}_i, \boldsymbol{x}'_i)$, $1 \leq i \leq n$:

$$\mathcal{E}_{\mathrm{F1}}(F) = \sum_{i=1}^{n} \left(d^2(\boldsymbol{x}_{\mathrm{h}i}, F^\top \boldsymbol{x}'_{\mathrm{h}i}) + d^2(\boldsymbol{x}'_{\mathrm{h}i}, F\boldsymbol{x}_{\mathrm{h}i}) \right) \quad . \tag{4.43}$$

Here, $d(\boldsymbol{x}_{\mathrm{h}}, \boldsymbol{l})$ denotes the Euclidean (orthogonal) distance between a point $\boldsymbol{x}$ and a line $\boldsymbol{l}$ in the image plane. The epipolar distance measures how far a point $\boldsymbol{x}$ lies from the epipolar line of the corresponding point $\boldsymbol{x}'$. Since the epipolar lines have to be considered in both the left and the right image, definition (4.43) ensures that the epipolar distance measure is symmetric with respect to $\boldsymbol{x}$ and $\boldsymbol{x}'$. The two parts of the epipolar distance are illustrated in Fig. 4.1. Ideally, all feature points should satisfy the epipolar constraint, but due to measurement noise these points do not lie exactly on each others epipolar lines. Minimising energy (4.43) expresses the objective of finding the fundamental matrix for which the corresponding point-line distance is smallest.

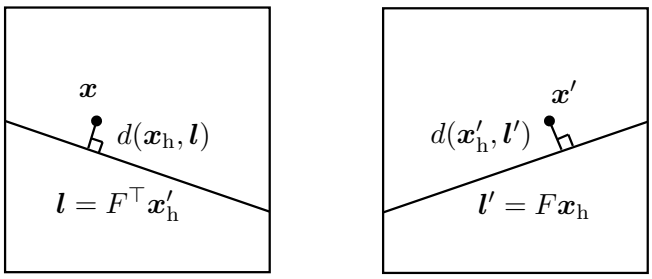

Fig. 4.1: The distance to the epipolar line in the left image and the right image.

Solution by Reweighted Least Squares. To minimise energy (4.43) with respect to the entries of F, we rewrite the epipolar distance as a local weighing of the algebraic distance. We minimise the reformulated energy

$$\mathcal{E}_{\mathrm{F1}}(\boldsymbol{f}) = \sum_{i=1}^{n} w_i^2(\boldsymbol{f})\,(\boldsymbol{s}_i^\top \boldsymbol{f})^2, \tag{4.44}$$

with the weight $w(\boldsymbol{f})$ defined as

$$w(\boldsymbol{f}) := \left(\frac{1}{a^2+b^2} + \frac{1}{a'^2+b'^2}\right)^{1/2} \quad . \tag{4.45}$$

In this definition, the epipolar line coefficients a and b are the first two coordinates of the epipolar line $\boldsymbol{l} = F^\top \boldsymbol{x}'_{\mathrm{h}}$ in the left image and the coefficients a' and b' the first two coordinates of the epipolar line $\boldsymbol{l}' = F\boldsymbol{x}_{\mathrm{h}}$ in the right image:

$$a := (F^\top \boldsymbol{x}'_{\mathrm{h}})_1, \quad b := (F^\top \boldsymbol{x}'_{\mathrm{h}})_2, \quad a' := (F\boldsymbol{x}_{\mathrm{h}})_1 \quad \text{and} \quad b' := (F\boldsymbol{x}_{\mathrm{h}})_2 \ . \tag{4.46}$$

It is clear from expression (4.44), that the epipolar distance does not depend on the scaling of F, because the same scaling factor is contained in the coefficients a, b, a' and b'.

As proposed by Torr and Murray [TM97], we minimise the energy (4.44) subject to the constraint $\|\boldsymbol{f}\|^2 = 1$. By keeping the weights $w_i(\boldsymbol{f})$, $1 \le i \le n$, constant for a given $\boldsymbol{f}$, it can be easily verified that we obtain a familiar RTLS problem, as discussed in Sec. 4.1.2. The solution thus comes down to iteratively solving the eigenvalue problem

$$\left(S^\top W(\boldsymbol{f})\, S - \lambda I\right) \boldsymbol{f} \;=\; \mathbf{0}^9, \tag{4.47}$$

$$1 - \|\boldsymbol{f}\|^2 \;=\; 0 \ , \tag{4.48}$$

where the $n \times n$ diagonal matrix $W(\boldsymbol{f})$ contains the weights $w_i^2(\boldsymbol{f})$.

We further reduce the effects of remaining outliers in the data by including a statistical weighing of the epipolar distance by the tri-weight function as proposed by Huber in the context of M-estimators [Hub81, FLP01, TM97, Ste99]. To this end, we multiply each weight w^2 in $W(\boldsymbol{f})$ by the additional M-estimator weight

$$h\left(w\,\boldsymbol{s}^\top \boldsymbol{f}\right) \quad . \tag{4.49}$$

The tri-weight function h is defined as

$$h(s) = \begin{cases} 1 & |s| \le \sigma_{\mathrm{r}} \\ \sigma_{\mathrm{r}}/|s| & \sigma_{\mathrm{r}} < |s| \le 3\sigma_{\mathrm{r}} \\ 0 & 3\sigma_{\mathrm{r}} < |s| \end{cases} \ , \tag{4.50}$$

where the robust standard deviation σ_{r} is estimated via equation (4.40). Since the M-estimator requires a good estimate of the fundamental matrix to start with, we initialise our implementation of the feature based method F1 with the estimate provided by the random sampling algorithm. The rank of the final solution is enforced by SVD.

4.2.3.2 Method Class F2: Minimisation of the Reprojection Error.

Basic Idea. As a second important geometrical distance measure for the estimation of the fundamental matrix, Hartley and Zisserman [HZ00] propose to minimise the *reprojection error* over all n inliers:

$$\mathcal{E}_{\text{F2}}(P', \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \sum_{i=1}^{n} \left(d^2(\boldsymbol{x}_{\text{h}i}, P\boldsymbol{X}_{\text{h}i}) + d^2(\boldsymbol{x}'_{\text{h}i}, P'\boldsymbol{X}_{\text{h}i}) \right) \quad , \tag{4.51}$$

where $d(\boldsymbol{x}_{\text{h}}, \boldsymbol{x}'_{\text{h}})$ denotes the Euclidean distance between two inhomogeneous points $\boldsymbol{x}$ and $\boldsymbol{x}'$ in the image plane. In the above definition, P and P' denote the 3×4 camera projection matrices for the left and the right image and $\boldsymbol{X}_i$, $1 \leq i \leq n$, are the 3D points that are reconstructed from the matched feature pairs $(\boldsymbol{x}_i, \boldsymbol{x}'_i)$. The reprojection error describes the distance of a point $\boldsymbol{x}$ in the left image to a *corrected* point location $P\boldsymbol{X}_{\text{h}}$ that has a perfect match $P'\boldsymbol{X}_{\text{h}}$ in the right image. Definition (4.51) further ensures that the distance measure is symmetric with respect to $\boldsymbol{x}$ and $\boldsymbol{x}'$. The two parts of the reprojection error are illustrated in Fig. 4.2. The projections $P\boldsymbol{X}_{\text{h}}$ and $P'\boldsymbol{X}_h$ can be regarded as the most likely true positions of the feature points $\boldsymbol{x}$ and $\boldsymbol{x}'$ if the errors in the localisation are assumed to be independent and Gaussian distributed [WAH93]. The Maximum Likelihood estimate of the fundamental matrix is then given by the rank 2 matrix F for which the corrected point locations satisfy the epipolar constraint exactly:

$$(P'\boldsymbol{X}_{\text{h}i})^\top F\,(P\boldsymbol{X}_{\text{h}i}) = 0 \quad \text{for} \quad 1 \leq i \leq n \ . \tag{4.52}$$

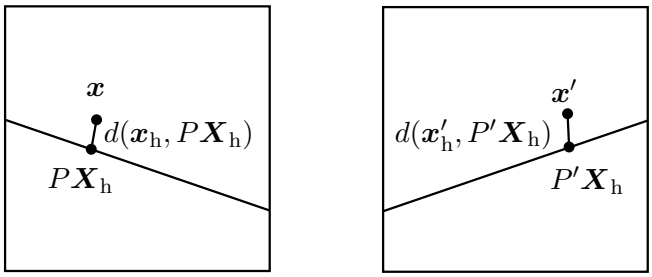

Fig. 4.2: The distance to the corrected point locations in the left image and the right image.

Solution via Non-linear Optimisation. We recall that the fundamental matrix can be written in terms of a pair of canonical camera matrices (2.39) via the equality (2.20). As a results, we can fix the left camera matrix in the world origin as $P = (I^3, \mathbf{0}^3)$ and parameterise F by the 12 entries of the right camera matrix P'. By construction, this parameterisation of F ensures the singularity constraint. Together with the 3 degrees of freedom of the reconstructed 3D points, it brings the total number of variables to $3n+12$.

To minimise the energy (4.51), we simplify the notation by collecting all $3n+12$ unknowns in a parameter vector $\boldsymbol{p}$. which is composed of the 12 motion parameters P' and the $3n$ structure parameters $\boldsymbol{X}_i$, $1 \leq i \leq n$. We then consider the function $f : \mathbb{R}^{(3n+12)} \rightarrow \mathbb{R}^{4n}$ that maps the parameter vector onto the inhomogeneous corrected point coordinates

$$f : \boldsymbol{p} \rightarrow \left((P\boldsymbol{X}_{\text{h}1})_1, (P\boldsymbol{X}_{\text{h}1})_2, (P'\boldsymbol{X}_{\text{h}1})_1, (P'\boldsymbol{X}_{\text{h}1})_2, \ldots, (P'\boldsymbol{X}_{\text{h}n})_1, (P'\boldsymbol{X}_{\text{h}n})_2 \right) \ . \tag{4.53}$$

If we define the $4n \times 1$ vector $\boldsymbol{b} = (\boldsymbol{x}_1, \boldsymbol{x}'_1, \ldots, \boldsymbol{x}_n, \boldsymbol{x}'_n)^\top$ of feature measurements, we can rewrite the energy (4.51) as a squared error norm

$$\mathcal{E}_{\text{F2}}(\boldsymbol{p}) = \|f(\boldsymbol{p}) - \boldsymbol{b}\|^2 \ . \tag{4.54}$$

In contrast to the epipolar distance, this error norm is highly nonlinear in the unknowns and its minimisation can not be formulated as a reweighted least squares problem. Instead we have to proceed by means of a non-linear optimisation method such as the iterative Levenberg-Marquardt (LM) technique [Lev44, Mar63].

The Levenberg-Marquardt algorithm smoothly shifts between a gradient descent method, that can always be applied far from the minimum, and a Gauss-Newton method, that assures fast convergence in a small neighbourhood. This is achieved by *augmenting* the Hessian of $f(\boldsymbol{p})$ with a factor λ that controls the transition between these two extremes. More concretely, we iteratively solve an augmented normal equation [PTVF92]

$$(A^\top A + \lambda\, I^{3n+12})\, d\boldsymbol{p} = -A^\top (f(\boldsymbol{p}) - \boldsymbol{b}) \ , \tag{4.55}$$

for an increment $d\boldsymbol{p}$, where A denotes the Jacobian of $f(\boldsymbol{p})$ and $A^\top A$ an approximation of the Hessian. After computing the increment $d\boldsymbol{p}$, the energy $\mathcal{E}_{\text{F2}}(\boldsymbol{p}+d\boldsymbol{p})$ is evaluated and λ is increased or decreased by a factor of 10 depending on whether the energy has increased or not. The method terminates when the energy has dropped by some small fractional amount like 10^{-3}. In the case of the reprojection error, the Jacobian A has a sparse block structure consisting of a 4×12 and a 4×3 block for each of the n corrected point pairs. This is due to the fact that a corrected point pair $(P\boldsymbol{X}_{\text{h}i}, P'\boldsymbol{X}_{\text{h}i})$ only depends on the parameters P' and $\boldsymbol{X}_{\text{h}i}$, but not on the $n-1$ other 3D points. This block structure can be exploited to speed up the computation by means of a partitioned variant of the Levenberg-Marquardt algorithm. For our specific implementation we have based ourselves on the fast sparse Levenberg-Marquardt algorithm outlined in detail in [HZ00].

As we did in the case of method F1, we reduce the effects of remaining outliers in the correspondences, by including an additional weighting by the tri-weight function (4.50). To this end, we multiply the blocks in $A^\top A$ that correspond to $(P\boldsymbol{X}_{\text{h}i}, P'\boldsymbol{X}_{\text{h}i})$ with

$$h\left(\left(d^2(\boldsymbol{x}_{\text{h}i}, P\boldsymbol{X}_{\text{h}i}) + d^2(\boldsymbol{x}'_{\text{h}i}, P'\boldsymbol{X}_{\text{h}i})\right)^{1/2}\right) \ . \tag{4.56}$$

A First-Order approximation to the Reprojection Error. An approximation to the reprojection error was introduced by Weng *et al.* [WHA89], adapted from the work of Sampson on conic fitting [Sam82]. This distance measure is known as the Sampson error and it can be written as a weighing of the algebraic distance by

$$w(\boldsymbol{f}) := \left(\frac{1}{a^2 + b^2 + a'^2 + b'^2}\right)^{1/2} \ . \tag{4.57}$$

The epipolar line coefficients a, b, a' and b' are defined as in Eq. (4.46). The Sampson error is a well-established distance measure in computer vision, but it will not be considered in this chapter since we have experienced that its performance is mostly bounded by that of methods F1 and F2. It is, however, mentioned at this point because it will be derived from the epipolar constraint via a novel normalisation strategy in Chapter 6.

Tab. 4.1: Overview of the distance measures discussed in this chapter.

Method	distance measure d
optical flow based	$\boldsymbol{x}_{\mathrm{h}}'^{\top} F \boldsymbol{x}_{\mathrm{h}}$
F1 [LF96]	$w(F)\left(\boldsymbol{x}_{\mathrm{h}}'^{\top} F \boldsymbol{x}_{\mathrm{h}}\right)$ with $w(F) := \left(\frac{1}{a^2+b^2} + \frac{1}{a'^2+b'^2}\right)^{1/2}$
F2 [HZ00]	$\left(d^2(\boldsymbol{x}_{\mathrm{h}}, \tilde{\boldsymbol{x}}_{\mathrm{h}}) + d^2(\boldsymbol{x}_{\mathrm{h}}', \tilde{\boldsymbol{x}}_{\mathrm{h}}')\right)^{1/2}$ with $\tilde{\boldsymbol{x}}_{\mathrm{h}} := P\boldsymbol{X}_{\mathrm{h}}$ and $\tilde{\boldsymbol{x}}_{\mathrm{h}}' := P'\boldsymbol{X}_{\mathrm{h}}$
Sampson [WHA89]	$w(F)\left(\boldsymbol{x}_{\mathrm{h}}'^{\top} F \boldsymbol{x}_{\mathrm{h}}\right)$ with $w(F) := \left(\frac{1}{a^2+b^2+a'^2+b'^2}\right)^{1/2}$

4.2.4 Inlier Refinement

After the fundamental matrix has been estimated from the set of inliers via method F1 or F2, we reclassify the correspondences in inliers and outliers by a similar selection criterion as used by RANSAC and LMedS. This is done by applying criterion (4.39) for the current fundamental matrix estimate. The threshold t is chosen to be $\sqrt{3.84}\,\sigma_{\mathrm{r}}$ [HZ00, TM97], where the robust standard deviation σ_{r} is computed via (4.40). Re-estimation of the fundamental matrix from the new set of inliers and reclassification of the correspondences is repeated until the final number of inliers converges.

4.3 Evaluation of the Optical Flow Based Method

In this experimental section we will compare the performance of our dense optical flow based method for the estimation of the fundamental matrix with the two sparse feature based method classes F1 and F2. In six different tests we compute the epipolar geometry of real-world image pairs that have been selected from two online multiview stereo databases. All images in these databases have been calibrated in advance by conventional techniques (based on a planar calibration grid and manually selected feature points) such that the ground truth fundamental matrix is known for all image pairs.

Error Measure. To assess the quality of our results, we evaluate the symmetric error between the estimated fundamental matrix F_{e} and the ground truth fundamental matrix F_{g} according to [Zha98] and [FLP01]. This error measure is computed as follows:

1. In the left image: Choose a random point $\boldsymbol{x}$ as depicted in Fig. 4.3.

2. In the right image: Determine the epipolar line $\boldsymbol{l}_{\mathrm{e}}'$ of $\boldsymbol{x}$ using F_{e} and the epipolar line $\boldsymbol{l}_{\mathrm{g}}'$ of $\boldsymbol{x}$ using F_{g}. Choose a random point $\boldsymbol{x}_{\mathrm{e}}'$ on $\boldsymbol{l}_{\mathrm{e}}'$ and a random point $\boldsymbol{x}_{\mathrm{g}}'$ on $\boldsymbol{l}_{\mathrm{g}}'$.

3. In the left image: Reverse the roles of F_{e} and F_{g} by computing the epipolar line $\boldsymbol{l}_{\mathrm{e}}$ of $\boldsymbol{x}_{\mathrm{g}}'$ using F_{e} and the epipolar line $\boldsymbol{l}_{\mathrm{g}}$ of $\boldsymbol{x}_{\mathrm{e}}'$ using F_{g}.

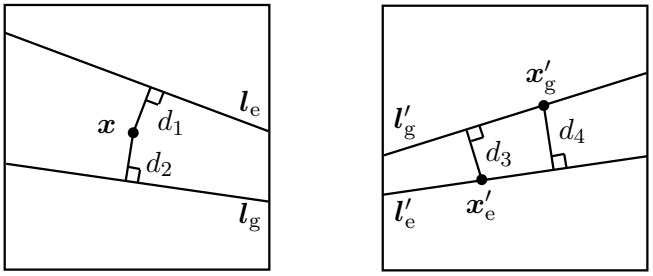

Fig. 4.3: The distances used in the computation of d_F.

4. Compute the four distances depicted in Fig. 4.3: $d_1 = d(\boldsymbol{x}, \boldsymbol{l}_\mathrm{e})$, $d_2 = d(\boldsymbol{x}, \boldsymbol{l}_\mathrm{g})$, $d_3 = d(\boldsymbol{x}'_\mathrm{e}, \boldsymbol{l}'_\mathrm{g})$ and $d_4 = d(\boldsymbol{x}'_\mathrm{g}, \boldsymbol{l}'_\mathrm{e})$
5. Repeat step 1 to 4 for a large number (100000) of times and compute the average of all distances.

This error measure, which we will denote by d_F from now on, is a symmetric distance that describes the average deviation of the estimated epipolar geometry from the ground truth geometry in pixel units. Because it is defined in image space, d_F is more meaningful than comparing for instance the Frobenius norms of F_e and F_g [FLP01].

Feature Based Implementation. All together, we consider for the feature based approaches two feature matching algorithms (KLT and SIFT), three random sampling algorithms (LMedS, LORANSAC and DEGENSAC), and two different distance measures (the epipolar distance of method class F1 and the reprojection error of method class F2). *This comes down to twelve different variants of feature based methods.* The implementational details for the feature extraction and random sampling are given below.

- *Feature Extraction.* For the SIFT feature extraction we use the implementation of David Lowe [3]. We only consider effective SIFT matches by removing feature pairs that have the same image locations but different histogram orientations. For the extraction of KLT features, we use publicly available code [4] that is based on the affine tracking algorithm described in [ST94]. SIFT and KLT are applied to all image test pairs, but we only list errors for the feature type for which we obtain the best result in combination with the distance measures minimised by F1 and F2.
- *Random Sampling.* Similarly, we either apply RANSAC or LMedS for the inlier selection, depending on which random sampling technique gives the best result in combination with SIFT or KLT and in combination with the distance measures minimised by F1 and F2. For both random sampling algorithms, we choose a minimal sample size of 7 and use the 7-point algorithm to generate hypotheses for F [HZ00]. By taking into account a certain radius around the feature points during sampling, we additionally assure that all points in a sample lie scattered enough over the whole image such that unstable estimates are avoided. For RANSAC, we estimate the number

3. available at `http://www.cs.ubc.ca/~lowe/keypoints/`
4. available at `http://www.ces.clemson.edu/~stb/klt/`

of samples adaptively, as discussed in Sec. 4.2.2.1, and use a fixed inlier threshold t between 0.5 and 1 pixels. This threshold works well in practice because the distance measures that are minimised by F1 and F2 can be interpreted as geometrical distances in the image plane. If degeneracy is suspected, we apply the DEGENSAC variant of RANSAC. Both standard RANSAC and DEGENSAC are equipped with a LO-step for local model optimisation. The number of samples taken by LMedS normally requires an estimate of the proportion of outliers. Following [FLP01], choosing a number of 2000 samples assures us that we can deal with the maximum percentage of outliers in LMedS [5]. Due to the random nature of both RANSAC and LMedS, we run the feature based method classes F1 and F2 for 100 consecutive times and present the minimum and mean for the error d_F over all test runs.

Optical Flow Based Settings. Our evaluation includes indoor sequences that have been captured under lab conditions, as well as outdoor sequences with varying illumination and large relative motion. All images are in colour and have been corrected for radial distortion. The indoor image pairs depict a bright object against a homogeneous black background. Because the optical flow is less reliable in the background, we exclude it from our optical flow based estimation method by detecting the object silhouette with a region-based Chan-Vese segmentation technique [CV01]. For a fair comparison with the feature based methods, we run our optical flow based method for a set of optimised parameters and for a set of *default* parameters that are given by $\alpha = 20.0$, $\gamma = 20.0$ and $\sigma = 0.9$. The optimal parameters are chosen for each image pair separately to give the lowest error d_F over a wide range of settings. To visualise the estimated epipolar geometries, we draw the epipolar lines for a set of meaningful points in the left and the right image. For the left image these are 8 feature points from the set of inliers that corresponds to the best feature based result. For the right image these are either the corresponding features or the pixel locations warped by the optical flow.

4.3.1 Low Texture

In our first experiment we investigate the influence of texture on the estimation of the epipolar geometry from image correspondences. This is done by the example of frames 24 and 25 of the *DinoRing* sequence [6] from the Middlebury multiview evaluation data set [SCD+06]. Both images have a resolution of 640×480, with a maximum displacement of 26 pixels. The camera rotates around the model of a small dinosaur set against a black background and the scene is characterised by the absence of texture. The results that are obtained by our optical flow based method and by the feature based methods F1 and F2 are summarised in Table 4.4. For the optical flow based estimation we list the results for both the RTLS solution and the ROLS solution. In the following we examine these results in more detail:

5. The maximum fraction of outliers for LMedS is 50%, because for a larger fraction the median distance would belong to an outlier. That 2000 iterations is more than enough to deal with the maximum fraction of outliers can be easily verified by means of formula (4.41) by setting $p = 0.99$ and $o = 0.5$.

6. available at `http://vision.middlebury.edu/mview/data/`

Tab. 4.2: Overview of the settings for the feature extraction corresponding to the best result for each image pair. We list the type of feature used (SIFT/KLT), the ratio of distances of the first and second neighbour in SIFT descriptor space (*ratio*) and the total number of matches (*# match*). For KLT we used the standard settings of the publicly available code.

Image Pair		Feature Extraction		
sequence	*frames*	*type*	*ratio*	*# match*
DinoRing	24 - 25	KLT	-	919
Entry-P10	1 - 0	SIFT	0.80	979
TempleRing	13 - 14	SIFT	0.90	627
Fountain-P11	1 - 2	SIFT	0.90	667
Herz-Jesu-P25	5 - 6	SIFT	0.90	945
City-Hall	1 - 2	SIFT	0.90	1502

Tab. 4.3: Overview of the settings for the feature based method classes F1 and F2 corresponding to the best result for each image pair. We list the type of random sampling algorithm (*randsam*), the applied RANSAC threshold t (*thresh*) and the number of inliers (# inl) corresponding to the best result over 100 test runs.

Image Pair		F1			F2		
sequence	*frames*	*randsam*	*thresh*	*# inl*	*randsam*	*thresh*	*# inl*
DinoRing	24 - 25	LORANSAC	0.5	697	LORANSAC	0.5	688
Entry-P10	1 - 0	DEGENSAC	1.0	733	DEGENSAC	1.0	728
TempleRing	13 - 14	LORANSAC	0.8	464	LORANSAC	0.8	463
Fountain-P11	1 - 2	LORANSAC	1.0	445	LORANSAC	1.0	446
Herz-Jesu-P25	5 - 6	LMedS	-	662	LMedS	-	663
City-Hall	1 - 2	LORANSAC	1.0	1100	LORANSAC	1.0	1094

Tab. 4.4: Overview of the error d_F for our optical flow based method and for the feature based method classes F1 and F2 over 100 test runs. The best results are highlighted.

Image Pair	Optical Flow Based Method		Feature Based Methods	
	RTLS	ROLS	F1	F2
	default (opt)	*default (opt)*	*mean (min)*	*mean (min)*
DinoRing	**0.717 (0.245)**	0.724 (0.348)	4.398 (0.422)	3.928 (0.451)
Entry-P10	2.448 **(0.762)**	**1.491** (0.984)	3.530 (1.058)	4.611 (0.945)
TempleRing	**0.151** (0.080)	**0.151 (0.078)**	0.810 (0.089)	0.881 (0.371)
Fountain-P11	2.060 (0.266)	2.057 **(0.232)**	**0.682** (0.373)	0.888 (0.312)
Herz-Jesu-P25	3.227 (1.040)	3.302 (1.318)	**1.139** (0.725)	3.021 **(0.446)**
City-Hall	7.349 (0.986)	7.105 (5.641)	1.236 (0.910)	**1.159 (0.524)**

- *Optical Flow based estimation.* For the default setting of our optical flow based method we obtain an error d_F of 0.717 for the RTLS solution. This means that the deviation of the estimated epipolar lines from the ground truth lines is well within sub-pixel precision. For optimised parameters we reduce the error further to 0.245, while similar results are obtained for the ROLS solution. The correspondence estimation by means of optical flow clearly benefits from the filling-in effect in homogeneous image regions that is caused by the global smoothness constraint. The flow field within the silhouette of the model is shown in Fig. 4.4 (a) for optimised optical flow settings. For visualisation we use the colour code shown in Fig. 4.5 (a), where colour encodes the direction of the flow and brightness the magnitude. We can distinguish some occlusion artifacts where the tail of the dinosaur covers parts of the base, but their influence on the fundamental matrix estimation is reduced by the proposed robust L_1 penalisation. Fig. 4.4 (b) and (c) further show the epipolar geometry that is estimated from the optical flow. The corresponding points in the left and right image for which the epipolar lines are drawn are marked as red crosses. The estimated epipolar lines are depicted as full white lines, while the ground truth lines are dotted. We observe that both sets of lines are aligned quite well.

- *Feature based Estimation.* The settings that are used for the feature extraction in all image pairs of this section are listed in Table 4.2 and Table 4.3. Whereas optical flow based methods benefit from the filling-in effect of the smoothness term in homogeneous regions, insufficient texture often poses a challenge to feature extraction. SIFT, for instance is unable to provide a sufficient amount of features in the DinoRing images (≈ 180) for sub-pixel performance. The number of features tracked by the KLT algorithm for this sequence is larger and renders the best results of F1 and F2 sub-pixel precise. We can conclude from Table 4.4, however, that the KLT features suffer from poor localisation, as the average performance of the feature based techniques is worse than our method. It is well-known that features may suffer from localisation errors due to their computation in scale-space; see e.g. [WB94, ZGS$^+$09]. Fig. 4.4 shows the estimated epipolar geometries and the set of inliers that correspond to the best feature based result.

4.3.2 Near-Degeneracy and Repetitive Structures

The epipolar geometry of two images taken from different viewpoints is unique. A set of established image correspondences, on the other hand, does not always uniquely define the epipolar geometry. Such a configuration is called *degenerate* with respect to the computation of the fundamental matrix. A scenario that frequently occurs in stereo vision is that the majority of correspondences lie in the same plane, such as in the case of a dominant plane or when features are primarily extracted on a planar surface. Then there exists a 2D projective transformation, a homography H^3, that relates an in-plane point $\boldsymbol{x}$ in the left image to its corresponding in-plane point $\boldsymbol{x}' = H^3\boldsymbol{x}$ in the right image. The epipolar constraint between $\boldsymbol{x}$ and $\boldsymbol{x}'$ can then be written as

$$\boldsymbol{x}_{\mathrm{h}}'^{\top} F\, \boldsymbol{x}_{\mathrm{h}} = \boldsymbol{x}_{\mathrm{h}}'^{\top} F H^{3^{-1}}\, \boldsymbol{x}_{\mathrm{h}}' = 0 \ . \tag{4.58}$$

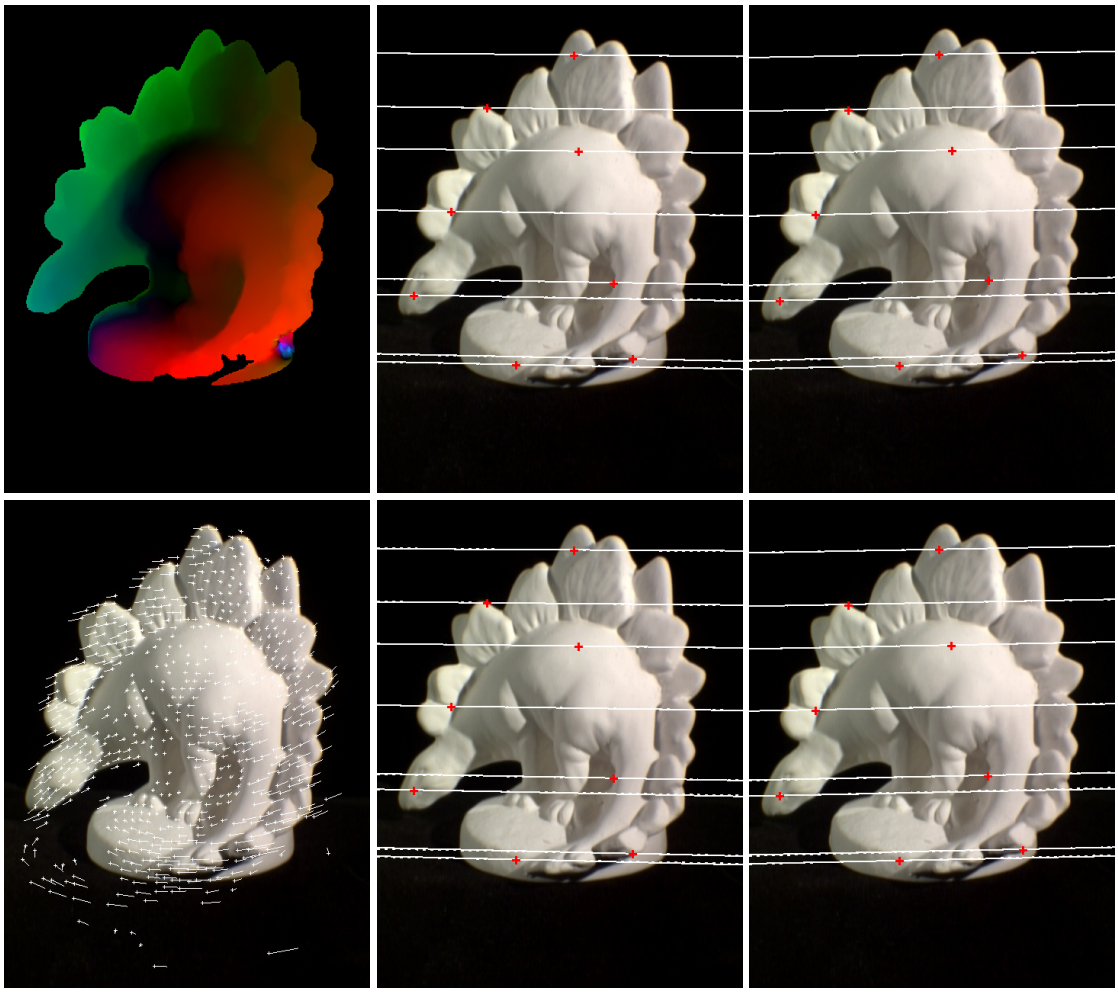

Fig. 4.4: Results for DinoRing. **Top Row: (a)** The optical flow between frames 24 and 25 within the object silhouette (optimal settings: $\alpha = 31.6$, $\gamma = 53.0$ and $\sigma = 0.9$). **(b) + (c)** The epipolar geometry estimated from the optical flow for frames 24 and 25. The points are depicted as red crosses, the corresponding estimated epipolar lines as full white lines and the corresponding ground truth lines as dotted lines. **Bottom Row: (d)** The inliers for the best result of F1. The correspondences are drawn on frame 24 as lines connecting the matched features. **(e) + (f)** The epipolar geometry estimated from these inliers.

This equation is satisfied for all in-plane points $\boldsymbol{x}'$ whenever the matrix $FH^{3^{-1}}$ is skew-symmetric. A solution for the fundamental matrix F is thus given by any matrix of the form ZH^3, where Z is a 3×3 skew-symmetric matrix. Such a solution is clearly not unique, and as a result the system matrix of equation system (4.5) will not provide enough constraints to uniquely determine the fundamental matrix [HZ00].

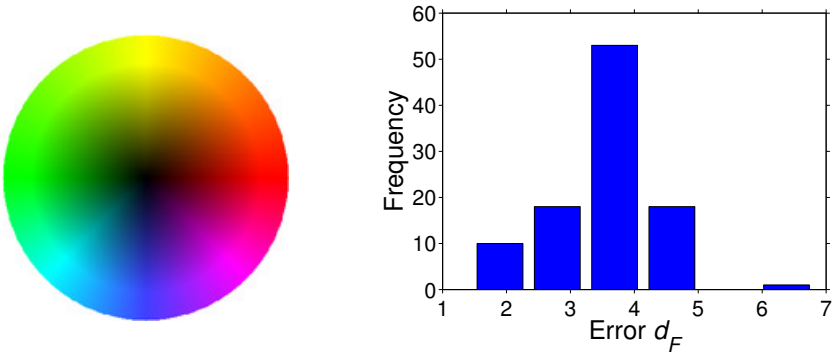

Fig. 4.5: **Left: (a)** Colour circle. **Right: (b)** Histogram for the Entry-P10 data set for F1 in combination with SIFT–LORANSAC.

A random sampling algorithm that is based on the 7- or 8-point algorithm is especially sensitive to planar degeneracies and can easily produce a consensus set of coplanar inliers. RANSAC, in particular, will return the solution that is consistent with the majority of the data, but it does not verify its uniqueness [CWM05, FP06]. The extra constraints necessary to overcome the ambiguity posed by a dominant plane must be provided by the *small number of out-of-plane inliers*. These are, however, often discarded as outliers in the sampling process. Moreover, actual outliers pose a treat: They are often regarded as inliers when they lie in the degenerate plane or they unwillingly disambiguate the solution towards a wrong estimate, thereby masking the degeneracy [TZM95].

For near-degenerate configurations, it is crucial that sufficient out-of-plane inliers are selected to overcome the estimation ambiguity, While this requires special care in the case of sparse feature based methods [TZM95, CWM05, FP06], optical flow will likely include a number of out-of-plane correspondences due to its dense and global nature. Optical flow thus provides the necessary constraints for the 8-point algorithm in a natural way. We illustrate this intrinsic robustness by estimating the epipolar geometry of frames 1 and 0 of the *Entry-P10* [7] multiview data set [SvHV+08]. This image pair depicts the facade of a building with a balcony as the only out-of-plane element. To ensure realistic run times for the parameter optimisation of our optical flow based technique, we tested all the methods on 640×427 versions of the original 3072×2048 images. A scalability test on the full resolution images will be presented in a later experiment in Chapter 5. The largest motion that is present in the downsampled images is 44 pixels.

- *Optical flow based estimation.* For our optical flow based technique we achieve a sub-pixel accuracy of 0.762 for optimised settings. The optical flow and the estimated epipolar geometry corresponding to these settings are shown in the top row of Fig. 4.6. The balcony is clearly visible in the flow field and the estimated epipolar lines are close to ground truth. For the standard settings we obtained a larger error because of the disturbing occlusion caused by a vehicle wheel disappearing in the lower right corner.

7. available at `http://cvlab.epfl.ch/~strecha/multiview/denseMVS.html`

- *Feature based Estimation.* If we take a look at the histogram of d_F for method F1 in Fig. 4.5, we can clearly distinguish a pronounced mode centred between 3 and 4. This mode corresponds to about 50% of the 100 LORANSAC test runs that predominantly select the inliers within the plane of the facade. The middle row of Fig. 4.6 shows such a degenerate set of inliers and the corresponding epipolar lines. These are wrongly estimated as the vanishing lines of the facade. A similar sensitivity to degeneracy was also observed for LMedS. Applying the robust DEGENSAC algorithm improves the performance of the feature based methods, but the results hardly become sub-pixel, as can be seen in Table 4.4. The reason for this is the large amount of in-plane outliers that arise from mismatched repetitive structures such as windows. These outnumber the out-of-plane inliers and model verification based on the amount of support tends to fail, even when the planar homography is estimated correctly. The inliers for the best result of F2 are depicted in the bottom row of Fig. 4.6, together with the outliers and the homography estimated by DEGENSAC.

4.3.3 Sufficient Texture and No Degeneracy

For the remainder of our comparison we have selected 4 image pairs that do not suffer from a lack of texture or degeneracy. First we compute the fundamental matrix for frames 13 and 14 of the Middlebury *TempleRing* [8] sequence. Both images have a resolution of 640×480 and a maximum displacement of approximately 20 pixels. The depicted temple model is much more textured than the DinoRing model. In Table 4.4 we observe that all estimation methods achieve sub-pixel precision, but that our optical flow based method performs best with an error of only 0.078 pixels for optimised parameters. The error for the default settings lies well below the average of both feature based methods. Fig. 4.7 shows the optimised optical flow within the model silhouette and the corresponding epipolar geometry. It can be observed that the epipolar lines practically coincide with the ground truth. The set of inliers for the best feature based result is also shown.

For our next experiments we again select image pairs from the Strecha multiview data base. All images have been scaled down to a resolution of 640×427 for parameter tuning purposes. We start with frames 1 and 2 of the *Fountain-P11* data set. The results in Table 4.4 show that our optical flow based method is the most accurate for optimised parameters, but that the feature based method F1 has a better overall performance with an average sub-pixel precision. The optical flow, depicted in the first column of Fig. 4.8, is not well estimated in the upper right corner. This results in a certain sensitivity to the parameter settings, explaining the difference between the optimised and the default error. The best feature based result is shown in the second column of Fig. 4.8.

We continue with 640×427 versions of frames 5 and 6 of the *Herz-Jesu-P25* [8] data set. It depicts the ornamented front view of a building, but contrary to the Entry-P10 data set, the entrances of the building provide sufficient out-of-plane correspondences to avoid degeneracy. The maximum displacement is approximately 53 pixels. In Table 4.4 we observe that our optical flow based method achieves an error of almost 1 pixel for optimal parameter settings but is outperformed by both feature based methods. The optimised flow and the corresponding epipolar geometry are shown in Fig. 4.9, together with the set of

8. available at `http://vision.middlebury.edu/mview/data/`

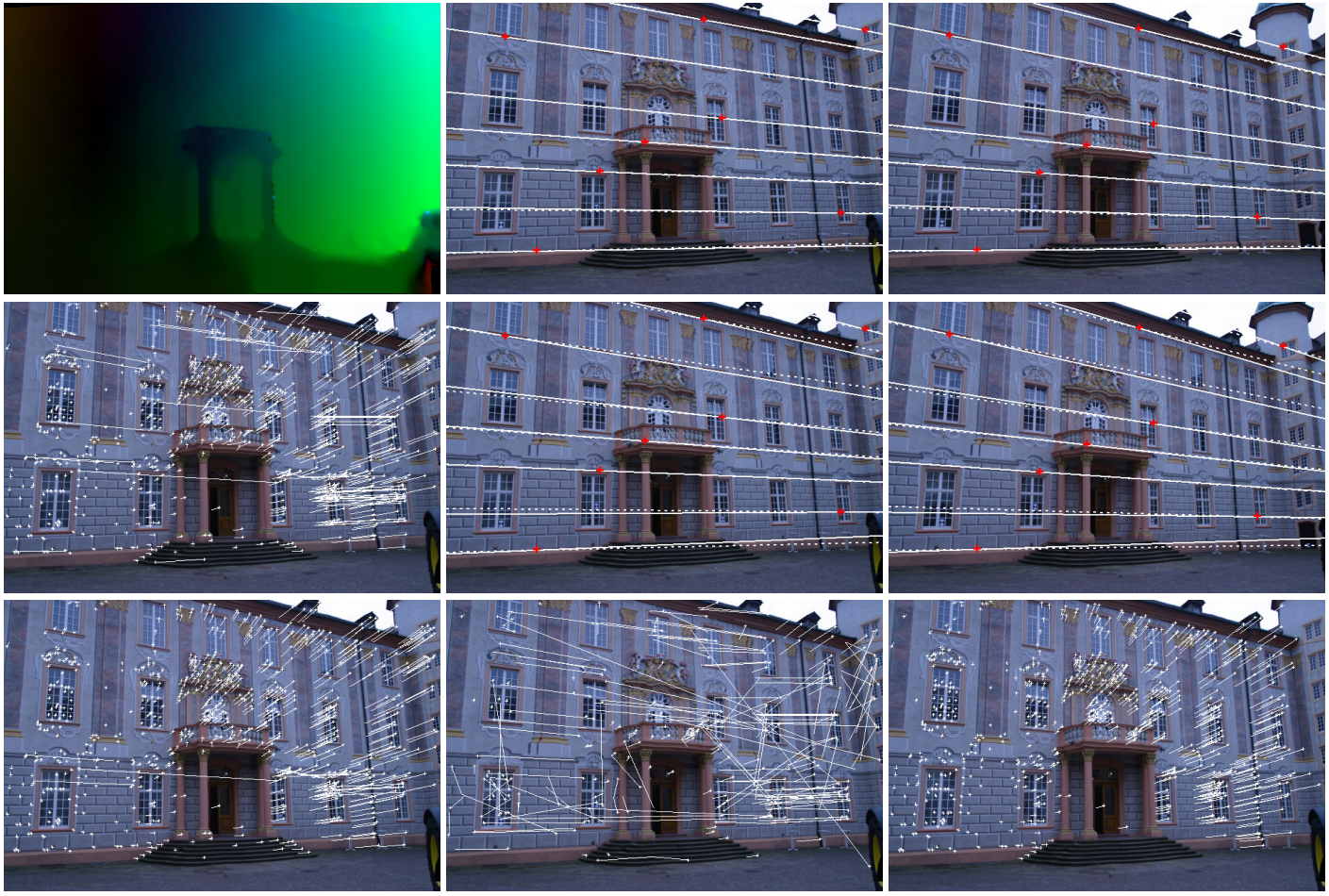

Fig. 4.6: Results for Entry-P10. **Top Row: (a)** The optical flow between frames 1 and 0 (optimal settings: $\alpha = 17.2$, $\gamma = 1.2$ and $\sigma = 0.6$). Pixels that are warped outside the image by the optical flow are coloured black. **(b) + (c)** The epipolar geometry estimated from the optical flow for frame 1 and 0. **Middle Row: (d)** A set of 763 degenerate inliers for F1. **(e) + (f)** The epipolar geometry estimated from these inliers for frame 1 and 0. **Bottom Row: (g)** A set of 728 non-degenerate inliers 728 inliers corresponding to the best result of F2. **(h)** The corresponding outliers. **(i)** The 706 inliers with respect to the planar homography.

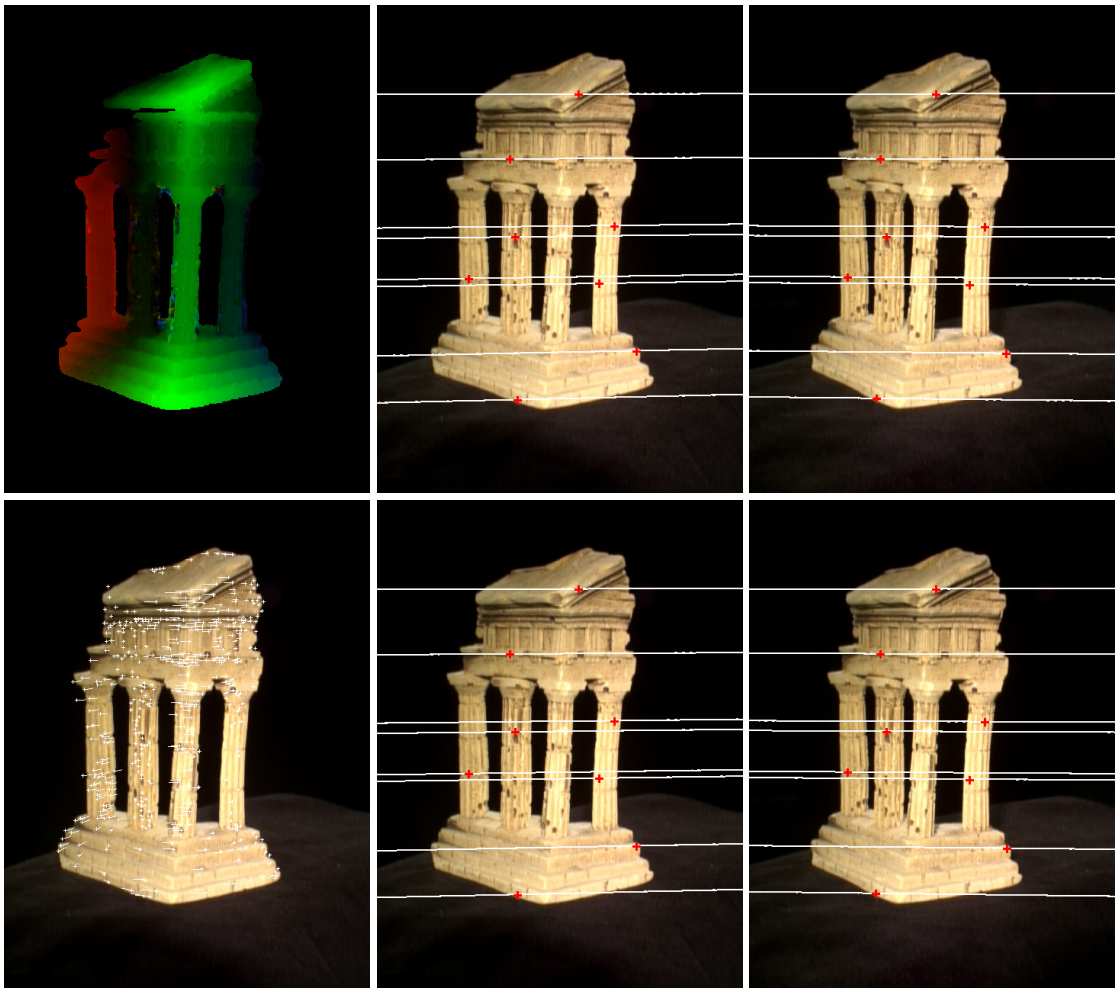

Fig. 4.7: Results for TempleRing. **Top Row: (a)** The optical flow between frames 13 and 14 within the silhouette (optimal settings: $\alpha = 25.5$, $\gamma = 43.8$ and $\sigma = 0.3$). **(b) + (c)** The epipolar geometry estimated from the optical flow. **Bottom Row: (d)** The inliers for the best result of F1. **(e) + (f)** The epipolar geometry estimated from the inliers.

inliers for the best feature based result. For outdoor sequences like this one, SIFT correspondences were overall more accurate than the tracked KLT features due to the large apparent motion. It can be seen in the cobbled stone region of the scene that the large change in viewpoint also makes matching more difficult for optical flow, causing a deterioration in the flow field at the bottom of the image. This leads to an undesirable parameter sensitivity and explains the error difference between the optimised and the default settings.

We conclude this section with the recovery of the epipolar geometry of frames 1 and 2 of the *City-Hall* [9] sequence [STV03]. Despite being scaled down to a resolution of 640×427, the image pair contains displacements of more than 85 pixels. The large apparent motion and occlusion on the left side of the building distort the optical flow for a wide range of settings. This is reflected by the large error of 7.3 for the default parameters of our method. For optimised settings we nevertheless obtain sub-pixel precision. Both feature

9. available at `http://cvlab.epfl.ch/data/strechamvs/`

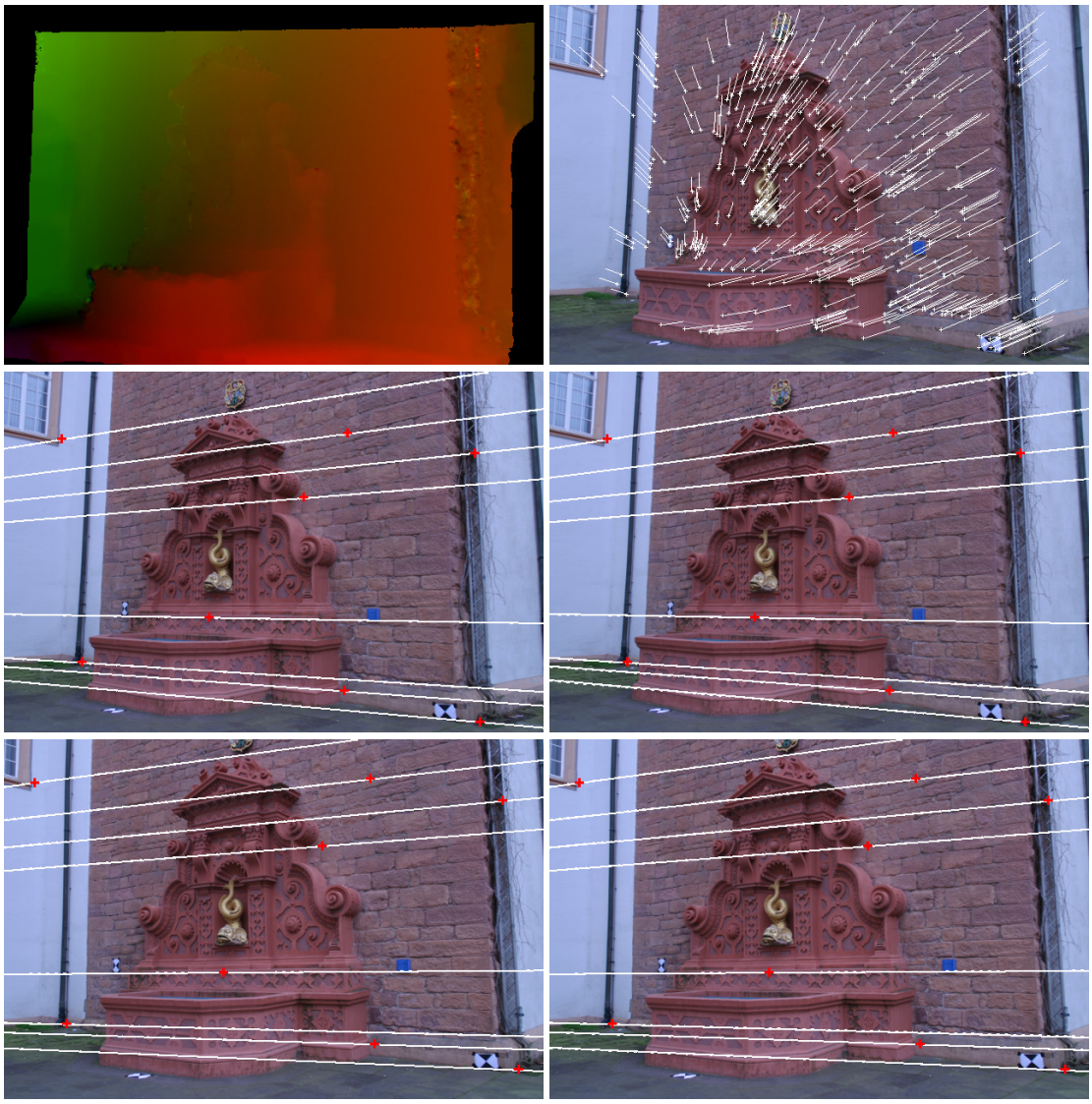

Fig. 4.8: Results for Fountain-P11. **Left Column: (a)** The optical flow between frames 1 and 2 (optimal settings: $\alpha = 6.2$, $\gamma = 7.2$ and $\sigma = 0.4$). **(c) + (e)** The epipolar geometry estimated from the optical flow. **Right Column:(b)** The inliers for the best result of F2. **(d) + (f)** The epipolar geometry estimated from the inliers.

based methods have a similar performance with an average error close to one pixel and a sub-pixel minimum. The optimised optical flow, the corresponding epipolar geometry and the set of inliers for the best feature based result are shown in Fig. 4.10.

4.4 Summary

In this chapter we have explored a new application field for dense optical flow techniques: the robust estimation of the fundamental matrix. Variational optical flow methods incorporate a global smoothness constraint that ensures filling-in in the absence of texture and dense correspondences in the case of degeneracy. Our experiments demonstrate that in

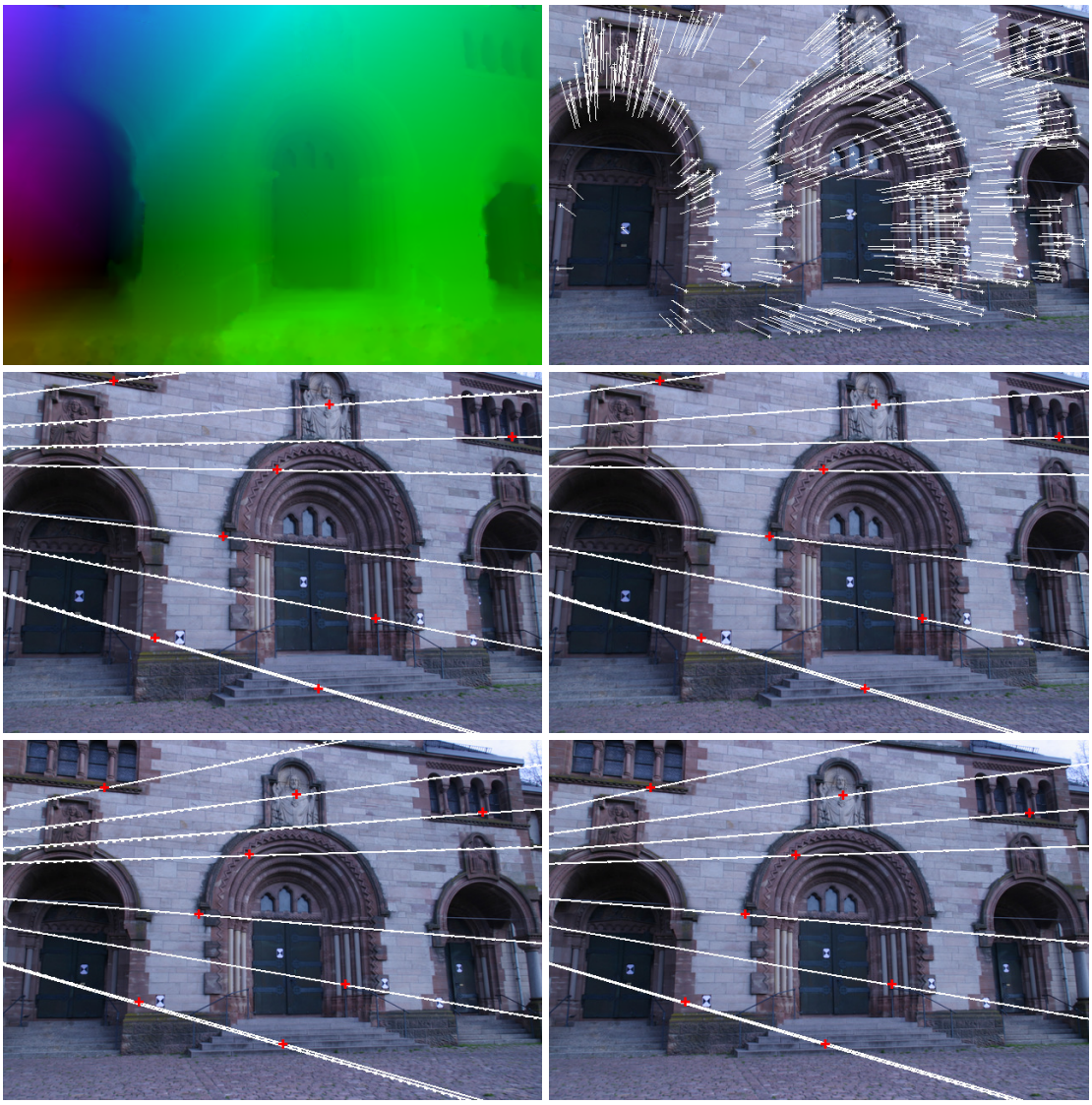

Fig. 4.9: Results for Herz-Jesu-P25. **Left Column: (a)** The optical flow between frames 5 and 6 (optimal settings: $\alpha = 22.9$, $\gamma = 10.5$ and $\sigma = 0.5$). **(c) + (e)** The epipolar geometry estimated from the optical flow. **Right Column:(b)** The inliers for the best result of F2. **(d) + (f)** The epipolar geometry estimated from the inliers.

these scenarios optical flow based fundamental matrix estimation clearly outperforms the widely-used feature based methods. In these scenarios we recommend to favour dense over sparse methods for estimating the fundamental matrix.

From our quantitative experiments it seems inconclusive if we should use a total least squares estimation or an ordinary least squares fit to compute the fundamental matrix. With this respect we give preference to the former, as it is based on the consistent modelling by a single energy and an extra constraint on the fundamental matrix norm. In addition, in a total least squares setting all entries of the fundamental matrix are treated with equal weight, which automatically leads to their simultaneous optimisation.

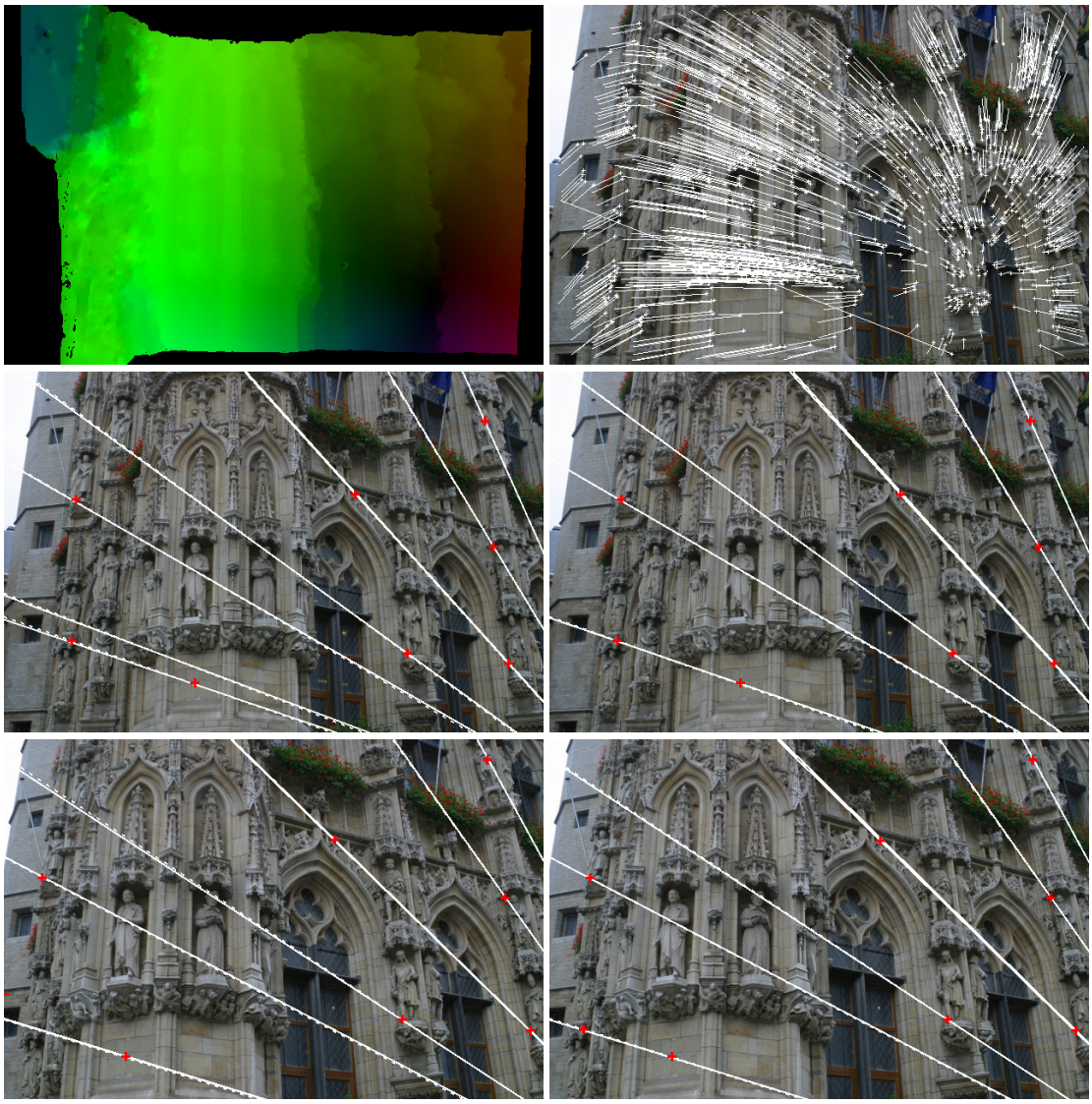

Fig. 4.10: Results for City-Hall. **Left Column: (a)** The optical flow between frames 1 and 2. (optimal settings: $\alpha = 27.7$, $\gamma = 51.9$ and $\sigma = 1.0$). **(c) + (e)** The epipolar geometry estimated from the optical flow. **Right Column:(b)** The inliers for the best result of F2. **(d) + (f)** The epipolar geometry estimated from the inliers.

It is interesting to analyse the reasons why a dense approach that incorporates *all* correspondences can be competitive with fairly sophisticated strategies that single out only the *very best* correspondences. Our explanation for this observation is that in those cases where feature based methods produce a mismatch, its influence on the final result is severe. Hence they require involved robustification methods such as RANSAC and its numerous variants. Dense methods, on the other hand, incorporate smoothness terms that prevent individual outliers. Furthermore, the accuracy of the fundamental matrix estimation benefits from the error averaging when exploiting thousands of correspondences. This accuracy will improve even further with ongoing, rapid progress in dense optical flow estimation.

5 Optical Flow for Uncalibrated Stereo Images

In the previous chapter we have seen that the estimation of the fundamental matrix from dense optical flow can be competitive with classical estimation techniques that rely on sparse image features. Due to the symmetry in the epipolar constraint, however, not only the estimation of the epipolar geometry can benefit from a known optical flow field, but knowledge about the epipolar geometry could at the same time have an advantageous effect on the computation of the optical flow. As a matter of fact, while reliable stereo correspondences are crucial for a precise estimation of the fundamental matrix, an accurate fundamental matrix might help us establish better correspondences to start with.

In this chapter we will solve this chicken and egg problem by *simultaneously* estimating the fundamental matrix and the optical flow of an uncalibrated stereo pair. We do this by feeding information about the computed epipolar geometry back into the optical flow estimation, thereby replacing the traditionally unconstrained estimation of the optical flow by one that takes into account the stereo geometry and the rigidity of the relative scene motion. Besides yielding better optical flow results than approaches that do not estimate the epipolar geometry in the process, optical flow methods for uncalibrated stereo further improve the fundamental matrix estimation of the two-step method that was presented in the previous chapter. Last but not least, a joint estimation of the fundamental matrix and the optical flow fuses the two steps of classical projective reconstruction: We solve for both the camera motion and the scene structure in a single variational framework.

We introduce our method for the joint estimation of the fundamental matrix and the optical flow in Sec. 5.1, where we discuss modelling aspects, minimisation strategies and the numerical solution. In Sec. 5.2 we evaluate the estimation accuracy of our joint method for both the epipolar geometry and the optical flow. Here, we also present results with respect to one of the main goals of this thesis: the 3D reconstruction of the depicted scene.

5.1 A Joint Variational Model

So far we have fed a given dense optical flow field into a classical approach for estimating the epipolar geometry. On the other side of the spectrum, there exists a large number of variational methods that use a given epipolar geometry to guide the estimation of the correspondences [ADSW02, SBW05, BAS07, ZBV+08]. Let us now investigate how we can achieve further improvements by coupling optical flow computation and fundamental matrix estimation in a joint model where they influence each other in a beneficial way. To this end we look at the epipolar constraint as a means of fitting a fundamental matrix to a set of correspondences and at the same time as a restriction on the correspondence search. In this section we present an intuitive way of coupling the computation of fundamental matrix and optical flow by minimising a single functional for both unknowns. Their simultaneous solution will ensure a scene structure that is most consistent with the camera motion, and vice versa, resulting in a higher overall accuracy and a lower parameter sensitivity.

5.1.1 The Epipolar Constraint as a Soft Constraint

To obtain a method for the joint estimation of the optical flow and the fundamental matrix, we propose to extend our general optical flow functional (3.3) with an extra term:

$$\mathcal{E}(\boldsymbol{w}, F) = \int_{\Omega} \left(\mathcal{E}_{\mathrm{D}}(\boldsymbol{w}) + \alpha\, \mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w}) + \beta\, \mathcal{E}_{\mathrm{E}}(\boldsymbol{w}, F) \right) \mathrm{d}\boldsymbol{x} \ . \tag{5.1}$$

While the first two terms in $\mathcal{E}(\boldsymbol{w}, F)$ correspond to the data and smoothness term of a standard optical flow model, the third term has been newly introduced to penalise deviations from the epipolar constraint $(\boldsymbol{x} + \boldsymbol{w})_{\mathrm{h}}^{\top} F\, \boldsymbol{x}_{\mathrm{h}} = 0$. We will call this term the *epipolar term*. Similar to the smoothness weight α, the weight β determines the relative importance of the epipolar term and to what extent the epipolar constraint will be satisfied in all points.

As opposed to methods that impose the epipolar constraint as a hard constraint by reducing the correspondence problem to a one-dimensional line search [ADSW02, SBW05], our joint model imposes the epipolar constraint as a *soft constraint*. This has two main advantages: (i) Contrary to methods that impose a hard epipolar constraint, our method is not restricted to image pairs with a known epipolar geometry (i.e. calibrated stereo images) and therefore offers a higher flexibility and usability. (ii) Additionally, the joint estimation of the optical flow and the fundamental matrix can benefit easily from progress in the design of new data and smoothness terms, since a soft constraint can be integrated directly in any future variational optical flow model that has a general form of (3.3).

5.1.1.1 A Model Prototype

If we integrate the epipolar term in our prototypical optical flow model (3.27), we obtain an energy of the form

$$\begin{aligned} \mathcal{E}(\boldsymbol{w}, F) = \int_{\Omega} \Bigg(& \Psi\left(|g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) - g_{\mathrm{l}}(\boldsymbol{x})|^2 + \gamma |\nabla g_{\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}) - \nabla g_{\mathrm{l}}(\boldsymbol{x})|^2 \right) \\ & + \alpha\, \Psi\left(|\nabla \boldsymbol{w}|^2 \right) + \beta\, \Psi\left(\left((\boldsymbol{x} + \boldsymbol{w})_{\mathrm{h}}^{\top} F\, \boldsymbol{x}_{\mathrm{h}} \right)^2 \right) \Bigg) \, \mathrm{d}x\, \mathrm{d}y \ . \end{aligned} \tag{5.2}$$

As in the case of the data and smoothness term, we choose for the penaliser Ψ in the epipolar term the regularised L_1 function (3.9). This reduces the influence of outliers in the computation of both F and $\boldsymbol{w}$. To avoid the trivial solution for the fundamental matrix F, we minimise energy (5.2) under the explicit constraint on the Frobenius norm

$$\|F\|_{\mathrm{Frob}}^2 = 1 \ . \tag{5.3}$$

In our experiments we will mainly make use of RGB colour images and we therefore need an extension to multiple image channels. This can be easily achieved by replacing the data term in the above model by its multi-channel variant (3.11).

While this model will serve as the baseline algorithm for the better part of our evaluation, some of our experiments will demonstrate the previously mentioned extensibility by integrating the epipolar term into a more recent optical flow method. For this we choose the method of Zimmer *et al.* [ZBW11], which has been described in more detail in Chap. 3. This optical flow method is characterised by a separate robustification of the brightness and gradient constancy assumption in the data term and it uses a flow- and image-driven anisotropic regulariser that works complementary to the data constraints.

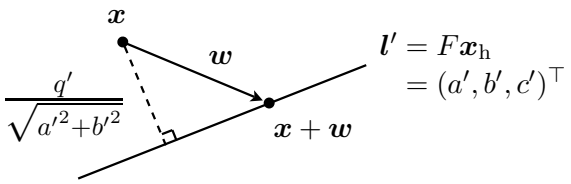

Fig. 5.1: The decomposition of the optical flow in its component along the epipolar line and its component perpendicular to the epipolar line.

5.1.1.2 Rewriting the Epipolar Constraint

Before turning to the minimisation of our prototypical energy (5.2), we provide two alternative formulations of the epipolar term that will be helpful in the process.

In terms of the Optical Flow. First of all, we note that we can rewrite the epipolar constraint between the points $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{w}$ as

$$(\boldsymbol{x}+\boldsymbol{w})_\mathrm{h}^\top F\,\boldsymbol{x}_\mathrm{h} = 0 \qquad \Longleftrightarrow \tag{5.4}$$

$$a'\,(x+u) + b'\,(y+v) + c' = 0\ . \tag{5.5}$$

This is actually the equation of the epipolar line $\boldsymbol{l}' = F\,\boldsymbol{x}_\mathrm{h}$ in the right image. We have defined the coefficients of the epipolar line $\boldsymbol{l}'$ as

$$a'(\boldsymbol{x}, F) = (F\,\boldsymbol{x}_\mathrm{h})_1 = F_{11}x + F_{12}y + F_{13}, \tag{5.6}$$

$$b'(\boldsymbol{x}, F) = (F\,\boldsymbol{x}_\mathrm{h})_2 = F_{21}x + F_{22}y + F_{23}, \tag{5.7}$$

$$c'(\boldsymbol{x}, F) = (F\,\boldsymbol{x}_\mathrm{h})_3 = F_{31}x + F_{32}y + F_{33}\ , \tag{5.8}$$

with F_{kl}, for $k, l \in \{1, 2, 3\}$, the kl-th entry of the fundamental matrix F. Note that the coefficients $(a', b', c')^\top$ are a function of $\boldsymbol{x}$ and that equation (5.5) forms a line constraint on the optical flow in every point of the left image. If we denote by q' the quadratic form

$$q'(\boldsymbol{x}, F) = \boldsymbol{x}_\mathrm{h}^\top F\,\boldsymbol{x}_\mathrm{h} = a'\,x + b'\,y + c'\ , \tag{5.9}$$

the epipolar constraint (5.5) can be further reduced to the expression

$$a'\,u + b'\,v + q' = 0\ . \tag{5.10}$$

An interesting remark is that q' has a geometrical meaning: Up to the normalisation factor $\sqrt{a'^2 + b'^2}$, it can be interpreted as the signed orthogonal distance of the point $\boldsymbol{x}$ to the corresponding epipolar line $\boldsymbol{l}'$. This is illustrated in Fig. 5.1, which shows the decomposition of the optical flow along and orthogonal to the epipolar line direction. The epipolar term can now be written explicitly in function of the optical flow as

$$\mathcal{E}_\mathrm{E}(\boldsymbol{w}, F) = \Psi\left(\big(a'(F)\,u + b'(F)\,v + q'(F)\big)^2\right)\ . \tag{5.11}$$

In terms of the Fundamental Matrix. Secondly, we recall from Chap. 4 that the epipolar constraint can be written in function of the fundamental matrix entries as

$$\boldsymbol{s}^\top \boldsymbol{f} = 0 \ . \tag{5.12}$$

Here, the 9×1 constraint vector $\boldsymbol{s}$ collects the coordinates of the corresponding image locations $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{w}$ as defined in Eq. (4.3) and the 9×1 vector $\boldsymbol{f}$ denotes the row major ordering of the fundamental matrix entries as defined in Eq. (4.4). The epipolar term can thus be written explicitly in terms of the fundamental matrix as

$$\mathcal{E}_{\mathrm{E}}(\boldsymbol{w}, \boldsymbol{f}) = \Psi\left(\left(\boldsymbol{s}^\top(\boldsymbol{w})\, \boldsymbol{f}\right)^2\right) \ . \tag{5.13}$$

The two alternative formulations given here additionally show that the epipolar constraint is linear in both the optical flow and the fundamental matrix. This has the advantage that the contribution of the epipolar term to the energy is convex in both unknowns.

5.1.2 Minimisation and Numerical Solution

Let us now turn to the minimisation of our prototypical model energy. Parameterising the fundamental matrix by a row major ordering of its entries, as discussed above, we aim to minimise the energy

$$\mathcal{E}(\boldsymbol{w}, \boldsymbol{f}) = \int_\Omega \left(\mathcal{E}_{\mathrm{D}}(\boldsymbol{w}) + \alpha\, \mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w}) + \beta\, \mathcal{E}_{\mathrm{E}}(\boldsymbol{w}, \boldsymbol{f})\right) \mathrm{d}\boldsymbol{x} \ , \tag{5.14}$$

with respect to the optical flow $\boldsymbol{w}$ and the fundamental matrix $\boldsymbol{f}$, subject to the constraint

$$\|\boldsymbol{f}\|^2 = \boldsymbol{f}^\top \boldsymbol{f} = 1 \ . \tag{5.15}$$

In this section we will solve this constrained optimisation problem by applying the method of the Lagrange multipliers, a strategy that we have used before in the previous chapter. We will show that the minimisation with respect to $\boldsymbol{f}$ will lead to an eigenvalue problem with matrix entries that depend on the optical flow. The minimisation with respect to $\boldsymbol{w}$ will lead to a system of Euler-Lagrange equations whose coefficients are a function of the estimated epipolar geometry. To solve for both $\boldsymbol{f}$ and $\boldsymbol{w}$ simultaneously, we will propose an alternating solution scheme that exhibits convergence in practice.

To allow for the estimation of large displacements and to avoid irrelevant local minima, we will formulate the minimisation in an incremental coarse-to-fine framework along the same line as for standard optical flow. To this end, we rewrite our joint energy (5.14) in terms of a small flow increment. In analogy with the motion tensor in the data term, this allows us to introduce a similar tensor notation for the epipolar term, which we will refer to as the *epipolar tensor*. The epipolar tensor will not only lead to a more compact notation that makes the convexity of the energy apparent, but will also prove important in the next chapter where we give a geometrical interpretation to the quantities being penalised.

5.1.2.1 Treatment of the Epipolar Term: The Epipolar Tensor

To express the epipolar term in function of an unknown flow increment $d\boldsymbol{w} = (du, dv)^\top$, we write the epipolar constraint (5.10) in terms of the total optical flow $\boldsymbol{w}+d\boldsymbol{w}$ as

$$a'\,(u + du) + b'\,(v + dv) + q' = 0 \ , \tag{5.16}$$

where $\boldsymbol{w} = (u, v)^\top$ is a given solution, stemming from a coarser resolution level of the coarse-to-fine warping pyramid. We further introduce the value

$$\tilde{q}' = a'\,u + b'\,v + q' \ . \tag{5.17}$$

Similar to q', $\tilde{q}'$ can be interpreted as a signed orthogonal distance to the epipolar line $\boldsymbol{l}' = F\,\boldsymbol{x}_\mathrm{h}$. This time, however, the distance is not measured between the point $\boldsymbol{x}$ and its epipolar line, but between the *warped* point $\boldsymbol{x} + \boldsymbol{w}$ and $\boldsymbol{l}'$ (again, up to the normalisation factor $\sqrt{a'^2 + b'^2}$). The epipolar constraint, written as

$$a'\,du + b'\,dv + \tilde{q}' = 0 \ , \tag{5.18}$$

now defines a line constraint on the flow increment in every point of the left image.

Using this notation, we can write the argument of the epipolar term as an inner product

$$\boldsymbol{e}^\top \boldsymbol{d}_\mathrm{h} = a'\,du + b'\,dv + \tilde{q}' \ , \tag{5.19}$$

where $\boldsymbol{d}_\mathrm{h} := (du, dv, 1)^\top$ is the homogeneous flow increment and $\boldsymbol{e}$ defined as

$$\boldsymbol{e} := (a', b', \tilde{q}')^\top \ . \tag{5.20}$$

The equation

$$\boldsymbol{e}^\top \boldsymbol{d}_\mathrm{h} = 0 \ , \tag{5.21}$$

can thus be regarded as a variant of the epipolar constraint for small perturbations in the right image position. Similar to the classical optical flow constraint (OFC) (3.32) of the data term, it expresses the restriction that the point $(du, dv)^\top$ has to lie on the line represented by $\boldsymbol{e}$. As in the case of the OFC, this equation does not tell us where exactly the increment has to lie on this line. Since every point on the epipolar line is a valid correspondence, the epipolar term alone does not provide a unique solution and additional data and smoothness constraints are necessary to resolve the ambiguity.

If we finally insert expression (5.19) as argument in the epipolar term, we can simplify the notation by writing the squared inner product as a quadratic form in the increment

$$\mathcal{E}_\mathrm{E}(d\boldsymbol{w}, \boldsymbol{f}) = \Psi\left((\,\boldsymbol{e}^\top \boldsymbol{d}_\mathrm{h}\,)^2\right), \tag{5.22}$$

$$= \Psi\left(\boldsymbol{d}_\mathrm{h}^\top E\,\boldsymbol{d}_\mathrm{h}\right) \ . \tag{5.23}$$

Here we have defined the symmetric 3×3 *epipolar tensor* as

$$E := \boldsymbol{e}\,\boldsymbol{e}^\top \ . \tag{5.24}$$

In contrast to the motion tensor, the epipolar tensor is always of rank one, which is a direct consequence of the fact that the epipolar constraint describes a single line constraint.

5.1.2.2 Minimisation of the Differential Form of the Energy

The differential forms of the data and smoothness term are the same as for pure optical flow and have been derived earlier in Sec. 3.2.1.1 and Sec. 3.2.1.2. If we combine these expressions with the epipolar term $\mathcal{E}_{\mathrm{E}}(d\boldsymbol{w}, \boldsymbol{f})$, we obtain the differential form of the total energy $\mathcal{E}(d\boldsymbol{w}, \boldsymbol{f})$. This energy will be minimised with respect to du, dv and $\boldsymbol{f}$ on each level of a coarse-to-fine approach. Taking into account the constraint $\|\boldsymbol{f}\|^2 = 1$, we apply the method of Lagrange multipliers and search for critical points of the Lagrangian

$$\mathcal{L}(d\boldsymbol{w}, \boldsymbol{f}, \lambda) = \mathcal{E}(d\boldsymbol{w}, \boldsymbol{f}) + \lambda(1 - \boldsymbol{f}^\top \boldsymbol{f}) \ . \tag{5.25}$$

These are tuples $(du^*, dv^*, \boldsymbol{f}^*, \lambda^*)^\top$ for which the functional derivatives of $\mathcal{L}$ with respect to du and dv and the derivatives of $\mathcal{L}$ with respect to $\boldsymbol{f}$ and λ vanish.

Solving for the Optical Flow. To solve for the optical flow, we express the differential energy as

$$\mathcal{E}(d\boldsymbol{w}, \boldsymbol{f}) = \int_\Omega \Big(\Psi\left(\boldsymbol{d}_{\mathrm{h}}^\top J\, \boldsymbol{d}_{\mathrm{h}}\right) + \alpha\, \Psi\left(\,|\nabla(\boldsymbol{w} + d\boldsymbol{w})|^2\,\right) + \beta\, \Psi\left(\boldsymbol{d}_{\mathrm{h}}^\top E\, \boldsymbol{d}_{\mathrm{h}}\right) \Big)\, \mathrm{d}\boldsymbol{x} \ , \tag{5.26}$$

where J is the symmetric motion tensor defined earlier in Eq. (3.42). By setting

$$\frac{\partial}{\partial\, du}\mathcal{L}(d\boldsymbol{w}, \boldsymbol{f}, \lambda) = 0 \quad \text{and} \quad \frac{\partial}{\partial\, dv}\mathcal{L}(d\boldsymbol{w}, \boldsymbol{f}, \lambda) = 0 \ , \tag{5.27}$$

we obtain the Euler-Lagrange equations of the optical flow increments du and dv. They are given by the following system of coupled partial differential equations

$$\begin{aligned} 0 \ = \ & \Psi'\left(\boldsymbol{d}_{\mathrm{h}}^\top J\, \boldsymbol{d}_{\mathrm{h}}\right)(J_{11}\, du + J_{12}\, dv + J_{13}) \\ & -\alpha\, \mathrm{div}\Big(\Psi'\big(\,|\nabla(\boldsymbol{w} + d\boldsymbol{w})|^2\,\big)\nabla(u + du)\Big) \\ & +\beta\, \Psi'\left(\boldsymbol{d}_{\mathrm{h}}^\top E\, \boldsymbol{d}_{\mathrm{h}}\right)(E_{11}\, du + E_{12}\, dv + E_{13})\,, \end{aligned} \tag{5.28}$$

$$\begin{aligned} 0 \ = \ & \Psi'\left(\boldsymbol{d}_{\mathrm{h}}^\top J\, \boldsymbol{d}_{\mathrm{h}}\right)(J_{12}\, du + J_{22}\, dv + J_{23}) \\ & -\alpha\, \mathrm{div}\Big(\Psi'\big(\,|\nabla(\boldsymbol{w} + d\boldsymbol{w})|^2\,\big)\nabla(v + dv)\Big) \\ & +\beta\, \Psi'\left(\boldsymbol{d}_{\mathrm{h}}^\top E\, \boldsymbol{d}_{\mathrm{h}}\right)(E_{12}\, du + E_{22}\, dv + E_{23}) \ , \end{aligned} \tag{5.29}$$

where J_{kl} and E_{kl}, for $k, l \in \{1, 2, 3\}$, denote the kl-th entry of J and E. These equations are non-linear in du and dv because of the penaliser function Ψ'. The strict convexity of the regularised L_1-norm, however, guarantees a unique solution.

Solving for the Fundamental Matrix. To solve for the fundamental matrix, we express the differential energy as

$$\mathcal{E}(d\boldsymbol{w}, \boldsymbol{f}) = \int_\Omega \Big(\Psi\left(\boldsymbol{d}_{\mathrm{h}}^\top J\, \boldsymbol{d}_{\mathrm{h}}\right) + \alpha\, \Psi\left(\,|\nabla(\boldsymbol{w} + d\boldsymbol{w})|^2\,\right) + \beta\, \Psi\left(\boldsymbol{s}^\top \boldsymbol{f}\right) \Big)\, \mathrm{d}\boldsymbol{x} \ , \tag{5.30}$$

and set

$$\nabla_{\boldsymbol{f}} \mathcal{L}(d\boldsymbol{w}, \boldsymbol{f}, \lambda) = \mathbf{0} \quad \text{and} \quad \frac{\partial}{\partial \lambda} \mathcal{L}(d\boldsymbol{w}, \boldsymbol{f}, \lambda) = 0 \ , \tag{5.31}$$

where $\nabla_{\boldsymbol{f}}$ stands for the gradient operator $(\partial_{f_1}, \ldots, \partial_{f_9})^\top$. To differentiate the Lagrangian $\mathcal{L}$ with respect to $\boldsymbol{f}$ we only have to consider the newly introduced epipolar term, since neither the data term nor the smoothness term depends on the fundamental matrix. We have encountered this type of minimisation before in Sec. 4.1.2, with the difference that the energy is now an integral over the continuous image domain Ω and not a sum over the discrete pixels. As in the discrete case, however, differentiation of the Lagrangian with respect to $\boldsymbol{f}$ and λ will give rise to an eigenvalue problem. It has the form

$$\left(\int_\Omega \Psi' \left((\boldsymbol{s}^\top \boldsymbol{f})^2 \right) \boldsymbol{s}\boldsymbol{s}^\top \, \mathrm{d}x \, \mathrm{d}y - \lambda I \right) \boldsymbol{f} =: (M - \lambda I) \, \boldsymbol{f} \ = \ \mathbf{0}^9 \tag{5.32}$$

$$1 - \|\boldsymbol{f}\|^2 \ = \ 0 \ . \tag{5.33}$$

Note that we were able to switch the order of differentiation and integration because $\boldsymbol{f}$ is constant over the image domain Ω. The system matrix M is symmetric positive definite and its entries are formed by the integral expressions

$$m_{i,j} = \int_\Omega \Psi' \left((\boldsymbol{s}^\top \boldsymbol{f})^2 \right) s_i s_j \, \mathrm{d}x \, \mathrm{d}y \ , \tag{5.34}$$

with $1 \leq i, j \leq 9$ and s_i being the i-th component of $\boldsymbol{s}$.

5.1.2.3 Solution of the System of Equations

The coefficients of the Euler-Lagrange equations (5.28) – (5.29) depend on the fundamental matrix entries via the epipolar tensor E. The system matrix M of the eigenvalue problem (5.32) – (5.33) in turn depends on the optical flow via the constraint vector $\boldsymbol{s}$. To solve for $\boldsymbol{w}$ and $\boldsymbol{f}$ simultaneously, two strategies can be considered:

A Fully Coupled Solution in a Coarse-to-Fine Framework. As a first approach, the estimation of the optical flow and the fundamental matrix can be embedded in the same coarse-to-fine framework. In this case we associate a variable k with the current image scale and assume that solutions $\boldsymbol{w}^k$ and $\boldsymbol{f}^k$ are already available from the previous image scale. Using these values, we compute the motion tensor J^k and the epipolar tensor E^k for the current level and solve the Euler-Lagrange equations for the flow increment $d\boldsymbol{w}^k$. Now, the total flow update $\boldsymbol{w}^{k+1} = \boldsymbol{w}^k + d\boldsymbol{w}^k$ can be used to compose the system matrix M^k and to solve the eigenvalue problem for $\boldsymbol{f}^{k+1}$. The updated flow and the updated fundamental matrix can then be used as an initialisation on the next finer image scale.

While this approach provides a maximal coupling between the optical flow and the fundamental matrix at each level of the multi-resolution pyramid, we have experienced that a solution of this kind tends to be unstable. The reason for this is that the initial fundamental matrix is estimated on a coarse image scale and is therefore mostly of bad quality due to the small amount of correspondences and due to the less reliable coarse flow. Any errors that appear in the epipolar geometry at a coarse scale then easily propagate throughout the coarse-to-fine pyramid, where they may be exacerbated further. Because of this observation, we will consider a second, alternative strategy in practice.

An Alternating Solution of the Optical Flow and the Fundamental Matrix. Instead of initialising the fundamental matrix at a coarse grid level, we kick-start its estimation by a more robust method, such as the optical flow based estimation method of the previous chapter. Assuming that we have an accurate estimate $\boldsymbol{f}$ of the fundamental matrix to start from, we can then solve the Euler-Lagrange equations in a coarse-to-fine framework where we keep $\boldsymbol{f}$ constant over all image scales. After computing the optical flow $\boldsymbol{w}$, we compose the system matrix M and solve the eigenvalue problem for a new fundamental matrix. This updated epipolar geometry will in turn be used to compute the optical flow again by solving the Euler-Lagrange equations. This alternating process of fundamental matrix estimation and optical flow computation is repeated until convergence.

In practice, we initialise the fundamental matrix with the zero matrix. This disables the epipolar term in the first iteration step, which automatically comes down to an estimation of the fundamental matrix from pure optical flow, as described in the previous chapter. We further exclude points from the estimation process that are warped outside the image by the optical flow, because the flow vectors are less reliable there. Our joint model does not explicitly enforce the singularity constraint on the fundamental matrix and therefore its rank is not enforced during the iterative process. The rank is, however, enforced on the final estimate. Since the estimation of the fundamental matrix proceeds by means of a familiar reweighted total least squares (RTLS) fit, as described in detail in Sec. 4.1.2, only the discretisation of the Euler-Lagrange equations remains to be discussed.

5.1.2.4 The Discrete Euler-Lagrange Equations

As we did for optical flow in Chap. 3, we denote by $[\,du\,]_{i,j}$ and $[\,dv\,]_{i,j}$ the approximations of du and dv at the pixel location (i,j), with $i = 1,\dots,n_x$ and $j = 1,\dots,n_y$. The image size on the current resolution level of the coarse-to-fine pyramid is assumed to be $n_x \times n_y$.

Point-Based Notation. Following the discretisation rules outlined in Sec. 3.2.2, we can discretise the equations (5.28) – (5.29) in all $n = n_x n_y$ pixel locations as

$$
\begin{aligned}
0 \;=\;& [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{11}]_{i,j}\,[\,du\,]_{i,j} + [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{12}]_{i,j}\,[\,dv\,]_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]_{i,j}\right)}{2} \frac{\left([\,du\,]_{\tilde{i},\tilde{j}} - [\,du\,]_{i,j}\right)}{h^2} \\
& + \beta\,[\Psi'_{\mathrm{E}}]_{i,j}\,[E_{11}]_{i,j}\,[\,du\,]_{i,j} + \beta\,[\Psi'_{\mathrm{E}}]_{i,j}\,[E_{12}]_{i,j}\,[\,dv\,]_{i,j} \\
& + \quad [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{13}]_{i,j} + \beta\,[\Psi'_{\mathrm{E}}]_{i,j}\,[E_{13}]_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]_{i,j}\right)}{2} \frac{\left([\,u\,]_{\tilde{i},\tilde{j}} - [\,u\,]_{i,j}\right)}{h^2} \qquad (5.35) \\
0 \;=\;& [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{12}]_{i,j}\,[\,du\,]_{i,j} + [\Psi'_{\mathrm{D}}]_{i,j}\,[J_{22}]_{i,j}\,[\,dv\,]_{i,j} \\
& - \alpha \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]_{i,j}\right)}{2} \frac{\left([\,dv\,]_{\tilde{i},\tilde{j}} - [\,dv\,]_{i,j}\right)}{h^2}
\end{aligned}
$$

$$
\begin{aligned}
&+ \beta\, [\Psi'_{\mathrm{E}}]_{i,j}\, [E_{12}]_{i,j}\, [\, du\,]_{i,j} + \beta\, [\Psi'_{\mathrm{E}}]_{i,j}\, [E_{22}]_{i,j}\, [\, dv\,]_{i,j} \\
&+ \quad [\Psi'_{\mathrm{D}}]_{i,j}\, [J_{23}]_{i,j} + \beta\, [\Psi'_{\mathrm{E}}]_{i,j}\, [E_{23}]_{i,j} \\
&- \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}}]_{i,j}\big)}{2} \frac{\big([\, v\,]_{\tilde{i},\tilde{j}} - [\, v\,]_{i,j}\big)}{h^2} \quad , \qquad (5.36)
\end{aligned}
$$

for $i = 1, \ldots, n_x$ and $j = 1, \ldots, n_y$. Here, we introduced the additional abbreviation

$$
[\Psi'_{\mathrm{E}}]_{i,j} \;=\; \frac{1}{2\sqrt{\big([\, du\,]_{i,j}, [\, dv\,]_{i,j}, 1\big)[E\,]_{i,j}\big([\, du\,]_{i,j}, [\, dv\,]_{i,j}, 1\big)^{\top} + \epsilon^2}} \quad , \qquad (5.37)
$$

for the approximation of $\Psi'\left(\boldsymbol{d}_{\mathrm{h}}^{\top} E\, \boldsymbol{d}_{\mathrm{h}}\right)$ in the pixel (i, j), with Ψ' the TV diffusivity (3.25). As before, $\mathcal{N}_l(i, j)$ denotes the set of the two neighbours of the pixel (i, j) in the axis direction $l \in \{x, y\}$ and $[\Psi'_{\mathrm{D}}]_{i,j}$ and $[\Psi'_{\mathrm{S}}]_{i,j}$ denote the approximations of $\Psi'\left(\boldsymbol{d}_{\mathrm{h}}^{\top} J\, \boldsymbol{d}_{\mathrm{h}}\right)$ and $\Psi'\big(\, |\nabla(\boldsymbol{w} + d\boldsymbol{w})|^2\,\big)$ in the pixel (i, j).

General Structure. The structure of the discrete Euler-Lagrange equation system can be made clear by writing it as

$$
A(\boldsymbol{p})\, \boldsymbol{p} + \boldsymbol{c}(\boldsymbol{p}) = \mathbf{0}^{2n} \quad , \qquad (5.38)
$$

where the $2n \times 1$ parameter vector is defined as $\boldsymbol{p} := (\mathbf{du}^{\top}, \mathbf{dv}^{\top})^{\top}$ and the $2n \times 2n$ matrix $A(\boldsymbol{p})$ and the $2n \times 1$ vector $\boldsymbol{c}(\boldsymbol{p})$ are defined as

$$
\begin{aligned}
A(\boldsymbol{p}) := \quad & \begin{pmatrix} \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\, \mathbf{J}_{11} & \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\, \mathbf{J}_{12} \\ \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\, \mathbf{J}_{12} & \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\, \mathbf{J}_{22} \end{pmatrix} - \alpha \begin{pmatrix} \mathbf{L}(\mathbf{du}, \mathbf{dv}) & 0^n \\ 0^n & \mathbf{L}(\mathbf{du}, \mathbf{dv}) \end{pmatrix} \\
& + \beta \begin{pmatrix} \boldsymbol{\Psi}'_{\mathrm{E}}(\mathbf{du}, \mathbf{dv})\, \mathbf{E}_{11} & \boldsymbol{\Psi}'_{\mathrm{E}}(\mathbf{du}, \mathbf{dv})\, \mathbf{E}_{12} \\ \boldsymbol{\Psi}'_{\mathrm{E}}(\mathbf{du}, \mathbf{dv})\, \mathbf{E}_{12} & \boldsymbol{\Psi}'_{\mathrm{E}}(\mathbf{du}, \mathbf{dv})\, \mathbf{E}_{22} \end{pmatrix}, \qquad (5.39) \\
\boldsymbol{c}(\boldsymbol{p}) := \quad & \begin{pmatrix} \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\, \mathbf{j}_{13} \\ \boldsymbol{\Psi}'_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})\, \mathbf{j}_{23} \end{pmatrix} - \alpha \begin{pmatrix} \mathbf{L}(\mathbf{du}, \mathbf{dv}) & 0^n \\ 0^n & \mathbf{L}(\mathbf{du}, \mathbf{dv}) \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \\
& + \beta \begin{pmatrix} \boldsymbol{\Psi}'_{\mathrm{E}}(\mathbf{du}, \mathbf{dv})\, \mathbf{e}_{13} \\ \boldsymbol{\Psi}'_{\mathrm{E}}(\mathbf{du}, \mathbf{dv})\, \mathbf{e}_{23} \end{pmatrix} \quad . \qquad (5.40)
\end{aligned}
$$

Here we have used the same compact notations $\mathbf{du}$, $\mathbf{dv}$, $\mathbf{u}$, $\mathbf{v}$, $\mathbf{j}_{kl}$, $\mathbf{J}_{kl}$, $\boldsymbol{\Psi}_{\mathrm{D}}(\mathbf{du}, \mathbf{dv})$ and $\mathbf{L}(\mathbf{du}, \mathbf{dv})$ as in Sec. 3.2.2.3. We further defined $\mathbf{e}_{kl}$, $\mathbf{E}_{kl}$ and $\boldsymbol{\Psi}'_{\mathrm{E}}(\mathbf{du}, \mathbf{dv})$ as

$$
\begin{aligned}
\mathbf{e}_{kl} \quad &:= \quad \big([E_{kl}]_{1,1}, \ldots, [\, E_{kl}\,]_{n_x,n_y}\big)^{\top} \quad \text{for } k, l \in \{1, 2, 3\}, \qquad (5.41) \\
\mathbf{E}_{kl} \quad &:= \quad \mathrm{diag}\,\big([E_{kl}]_{1,1}, \ldots, [\, E_{kl}\,]_{n_x,n_y}\big) \quad \text{for } k, l \in \{1, 2, 3\}, \qquad (5.42) \\
\boldsymbol{\Psi}'_{\mathrm{E}}(\mathbf{du}, \mathbf{dv}) \quad &:= \quad \mathrm{diag}\,\big([\Psi'_{\mathrm{E}}]_{1,1}, \ldots, [\Psi'_{\mathrm{E}}]_{n_x,n_y}\big) \quad . \qquad (5.43)
\end{aligned}
$$

The matrix $-\mathbf{L}(\mathbf{du}, \mathbf{dv})$ and the matrices arising from the data term and the epipolar term are symmetric positive definite, such that the matrix A is symmetric positive definite as well. The matrix A has a sparse structure and provides a point-wise coupling between $\mathbf{du}$ and $\mathbf{dv}$ via the off-diagonal blocks $\mathbf{J}_{12}$ and $\mathbf{E}_{12}$ and a neighbourhood coupling via the non-zero diagonals of $\mathbf{L}(\mathbf{du}, \mathbf{dv})$. Because of this structure, we will be able to solve the non-linear system by means of the Gauß-Seidel method with frozen coefficients.

5.1.2.5 Solution by Coupled Point Gauß-Seidel Relaxation

To solve the discrete Euler-Lagrange equations, we introduce a lagged diffusivity index k, and assume that a solution $([\,du\,]_{i,j}^k, [\,du\,]_{i,j}^k)$ is known from a previous iteration step. We then update the flow increments via a coupled point Gauß-Seidel relaxation step, whose update rule reads in point-based notation:

$$\begin{pmatrix} [\,du\,]_{i,j}^{k+1} \\ [\,dv\,]_{i,j}^{k+1} \end{pmatrix} = \begin{pmatrix} [M_{11}]_{i,j}^k & [M_{12}]_{i,j}^k \\ [M_{12}]_{i,j}^k & [M_{22}]_{i,j}^k \end{pmatrix}^{-1} \begin{pmatrix} [r_1]_{i,j}^k \\ [r_2]_{i,j}^k \end{pmatrix} \quad , \tag{5.44}$$

for $i = 1, \ldots, n_x$ and $j = 1, \ldots, n_y$. The matrix entries are given by

$$\begin{aligned} [M_{11}]_{i,j}^k &= [\Psi'_{\mathrm{D}}]_{i,j}^k \, [J_{11}]_{i,j} + \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}}^k + [\Psi'_{\mathrm{S}}]_{i,j}^k \right)}{2h^2} \\ &\quad + \beta \, [\Psi'_{\mathrm{E}}]_{i,j}^k \, [E_{11}]_{i,j}, \end{aligned} \tag{5.45}$$

$$\begin{aligned} [M_{22}]_{i,j}^k &= [\Psi'_{\mathrm{D}}]_{i,j}^k \, [J_{22}]_{i,j} + \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}}^k + [\Psi'_{\mathrm{S}}]_{i,j}^k \right)}{2h^2} \\ &\quad + \beta \, [\Psi'_{\mathrm{E}}]_{i,j}^k \, [E_{22}]_{i,j}, \end{aligned} \tag{5.46}$$

$$[M_{12}]_{i,j}^k = [\Psi'_{\mathrm{D}}]_{i,j}^k \, [J_{12}]_{i,j} + \beta \, [\Psi'_{\mathrm{E}}]_{i,j}^k \, [E_{12}]_{i,j} \quad , \tag{5.47}$$

and the right hand side by

$$\begin{aligned} [r_1]_{i,j}^k &= - [\Psi'_{\mathrm{D}}]_{i,j}^k \, [J_{13}]_{i,j} - \beta \, [\Psi'_{\mathrm{E}}]_{i,j}^k \, [E_{13}]_{i,j} \\ &\quad + \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l^-(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}}^k + [\Psi'_{\mathrm{S}}]_{i,j}^k \right)}{2h^2} [\,du\,]_{\tilde{i},\tilde{j}}^{k+1} \\ &\quad + \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l^+(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}}^k + [\Psi'_{\mathrm{S}}]_{i,j}^k \right)}{2h^2} [\,du\,]_{\tilde{i},\tilde{j}}^k \\ &\quad + \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}}^k + [\Psi'_{\mathrm{S}}]_{i,j}^k \right)}{2h^2} \left([\,u\,]_{\tilde{i},\tilde{j}} - [\,u\,]_{i,j} \right), \end{aligned} \tag{5.48}$$

$$\begin{aligned} [r_2]_{i,j}^k &= - [\Psi'_{\mathrm{D}}]_{i,j}^k \, [J_{23}]_{i,j} - \beta \, [\Psi'_{\mathrm{E}}]_{i,j}^k \, [E_{23}]_{i,j} \\ &\quad + \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l^-(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}}^k + [\Psi'_{\mathrm{S}}]_{i,j}^k \right)}{2h^2} [\,dv\,]_{\tilde{i},\tilde{j}}^{k+1} \\ &\quad + \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l^+(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}}^k + [\Psi'_{\mathrm{S}}]_{i,j}^k \right)}{2h^2} [\,dv\,]_{\tilde{i},\tilde{j}}^k \\ &\quad + \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}}]_{\tilde{i},\tilde{j}}^k + [\Psi'_{\mathrm{S}}]_{i,j}^k \right)}{2h^2} \left([\,v\,]_{\tilde{i},\tilde{j}} - [\,v\,]_{i,j} \right) \quad , \end{aligned} \tag{5.49}$$

where $\mathcal{N}^-(i,j)$ is the set of neighbours of pixel (i,j) in the direction of the l-axis that have already been updated and $\mathcal{N}^+(i,j)$ the set of neighbours that still need to be updated. Since the flow increments are assumed to be small, they are initialised by zero.

5.1.3 A Joint Model with Data Normalisation

Normalisation of the correspondence data, as proposed by Hartley [Har97], dramatically improves the condition number of the eigenvalue problem (5.32) – (5.33) and is essential for obtaining an accurate estimate of the fundamental matrix. It generally consists of replacing each point $\boldsymbol{x}_\mathrm{h} = (x, y, 1)^\top$ in the left image and its corresponding point $\boldsymbol{x}'_\mathrm{h} = (x+u, y+v, 1)^\top$ in the right image by the transformed points $T\tilde{\boldsymbol{x}}$ and $T'\tilde{\boldsymbol{x}}'$. The normalisation transformations T and T' are typically composed of a translation and a rotation and are chosen such that the normalised projective coordinates become $(1,1,1)^\top$ on average. We refer back to Sec. 4.1.3 for a more detailed description.

To obtain a closed model formulation, it is highly desirable that this normalisation step can be integrated into our prototypical energy (5.2). With the help of the epipolar constraint for normalised correspondence data (4.34), we can easily express the epipolar term in function of the normalised fundamental matrix $\hat{\boldsymbol{f}}$. This leads then to a joint variational model with data normalisation

$$\begin{aligned}\mathcal{E}(\boldsymbol{w}, \hat{\boldsymbol{f}}) = \int_\Omega \Big(\Psi\left(|g_\mathrm{r}(\boldsymbol{x}+\boldsymbol{w}) - g_\mathrm{l}(\boldsymbol{x})|^2 + \gamma |\nabla g_\mathrm{r}(\boldsymbol{x}+\boldsymbol{w}) - \nabla g_\mathrm{l}(\boldsymbol{x})|^2 \right) \\ + \alpha\, \Psi\left(|\nabla \boldsymbol{w}|^2 \right) + \beta\, \Psi\left((\hat{\boldsymbol{s}}^\top\ \hat{\boldsymbol{f}}\,)^2 \right) \Big) \,\mathrm{d}x\,\mathrm{d}y\ . \end{aligned} \tag{5.50}$$

Choosing the Normalisation Transformations. By imposing the constraint $\|\hat{\boldsymbol{f}}\|^2 = 1$ and by applying the method of the Lagrange multipliers to find a minimiser $\hat{\boldsymbol{f}}$ of (5.50), we obtain an eigenvalue problem that is similar to that of Eq. (5.32) – (5.33). To minimise $\mathcal{E}(\boldsymbol{w}, \hat{\boldsymbol{f}})$ with respect to the optical flow $\boldsymbol{w}$, however, we have to take into consideration that the epipolar line coefficients a', b' and q' now not only depend on the fundamental matrix entries, but also on the normalisation transformations T and T'. In the approach of Hartley, both normalisation transformations depend in turn on the set of points to be normalised. The correspondences in the left image consists of all pixels in the rectangular image domain Ω, so the transformation T is a constant mapping that only depends on the image size. The transformation T', on the other hand, normalises the warped pixel coordinates and thus depends on the optical flow $\boldsymbol{w}$. To avoid derivatives of T' with respect to u and v in the Euler-Lagrange equations, we have to replace it by a constant mapping. This way, the Euler-Lagrange equations of u and v do not change under the normalisation step and remain equal to those presented in Eq. (5.28) - (5.29). For simplicity, we further assume that the normalising transformations for the left and right image are similar, such that we can choose $T' = T$. Experiments have shown that this approximation has only a minor influence on the results compared with the original approach of Hartley.

The solution of the resulting system of equations is done iteratively: As for the joint model without data normalisation, the eigenvalue problem is first solved for $\hat{\boldsymbol{f}}$. After the fundamental matrix is denormalised via $F = T^\top \hat{F}\, T$ (see Sec. 4.1.3), it is used to solve the Euler-Lagrange equations for $\boldsymbol{w}$. This process is repeated until convergence.

5.2 Evaluation of the Joint Variational Method

In this experimental section we assess the performance of our joint variational method by separately evaluating the accuracy of the fundamental matrix estimation and the optical flow computation. For the recovery of the epipolar geometry, we use the same image pairs as for the evaluation of our optical flow based method in Sec. 4.3. Equivalently, we compare our results with those of the sparse estimation techniques F1 and F2. For the optical flow evaluation, we use stereo pairs from the Middlebury optical flow training data base with publicly available ground truth. For both types of experiments we present results for a set of optimised parameters and results for the default settings of $\alpha = 20.0$, $\gamma = 20.0$ and $\sigma = 0.9$. For the quality of the fundamental matrix estimation, we will use the error measure d_F of Sec. 4.3, which represents the deviation of the estimated epipolar geometry from ground truth in pixels units. The estimated optical flow will be assessed by means of the classical *average angular error* (AAE), which is defined as [BFB94]

$$\mathrm{AAE}\left(\boldsymbol{w}_{\mathrm{e}}, \boldsymbol{w}_{\mathrm{g}}\right) = \frac{1}{|\Omega|} \int\limits_{\Omega} \arccos\left(\frac{u_{\mathrm{e}} u_{\mathrm{g}} + v_{\mathrm{e}} v_{\mathrm{g}} + 1}{(u_{\mathrm{e}}^2 + v_{\mathrm{e}}^2 + 1)^{1/2} (u_{\mathrm{g}}^2 + v_{\mathrm{g}}^2 + 1)^{1/2}}\right) d\boldsymbol{x} \ , \tag{5.51}$$

with $|\Omega| = \int_{\Omega} d\boldsymbol{x}$ the size of the image domain, $\boldsymbol{w}_{\mathrm{e}} = (u_{\mathrm{e}}, v_{\mathrm{e}})$ the estimated flow field and $\boldsymbol{w}_{\mathrm{g}} = (u_{\mathrm{g}}, v_{\mathrm{g}})$ the ground truth flow field. We additionally evaluate the *average endpoint error* (AEE) of the estimated optical flow to include an error measure that is computed in pixel units. It is defined as [BSL+09]

$$\mathrm{AEE}\left(\boldsymbol{w}_{\mathrm{e}}, \boldsymbol{w}_{\mathrm{g}}\right) = \frac{1}{|\Omega|} \int\limits_{\Omega} \left((u_{\mathrm{e}} - u_{\mathrm{g}})^2 + (v_{\mathrm{e}} - v_{\mathrm{g}})^2\right)^{1/2} d\boldsymbol{x} \ . \tag{5.52}$$

5.2.1 Fundamental Matrix

The experiments for assessing the fundamental matrix estimation are divided into two parts. In a first part we investigate the convergence behaviour of our alternating minimisation strategy. In a second part we evaluate the scalability of our results under increasing image sizes and discuss run time issues.

5.2.1.1 Convergence of the Alternating Minimisation Strategy

We first demonstrate the convergence behaviour of our alternating minimisation strategy. To this end we recover the epipolar geometry of the Herz-Jesu-P25 and City-Hall image pairs with our joint estimation method for the default parameter settings. The first column of Fig. 5.2 and Fig. 5.3 shows the estimated epipolar lines after the first iteration step. We remind that this step comes down to an estimation of the fundamental matrix from pure optical flow and that these geometries correspond to the error d_F listed before in Table 4.4 for the respective image pairs (under the default settings for RTLS). The second column of these figures shows how these initial estimates are readjusted over 30 iterations to almost coincide with the ground truth. These geometries correspond to the default errors that can be found in Table 5.1. We additionally observe that the simultaneous recovery of the optical flow and the epipolar geometry has led to a visual improvement of both flow fields. This is most apparent in the occluding side of the building in the City-Hall sequence.

Tab. 5.1: Overview of the error d_F for 30 iterations of our joint variational method and for the feature based method classes F1 and F2 over 100 test runs. The best results are highlighted.

Image Pair	Our Method *default (opt)*	F1 *mean (min)*	F2 *mean (min)*
DinoRing	**1.175 (0.374)**	4.398 (0.422)	3.928 (0.451)
Entry-P10	**0.645 (0.476)**	3.530 (1.058)	4.611 (0.945)
TempleRing	**0.274** (0.168)	0.810 **(0.089)**	0.881 (0.371)
Fountain-P11	0.737 **(0.280)**	**0.682** (0.373)	0.888 (0.312)
Herz-Jesu-P25	**0.502 (0.396)**	1.139 (0.725)	3.021 (0.446)
City-Hall	**1.002** (0.870)	1.236 (0.910)	1.159 **(0.524)**

We found empirically that the value of the weight β has mainly an influence on the convergence speed and to a much lesser extend on the final error. Fig. 5.4 shows how d_F decreases as a function of the number of iterations, and eventually converges for different choices of β. In combination with the default settings, we choose $\beta = 40$, for which we obtained convergence within 10 iteration steps for all image pairs. If we compare the results of Table 5.1 with those of Table 4.4 for the default settings, we see that our joint estimation method has improved the accuracy substantially for all outdoor sequences. At the same time our method performs better on average than both feature based methods for all but the Fountain-P11 image pair. For 4 out of 6 data sets our results are within sub-pixel precision. Table 5.1 also lists the error for optimised parameters, including β. All results are within sub-pixel precision for the tested image pairs and for 4 out of 6 data sets we outperform both feature based methods. From the small deviation in error between the default and the optimised parameters, we conclude that our joint estimation method exhibits a lower sensitivity with respect to the parameter settings.

We make the remark here that our joint method converges towards an optical flow that is most consistent with the estimated fundamental matrix. While this will generally result in a more accurate epipolar geometry than the purely optical flow based estimates of the previous chapter, a comparison of the errors for the DinoRing and TempleRing sequence in Table 4.4 and Table 5.1, for instance, illustrate that this is not always guaranteed.

5.2.1.2 Scalability and Run Times

To demonstrate the scalability of our results, we computed the error d_F for the full (3072×2048) and half (1536×1024) resolution images of the Strecha data base. These are shown in Table 5.2, together with the outcome for the quarter size (640×427) versions. Due to the large run times of our joint method for larger image data, we did not search for optimised parameters and only list results for the default settings. Consequently, we only compare with the average results achieved by the feature based methods, for which we chose the same settings as in Table 4.3. We see that the errors for the full resolution images scale well for our method. For the half resolution versions we even obtain sub-pixel precision. For the feature based methods, the inlier ratios of the different sequences stayed roughly the same for all image sizes. The absolute number of inliers for Entry-P10

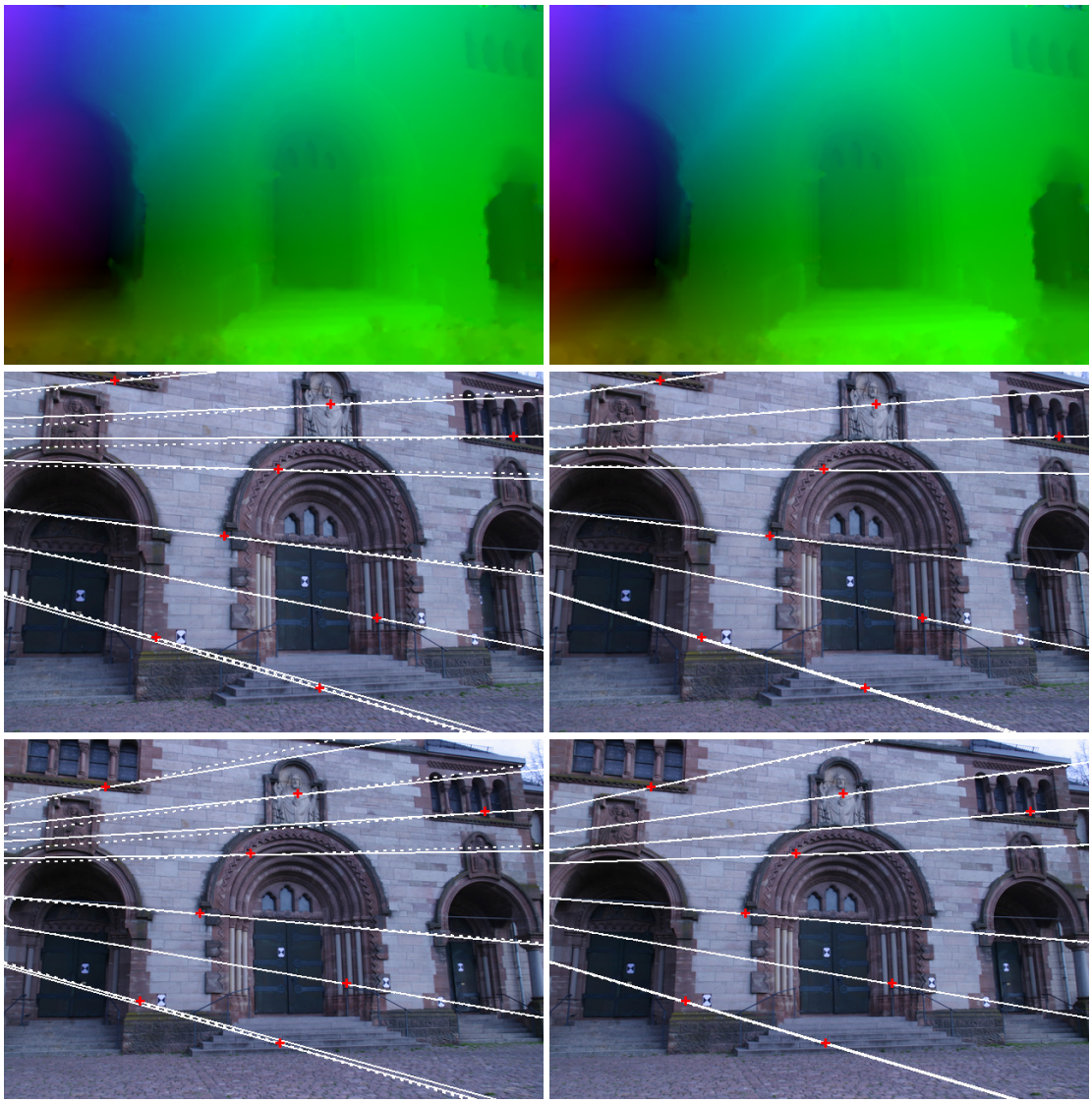

Fig. 5.2: Results for Herz-Jesu-P25 **Left Column: (a)** The optical flow between frames 5 and 6 after 1 iteration step (default settings). **(c) + (e)** The corresponding estimated epipolar geometry for frames 5 and 6. The points are depicted as red crosses, the corresponding estimated epipolar lines as full white lines and the corresponding ground truth lines as dotted lines. **Right Column: (b)** The optical flow between frames 5 and 6 after 30 iteration steps. **(d) + (f)** The corresponding epipolar geometry.

and Herz-Jesu-P25 ($\approx$ 3270 and $\approx$ 1520 respectively for full resolution) did, however, not scale with the image size. The good results of F1 for the full resolution Herz-Jesu-P25 sequence can therefore mainly be attributed to an increased precision of the feature locations. For the half resolution images of the City-Hall sequence, the absolute inlier count scaled significantly with the image size ($\approx$ 4304) and this helped to achieve a lower feature based error.

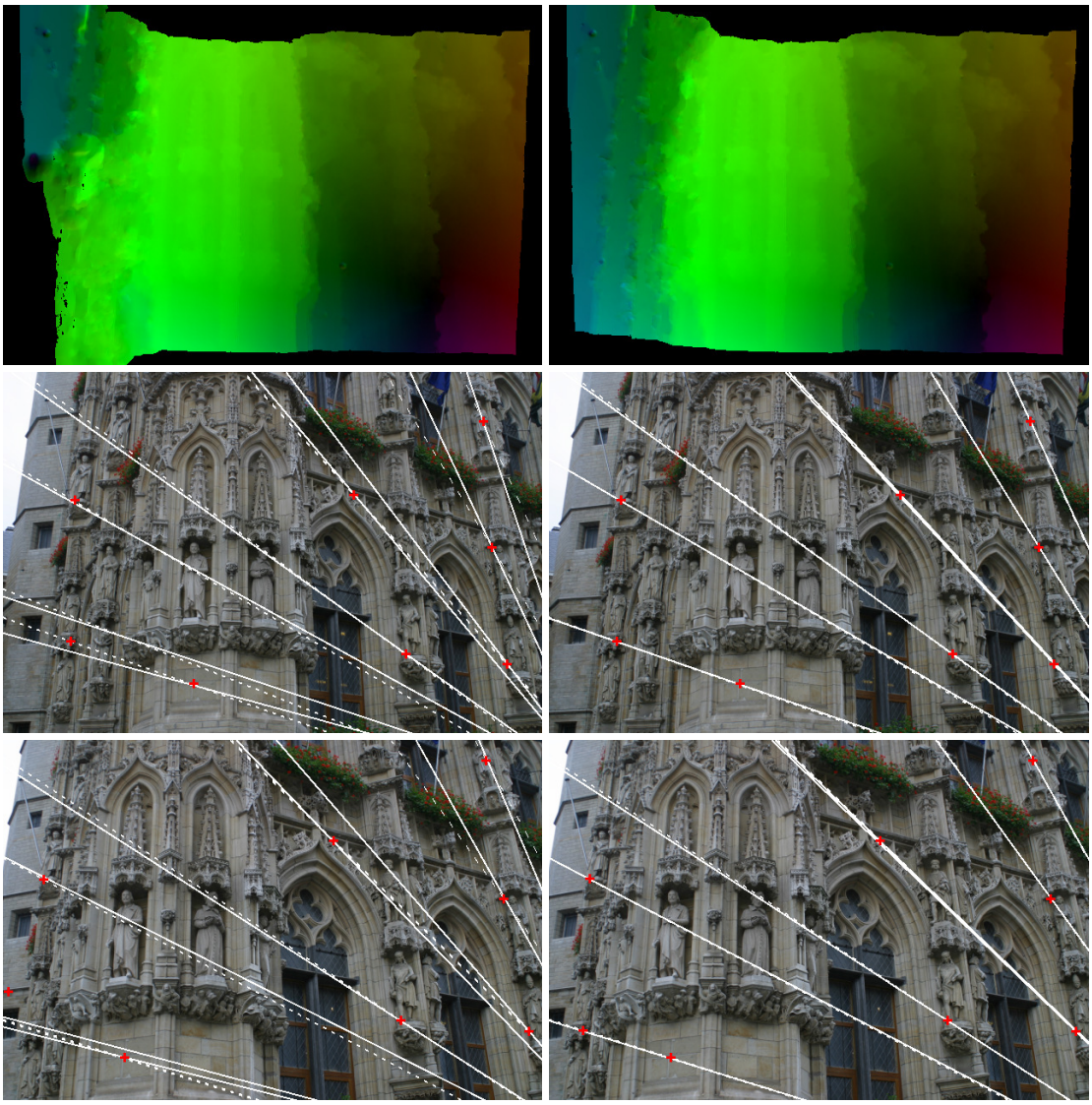

Fig. 5.3: Results for City-Hall. **Left Column: (a)** The optical flow between frames 1 and 2 after 1 iteration step (default settings). **(c) + (e)** The corresponding epipolar geometry. **Right Column: (b)** The optical flow between frames 1 and 2 after 30 iteration steps. **(d) + (f)** The corresponding epipolar geometry.

Table 5.2 additionally lists run time information for our joint method. These values refer to an implementation in standard ANSI C on a machine with a 1862MHz Intel Core2 CPU. For the quarter size images, *one iteration* of our alternating minimisation requires about 30 s, which is almost exclusively spent on the optical flow computation. For the half size images this grows to approximately 170 s per iteration. The total run time of the feature based methods with adaptive RANSAC ranges from less than 10 s for the quarter size images to more than 150 s for the half size images, depending on the amount of features and the distance measure being minimised. Here the run time is mostly dominated by feature extraction and matching.

Tab. 5.2: Overview of the error d_F for the quarter (640×427), half (1536×1024) and full (3072×2048) resolution versions of the outdoor image pairs for 30 iteration steps of our joint variational method and for 100 test runs of the feature based methods F1 and F2. The best results are highlighted in bold face. The corresponding run time (in seconds) of our method on a machine with a 1862MHz Intel Core2 CPU is given in the last column.

Image Size	Image Pair	Our Method	F1	F2	Run Time
quarter	Entry-P10	**0.645**	3.530	4.611	880 s
	Fountain-P11	0.737	**0.682**	0.888	
	Herz-Jesu-P25	**0.502**	1.139	3.021	
	City-Hall	**1.002**	1.236	1.159	
half	Entry-P10	**0.400**	7.062	10.094	5190 s
	Fountain-P11	**0.445**	0.490	0.948	
	Herz-Jesu-P25	**0.977**	1.940	4.393	
	City-Hall	0.657	**0.591**	0.600	
full	Entry-P10	**1.957**	10.142	19.165	$>$ 8 h
	Fountain-P11	**1.250**	1.499	1.488	
	Herz-Jesu-P25	2.404	**2.149**	7.997	
	City-Hall	**1.305**	2.011	2.102	

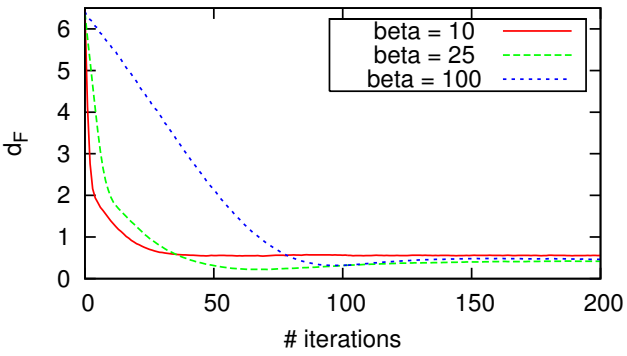

Fig. 5.4: Influence of the parameter β on the progression of d_F. The graph only shows the first 200 iterations of the 1000 that were carried out in total.

Tab. 5.3: Results for the Yosemite sequence without clouds compared with other 2D methods.

Method	AAE
Brox *et al.* [BBPW04]	1.59°
Mémin and Pérez [MP02]	1.58°
Roth and Black [RB05]	1.47°
Bruhn *et al.* [BWS05]	1.46°
Amiaz *et al.* [ALK07]	1.44°
Nir *et al.* [NBK08]	1.15°
Our method	**1.15°**

In order to speed up the run time of our method, several options exist. First of all, we require significantly less than 30 iterations to converge to an accurate solution. This allows us to reduce the number of iterations for practical applications. Secondly, variational optical flow computation can be parallelised efficiently on recent graphical hardware. According to Gwosdek *et al.* [GZG+10], the run time for a GPU implementation of the advanced optical flow method of [ZBW+09] is less than 1s for quarter size images and less than 2.5s for images of half size. If we assume convergence within 10 iterations, the total run time of a parallel implementation of our baseline method would thus be around 10s and 30s, respectively. For the full size images, run times are prohibitively large and memory requirements make the execution on current GPUs infeasible.

5.2.2 Optical Flow

In a second set of experiments we provide evidence that the concept of simultaneously recovering the correspondences and the epipolar geometry can improve the optical flow computation. First we use our joint estimation method to compute the optical flow between frames 8 and 9 of the *Yosemite sequence* [1] without clouds. This classical sequence of size 316×252 actually depicts a static scene captured by a moving camera and therefore forms a stereo pair. Table 5.3 shows that we are able to improve the AAE from 1.59° for standard optical flow to 1.15°. This ranks us among the best methods with spatial regularisation published so far [2]. For this experiment all parameters have been optimised with respect to the AAE of the optical flow in the first iteration step and β has been set to 50. The sky region has not been excluded from the computation, and pixels that are warped outside the image are included in the evaluation of the AAE. Our result is similar to the one presented by [NBK08], which is not surprising since a rigid motion model enters the functional of both methods. It has to be noted that methods with spatio-temporal smoothness terms may give lower errors. In Fig. 5.5 we show the results for the estimated optical flow and the corresponding epipolar geometry for 15 iteration steps.

In a final experiment we evaluate our methodology on four image pairs of the Middlebury optical flow benchmark ([BRS+07]). Frames 10 and 11 of the synthetic *Urban2*, *Urban3*, *Grove2* and *Grove3* [3] training set deal with rigid stereo motion for which the ground truth is publicly available. In Table 5.4 we show the influence of including the epipolar constraint in pure optical flow by collecting the AAE and the AEE of the estimated flow fields. The first column shows the errors for the optical flow method of [BBPW04], while the second column lists the errors for our joint model that adds the epipolar term to the functional. The results for both methods are obtained for the default settings with $\beta = 40$. Our joint method outperforms pure optical flow for all tested image pairs, even for Urban2 where the motion of a small car does not fulfil the epipolar constraint. To illustrate that our idea can even improve the performance of more recent state of the art optical flow techniques, we additionally provide results for the method of Zimmer *et al.* [ZBW11] without and with our additional epipolar term. The original method without epipolar term is currently one of the top ranking methods in the Middlebury benchmark and has been briefly sketched at the end of Sec. 3.1.3. For *fixed* settings ($\alpha = 400.0$, $\gamma = 20.0$, $\sigma = 0.5$ and $\beta = 5.0$), the joint variant of Zimmer *et al.* improves the AAE and AEE over pure optical flow for all four sequences, leading to some of the best results published so far. The corresponding estimated flow fields and epipolar lines are shown in Fig. 5.6.

5.2.3 Automatic Reconstruction

To conclude, we present reconstruction results for the outdoor image pairs that we used in our experiments and for image pairs that we recorded ourselves. To perform a dense reconstruction of the depicted scene, we extract the left and right camera projection matrices P and P' from the estimated fundamental matrix F and use them to triangulate the

1. available at http://www.cs.brown.edu/people/black/images.html
2. Better results are reported in the technical report [BRN+09] which did not yet undergo the process of peer reviewing.
3. all available at http://vision.middlebury.edu/flow/data/

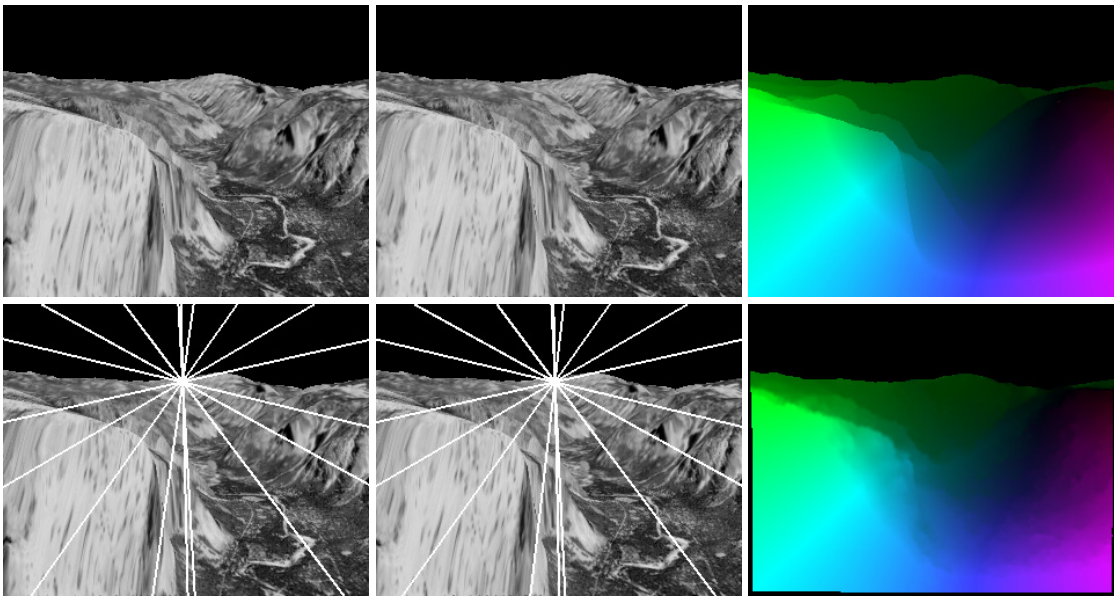

Fig. 5.5: Results for the Yosemite sequence without clouds. **Top Row: (a)** Frame 8. **(b)** Frame 9. **(c)** Ground truth optical flow. **Bottom Row: (c)** Estimated eipolar lines in frame 8. **(d)** Estimated eipolar lines in frame 9. **(f)** Estimated optical flow (settings: $\alpha = 19.1$, $\gamma = 2.1$, $\sigma = 0.9$ and $\beta = 50.0$). Pixels (apart from the sky region) that are warped outside the image are coloured black.

Tab. 5.4: Influence of including the epipolar term (+ET) in optical flow for the four stereo image pairs of the Middlebury optical flow training set. The results in the first two columns correspond to the baseline method of Brox *et al.* [BBPW04] and are presented for default settings ($\alpha = 20.0$, $\gamma = 20.0$, $\sigma = 0.9$ and $\beta = 40.0$). The results in the last two columns correspond to the method of Zimmer *et al.* [ZBW11] and are presented for the fixed settings given in the corresponding paper ($\alpha = 400.0$, $\gamma = 20.0$, $\sigma = 0.5$ and $\beta = 5.0$).

Image Pair	Brox *et al.* AAE / AEE	Brox *et al.* + ET AAE / AEE	Zimmer *et al.* AAE / AEE	Zimmer *et al.* + ET AAE / AEE
Grove2	2.67 / 0.19	2.53 / 0.17	2.19 / 0.16	2.13 / 0.14
Grove3	6.78 / 0.69	5.89 / 0.64	5.84 / 0.59	5.61 / 0.57
Urban2	2.66 / 0.32	2.20 / 0.29	2.46 / 0.26	2.15 / 0.24
Urban3	5.26 / 0.61	4.96 / 0.56	3.40 / 0.44	3.11 / 0.39

back-projected rays for each pixel correspondence. While doing so, we assume that the internal camera parameters, such as the focal length and the principal point, are known. As explained in Sec. 2.4, this is valid assumption in practice and allows us to extract the essential matrix E from F to determine the relative pose and orientation of the second camera with respect to the first one. The resulting reconstruction is then up to scale.

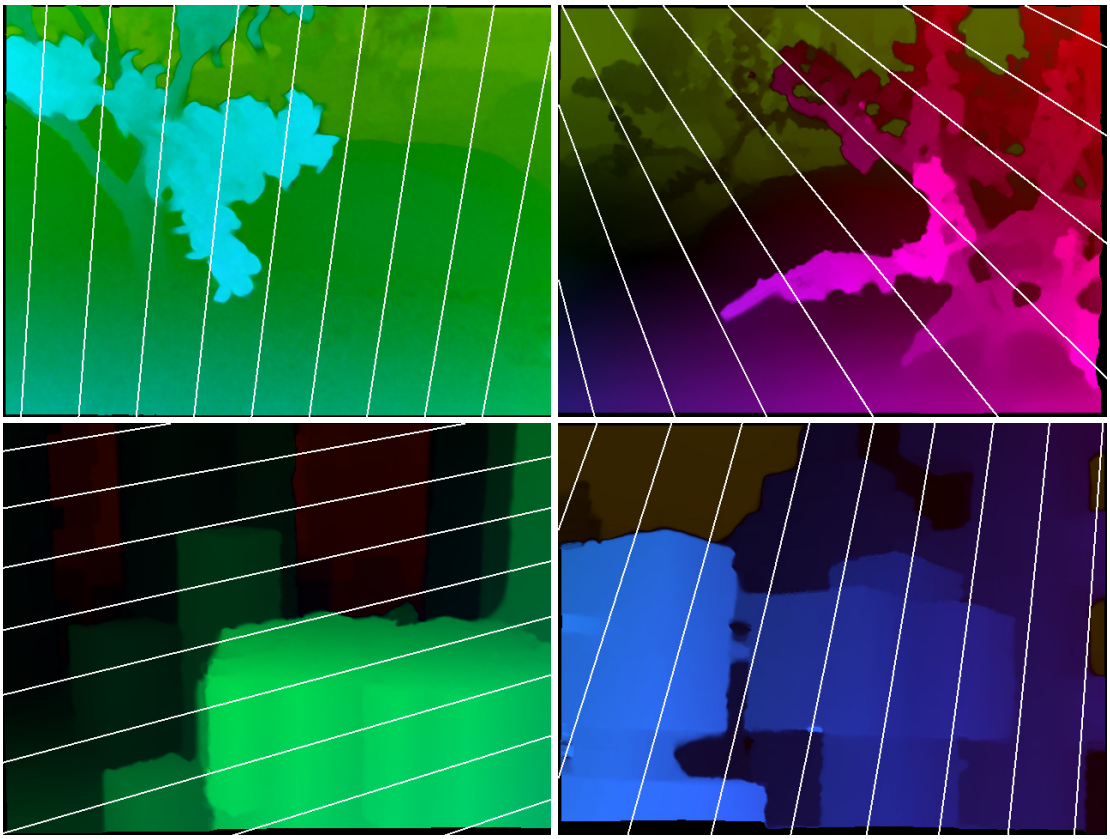

Fig. 5.6: The flow fields between frames 10 and 11 of the Middlebury training sequences. The estimated epipolar lines for frame 10 have been overlaid. **Top Left: (a)** Grove2. **Top Right: (b)** Grove3. **Bottom Left: (c)** Urban2. **Bottom Right: (d)** Urban3.

5.2.3.1 Simultaneous Recovery of Camera Motion and Scene Structure

By simultaneously solving for the dense optical flow and the epipolar geometry of two images, we can associate with each pixel of the left image a 3D point $\boldsymbol{X}$ in space. As seen in Sec. 2.4.3.3, this can be easily achieved by solving the triangulation equations

$$\begin{cases} P(F)\boldsymbol{X}_{\mathrm{h}} &= \boldsymbol{x}_{\mathrm{h}} \\ P'(F)\boldsymbol{X}_{\mathrm{h}} &= (\boldsymbol{x}+\boldsymbol{w})_{\mathrm{h}} \end{cases}, \tag{5.53}$$

in a total least squares sense [HZ00]. Unlike 3D reconstruction in a calibrated (or rectified) setting, the camera matrices P and P' are not given, but are estimated together with the optical flow in the form of the fundamental matrix. In this sense, *our joint variational method fuses the two steps of traditional projective reconstruction by simultaneously solving for the camera motion and scene structure in a single optimisation framework.*

Contrary to multiview stereo, we do not obtain a complete representation of the depicted scene. Instead, our reconstruction is viewpoint specific and only contains points that are visible in the first image. Such a reconstruction is often referred to as a *range image*. Because an assessment of the reconstruction quality encompasses an evaluation of the reconstruction accuracy together with the reconstruction completeness against ground truth, it is

difficult to compare a range image in a multiview ranking such as [SCD+06, SvHV+08]. We will therefore restrict ourselves here to a visual assessment of the results. Nevertheless, the methodology presented in this chapter could form the basis for an uncalibrated multiview system that integrates range images from multiple stereo pairs [ZPB07b].

5.2.3.2 Reconstruction Results

In Fig. 5.7 we present 3D reconstructions from frames 1 and 0 of the Entry-P10 data set, frames 1 and 2 of the Fountain-P11 data set and frames 5 and 6 of the Herz-Jesu-P25 data set. These reconstructions have been obtained by our baseline joint estimation method. For all image pairs we used the 1536×1024 versions of the original images, which comes down to more than 1.5 million reconstructed points. The settings for our method were the default ones, which corresponds to an accuracy of the epipolar geometry listed in Table 5.2 for the half resolution images. To obtain a reconstruction up to scale, we used the internal camera parameters that are provided with the images. The reconstructions are displayed in OpenGL [WND99] as an untextured mesh of triangles connecting the 3D points. The ground part of the results is left out for visualisation purposes. As one can observe, many details are visible and discontinuities in the depth are accurately recovered. The same is true for the reconstruction from frames 1 and 2 of the City-Hall sequence shown in Fig. 5.8. Here, we do not display points that are warped outside the image by the optical flow. The close-ups in Fig. 5.8 (b) and (c) show a detailed view of the middle section of the building with and without texture. Fine details can be clearly distinguished and the statues in the close-up are easily recognisable in the untextured surface.

In a final experiment we reconstruct a human face from a stereo pair that we have recorded ourselves with two Point Grey [4] Flea cameras. The images are shown in Fig. 5.9 (a) and (b) and are of size 280×430. We do not perform an internal calibration of the cameras, but instead use the provided focal length and approximate the principal point by the image centre. To demonstrate the flexibility of our approach in different problem settings, we replace the TV-regularisation of our baseline method with a smoothing strategy that allows a stronger smoothing between flow discontinuities and at the same time preserves meaningful features by not smoothing across them. This will result in a naturally smooth reconstruction while preserving a realistic facial expression. To achieve this, we opt for the PDE based anisotropic flow-driven strategy discussed in Sec. 3.1.2.2. This type of smoothing is easily obtained by replacing the scalar valued TV-diffusivity in the Euler-Lagrange equations by the diffusion tensor (3.24). For the function Ψ' we choose the Perona-Malik diffusivity (3.26) with a contrast parameter of $c = 0.1$. This diffusivity is known to make backward diffusion possible and thereby enhance edges even more. The remaining model parameters where optimised individually. The resulting reconstruction is shown in Fig. 5.9 (c)–(e). We see that the facial expression is captured very well with only a slight degeneration of the reconstruction near specularities on the nose and the eyes. For this example the background was excluded from the fundamental matrix estimation.

4. `http://www.ptgrey.com/`

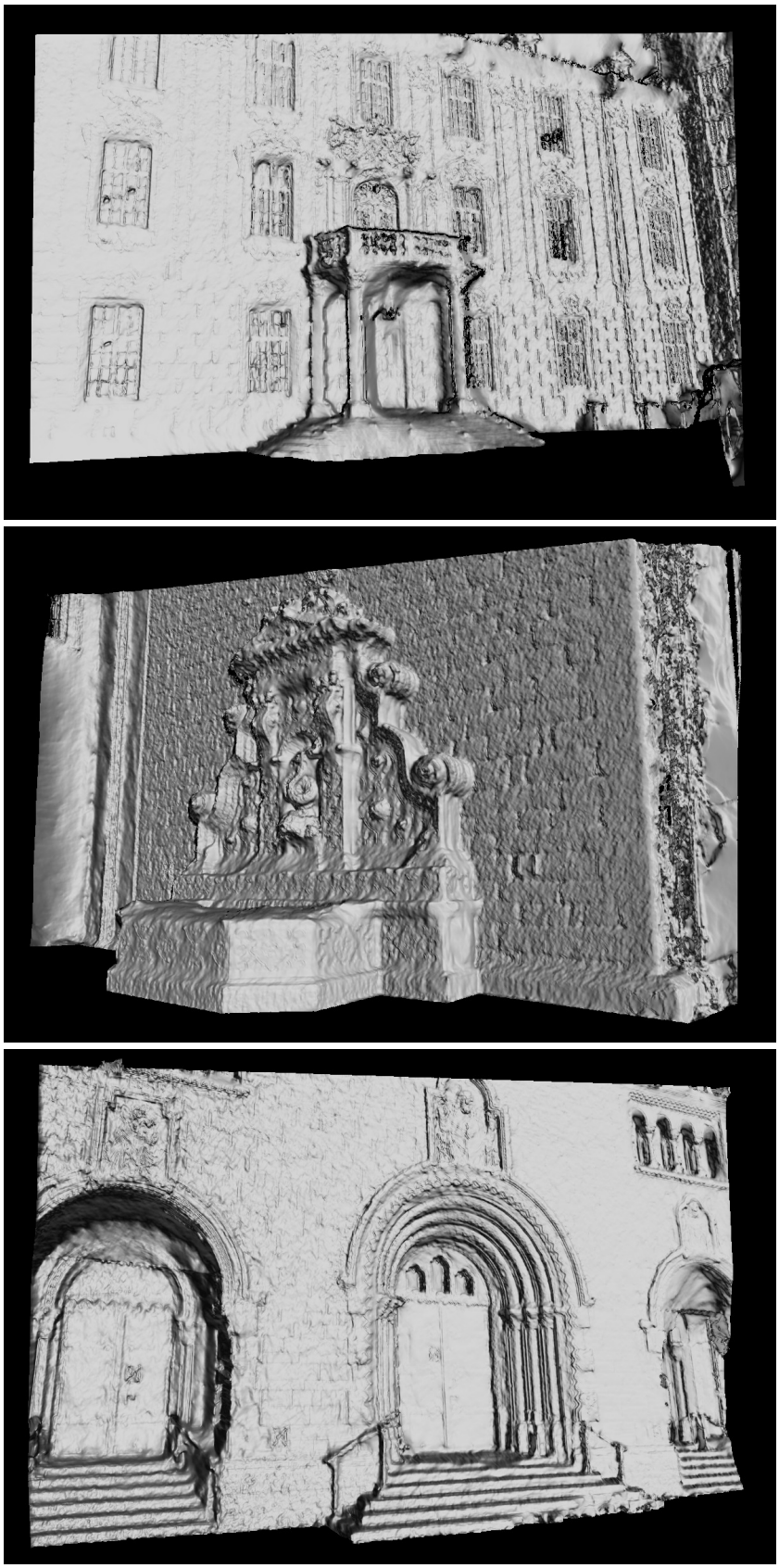

Fig. 5.7: Untextured reconstruction results for the Entry-P10 data set (**Top**), the Fountain-P11 data set (**Middle**) and the Herz-Jesu-P25 data set (**Bottom**).

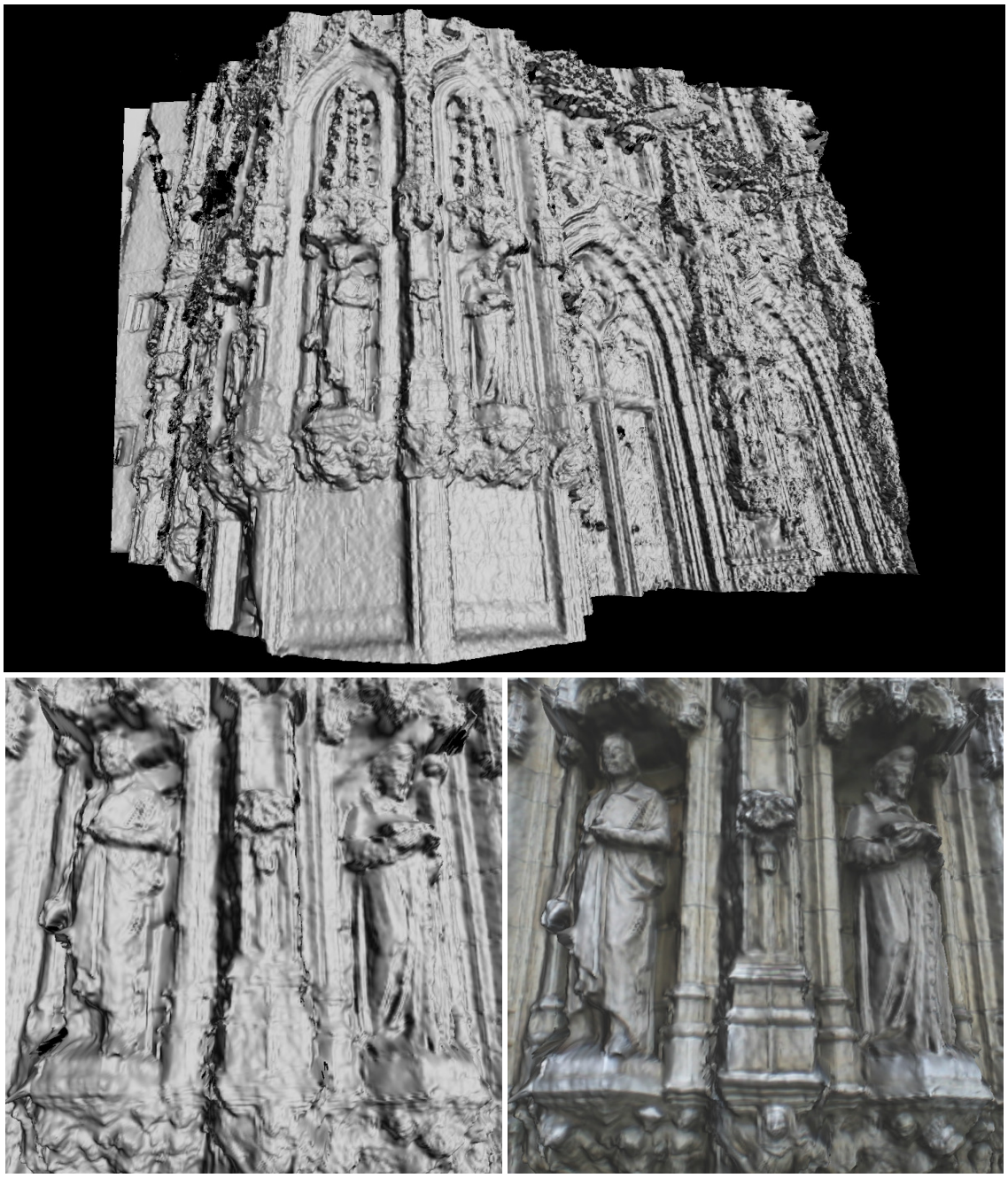

Fig. 5.8: Reconstruction result for the City-Hall data set. **Top: (a)** Untextured reconstruction. **Bottom Left: (b)** Untextured close-up. **Bottom Right: (c)** Textured close-up.

5.2.4 Limitations of Variational Methods

While the previous chapter showed that the estimation of the fundamental matrix from pure optical flow was promising, particularly in challenging cases with low-texture or near-degenerate configurations, we have demonstrated in this chapter that an intricate coupling of both unknowns can improve results even further.

Fig. 5.9: Face reconstruction from 2 frames using anisotropic smoothing. **Top Left: (a)** Left image. **Top Middle: (b)** Right image. **Top Right: (c)** Untextured frontal view. **Bottom Left: (d)** Untextured side view. **Bottom Right: (e)** Textured side view.

From the experiments, such as City-Hall, however, it is clear that large changes in view point can pose a problem to dense matching algorithms due to the large displacements and induced occlusions. While some of these effects are counterbalanced by the coupling between the optical flow and the fundamental matrix, sparse feature based methods still offer advantages for very wide baseline image pairs. Recently, however, optical flow methods have been proposed to overcome this limitation by either integrating feature matches [BBM09] or by applying a more global search [SPC09b]. As demonstrated, such methods could be extended by an epipolar term as well to improve accuracy in wide baseline scenarios. Also motion parallax induced by sudden and large changes in depth, forms a problem for traditional optical flow, since large jumps in the displacement field are difficult to capture in a global framework. Feature matching, in contrast, does not suffer from

parallax because it is essentially a local process that does not enforce spatial consistency.

Occlusions can form a second challenge. They are usually present in dense flow fields, but hardly pose a problem in feature matching where disappearing interest points will generally not be matched in the next frame. For our joint method we experienced that small occlusions only have a very limited influence because they are down weighted by the robust epipolar term. Additionally these regions are filled in by the smoothness term in accordance with the estimated stereo geometry. In wide baseline scenarios, occlusions will have a larger impact such that their explicit detection should be considered in combination with the afore mentioned techniques for large displacement optical flow. A strategy for occlusion handling for large displacements will be proposed in the next chapter.

In contrast to wide baseline images and occlusions, illumination changes are less problematic for dense estimation methods. While the SIFT descriptor is invariant under multiplicative and additive illumination changes by using normalised gradient information, similar concepts can be used by optical flow methods, e.g. by using photometric invariants ([MBW07]) or normalised cross correlation ([SPC09a, WPB10]).

Since our joint variational model relies on image sequences that allow a stable estimation of the fundamental matrix, the application is of course restricted to rigid scenes which are not dominated by moving objects.

5.3 Summary

In this chapter we have proposed a variational method for establishing image correspondences between two uncalibrated stereo images. In contrast to the large corpus of existing methods that make use of the given epipolar geometry in a calibrated stereo setting, we estimate the optical flow together with the unknown fundamental matrix. To this end, we introduced a novel joint energy formulation that allows us to estimate the fundamental matrix and the optical flow simultaneously, thereby ensuring that the set of recovered image correspondences is most consistent with the estimated stereo geometry, and vice versa. Concretely, we added an extra term to a classical optical flow functional which imposes the epipolar constraint in two directions: In one way it forces the optical flow to lie in accordance to the estimated epipolar lines and in the other way it fits the fundamental matrix to the current set of image correspondences. To minimise this joint energy, we proposed an alternating solution strategy that exhibits convergence in practice.

In our experiments we demonstrated that our joint model yields better optical flow results than approaches that do not estimate the epipolar geometry in the process. At the same time, a coupled solution of the fundamental matrix and the optical flow further improved the fundamental matrix estimation of the two-step method proposed in the previous chapter. Besides yielding a higher overall accuracy, a joint estimation of the fundamental matrix and the optical flow further exhibits a better parameter stability than their separate computation. From a conceptual point of view, we fuse the two steps of classical projective reconstruction in a single variational framework by simultaneously solving for both the camera motion and the scene structure. This was demonstrated by providing results for the 3D reconstruction of the depicted scene. We ended the chapter with a discussion of the limitations of variational stereo methods that are based on optical flow computation.

6 Scene Flow for Uncalibrated Stereo Sequences

Until now we have extensively studied variational methods for computing the optical flow and the epipolar geometry from two stereo images, i.e. approaches that recover the scene structure and the camera motion at a certain time instance. For many tasks in computer vision, however, it is essential to not only recover the three-dimensional structure of a scene, but also its three-dimensional displacement field, the so called *scene flow*. Scene flow represents the real, possibly non-rigid, 3D motion of objects in the scene, while optical flow only describes the projection of this motion on the 2D image plane. Since time varying depth information is necessary to determine 3D motion, the estimation of scene flow requires a *stereo sequence* that provides two views per time instance.

In this chapter we design a variational scene flow method for uncalibrated stereo sequences. We do this by integrating the spatial and temporal information from two stereo pairs in a global energy functional while jointly estimating the unknown stereo geometry at consecutive time steps. To this end, we extend the two-frame models for pure optical flow and uncalibrated stereo correspondence from Chapters 3 and 5 to a four-frame model that simultaneously solves for the scene flow and the scene structure.

Apart from this novel generalised model, that will be presented in Sec. 6.1, we make several additional contributions: First, we propose a regularisation strategy that penalises discontinuities in the different displacement fields separately. This makes sense, since motion and depth continuities do not necessarily coincide. Secondly, within the multi-resolution framework required to handle large displacements, we introduce a multi-dimensional tensor based notation for linearised constraints in Sec. 6.2. This notation allows to normalise the data and epipolar constraints such that deviations from model assumptions can be interpreted as geometrical distances. In case of the epipolar constraint, we show that this approach is equivalent to the minimisation of the epipolar distance or the Sampson error, two distance measures that have so far only been encountered in the context of feature based methods. Finally, we tackle some of the limitations of variational methods that were mentioned at the end of the previous chapter. By explicitly detecting occlusions we can exclude regions from the estimation process where the constancy assumptions do not hold, while the initialisation of the stereo correspondences with a dedicated method for large displacements renders our method applicable in large baseline scenarios. After discussing the minimisation of our functional in Sec. 6.3, we conclude this chapter with the experimental section 6.4 where we clearly demonstrate the benefits of our contributions. We show the favourable performance of our method in comparison with recent techniques for the rectified case and present several real-world examples.

6.1 A Scene Flow Model for Uncalibrated Stereo

To compute the scene flow, we consider the classical four-frame case depicted in Fig. 6.1. It consists of two consecutive image pairs of a synchronised stereo sequence: the left image $g_{1\mathrm{l}}(\boldsymbol{x})$ and the right image $g_{1\mathrm{r}}(\boldsymbol{x})$ at time t and the left image $g_{2\mathrm{l}}(\boldsymbol{x})$ and the right image $g_{2\mathrm{r}}(\boldsymbol{x})$ at time $t+1$. Here $\boldsymbol{x} = (x, y)^\top$ denotes the location in a rectangular image domain $\Omega \subset \mathbb{R}^2$ that is assumed to be the same for all images. In the following we assume that the sequence has been recorded by a single fixed stereo rig, i.e. there exists a single fundamental matrix F that describes the epipolar geometry of the stereo pairs at time t and $t+1$. Although the assumption of a common fundamental matrix for both stereo pairs is not a prerequisite for our method, it makes the subsequent modelling easier and more clear. Permitting a time-variant fundamental matrix would lead to a more general model and allows to extend the scene flow computation to arbitrarily moving unconstrained cameras.

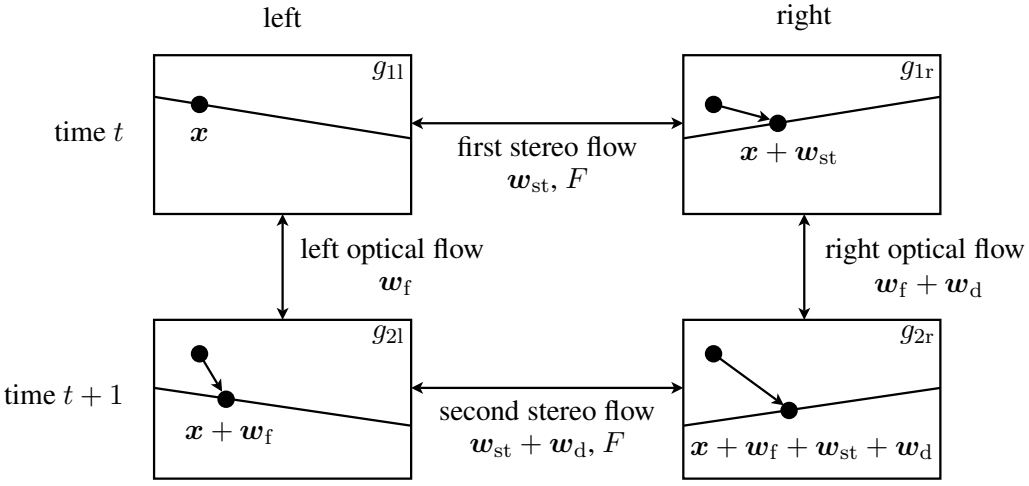

Fig. 6.1: The correspondences between the four frames of a binocular stereo sequence.

Parameterisation of Scene Flow, Scene Structure and Camera Geometry. In contrast to existing variational scene flow methods that start out from a rectified stereo sequence [HD07, WRV$^+$08], our approach assumes a general stereo geometry with unknown fundamental matrix. As a consequence, the stereo correspondences do not take on the form of a scalar valued disparity, but of a 2-dimensional displacement. To distinguish this displacement field from the unconstrained optical flow in temporal direction, we will refer to it from now on as *stereo flow*. All together, we consider four types of correspondences in our model: two optical flows between consecutive frames of the same camera (left and right camera) and two stereo flows between the left and right frame at the same time instance (time t and $t+1$). Exploiting the dependencies that exist between the images in Fig. 6.1, these correspondences can be parameterised by six unknown functions with respect to the reference image $g_{1\mathrm{l}}$: the first stereo flow $\boldsymbol{w}_{\mathrm{st}} = (u_{\mathrm{st}}, v_{\mathrm{st}})^\top$, the left optical flow $\boldsymbol{w}_{\mathrm{f}} = (u_{\mathrm{f}}, v_{\mathrm{f}})^\top$ and the difference flow $\boldsymbol{w}_{\mathrm{d}} = (u_{\mathrm{d}}, v_{\mathrm{d}})^\top$. This is a result of the fact that there are two different paths from the first frame $g_{1\mathrm{l}}$ to the fourth frame $g_{2\mathrm{r}}$: One goes over $g_{1\mathrm{r}}$ via the first stereo flow and the right optical flow and the other goes over $g_{2\mathrm{l}}$ via the left

optical flow and the second stereo flow. This redundancy becomes clear if we write down the corresponding image location in the fourth frame and simply reorder the terms as

$$\boldsymbol{x} \quad + \quad \underbrace{\boldsymbol{w}_{\mathrm{st}}}_{\substack{\text{first}\\ \text{stereo flow}}} \quad + \quad \underbrace{(\boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{d}})}_{\substack{\text{right}\\ \text{optical flow}}} = \boldsymbol{x} \quad + \quad \underbrace{\boldsymbol{w}_{\mathrm{f}}}_{\substack{\text{left}\\ \text{optical flow}}} \quad + \quad \underbrace{\boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}}_{\substack{\text{second}\\ \text{stereo flow}}} \quad . \tag{6.1}$$

The difference flow can thus be interpreted as either a change in optical flow or as a change in stereo flow, depending on whether one regards it as a correction in temporal or spatial direction. In our model we have furthermore seven degrees of freedom from the unknown fundamental matrix F, which restricts points in the left and right images to lie on corresponding epipolar lines, as shown in Fig. 6.1. These seven extra degrees of freedom arise from the fact that F is a 3×3 matrix of rank 2 that is defined up to a scale factor.

To demonstrate that the above parameterisation indeed forms a complete description of the three-dimensional scene flow, we recall that knowing the fundamental matrix is sufficient to recover a pair of camera projection matrices (P, P') for the left and the right image sequence (see Chapter 2). Together with the stereo flow $\boldsymbol{w}_{\mathrm{st}}$ at time t, these matrices allow us to reconstruct an image point from the reference frame as:

$$\begin{cases} P(F)\boldsymbol{X}_{\mathrm{h}}(t) &= \boldsymbol{x}_{\mathrm{h}} \\ P'(F)\boldsymbol{X}_{\mathrm{h}}(t) &= (\boldsymbol{x} + \boldsymbol{w}_{\mathrm{st}})_{\mathrm{h}} \end{cases} . \tag{6.2}$$

To obtain the same reconstruction at time $t + 1$, and thus the scene flow relative to the cameras, the left optical flow $\boldsymbol{w}_{\mathrm{f}}$ and the flow change $\boldsymbol{w}_{\mathrm{d}}$ have to be known additionally:

$$\begin{cases} P(F)\boldsymbol{X}_{\mathrm{h}}(t+1) &= (\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}} \\ P'(F)\boldsymbol{X}_{\mathrm{h}}(t+1) &= (\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}} \end{cases} . \tag{6.3}$$

Both systems above are easily solved in a total least squares sense [HZ00] and the scene flow can then be recovered as the difference in 3D location

$$\boldsymbol{V}(\boldsymbol{X}, t) = \boldsymbol{X}(t+1) - \boldsymbol{X}(t) \ . \tag{6.4}$$

In this chapter we moreover assume that the intrinsic camera parameters are given, such that the scene reconstruction and the scene flow can be recovered up to a scale in the first camera coordinate system via the essential matrix. In such case, the scene flow is not only a result of movement in the scene, but of the ego-motion of the camera system as well.

Energy Formulation. Since we are interested in a joint computation of the 3D motion, structure and geometry, that are parameterised by $(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}})^{\top}$ and F, we propose to minimise a global energy functional that combines the spatial and temporal information of the different views while imposing geometric consistency. This functional has the form

$$\mathcal{E}(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}}, F) =$$

$$\int_{\Omega} (\mathcal{E}_{\mathrm{D}}(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}}) + \mathcal{E}_{\mathrm{S}}(\nabla \boldsymbol{w}_{\mathrm{f}}, \nabla \boldsymbol{w}_{\mathrm{st}}, \nabla \boldsymbol{w}_{\mathrm{d}}) + \mathcal{E}_{\mathrm{E}}(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}}, F)) \ \mathrm{d}\boldsymbol{x} \ , \tag{6.5}$$

where $\mathcal{E}_{\mathrm{D}}$ is the data term that models the assumption that certain image features remain constant between the four frames, $\mathcal{E}_{\mathrm{E}}$ is the epipolar term that relates the stereo views by the unknown epipolar geometry, and $\mathcal{E}_{\mathrm{S}}$ is the smoothness term that assumes the solution to be piecewise smooth. In the following we will detail on the different terms.

6.1.1 The Data Constraints

Let us now derive the data constraints that model the relations between the four input images w.r.t. the reference image. The different connections that can be exploited between the four input frames are depicted in Fig. 6.2. They have been numbered according to the four main spatial and temporal directions, and the two transversal directions between the left and right images at different time steps.

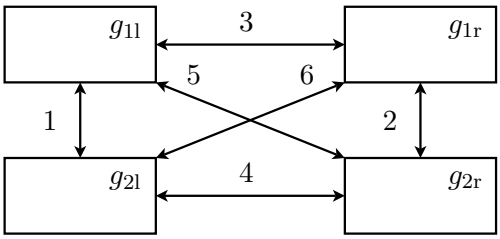

Fig. 6.2: The relations between the four frames of a binocular stereo sequence.

For simplicity, let us assume for now that the brightness of corresponding image points remains constant between the four frames [HS81]. Following the enumeration of the constraints in Fig. 6.2, we obtain the following expressions for the data constancy

$$\mathcal{E}_{\mathrm{D1}}(\boldsymbol{w}_{\mathrm{f}}) = \Psi\left(|g_{2\mathrm{l}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}) - g_{1\mathrm{l}}(\boldsymbol{x})|^2\right), \tag{6.6}$$

$$\mathcal{E}_{\mathrm{D2}}(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}}) = \Psi\left(|g_{2\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}}) - g_{1\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{st}})|^2\right), \tag{6.7}$$

$$\mathcal{E}_{\mathrm{D3}}(\boldsymbol{w}_{\mathrm{st}}) = \Psi\left(|g_{1\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{st}}) - g_{1\mathrm{l}}(\boldsymbol{x})|^2\right), \tag{6.8}$$

$$\mathcal{E}_{\mathrm{D4}}(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}}) = \Psi\left(|g_{2\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}}) - g_{2\mathrm{l}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}})|^2\right), \tag{6.9}$$

$$\mathcal{E}_{\mathrm{D5}}(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}}) = \Psi\left(|g_{2\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}}) - g_{1\mathrm{l}}(\boldsymbol{x})|^2\right), \tag{6.10}$$

$$\mathcal{E}_{\mathrm{D6}}(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}) = \Psi\left(|g_{2\mathrm{l}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}) - g_{1\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{st}})|^2\right) \quad . \tag{6.11}$$

The first two terms correspond to an optical flow constraint between two time instances, while the third and fourth term arise from a stereo correspondence at consecutive time steps. The last two terms express the brightness constancy for a mixture of optical flow and stereo flow. We choose to penalise all constraints separately since outliers for optical flow and stereo do not necessarily occur in the same location. As penalty a function Ψ we choose the regularised L_1 norm $\Psi(s^2) = \sqrt{s^2+\epsilon^2}$, with $\epsilon = 0.001$ as proposed e.g. in [BBPW04, PBB$^+$06]. In our final model we include the gradient constancy assumption to cope with varying illumination and extend the expressions above to RGB colour images. With these extensions, the second term (6.7), for example, becomes

$$\mathcal{E}_{\mathrm{D2}} = \Psi\left(\sum_{i=1}^{3}\Big(\quad |g_{2\mathrm{r}\,i}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}}) - g_{1\mathrm{r}\,i}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{st}})|^2 \right.\right.$$
$$\left.\left. + \gamma\,|\nabla g_{2\mathrm{r}\,i}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}}) - \nabla g_{1\mathrm{r}\,i}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{st}})|^2\Big)\right), \tag{6.12}$$

where $\gamma \geq 0$ is a weighting factor and g_1, g_2, and g_3 represent the three RGB colour channels. The constraints $\mathcal{E}_{\mathrm{D1}}$, $\mathcal{E}_{\mathrm{D3}}$, $\mathcal{E}_{\mathrm{D4}}$, $\mathcal{E}_{\mathrm{D5}}$ and $\mathcal{E}_{\mathrm{D6}}$ are extended in the same way.

It is clear that the six constraints listed above are redundant, since only three of them are required to compute the three unknown flows. Adding extra constraints, on the other hand, might render the estimation of the scene flow more robust. As in [ZK01, HD07, MS06], we exploit in our baseline method the first four image relations that correspond to the main correspondence directions. This gives rise to the data term

$$\mathcal{E}_{\mathrm{D}} = \mathcal{E}_{\mathrm{D1}} + \mathcal{E}_{\mathrm{D2}} + \mathcal{E}_{\mathrm{D3}} + \mathcal{E}_{\mathrm{D4}} \ . \tag{6.13}$$

Adding or dropping constraints from this data term is straightforward and will not influence the modelling steps that are subsequently explained. Although the evaluation of our scene flow method will be based on the four-constraint data term $\mathcal{E}_{\mathrm{D}}$, different combinations for the data constraints could be considered as well.

6.1.2 Occlusion Handling

The explicit detection of occluded regions should be an integral part of any state-of-the-art flow-based method and its effect on the estimation quality has already been the topic of discussion in the previous chapter. Occlusion handling will generally play a more important role in stereo methods than in optical flow methods because of the possibility of large displacements (wide baseline). At this stage we wish to take occlusions into account by detecting those points in the flow field for which there exists no correspondence in the other images, and thereby excluding them from the matching process.

In order to handle situations where parts of the scene become occluded due to motion or a change of camera viewpoint, we introduce *occlusions scores*. For instance, the score $o_{1\mathrm{r}}(\boldsymbol{x}) : \Omega \to \{0, 1\}$ takes on the value 1 for points in the reference image $g_{1\mathrm{l}}$ that are visible in $g_{1\mathrm{r}}$, and 0 otherwise. One of the most established techniques for computing occlusions scores between two images is a comparison of the forward and backward correspondences [Sha93, PPV94, ADPS02, BAS07]. The disadvantage of this approach, however, is that it is computationally expensive since it requires the double amount of unknown flow fields to be computed, namely an extra set of backward optical flows in addition to the three forward optical flows. Instead we will follow a more 3D centred approach and make direct use of the reconstructed scene. This is possible, since once the fundamental matrix has been estimated, the left and right camera matrices (P, P') can be recovered and each point in the reference frame can be reconstructed in 3D space.

Basic Idea. Making use of the reconstructed surface, we can label the values of $o_{1\mathrm{r}}$ by reprojecting all 3D points at time t back on the image $g_{1\mathrm{r}}$ using P': Of all the 3D points that project onto the same image location, only the point closest to the camera centre of P' will then be marked as visible. The corresponding entry in $o_{1\mathrm{r}}$ is set to 1. The 3D points that lie on the same optical ray, but are covered by the point closest to the camera centre, will then be marked as occluded. To recover the camera centre from a camera matrix, we can make use of formula (2.12). The occlusion scores $o_{2\mathrm{l}}(\boldsymbol{x})$ and $o_{2\mathrm{r}}(\boldsymbol{x})$ for the two remaining images are determined analogously by reprojecting all reconstructed 3D points at time $t + 1$ on $g_{2\mathrm{l}}$ and $g_{2\mathrm{r}}$ using P and P', respectively. Note that the occlusion score $o_{1\mathrm{l}}(\boldsymbol{x})$ of $g_{1\mathrm{l}}$ is 1 everywhere since all 3D points are visible in the first camera.

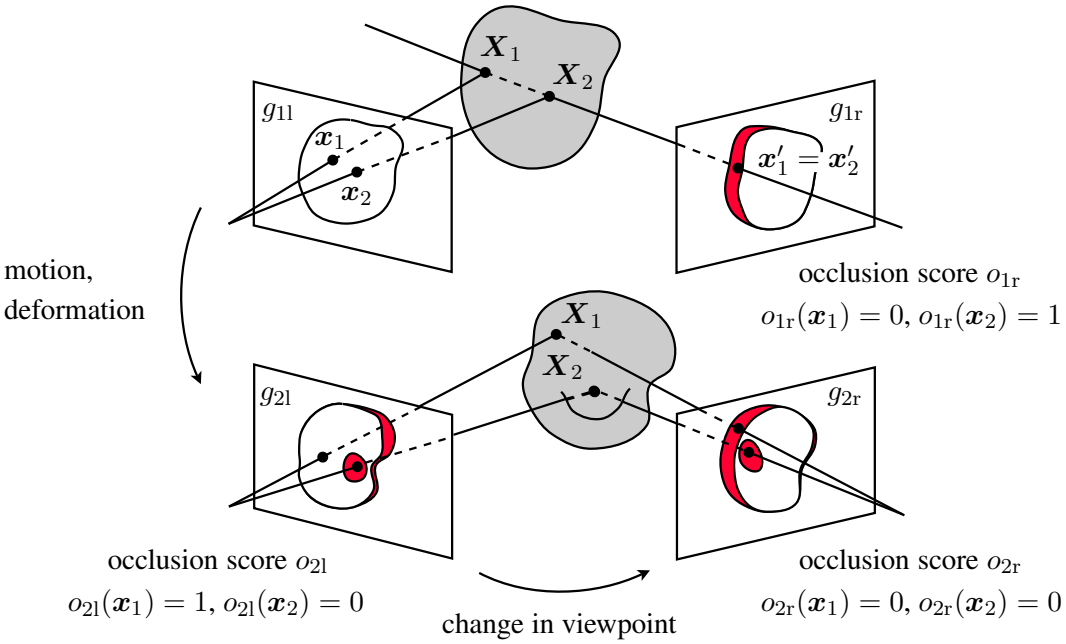

Fig. 6.3: The computation of the occlusion scores via Z-buffering.

The technique described here is also known as *Z-buffering* and its basic idea is illustrated in Fig. 6.3. Depicted are the projections of two reconstructed surface points on the four images. Regions that receive multiple projections are indicated in red. We can see that the reconstructed points $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are projected to the same location in the image $g_{1\mathrm{r}}$. Since $\boldsymbol{X}_2$ lies closest to the right camera centre, the occlusion score $o_{1\mathrm{r}}(\boldsymbol{x}_2)$ is set to 1 and $o_{1\mathrm{r}}(\boldsymbol{x}_1)$ is set to 0. The point $\boldsymbol{X}_1$ is thus occluded by the change in camera viewpoint. Due to object motion and surface deformation, the point $\boldsymbol{X}_2$ becomes covered at time $t+1$ and invisible to both cameras. This is depicted in the figure by a double folding of the surface. The occlusion scores $o_{2\mathrm{l}}(\boldsymbol{x}_2)$ and $o_{2\mathrm{r}}(\boldsymbol{x}_2)$ will therefore be 0. The point $\boldsymbol{X}_1$ remains occluded in $g_{2\mathrm{r}}$ by the change in camera viewpoint. All summed up, occlusions in $g_{1\mathrm{r}}$ will be due to a change in viewpoint, occlusions in $g_{2\mathrm{l}}$ due to scene motion, while occlusions in can $g_{2\mathrm{r}}$ arise from a combination of both. Note that there are no multiple projections in $g_{1\mathrm{l}}$ because the first image serves as the reference for the 3D reconstruction. We also make the remark that the regions depicted in red are not the actual occlusion scores $o_{1\mathrm{r}}$, $o_{2\mathrm{l}}$ and $o_{2\mathrm{r}}$, but the scores warped on $g_{1\mathrm{r}}$, $g_{2\mathrm{l}}$ and $g_{2\mathrm{r}}$ via the reconstructed surface.

A Data Term with Occlusion Scores. To incorporate occlusion handling in our model, the data terms (6.13) are multiplied by the occlusion scores to switch them off where the constancy assumptions can not be fulfilled. This yields the final data term

$$\mathcal{E}_{\mathrm{D}} = o_{2\mathrm{l}}\,\mathcal{E}_{\mathrm{D1}} + o_{1\mathrm{r}}\,o_{2\mathrm{r}}\,\mathcal{E}_{\mathrm{D2}} + o_{1\mathrm{r}}\,\mathcal{E}_{\mathrm{D3}} + o_{2\mathrm{l}}\,o_{2\mathrm{r}}\,\mathcal{E}_{\mathrm{D4}} \quad . \tag{6.14}$$

We point out that each term has to be multiplied by the occlusion scores of the images that appear in the according data constraint. This is required since the reappearance of a point in $g_{2\mathrm{r}}$ that is occluded in $g_{1\mathrm{r}}$ or $g_{2\mathrm{l}}$ goes unnoticed by the reference image. The same reasoning also applies to the data terms $\mathcal{E}_{\mathrm{D5}}$ and $\mathcal{E}_{\mathrm{D6}}$.

6.1.3 The Smoothness Constraints

Next we detail on the design of the smoothness term in our model. The task of the smoothness term is to regularise the problem in locations where the remaining terms do not guarantee a unique solution (the aperture problem) or to fill in information in the presence of outliers, such as noise and occlusions. Because in practical situations there often exists an overlap between the discontinuities of the three flow fields $\boldsymbol{w}_{\mathrm{f}}$, $\boldsymbol{w}_{\mathrm{st}}$ and $\boldsymbol{w}_{\mathrm{d}}$, some authors [HD07, WRV+08] suggest a joint piecewise smoothness assumption on all flow variables. With our method, however, we want to cover the general scenario where flow and stereo discontinuities do not necessarily coincide. This can be the case for example when in-plane motions occur, such as a the cracking of a smooth surface. Therefore we propose a separate penalisation of the flow gradients:

$$\mathcal{E}_{\mathrm{S1}}(\nabla \boldsymbol{w}_{\mathrm{f}}) = \Psi\left(|\nabla \boldsymbol{w}_{\mathrm{f}}|^2\right), \mathcal{E}_{\mathrm{S2}}(\nabla \boldsymbol{w}_{\mathrm{st}}) = \Psi\left(|\nabla \boldsymbol{w}_{\mathrm{st}}|^2\right), \mathcal{E}_{\mathrm{S3}}(\nabla \boldsymbol{w}_{\mathrm{d}}) = \Psi\left(|\nabla \boldsymbol{w}_{\mathrm{d}}|^2\right) \quad , \tag{6.15}$$

The penalisation via the subquadratic function Ψ, as defined before, equals a total variation (TV) regularisation. Combining the above terms gives rise to the smoothness term

$$\mathcal{E}_{\mathrm{S}} = \alpha_1 \mathcal{E}_{\mathrm{S1}} + \alpha_2 \mathcal{E}_{\mathrm{S2}} + \alpha_3 \mathcal{E}_{\mathrm{S3}} \quad , \tag{6.16}$$

where α_1, α_2, α_3 are positive weights that balance the smoothness assumptions for the three displacement fields. Although $\mathcal{E}_{\mathrm{S}}$ will serve as the smoothness term for our baseline scene flow algorithm, other choices for the regulariser exist. Since our method is based on variational optical flow, $\mathcal{E}_{\mathrm{S1}}$, $\mathcal{E}_{\mathrm{S2}}$ and $\mathcal{E}_{\mathrm{S3}}$ can be replaced by any regulariser from Tab. 3.1.

6.1.4 The Epipolar Constraints

Let us finally model the geometric relation between the left and right images of the stereo pairs $(g_{1\mathrm{l}}, g_{1\mathrm{r}})$ and $(g_{2\mathrm{l}}, g_{2\mathrm{r}})$. To this end we introduce two terms that relate the unknown flows and the fundamental matrix F via the respective epipolar constraints:

$$\mathcal{E}_{\mathrm{E1}}(\boldsymbol{w}_{\mathrm{st}}, F) = \Psi\left(\left((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{st}})_{\mathrm{h}}^{\top} F\, \boldsymbol{x}_{\mathrm{h}}\right)^2\right), \tag{6.17}$$

$$\mathcal{E}_{\mathrm{E2}}(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}}, F) = \Psi\left(\left((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^{\top} F\, (\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}}\right)^2\right) \quad . \tag{6.18}$$

Both terms $\mathcal{E}_{\mathrm{E1}}$ and $\mathcal{E}_{\mathrm{E2}}$ are soft constraints that penalise deviations of a point from its epipolar line. The use of the regularised L_1 norm Ψ increases the robustness of the estimation of F with respect to outliers. While the first epipolar term can be modelled completely in accordance with the soft constraint introduced in Sec. 5.1.1, the second epipolar constraint is much more complicated: Although it is linear in $\boldsymbol{w}_{\mathrm{st}}$ and $\boldsymbol{w}_{\mathrm{d}}$, it is quadratic with respect to the left optical flow $\boldsymbol{w}_{\mathrm{f}}$. This leads to higher order terms in the corresponding energy that make the minimisation difficult and raise questions about the convexity of the problem. To nevertheless obtain a linear expression in all flows, we propose to introduce an auxiliary flow variable $\boldsymbol{w}_{\mathrm{a}} = (u_{\mathrm{a}}, v_{\mathrm{a}})^{\top}$, which is assumed to be close to $\boldsymbol{w}_{\mathrm{f}}$, and split up the second epipolar constraint such that $\boldsymbol{w}_{\mathrm{f}}$ and $\boldsymbol{w}_{\mathrm{a}}$ take on symmetric roles. In this

way we can approximate the second term (6.18) by the expression

$$\begin{aligned}\mathcal{E}_{\mathrm{E2}}(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}}, \boldsymbol{w}_{\mathrm{a}}, F) \;\; = \;\; & \Psi\Big(\;\; \tfrac{1}{2}\big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^{\top} F\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}})_{\mathrm{h}}\big)^2 \\ & \qquad + \tfrac{1}{2}\big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^{\top} F\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}}\big)^2\Big) \\ & + \mu\left(|\boldsymbol{w}_{\mathrm{f}} - \boldsymbol{w}_{\mathrm{a}}|^2\right) \;\; , \end{aligned} \tag{6.19}$$

where μ is the weight of the additional similarity term that is required to couple $\boldsymbol{w}_{\mathrm{a}}$ and $\boldsymbol{w}_{\mathrm{f}}$. Introducing the weights β_1 and β_2 we obtain the final epipolar term

$$\mathcal{E}_{\mathrm{E}} = \beta_1\, \mathcal{E}_{\mathrm{E1}} + \beta_2\, \mathcal{E}_{\mathrm{E2}} \;\; . \tag{6.20}$$

As we did previously in connection with the soft epipolar constraint, we avoid the trivial solution by additionally imposing the constraint $\|F\|^2_{\mathrm{Frob}} = 1$ on the Frobenius norm of the fundamental matrix F. In practice, we will mostly choose $\beta_1 = \beta_2$, which gives an equal weight to both contributions in the epipolar term.

6.2 Linearisation and Constraint Normalisation

Substituting all data, epipolar and smoothness terms in the total energy (6.5), we obtain a functional that is rather complicated. Moreover, it is non-convex, since the unknown flows appear implicitly in the arguments of the data term. To resolve this problem, it is common to perform an incremental computation of the unknowns within a coarse-to-fine multi-scale approach. As for all methods in this thesis, this is done by a series of energies that approximate the original model on every resolution level. In the following we discuss how the corresponding energy for each level can be derived. Assuming that solutions $\boldsymbol{w}_{\mathrm{f}}$, $\boldsymbol{w}_{\mathrm{st}}$, $\boldsymbol{w}_{\mathrm{d}}$ and $\boldsymbol{w}_{\mathrm{a}}$ are available from a coarser scale, we aim at expressing the total energy in terms of the increments $d\boldsymbol{w}_{\mathrm{f}} = (du_{\mathrm{f}}, dv_{\mathrm{f}})$, $d\boldsymbol{w}_{\mathrm{st}} = (du_{\mathrm{st}}, dv_{\mathrm{st}})$, $d\boldsymbol{w}_{\mathrm{d}} = (du_{\mathrm{d}}, dv_{\mathrm{d}})$, and $d\boldsymbol{w}_{\mathrm{a}} = (du_{\mathrm{a}}, dv_{\mathrm{a}})$. This allows us to introduce a tensor notation in the data and epipolar terms which offers two advantages: (i) The convexity of the resulting energy functional in the flow increments becomes explicit, and (ii) a normalisation strategy can be applied that makes deviations from the model assumptions interpretable in a geometric way.

6.2.1 Linearisation in the Data Term

Let us discuss the differential form of the data term starting from the simplified data constraint of expression (6.7). Using a first order Taylor expansion to linearise this expression with respect to all flow increments we obtain the approximation

$$\begin{aligned} & g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}}) - g_{1\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}}) \\ \approx \;\; & g_{2\mathrm{r}} + \partial_x g_{2\mathrm{r}} \cdot (du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \partial_y g_{2\mathrm{r}} \cdot (dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) \\ & - g_{1\mathrm{r}} - \partial_x g_{1\mathrm{r}} \cdot (du_{\mathrm{st}}) - \partial_y g_{1\mathrm{r}} \cdot (dv_{\mathrm{st}}) \;\; . \end{aligned} \tag{6.21}$$

Rearranging the terms and using the following abbreviations

$$g_{2z} = g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) - g_{1\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{st}}) \,, \tag{6.22}$$

$$g_{2\mathrm{r}x} = \partial_x g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) \,, \qquad g_{2xz} = \partial_x g_{2z} \,, \tag{6.23}$$

$$g_{2\mathrm{r}y} = \partial_y g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) \,, \qquad g_{2yz} = \partial_y g_{2z} \;\; , \tag{6.24}$$

we can rewrite the linearised brightness difference in (6.21) as inner product

$$\boldsymbol{j}_2^\top \boldsymbol{d}_{\mathrm{h}} = g_{2\mathrm{r}x} du_{\mathrm{f}} + g_{2\mathrm{r}y} dv_{\mathrm{f}} + g_{2xz} du_{\mathrm{st}} + g_{2yz} dv_{\mathrm{st}} + g_{2\mathrm{r}x} du_{\mathrm{d}} + g_{2\mathrm{r}y} dv_{\mathrm{d}} + g_{2z} \;\; , \tag{6.25}$$

where the homogeneous motion increment and the extended gradient are defined as

$$\boldsymbol{d}_{\mathrm{h}} \;\; := \;\; (du_{\mathrm{f}}, dv_{\mathrm{f}}, du_{\mathrm{st}}, dv_{\mathrm{st}}, du_{\mathrm{d}}, dv_{\mathrm{d}}, 1)^\top \tag{6.26}$$

$$\boldsymbol{j}_2 \;\; := \;\; (g_{2\mathrm{r}x}, g_{2\mathrm{r}y}, g_{2xz}, g_{2yz}, g_{2\mathrm{r}x}, g_{2\mathrm{r}y}, g_{2z})^\top \tag{6.27}$$

The equation $\boldsymbol{j}_2^\top \boldsymbol{d}_{\mathrm{h}} = 0$ can be seen as a multi-dimensional extension of the classical optical flow constraint (OFC) (3.30) and expresses that the point $(du_{\mathrm{f}}, dv_{\mathrm{f}}, du_{\mathrm{st}}, dv_{\mathrm{st}}, du_{\mathrm{d}}, dv_{\mathrm{d}})$ has to lie on the hyperplane represented by $\boldsymbol{j}_2$. As in the case of two-dimensional optical flow, the multi-dimensional OFC does not suffice to determine the flows uniquely and imposing a constancy on the gradient, as in expression (6.12), can provide extra constraints to disambiguate the solution. By introducing the additional abbreviations

$$g_{2\mathrm{r}xx} = \partial_{xx} g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) \,, \qquad g_{2xxz} = \partial_{xx} g_{2z} \,, \tag{6.28}$$

$$g_{2\mathrm{r}xy} = \partial_{xy} g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) \,, \qquad g_{2xyz} = \partial_{xy} g_{2z} \,, \tag{6.29}$$

$$g_{2\mathrm{r}yy} = \partial_{yy} g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) \,, \qquad g_{2yyz} = \partial_{yy} g_{2z} \;\; , \tag{6.30}$$

and defining the vectors

$$\boldsymbol{j}_{2x} \;\; := \;\; \partial_x \, \boldsymbol{j}_2 = (g_{2\mathrm{r}xx}, g_{2\mathrm{r}xy}, g_{2xxz}, g_{2xyz}, g_{2\mathrm{r}xx}, g_{2\mathrm{r}xy}, g_{2xz})^\top, \tag{6.31}$$

$$\boldsymbol{j}_{2y} \;\; := \;\; \partial_y \, \boldsymbol{j}_2 = (g_{2\mathrm{r}xy}, g_{2\mathrm{r}yy}, g_{2xyz}, g_{2yyz}, g_{2\mathrm{r}xy}, g_{2\mathrm{r}yy}, g_{2zy})^\top \;\; , \tag{6.32}$$

the two gradient component differences can also be approximated by inner products as

$$\partial_x g_{2\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+d\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+d\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}}+d\boldsymbol{w}_{\mathrm{d}}) - \partial_x g_{1\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{st}}+d\boldsymbol{w}_{\mathrm{st}}) \approx \boldsymbol{j}_{2x}^\top \boldsymbol{d}_{\mathrm{h}}, \tag{6.33}$$

$$\partial_y g_{2\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+d\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+d\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}}+d\boldsymbol{w}_{\mathrm{d}}) - \partial_x g_{1\mathrm{r}}(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{st}}+d\boldsymbol{w}_{\mathrm{st}}) \approx \boldsymbol{j}_{2y}^\top \boldsymbol{d}_{\mathrm{h}}. \tag{6.34}$$

The linearised gradient constancy assumption can thus be expressed by the two linear equations $\boldsymbol{j}_{2x}^\top \boldsymbol{d}_{\mathrm{h}} = 0$ and $\boldsymbol{j}_{2y}^\top \boldsymbol{d}_{\mathrm{h}} = 0$.

Inserting the three linearised constraints as squared arguments into the penaliser Ψ of expression (6.12), yields the robustified quadratic form

$$\mathcal{E}_{\mathrm{D2}}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}) \;\; = \;\; \Psi\left((\, \boldsymbol{j}_2^\top \boldsymbol{d}_{\mathrm{h}}\,)^2 + \gamma\, (\boldsymbol{j}_{2x}^\top \, \boldsymbol{d}_{\mathrm{h}}\,)^2 + \gamma\, (\boldsymbol{j}_{2y}^\top \, \boldsymbol{d}_{\mathrm{h}}\,)^2 \right), \tag{6.35}$$

$$\;\; = \;\; \Psi\left(\boldsymbol{d}_{\mathrm{h}}^\top J_2 \, \boldsymbol{d}_{\mathrm{h}} \right) \;\; , \tag{6.36}$$

with a 7×7 matrix that provides coupling between all increments

$$J_2 := \boldsymbol{j}_2 \boldsymbol{j}_2^\top + \gamma \, \boldsymbol{j}_{2x} \boldsymbol{j}_{2x}^\top + \gamma \, \boldsymbol{j}_{2y} \boldsymbol{j}_{2y}^\top \ , \tag{6.37}$$

By analogy to the motion tensor notation in optical flow estimation, we denote by J_2 the *scene flow tensor*. The linearisation of the three remaining data constraints is carried out accordingly, and results in the 7×7 scene flow tensors J_1, J_3 and J_4. Missing dependencies between the variables give rise to zero tensor entries. Including the RGB colour channels is straightforward and leads to a weighted sum of the corresponding tensors.

Overcoming the Aperture Problem. As for the motion tensor in optical flow, the rank of the scene flow tensor will determine the type and the uniqueness of the solution. In the light of the higher dimensionality of the scene flow problem, however, a combination of the brightness and the gradient constancy will not as easily provide a full tensor rank as in the case of optical flow. Contrary to optical flow, on the other hand, scene flow estimation can exploit the constancy of image features between more than two frames. This results in a number of independent scene flow tensors, each one containing the data constraints that belong to the corresponding image pair. To overcome the aperture problem for scene flow, we can thus not only rely on the combined effects of the different constancy assumptions, but also on the integration of information from multiple scene flow tensor.

6.2.2 Treatment of the Epipolar Term

As pointed out in the previous chapter in relation to correspondence methods for uncalibrated stereo, the argument of the first epipolar term $(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}})_{\mathrm{h}}^\top F \boldsymbol{x}_{\mathrm{h}}$ is already linear in the increment $d\boldsymbol{w}_{\mathrm{st}}$. We can thus define the vector $\boldsymbol{d}_{1\mathrm{h}} = (du_{\mathrm{st}}, dv_{\mathrm{st}}, 1)^\top$ and write the argument of the first epipolar term (6.17) as a quadratic form

$$\mathcal{E}_{\mathrm{E1}}(d\boldsymbol{w}_{\mathrm{st}}, F) = \Psi \left(\boldsymbol{d}_{1\mathrm{h}}^\top E_1 \, \boldsymbol{d}_{1\mathrm{h}} \right) \ . \tag{6.38}$$

The corresponding epipolar tensor E_1 of size 3×3 is defined as in Eq. (5.24) and has entries that are related to the epipolar line $F \, \boldsymbol{x}_{\mathrm{h}}$ in the first right image $g_{1\mathrm{r}}$.

Although the treatment of the first epipolar term proceeds along a familiar line, care has to be taken with respect to symmetry when introducing the flow increments in the second epipolar term (6.19). Introducing the flow increments on both sides of the fundamental matrix will lead to mixed multiplicative terms that jeopardise the convexity of the energy and cause difficulties for the constraint normalisation in the next section. As a remedy, we apply an expansion of the argument such that the differential variant reads

$$\begin{aligned}
&\tfrac{1}{4} \big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^\top \, F \, (\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}})_{\mathrm{h}} \big)^2 \\
+&\tfrac{1}{4} \big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + d\boldsymbol{w}_{\mathrm{a}})_{\mathrm{h}}^\top \, F^\top \, (\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}} \big)^2 \\
+&\tfrac{1}{4} \big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + d\boldsymbol{w}_{\mathrm{a}} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^\top \, F \, (\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}} \big)^2 \\
+&\tfrac{1}{4} \big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}}^\top \, F^\top \, (\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}} \big)^2 \ .
\end{aligned} \tag{6.39}$$

Here we added the second and the fourth term with the transposed fundamental matrix to ensure a symmetrical treatment of the left and right flow increments. This is required since,

as opposed to the first epipolar constraint, variations can occur in both the left and the right image position. Since all terms of expression (6.39) are now linear in the increments, the second epipolar term can be written as

$$\begin{aligned}\mathcal{E}_{\mathrm{E2}}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, F) = \\ \Psi\left(\tfrac{1}{4}\,\boldsymbol{d}_{2\mathrm{h}}^{\top} E_2\,\boldsymbol{d}_{2\mathrm{h}} + \tfrac{1}{4}\,\boldsymbol{d}_{3\mathrm{h}}^{\top} E_3\,\boldsymbol{d}_{3\mathrm{h}} + \tfrac{1}{4}\,\boldsymbol{d}_{4\mathrm{h}}^{\top} E_4\,\boldsymbol{d}_{4\mathrm{h}} + \tfrac{1}{4}\,\boldsymbol{d}_{5\mathrm{h}}^{\top} E_5\,\boldsymbol{d}_{5\mathrm{h}}\right) \\ + \mu\left(|\boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} - \boldsymbol{w}_{\mathrm{a}} - d\boldsymbol{w}_{\mathrm{a}}|\right)^2 \quad , \end{aligned} \tag{6.40}$$

where we have defined the following homogeneous vectors:

$$\boldsymbol{d}_{2\mathrm{h}} = (du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}, dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}, 1)^{\top}\,, \quad \boldsymbol{d}_{3\mathrm{h}} = (du_{\mathrm{a}}, dv_{\mathrm{a}}, 1)^{\top}\,, \tag{6.41}$$

$$\boldsymbol{d}_{4\mathrm{h}} = (du_{\mathrm{a}} + du_{\mathrm{st}} + du_{\mathrm{d}}, dv_{\mathrm{a}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}, 1)^{\top}\,, \quad \boldsymbol{d}_{5\mathrm{h}} = (du_{\mathrm{f}}, dv_{\mathrm{f}}, 1)^{\top}\,. \tag{6.42}$$

As in the case of the first epipolar tensor, the entries of the four other epipolar tensors E_i, for $2 \leq i \leq 5$, are related to the corresponding epipolar lines. The second epipolar tensor, for example, can be written as

$$E_2 = (a', b', \tilde{q}')^{\top}(a', b', \tilde{q}')\ , \tag{6.43}$$

where a' and b' are the coefficients of the epipolar line $\boldsymbol{l}' = F\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}})_{\mathrm{h}}$ in the second right image $g_{2\mathrm{r}}$, and

$$\tilde{q}' = a'\,(u_{\mathrm{f}} + u_{\mathrm{st}} + u_{\mathrm{d}}) + b'\,(v_{\mathrm{f}} + v_{\mathrm{st}} + v_{\mathrm{d}}) + q'\ , \tag{6.44}$$

is the argument of the epipolar term in the warped image location $\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}}$. We further recall that q' is the scaled distance (5.9) of the point $\boldsymbol{x}+\boldsymbol{w}_{\mathrm{a}}$ to the line $\boldsymbol{l}'$. The third epipolar tensor, on the other hand, can be written as

$$E_3 = (a, b, \tilde{q})^{\top}(a, b, \tilde{q})\ , \tag{6.45}$$

where a and b are the coefficients of the epipolar line $\boldsymbol{l} = F^{\top}\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}$ in the second left image $g_{2\mathrm{l}}$, and

$$\tilde{q} = a\,u_{\mathrm{a}} + b\,v_{\mathrm{a}} + q\ , \tag{6.46}$$

with q the scaled distance of $\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}}$ to $\boldsymbol{l}$. The tensors E_4 and E_5 are defined identical to E_2 and E_3, but with the roles of $\boldsymbol{w}_{\mathrm{f}}$ and $\boldsymbol{w}_{\mathrm{a}}$ switched.

A Different Expansion in the Second Epipolar term. It is possible to expand the argument of the second epipolar term (6.19) in an alternative way. Although we will not consider this expansion for our final scene flow model, it will be of theoretical interest in the next section where we apply a normalisation to the epipolar constraint. This different expansion reads as follows

$$\begin{aligned}\tfrac{1}{2}\big(\quad & \tfrac{1}{2}\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^{\top}\, F\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}})_{\mathrm{h}} \\ & + \tfrac{1}{2}\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + d\boldsymbol{w}_{\mathrm{a}})_{\mathrm{h}}^{\top}\, F^{\top}\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}\big)^2 \\ +\tfrac{1}{2}\big(\quad & \tfrac{1}{2}\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + d\boldsymbol{w}_{\mathrm{a}} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^{\top}\, F\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}} \\ & + \tfrac{1}{2}\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}}^{\top}\, F^{\top}\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}\big)^2\ , \end{aligned} \tag{6.47}$$

where we have introduced the transposed terms under the squares this time. Since all terms are linear in the increments, the second epipolar term can now be written as

$$\begin{aligned}\mathcal{E}_{\mathrm{E2}}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, F) &= \Psi\left(\tfrac{1}{8}\,\boldsymbol{d}_{6\mathrm{h}}^{\top} E_6\,\boldsymbol{d}_{6\mathrm{h}} + \tfrac{1}{8}\,\boldsymbol{d}_{7\mathrm{h}}^{\top} E_7\,\boldsymbol{d}_{7\mathrm{h}}\right) \\ &\quad +\mu\left(|\boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} - \boldsymbol{w}_{\mathrm{a}} - d\boldsymbol{w}_{\mathrm{a}}|\right)^2 \quad , \end{aligned} \tag{6.48}$$

where we have defined the following homogeneous vectors:

$$\boldsymbol{d}_{6\mathrm{h}} = (du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}, dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}},\ du_{\mathrm{a}},\ dv_{\mathrm{a}}, 1)^{\top} , \tag{6.49}$$

$$\boldsymbol{d}_{7\mathrm{h}} = (du_{\mathrm{a}} + du_{\mathrm{st}} + du_{\mathrm{d}}, dv_{\mathrm{a}} + dv_{\mathrm{st}} + dv_{\mathrm{d}},\ du_{\mathrm{f}},\ dv_{\mathrm{f}}, 1)^{\top} . \tag{6.50}$$

The epipolar tensors E_6 and E_7 are now of size 5×5 and introduce an extra coupling between the unknowns. As before, the entries of these two tensors are related to the corresponding epipolar lines and E_6 can, for example, be written as

$$E_6 = (a', b', a, b, \tilde{q}' + \tilde{q})^{\top}(a', b', a, b, \tilde{q}' + \tilde{q}) \ , \tag{6.51}$$

where a, b, q, a', b' and q' have the same meaning as before. The tensor E_7 is defined identical to E_6, but with the roles of $\boldsymbol{w}_{\mathrm{f}}$ and $\boldsymbol{w}_{\mathrm{a}}$ exchanged.

6.2.3 Constraint Normalisation

In [SAH91, LV98, ZBW+09] it is demonstrated that the linearised brightness constancy assumption for optical flow can be interpreted geometrically as a weighted distance of the estimated flow to the line described by the OFC. Here we will extend this notion to scene flow and propose a normalisation of the constancy constraints that results in the minimisation of the actual geometrical distance. With the help of the tensor notation in the epipolar term, we apply the same normalisation strategy to the epipolar constraints and show the equivalence to two widely known geometrical error measures from computer vision.

6.2.3.1 Constraint Normalisation in the Data Term

Analogously to the OFC for optical flow, the multi-dimensional brightness constancy constraint (6.25) can be considered as a weighted distance of the scene flow to the hyperplane described by $\boldsymbol{j}_2^{\top}\boldsymbol{d}_{\mathrm{h}} = 0$. To see this, we rewrite $\boldsymbol{j}_2^{\top}\boldsymbol{d}_{\mathrm{h}}$ in terms of the scene flow increment $\boldsymbol{d} = (d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}})^{\top}$ and the hyperplane normal $\boldsymbol{n}$, which is given by the first six components of $\boldsymbol{j}_2$, i.e. $\boldsymbol{n} = (g_{2\mathrm{r}x}, g_{2\mathrm{r}y}, g_{2xz}, g_{2yz}, g_{2\mathrm{r}x}, g_{2\mathrm{r}y})^{\top}$. We then obtain

$$\boldsymbol{j}_2^{\top}\boldsymbol{d}_{\mathrm{h}} = \boldsymbol{n}^{\top}\boldsymbol{d} + g_{2z} \tag{6.52}$$

$$= \boldsymbol{n}^{\top}\left(\boldsymbol{d} + g_{2z}\frac{\boldsymbol{n}}{\|\boldsymbol{n}\|^2}\right) \tag{6.53}$$

$$= \|\boldsymbol{n}\| \underbrace{\frac{\boldsymbol{n}^{\top}}{\|\boldsymbol{n}\|}\left(\boldsymbol{d} - \boldsymbol{d}^{\perp}\right)}_{d} \ . \tag{6.54}$$

The term $\boldsymbol{d}^{\perp} = -g_{2z}\boldsymbol{n}/\|\boldsymbol{n}\|^2$ is called the *normal flow* and constitutes the scene flow component parallel to $\boldsymbol{n}$, i.e. normal to the hyperplane $\boldsymbol{j}_2$. For $\boldsymbol{n} \neq \boldsymbol{0}$[6], the normal flow

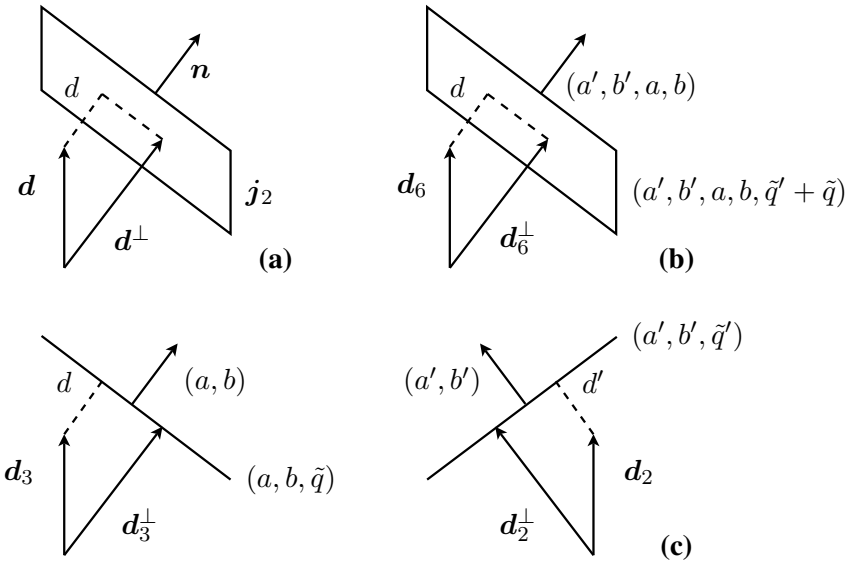

Fig. 6.4: The geometrical interpretation of the distance d for **(a)** the data constraint, **(b)** the Sampson error **(c)** the epipolar distance.

is the only scene flow component that can be determined from the brightness constancy constraint [1]. The value d is defined as the difference between the estimated scene flow $\boldsymbol{d}$ and the normal flow $\boldsymbol{d}^{\perp}$ in the direction of $\boldsymbol{n}$. In a geometrical sense, d can be interpreted as the distance of $\boldsymbol{d}$ to the hyperplane $\boldsymbol{j}_2$. This is depicted in Fig. 6.4 (a).

To penalise the actual distance to the hyperplane, we have to normalise the brightness constancy constraint by dividing it by the magnitude of the hyperplane normal. Now it becomes explicit why it is desirable to penalise the actual distance to the hyperplane: Unlike the original constraint, the distance d does not scale with the magnitude of the derivatives contained in $\boldsymbol{j}_2$. This prevents overweighting at unreliable structures such as noise or occlusions that typically manifest themselves in large image gradients. The same reasoning also applies to the linearised gradient constancy constraints (6.33) – (6.34), where we obtain a geometrically meaningful measure by normalising the constraints with the magnitudes of the hyperplane normals $\boldsymbol{n}_x$ and $\boldsymbol{n}_y$ of $\boldsymbol{j}_{2x}$ and $\boldsymbol{j}_{2y}$. The corresponding normalised quadratic form for the second data term $\mathcal{E}_{\mathrm{D2}}$ is now given by

$$\frac{1}{\|\boldsymbol{n}\|^2+\zeta^2}\,(\boldsymbol{j}_2^{\top}\boldsymbol{d}_{\mathrm{h}})^2 + \frac{\gamma}{\|\boldsymbol{n}_x\|^2+\zeta^2}\,(\boldsymbol{j}_{2x}^{\top}\boldsymbol{d}_{\mathrm{h}})^2 + \frac{\gamma}{\|\boldsymbol{n}_y\|^2+\zeta^2}\,(\boldsymbol{j}_{2y}^{\top}\boldsymbol{d}_{\mathrm{h}})^2 = \boldsymbol{d}_{\mathrm{h}}^{\top}\widehat{J}_2\,\boldsymbol{d}_{\mathrm{h}} \quad , \tag{6.55}$$

where $\zeta = 0.1$ is a constant that avoids division by zero, and $\widehat{J}_2$ denotes the normalised version of J_2. We apply the same normalisation strategy to the remaining data terms to produce the normalised scene flow tensors $\widehat{J}_1$, $\widehat{J}_3$ and $\widehat{J}_4$. The extension to colour images is straightforward and is done in accordance to [ZBW+09, ZBW11].

1. The normal flow for scene flow is an extension of the normal flow for optical flow, which is defined as the least squares solution of the OFC [LK81, BWS05].

6.2.3.2 Constraint Normalisation in the Epipolar Term

Our normalisation idea is, however, not restricted to the scene flow tensors only. Analogously to (6.55), we can derive normalisation factors for the epipolar tensors as well. Since each epipolar tensor only contains a single line constraint, the normalisation factors correspond to the two first terms of the tensor trace

$$\frac{1}{||\boldsymbol{n}_i||^2+\zeta^2}\,\boldsymbol{d}_{i\mathrm{h}}^\top E_i\,\boldsymbol{d}_{i\mathrm{h}} \;=\; \frac{1}{\sum_{j=1}^{2}(E_i)_{jj}+\zeta^2}\,\boldsymbol{d}_{i\mathrm{h}}^\top E_i\,\boldsymbol{d}_{i\mathrm{h}} \;=\; \boldsymbol{d}_{i\mathrm{h}}^\top \widehat{E}_i\,\boldsymbol{d}_{i\mathrm{h}}\;, \tag{6.56}$$

where $\widehat{E}_i$, for $1 \leq i \leq 5$ are the normalised epipolar tensors. More precisely, we can choose $\zeta = 0$ because the epipolar line coefficients, and thus the diagonal entries of the epipolar tensors, can not be zero at the same time. By normalising the quadratic forms in the epipolar terms this way, we obtain a widely used geometrical error measure from computer vision: the distance to the epipolar lines. This will be made clear next.

Equivalence to the Epipolar Distance. Writing out the normalised epipolar tensors $\widehat{E}_2$ and $\widehat{E}_3$, the first part of the second epipolar term reads (omitting the constants $1/4$)

$$\boldsymbol{d}_{2\mathrm{h}}^\top \widehat{E}_2\,\boldsymbol{d}_{2\mathrm{h}} + \boldsymbol{d}_{3\mathrm{h}}^\top \widehat{E}_3\,\boldsymbol{d}_{3\mathrm{h}} = \underbrace{\frac{1}{a'^2+b'^2}\,\boldsymbol{d}_{2\mathrm{h}}^\top E_2\,\boldsymbol{d}_{2\mathrm{h}}}_{d'^2} + \underbrace{\frac{1}{a^2+b^2}\,\boldsymbol{d}_{3\mathrm{h}}^\top E_3\,\boldsymbol{d}_{3\mathrm{h}}}_{d^2}\;. \tag{6.57}$$

By normalising the epipolar tensor E_2, we thus effectively penalise the orthogonal distance of the flow estimate $\boldsymbol{d}_2$ to the line in the right image that is described by $(a', b', \tilde{q}')$ and has as normal vector (a', b'). This is illustrated in Fig. 6.4 (c), where $\boldsymbol{d}_2^\perp$ denotes the normal flow orthogonal to $(a', b', \tilde{q}')$ and d' denotes the distance in the direction (a', b'). Normalisation of the tensor entries in E_3, results in the similar minimisation of the orthogonal distance of the flow estimate $\boldsymbol{d}_3$ to the line described by $(a, b, \tilde{q})$ in the left image. Its geometrical representation is given by the distance d on the left side of Fig. 6.4 (c).

If we now assume that the flow increments are infinitesimally small, the quadratic forms $\boldsymbol{d}_{2\mathrm{h}}^\top E_2\,\boldsymbol{d}_{2\mathrm{h}}$ and $\boldsymbol{d}_{3\mathrm{h}}^\top E_3\,\boldsymbol{d}_{3\mathrm{h}}$ will be equal and we can single out the normalisation factor

$$\frac{1}{a^2+b^2} + \frac{1}{a'^2+b'^2}\;. \tag{6.58}$$

This factor acts as a weighing of the original epipolar constraint and is equivalent to the squared weight (4.45) defined in the context of the feature based estimation method F1. With respect to the fundamental matrix, we are thus minimising the classical distance to the epipolar lines, which has been discussed at length in Sec. 4.2.3.1. The advantage of using the epipolar distance as an error measure, is that it is invariant to a scaling of the fundamental matrix and that it is independent of the actual image location.

By switching the roles of $\boldsymbol{w}_\mathrm{f}$ and $\boldsymbol{w}_\mathrm{a}$, the same weighing can also be applied to the part of the second epipolar term involving the tensors E_4 and E_5. As opposed to the second epipolar term, normalisation in the first epipolar term leads to a non-symmetric variant of the epipolar distance, since only variations in the right image position occur.

Equivalence to the Sampson Error. If we adopt the alternative expansion of the second epipolar constraint (6.47), we end up with two 5×5 epipolar tensors that can be normalised by the four first terms of their respective trace. The quadratic form involving the normalised epipolar tensor $\widehat{E}_6$ then reads

$$\boldsymbol{d}_{6\mathrm{h}}^{\top} \widehat{E}_6 \, \boldsymbol{d}_{6\mathrm{h}} = \underbrace{\frac{1}{a'^2 + b'^2 + a^2 + b^2}}_{d^2} \, \boldsymbol{d}_{6\mathrm{h}}^{\top} E_6 \, \boldsymbol{d}_{6\mathrm{h}} \quad . \tag{6.59}$$

Geometrically, the normalisation of the epipolar tensor E_6 is equivalent to minimising the orthogonal distance of the flow estimate $\boldsymbol{d}_6$ to the hyperplane described by the vector $(a', b', a, b, \tilde{q}' + \tilde{q})$. This distance d is sketched in Fig. 6.4 (b), where (a', b', a, b) denotes the hyperplane normal and $\boldsymbol{d}_6^{\perp}$ the normal flow.

In the above equation, the original quadratic constraint is weighed by the factor

$$\frac{1}{a^2 + b^2 + a'^2 + b'^2} \quad . \tag{6.60}$$

This factor is equivalent to the squared weight (4.57) of the Sampson error. The Sampson error is derived as a first order approximation to the geometric distance measure that is minimised by the feature based estimation method F2 discussed in Sec. 4.2.3.2. By applying this type of constraint normalisation in the epipolar term, we are thus able to linearly approximate the reprojection error, which has been shown to be optimal in a statistical sense [WAH93]. The advantages of the Sampson error are the same as those of the epipolar distance, but unlike the epipolar distance, it does not represent an actual distance measure in the image plane. Additionally, the minimisation of the Sampson error leads to an update of the complete 4D vector $\boldsymbol{d}_6$, whereas the increments in $\boldsymbol{d}_2$ and $\boldsymbol{d}_3$ are corrected separately in the left and the right image in case of the epipolar distance. The same normalisation strategy as discussed here can also be applied to the epipolar tensor E_7. This results in an identical weight as (6.60), but computed with a changed order of $\boldsymbol{w}_{\mathrm{f}}$ and $\boldsymbol{w}_{\mathrm{a}}$.

6.3 Minimisation and Numerical Solution

By combining all terms derived in Sec. 6.2, we obtain the following differential form of our energy that has to be minimised at each level of the coarse-to-fine approach:

$$\begin{aligned}
&\mathcal{E}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, F) = \\
&\int_{\Omega} \Bigg(o_{2\mathrm{l}} \, \Psi\Big(\boldsymbol{d}_{\mathrm{h}}^{\top} \widehat{J}_1 \, \boldsymbol{d}_{\mathrm{h}}\Big) + o_{1\mathrm{r}} \, o_{2\mathrm{r}} \, \Psi\Big(\boldsymbol{d}_{\mathrm{h}}^{\top} \widehat{J}_2 \, \boldsymbol{d}_{\mathrm{h}}\Big) + o_{1\mathrm{r}} \, \Psi\Big(\boldsymbol{d}_{\mathrm{h}}^{\top} \widehat{J}_3 \, \boldsymbol{d}_{\mathrm{h}}\Big) + o_{2\mathrm{l}} \, o_{2\mathrm{r}} \, \Psi\Big(\boldsymbol{d}_{\mathrm{h}}^{\top} \widehat{J}_4 \, \boldsymbol{d}_{\mathrm{h}}\Big) \\
&\quad + \alpha_1 \, \Psi\big(|\nabla(\boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}})|^2\big) + \alpha_2 \, \Psi\big(|\nabla(\boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}})|^2\big) + \alpha_3 \, \Psi\big(|\nabla(\boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}})|^2\big) \\
&\quad + \beta_1 \, \Psi\Big(\boldsymbol{d}_{1\mathrm{h}}^{\top} \widehat{E}_1 \, \boldsymbol{d}_{1\mathrm{h}}\Big) + \beta_2 \, \Psi\Big(\tfrac{1}{4} \left(\boldsymbol{d}_{2\mathrm{h}}^{\top} \widehat{E}_2 \, \boldsymbol{d}_{2\mathrm{h}} + \boldsymbol{d}_{3\mathrm{h}}^{\top} \widehat{E}_3 \, \boldsymbol{d}_{3\mathrm{h}} + \boldsymbol{d}_{4\mathrm{h}}^{\top} \widehat{E}_4 \, \boldsymbol{d}_{4\mathrm{h}} + \boldsymbol{d}_{5\mathrm{h}}^{\top} \widehat{E}_5 \, \boldsymbol{d}_{5\mathrm{h}}\right)\Big) \\
&\quad + \beta_2 \, \mu \left(|\boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} - \boldsymbol{w}_{\mathrm{a}} - d\boldsymbol{w}_{\mathrm{a}}|^2\right) \Bigg) \mathrm{d}\boldsymbol{x} \quad .
\end{aligned} \tag{6.61}$$

Note that this energy is convex in the flow increments $d\boldsymbol{w}_{\mathrm{f}}$, $d\boldsymbol{w}_{\mathrm{st}}$, $d\boldsymbol{w}_{\mathrm{d}}$ and the auxiliary variable $d\boldsymbol{w}_{\mathrm{a}}$, since only squared arguments and convex penaliser functions are used. In order to minimise it under the given constraint $\|F\|_{\mathrm{Frob}}^2 = 1$, we follow an approach that was already adopted in the previous chapter and use the method of the Lagrange multipliers. We thus obtain the Lagrangian

$$\mathcal{L}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, \boldsymbol{f}, \lambda) = \mathcal{E}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, \boldsymbol{f}) + \lambda(1 - \boldsymbol{f}^\top \boldsymbol{f}) \quad , \quad (6.62)$$

where λ is the Lagrangian multiplier and $\boldsymbol{f}$ the 9×1 vector that parameterises F via its entries in a row-major ordering. This formulation suggests an alternating minimisation with respect to the scene flow and the fundamental matrix in two steps:

(I) Solving for the Scene Flow. Minimising the Lagrangian with respect to the flow increments leads to the corresponding Euler-Lagrange equations. By discretising them via finite difference approximations, one ends up with a nonlinear system of equations due to the robust function Ψ. To ensure fast convergence, we solve this system in a bidirectional multigrid framework based on a nonlinear point coupled Gauß-Seidel solver. In the coarse-to-fine pyramid we use a downsampling factor of $\eta = 0.9$, while the images are warped onto the reference image using Coons patches based on bicubic interpolation [Coo67].

(II) Solving for the Fundamental Matrix. Differentiation of the Lagrangian with respect to the fundamental matrix results in an eigenvalue problem that is nonlinear due to the robust function Ψ and the normalisation weights (6.56). To solve this eigenvalue problem, we apply a reweighted total least squares method in which the weights and the arguments of Ψ are fixed iteratively. We point out that this step of the minimisation determines the fundamental matrix from the dense correspondences of *two* stereo pairs. This has the potential to improve the estimation from a single stereo pair even further, since a larger amount of correspondence information is utilised.

The alternating computation of the flow increments and the fundamental matrix works as follows: The Euler-Lagrange equations are solved with a current estimate of the fundamental matrix. Using the newly computed flows, the fundamental matrix is then updated by solving the eigenvalue problem. We extract a pair of camera matrices and perform a dense scene reconstruction at times t and $t + 1$ by triangulation according to Eq. (6.2) – (6.3). After recomputing the occlusion scores for all images, the Euler-Lagrange equations are solved again, this time with the updated fundamental matrix. This iterative process is repeated until convergence of the solutions. We initialise the occlusion scores with 1 and the fundamental matrix with the zero matrix. The latter initialisation comes down to computing the scene flow in the first iteration with disabled epipolar constraints.

After this general overview of the solution strategy, we will now focus in more detail on the two steps of the alternating minimisation.

6.3.1 Solving for the Scene Flow

Minimising the energy $\mathcal{E}$ with respect to the scene flow and the fundamental matrix under the constraint $\|F\|^2_{\mathrm{Frob}} = 1$, means that we are looking for critical points of the Lagrangian $\mathcal{L}$. Critical points are formed by the tuples $(du^*_{\mathrm{f}}, dv^*_{\mathrm{f}}, du^*_{\mathrm{st}}, dv^*_{\mathrm{st}}, du^*_{\mathrm{d}}, dv^*_{\mathrm{d}}, du^*_{\mathrm{a}}, dv^*_{\mathrm{a}}, \boldsymbol{f}^*, \lambda^*)^\top$ for which the functional derivatives of $\mathcal{L}$ with respect to du_{f}, dv_{f}, du_{st}, dv_{st}, du_{d}, dv_{d}, du_{a} and dv_{a} and the derivatives of $\mathcal{L}$ with respect to $\boldsymbol{f}$ and λ vanish. The Euler-Lagrange equations of the flow increments are obtained by setting

$$\frac{\partial}{\partial\, du_p}\mathcal{L}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, \boldsymbol{f}, \lambda) \;=\; 0, \tag{6.63}$$

$$\frac{\partial}{\partial\, dv_p}\mathcal{L}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, \boldsymbol{f}, \lambda) \;=\; 0\ , \tag{6.64}$$

where the subscript p stands for either f, st, d or a. The above functional derivatives form a system of coupled differential equations that has to be solved for all flow increments on each resolution level of the coarse-to-fine pyramid. In the following we will discuss the aspects of their derivation, discretisation and numerical solution in more depth.

6.3.1.1 The Euler-Lagrange Equations

As an example and in order to prevent overloading our presentation, we will restrict ourselves to providing the Euler-Lagrange equations for the left optical flow increments du_{f} and dv_{f} only. The Euler-Lagrange equations for the remaining six flow increments du_{st}, dv_{st}, du_{d}, dv_{d}, du_{a} and dv_{a} are derived in a similar fashion and can be found for completeness as Eq. (A.3) - (A.8) in Appendix A.

For du_{f} and dv_{f} the Euler-Lagrange equations are given by

$$\begin{aligned} 0 \;=\;& o_{2\mathrm{l}}\,\partial_{du_{\mathrm{f}}}\mathcal{E}_{\mathrm{D1}} + o_{1\mathrm{r}}\,o_{2\mathrm{r}}\,\partial_{du_{\mathrm{f}}}\mathcal{E}_{\mathrm{D2}} + o_{2\mathrm{l}}\,o_{2\mathrm{r}}\,\partial_{du_{\mathrm{f}}}\mathcal{E}_{\mathrm{D4}} \\ & -\alpha_1\left(\partial_x\left(\partial_{du_{\mathrm{f}x}}\mathcal{E}_{\mathrm{S1}}\right) + \partial_y\left(\partial_{du_{\mathrm{f}y}}\mathcal{E}_{\mathrm{S1}}\right)\right) - \beta_2\,\partial_{du_{\mathrm{f}}}\mathcal{E}_{\mathrm{E2}}, \end{aligned} \tag{6.65}$$

$$\begin{aligned} 0 \;=\;& o_{2\mathrm{l}}\,\partial_{dv_{\mathrm{f}}}\mathcal{E}_{\mathrm{D1}} + o_{1\mathrm{r}}\,o_{2\mathrm{r}}\,\partial_{dv_{\mathrm{f}}}\mathcal{E}_{\mathrm{D2}} + o_{2\mathrm{l}}\,o_{2\mathrm{r}}\,\partial_{dv_{\mathrm{f}}}\mathcal{E}_{\mathrm{D4}} \\ & -\alpha_1\left(\partial_x\left(\partial_{dv_{\mathrm{f}x}}\mathcal{E}_{\mathrm{S1}}\right) + \partial_y\left(\partial_{dv_{\mathrm{f}y}}\mathcal{E}_{\mathrm{S1}}\right)\right) - \beta_2\,\partial_{dv_{\mathrm{f}}}\mathcal{E}_{\mathrm{E2}}\ , \end{aligned} \tag{6.66}$$

where we only consider those terms in the energy $\mathcal{E}$ that depend on $(du_{\mathrm{f}}, dv_{\mathrm{f}})$. After expanding the different terms, we can write the Euler-Lagrange equations as

$$\begin{aligned} 0 \;=\;& \Psi'_{\mathrm{D1}}\cdot\left(\widehat{J}_{111}\,du_{\mathrm{f}} + \widehat{J}_{112}\,dv_{\mathrm{f}} + \widehat{J}_{113}\,du_{\mathrm{st}} + \widehat{J}_{114}\,dv_{\mathrm{st}} + \widehat{J}_{115}\,du_{\mathrm{d}} + \widehat{J}_{116}\,dv_{\mathrm{d}} + \widehat{J}_{117}\right) \\ +\;& \Psi'_{\mathrm{D2}}\cdot\left(\widehat{J}_{211}\,du_{\mathrm{f}} + \widehat{J}_{212}\,dv_{\mathrm{f}} + \widehat{J}_{213}\,du_{\mathrm{st}} + \widehat{J}_{214}\,dv_{\mathrm{st}} + \widehat{J}_{215}\,du_{\mathrm{d}} + \widehat{J}_{216}\,dv_{\mathrm{d}} + \widehat{J}_{217}\right) \\ +\;& \Psi'_{\mathrm{D4}}\cdot\left(\widehat{J}_{411}\,du_{\mathrm{f}} + \widehat{J}_{412}\,dv_{\mathrm{f}} + \widehat{J}_{413}\,du_{\mathrm{st}} + \widehat{J}_{414}\,dv_{\mathrm{st}} + \widehat{J}_{415}\,du_{\mathrm{d}} + \widehat{J}_{416}\,dv_{\mathrm{d}} + \widehat{J}_{417}\right) \\ -\;& \alpha_1\,\mathrm{div}\Big(\Psi'_{\mathrm{S1}}\cdot\nabla(u_{\mathrm{f}} + du_{\mathrm{f}})\Big) \\ +\;& \beta_2\,\Psi'_{\mathrm{E2}}\cdot\left(\widehat{E}_{211}\,(du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{212}\,(dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{213}\right) \end{aligned}$$

$$
\begin{aligned}
&+ \quad \beta_2\, \Psi'_{\mathrm{E2}} \cdot \big(\widehat{E}_{511}\, du_{\mathrm{f}} + \widehat{E}_{512}\, dv_{\mathrm{f}} + \widehat{E}_{513}\big) \\
&+ \quad \mu \;\; (u_{\mathrm{f}} + du_{\mathrm{f}} - u_{\mathrm{a}} - du_{\mathrm{a}}), \qquad (6.67) \\
0 \;\; &= \quad \Psi'_{\mathrm{D1}} \cdot \big(\widehat{J}_{112}\, du_{\mathrm{f}} + \widehat{J}_{122}\, dv_{\mathrm{f}} + \widehat{J}_{123}\, du_{\mathrm{st}} + \widehat{J}_{124}\, dv_{\mathrm{st}} + \widehat{J}_{125}\, du_{\mathrm{d}} + \widehat{J}_{126}\, dv_{\mathrm{d}} + \widehat{J}_{127}\big) \\
&+ \quad \Psi'_{\mathrm{D2}} \cdot \big(\widehat{J}_{212}\, du_{\mathrm{f}} + \widehat{J}_{222}\, dv_{\mathrm{f}} + \widehat{J}_{223}\, du_{\mathrm{st}} + \widehat{J}_{224}\, dv_{\mathrm{st}} + \widehat{J}_{225}\, du_{\mathrm{d}} + \widehat{J}_{226}\, dv_{\mathrm{d}} + \widehat{J}_{227}\big) \\
&+ \quad \Psi'_{\mathrm{D4}} \cdot \big(\widehat{J}_{412}\, du_{\mathrm{f}} + \widehat{J}_{422}\, dv_{\mathrm{f}} + \widehat{J}_{423}\, du_{\mathrm{st}} + \widehat{J}_{424}\, dv_{\mathrm{st}} + \widehat{J}_{425}\, du_{\mathrm{d}} + \widehat{J}_{426}\, dv_{\mathrm{d}} + \widehat{J}_{427}\big) \\
&- \quad \alpha_1\, \mathrm{div}\Big(\Psi'_{\mathrm{S1}} \cdot \nabla(v_{\mathrm{f}} + dv_{\mathrm{f}})\Big) \\
&+ \quad \beta_2\, \Psi'_{\mathrm{E2}} \cdot \big(\widehat{E}_{212}\,(du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{222}\,(dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{223}\big) \\
&+ \quad \beta_2\, \Psi'_{\mathrm{E2}} \cdot \big(\widehat{E}_{512}\, du_{\mathrm{f}} + \widehat{E}_{522}\, dv_{\mathrm{f}} + \widehat{E}_{523}\big) \\
&+ \quad \mu \;\; (v_{\mathrm{f}} + dv_{\mathrm{f}} - v_{\mathrm{a}} - dv_{\mathrm{a}}) \;\; , \qquad (6.68)
\end{aligned}
$$

where $\widehat{J}_{1kl}$, $\widehat{J}_{2kl}$ and $\widehat{J}_{4kl}$, for $k, l \in \{1, 2, \ldots, 7\}$, denote the kl-th entry of the scene flow tensors $\widehat{J}_1$, $\widehat{J}_2$ and $\widehat{J}_4$ and $\widehat{E}_{2kl}$ and $\widehat{E}_{5kl}$, for $k, l \in \{1, 2, 3\}$, denote the kl-th entry of the epipolar tensors $\widehat{E}_2$ and $\widehat{E}_5$. The same notation is used for the scene flow tensor $\widehat{J}_3$ and epipolar tensors $\widehat{E}_1$, $\widehat{E}_3$ and $\widehat{E}_4$ that appear in the Euler-Lagrange equations of the remaining flow increments in Appendix A. The sign $\cdot$ further indicates the standard multiplication operator, while the abbreviations

$$
\Psi'_{\mathrm{D1}} = o_{2\mathrm{l}} \quad \Psi'\Big(\boldsymbol{d}_{\mathrm{h}}^{\top}\, \widehat{J}_1\, \boldsymbol{d}_{\mathrm{h}}\Big), \qquad \Psi'_{\mathrm{S1}} = \Psi'\big(\, |\nabla(\boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}})|^2 \,\big), \qquad (6.69)
$$

$$
\Psi'_{\mathrm{D2}} = o_{1\mathrm{r}}\, o_{2\mathrm{r}}\, \Psi'\Big(\boldsymbol{d}_{\mathrm{h}}^{\top}\, \widehat{J}_2\, \boldsymbol{d}_{\mathrm{h}}\Big), \qquad \Psi'_{\mathrm{S2}} = \Psi'\big(\, |\nabla(\boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}})|^2 \,\big), \qquad (6.70)
$$

$$
\Psi'_{\mathrm{D3}} = o_{1\mathrm{r}} \quad \Psi'\Big(\boldsymbol{d}_{\mathrm{h}}^{\top}\, \widehat{J}_3\, \boldsymbol{d}_{\mathrm{h}}\Big), \qquad \Psi'_{\mathrm{S3}} = \Psi'\big(\, |\nabla(\boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}})|^2 \,\big), \qquad (6.71)
$$

$$
\Psi'_{\mathrm{D4}} = o_{2\mathrm{l}}\, o_{2\mathrm{r}}\, \Psi'\Big(\boldsymbol{d}_{\mathrm{h}}^{\top}\, \widehat{J}_4\, \boldsymbol{d}_{\mathrm{h}}\Big), \qquad \Psi'_{\mathrm{E1}} = \Psi'\Big(\boldsymbol{d}_{1\mathrm{h}}^{\top} \widehat{E}_1\, \boldsymbol{d}_{1\mathrm{h}}\Big), \qquad (6.72)
$$

$$
\Psi'_{\mathrm{E2}} = \Psi'\Big(\tfrac{1}{4}\,\big(\boldsymbol{d}_{2\mathrm{h}}^{\top} \widehat{E}_2\, \boldsymbol{d}_{2\mathrm{h}} + \boldsymbol{d}_{3\mathrm{h}}^{\top} \widehat{E}_3\, \boldsymbol{d}_{3\mathrm{h}} + \boldsymbol{d}_{4\mathrm{h}}^{\top} \widehat{E}_4\, \boldsymbol{d}_{4\mathrm{h}} + \boldsymbol{d}_{5\mathrm{h}}^{\top} \widehat{E}_5\, \boldsymbol{d}_{5\mathrm{h}}\big)\Big) \;\; , \qquad (6.73)
$$

allow for a more compact notation in all equations. For notational convenience we also assume that the factor $\frac{1}{4}$, that arises from deriving the argument of the second epipolar term $\mathcal{E}_{\mathrm{E2}}$ with respect to each flow increment, is absorbed by the epipolar weight β_2. For the same reason we assume that the weight μ has already been scaled by β_2, such that it can be regarded as the sole weighting factor of the similarity term. The values of the weights β_2 and μ that are given in the experimental section will with this respect refer to those of the Euler-Lagrange equations presented here and not to those of the original energy.

In interpreting the Euler-Lagrange equations, one has to keep in mind that some of the scene flow tensor entries might be zero. This is for instance the case for the entries 13, 14, 15 and 16 of the tensor $\widehat{J}_1$ in the equation of the optical flow increment du_{f}. These zero entries imply that there exists no coupling between the increment du_{f} and the increments du_{st}, dv_{st}, du_{d} and dv_{d} via the first scene flow tensor. Since the (absence of) coupling between the unknowns is reciprocal, the entries of $\widehat{J}_1$ will consequently not appear in the Euler-Lagrange equations of du_{st}, dv_{st}, du_{d} and dv_{d}.

It is important to note that the Euler-Lagrange equations (6.67) - (6.68) are non-linear in the unknowns du_{f} and dv_{f} due to the derivative of the non-quadratic penaliser Ψ in the expressions Ψ'_{D1}, Ψ'_{D2}, Ψ'_{D4}, Ψ'_{S1} and Ψ'_{E2}. These equations nevertheless have a unique solution because of the strict convexity of Ψ and can therefore be solved by any globally convergent algorithm. Before arriving at a suitable choice for the solver, however, we must first discuss the properties of the discrete equation system.

6.3.1.2 The Discrete Euler-Lagrange Equations

After deriving the set of continuous Euler-Lagrange equations for all flow increments, we now turn to the discretisation of the resulting system.

Point-Based Notation. Before writing out the discrete Euler-Lagrange equations in a point-based form, we will first agree on the following notational conventions. Let the subscript p stand for either f, st, d or a. Then we can denote by $[\,du_p\,]_{i,j}$ and $[\,dv_p\,]_{i,j}$ the approximations of the flow increments du_p and dv_p at the pixel location (i,j), with $i = 1,\ldots,n_x$ and $j = 1,\ldots,n_y$. We further represent by $[\,\boldsymbol{d}\,]_{i,j}$ and $[\,\boldsymbol{d}_m\,]_{i,j}$, for $m = 1,2,\ldots,5$, the discrete versions of the vectors $\boldsymbol{d}$ and $\boldsymbol{d}_m$. It thus holds for instance that

$$[\,\boldsymbol{d}\,]_{i,j} = \left([\,du_{\mathrm{f}}\,]_{i,j}, [\,dv_{\mathrm{f}}\,]_{i,j}, [\,du_{\mathrm{st}}\,]_{i,j}, [\,dv_{\mathrm{st}}\,]_{i,j}, [\,du_{\mathrm{d}}\,]_{i,j}, [\,dv_{\mathrm{d}}\,]_{i,j}\right)^{\top} \quad , \tag{6.74}$$

while $[\,\boldsymbol{d}_m\,]_{i,j}$ are defined on an equivalent basis.

We are now in a position to define the approximations of the non-linear expressions (6.69) - (6.73) in the pixel location (i,j). These are given by the following abbreviations

$$[\Psi'_{\mathrm{D}m}]_{i,j} = \frac{[\,o_m\,]_{i,j}}{2\sqrt{\left([\,\boldsymbol{d}\,]^{\top}_{i,j},1\right)[\widehat{J}_m\,]_{i,j}\left([\,\boldsymbol{d}\,]^{\top}_{i,j},1\right)^{\top}+\epsilon^2}} \tag{6.75}$$

$$\text{for } m = 1,2,3,4 \text{ and } \begin{cases} [\,o_1\,]_{i,j} := [\,o_{2\mathrm{l}}\,]_{i,j} \\ [\,o_2\,]_{i,j} := [\,o_{1\mathrm{r}}\,]_{i,j}[\,o_{2\mathrm{r}}\,]_{i,j} \\ [\,o_3\,]_{i,j} := [\,o_{1\mathrm{r}}\,]_{i,j} \\ [\,o_4\,]_{i,j} := [\,o_{2\mathrm{l}}\,]_{i,j}[\,o_{2\mathrm{r}}\,]_{i,j} \end{cases},$$

$$[\Psi'_{\mathrm{S}m}]_{i,j} = \frac{1}{2\sqrt{\left|\mathrm{D}^2\left([\,u_p\,]_{i,j}+[\,du_p\,]_{i,j}\right)\right|^2+\left|\mathrm{D}^2\left([\,v_p\,]_{i,j}+[\,dv_p\,]_{i,j}\right)\right|^2+\epsilon^2}} \tag{6.76}$$

$$\text{for } (m,p) \in \{(1,\mathrm{f}),(2,\mathrm{st}),(3,\mathrm{d})\},$$

$$[\Psi'_{\mathrm{E1}}]_{i,j} = \frac{1}{2\sqrt{\left([\,\boldsymbol{d}_1\,]^{\top}_{i,j},1\right)[\widehat{E}_1\,]_{i,j}\left([\,\boldsymbol{d}_1\,]^{\top}_{i,j},1\right)^{\top}+\epsilon^2}}, \tag{6.77}$$

$$[\Psi'_{\mathrm{E2}}]_{i,j} = \frac{1}{2\sqrt{\frac{1}{4}\sum_{m=2}^{5}\left([\,\boldsymbol{d}_m\,]^{\top}_{i,j},1\right)[\widehat{E}_m\,]_{i,j}\left([\,\boldsymbol{d}_m\,]^{\top}_{i,j},1\right)^{\top}+\epsilon^2}} \quad , \tag{6.78}$$

for Ψ' the TV diffusivity (3.25). Additionally, $[\widehat{J}_m\,]_{i,j}$, $[\widehat{E}_1\,]_{i,j}$ and $[\widehat{E}_m\,]_{i,j}$ stand for the discrete approximation of the tensors $\widehat{J}_m$, $\widehat{E}_1$ and $\widehat{E}_m$, for m the enumeration index of

the corresponding definition above. The discrete gradient magnitude operator $\left|\mathrm{D}^2\ \left([z]_{i,j}\right)\right|$ is defined as in Table 3.2. Using these abbreviations, the point-based Euler-Lagrange equations for $[\,du_{\mathrm{f}}\,]_{i,j}$ and $[\,dv_{\mathrm{f}}\,]_{i,j}$ read

$$
\begin{aligned}
0 \;=\; & \sum_{m\in\{1,2,4\}} \Big(\quad [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m11}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m12}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m13}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m14}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m15}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m16}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j}\Big) \\
& - \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}1}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}1}]_{i,j}\right)}{2} \frac{\left([\,du_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}} - [\,du_{\mathrm{f}}\,]_{i,j}\right)}{h^2} \\
& + \sum_{m\in\{2,5\}} \Big(\beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m11}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j}\Big) \\
& + \qquad \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{211}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& + \qquad \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{211}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \\
& + \mu\,[\,du_{\mathrm{f}}\,]_{i,j} - \mu\,[\,du_{\mathrm{a}}\,]_{i,j} \\
& + \sum_{m\in\{1,2,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}\,[J_{m17}]_{i,j} + \sum_{m\in\{2,5\}} \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m13}]_{i,j} \\
& - \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}1}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}1}]_{i,j}\right)}{2} \frac{\left([\,u_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}} - [\,u_{\mathrm{f}}\,]_{i,j}\right)}{h^2} \\
& + \mu\,[\,u_{\mathrm{f}}\,]_{i,j} - \mu\,[\,u_{\mathrm{a}}\,]_{i,j}, \qquad (6.79)
\end{aligned}
$$

$$
\begin{aligned}
0 \;=\; & \sum_{m\in\{1,2,4\}} \Big(\quad [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m12}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m22}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m23}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m24}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m25}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m26}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j}\Big) \\
& - \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}1}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}1}]_{i,j}\right)}{2} \frac{\left([\,dv_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}} - [\,dv_{\mathrm{f}}\,]_{i,j}\right)}{h^2} \\
& + \sum_{m\in\{2,5\}} \Big(\beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m22}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j}\Big) \\
& + \qquad \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{222}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& + \qquad \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{222}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \\
& + \mu\,[\,dv_{\mathrm{f}}\,]_{i,j} - \mu\,[\,dv_{\mathrm{a}}\,]_{i,j} \\
& + \sum_{m\in\{1,2,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}\,[J_{m27}]_{i,j} + \sum_{m\in\{2,5\}} \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m23}]_{i,j} \\
& - \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S}1}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}1}]_{i,j}\right)}{2} \frac{\left([\,v_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}} - [\,v_{\mathrm{f}}\,]_{i,j}\right)}{h^2}
\end{aligned}
$$

$$+ \mu \, [\, v_{\mathrm{f}} \,]_{i,j} - \mu \, [\, v_{\mathrm{a}} \,]_{i,j} \;\; , \tag{6.80}$$

for $i = 1, \ldots, n_x$ and $j = 1, \ldots, n_y$ and $\mathcal{N}_l(i,j)$ the set of the two neighbours of the pixel (i,j) in the axial direction $l \in \{x, y\}$. The discrete equations for $[\, du_{\mathrm{st}} \,]_{i,j}$, $[\, dv_{\mathrm{st}} \,]_{i,j}$, $[\, du_{\mathrm{d}} \,]_{i,j}$, $[\, dv_{\mathrm{d}} \,]_{i,j}$,$[\, du_{\mathrm{a}} \,]_{i,j}$, and $[\, dv_{\mathrm{a}} \,]_{i,j}$ can be found in Appendix A. In all these equations we have collected the terms that depend directly on the flow increments and those terms that do not depend on them, or only implicitly via the expressions (6.75) - (6.78). This ordering will make it easier to visualise the structure of the discrete system and the interdependencies between the unknowns, which will be highlighted next.

General Structure. Extending the compact notations of Sec. 3.2.2.3 to scene flow, we arrange the unknowns via a row-major ordering in eight $n \times 1$ vectors

$$\mathbf{du}_p \;:=\; \big([\, du_p \,]_{1,1}, \ldots, [\, du_p \,]_{n_x,n_y}\big)^\top \quad \text{for } p \in \{\mathrm{f}, \mathrm{st}, \mathrm{d}, \mathrm{a}\}, \tag{6.81}$$

$$\mathbf{dv}_p \;:=\; \big([\, dv_p \,]_{1,1}, \ldots, [\, dv_p \,]_{n_x,n_y}\big)^\top \quad \text{for } p \in \{\mathrm{f}, \mathrm{st}, \mathrm{d}, \mathrm{a}\} \;\; . \tag{6.82}$$

We will reserve the $6n \times 1$ vector $\mathbf{d}$ for collecting all scene flow components, i.e.

$$\mathbf{d} := \big(\mathbf{du}_{\mathrm{f}}^\top, \mathbf{dv}_{\mathrm{f}}^\top, \mathbf{du}_{\mathrm{st}}^\top, \mathbf{dv}_{\mathrm{st}}^\top, \mathbf{du}_{\mathrm{d}}^\top, \mathbf{dv}_{\mathrm{d}}^\top\big)^\top \;\; , \tag{6.83}$$

and the $8n \times 1$ parameter vector $\boldsymbol{p} := (\mathbf{d}^\top, \mathbf{du}_{\mathrm{a}}^\top, \mathbf{dv}_{\mathrm{a}}^\top)^\top$ for collecting all unknown flow increments. We also introduce the following short forms

$$\mathbf{u}_p \;:=\; \big([\, u \,]_{1,1}, \ldots, [\, u \,]_{n_x,n_y}\big)^\top \quad \text{for } p \in \{\mathrm{f}, \mathrm{st}, \mathrm{d}, \mathrm{a}\}, \tag{6.84}$$

$$\mathbf{v}_p \;:=\; \big([\, v \,]_{1,1}, \ldots, [\, v \,]_{n_x,n_y}\big)^\top \quad \text{for } p \in \{\mathrm{f}, \mathrm{st}, \mathrm{d}, \mathrm{a}\}, \tag{6.85}$$

$$\widehat{\mathbf{j}}_{mkl} \;:=\; \Big([\, \widehat{J}_{mkl} \,]_{1,1}, \ldots, [\, \widehat{J}_{mkl} \,]_{n_x,n_y}\Big)^\top \quad \text{for } \begin{cases} m & \in \{1,2,3,4\} \\ k,l & \in \{1,2,\ldots,7\} \end{cases} , \tag{6.86}$$

$$\widehat{\mathbf{e}}_{mkl} \;:=\; \Big([\, \widehat{E}_{mkl} \,]_{1,1}, \ldots, [\, \widehat{E}_{mkl} \,]_{n_x,n_y}\Big)^\top \quad \text{for } \begin{cases} m & \in \{1,2\} \\ k,l & \in \{1,2,3\} \end{cases} , \tag{6.87}$$

$$\widehat{\mathbf{J}}_{mkl} \;:=\; \mathrm{diag}\,\Big([\, \widehat{J}_{mkl} \,]_{1,1}, \ldots, [\, \widehat{J}_{mkl} \,]_{n_x,n_y}\Big) \quad \text{for } \begin{cases} m & \in \{1,2,3,4\} \\ k,l & \in \{1,2,\ldots,7\} \end{cases} , \tag{6.88}$$

$$\widehat{\mathbf{E}}_{mkl} \;:=\; \mathrm{diag}\,\Big([\, \widehat{E}_{mkl} \,]_{1,1}, \ldots, [\, \widehat{E}_{mkl} \,]_{n_x,n_y}\Big) \quad \text{for } \begin{cases} m & \in \{1,2\} \\ k,l & \in \{1,2,3\} \end{cases} , \tag{6.89}$$

$$\boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d}) \;:=\; \mathrm{diag}\,\big([\Psi'_{\mathrm{D}m}]_{1,1}, \ldots, [\Psi'_{\mathrm{D}m}]_{n_x,n_y}\big) \quad \text{for } m \in \{1,2,3,4\} \tag{6.90}$$

$$\boldsymbol{\Psi}'_{\mathrm{E}m}(\boldsymbol{p}) \;:=\; \mathrm{diag}\,\big([\Psi'_{\mathrm{E}m}]_{1,1}, \ldots, [\Psi'_{\mathrm{E}m}]_{n_x,n_y}\big) \quad \text{for } m \in \{1,2\} \;\; . \tag{6.91}$$

This allows us now to write the discrete Euler-Lagrange equations as a non-linear homogeneous system, given by

$$A(\boldsymbol{p})\,\boldsymbol{p} + \boldsymbol{c}(\boldsymbol{p}) = \mathbf{0}^{8n} \quad . \tag{6.92}$$

For this system, the partitioned $8n \times 8n$ matrix $A(\boldsymbol{p})$ and $8n \times 1$ vector $\boldsymbol{c}(\boldsymbol{p})$ are written out in full in Eq. (6.93) and Eq. (6.94) on pages 131 and 132.

Despite its larger size and the greater number of unknowns, this system is very reminiscent of the discrete Euler-Lagrange equations that we have already encountered for pure optical flow in Chapter 3 and for our joint estimation method in Chapter 5. The part of the matrix A that arises from the data terms is by construction symmetric positive semidefinite, because the robust weights and the diagonal entries of the scene flow tensors are per definition positive. The same reasoning can be applied to the parts of A that derive from the epipolar terms and the epipolar tensors. As defined before, $\mathbf{L}(\mathbf{du}_p, \mathbf{dv}_p)$, for $p \in \{\mathrm{f}, \mathrm{st}, \mathrm{d}\}$, is the discrete variant of the differential operator of the divergence expression and takes on the form of a sparse pentadiagonal $n \times n$ matrix for isotropic TV regularisation. Since this matrix is symmetric negative semidefinite, the overall system matrix A is a sparse symmetric positive semidefinite matrix as well. As for optical flow, one can even show that for non-constant images A turns out to be positive definite, which means that it is invertible and that a unique solution exists [Bru06, SBK10].

When taking a closer look at the dependencies between the different unknowns, we can distinguish a point-wise coupling between the different unknowns $\mathbf{du}_p$ and $\mathbf{dv}_p$ via the off-diagonal blocks in A. This type of coupling stems from the data terms and the epipolar terms in our model. A second type of dependency is formed by the neighbourhood coupling within the vectors $\mathbf{du}_p$ and $\mathbf{dv}_p$, which is induced by the differential operator $\mathbf{L}(\mathbf{du}_p, \mathbf{dv}_p)$. Finally, all unknowns are also linked implicitly in a point-wise manner via the joint arguments of the non-linear expressions in $\boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})$, $\boldsymbol{\Psi}'_{\mathrm{E}m}(\boldsymbol{p})$ and $\mathbf{L}(\mathbf{du}_p, \mathbf{dv}_p)$.

This discussion makes it clear that for the solution of the Euler-Lagrange system we can fall back on the block relaxation techniques that we have been using so far in the context of optical flow computation. By doing so we will make full use of the coupling between the unknowns and at the same time allow for the extension on a multigrid platform. The choice and implementation of the concrete solver will be discussed next.

$$
\begin{aligned}
A(\boldsymbol{p}) \coloneqq & \left(\begin{array}{cc|cc|c}
\sum\limits_{\{1,2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m11} & \sum\limits_{\{1,2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m12} & \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m13} & \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m14} & 0^{2n} \\
\sum\limits_{\{1,2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m12} & \sum\limits_{\{1,2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m22} & \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m23} & \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m24} & 0^{2n} \\
\hline
 & & \sum\limits_{\{2,3,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m33} & \sum\limits_{\{2,3,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m34} & 0^{2n} \\
 & & \sum\limits_{\{2,3,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m34} & \sum\limits_{\{2,3,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{J}}_{m44} & 0^{2n} \\
\hline
 & & & & 0^{2n}
\end{array}\right) \\
& - \left(\begin{array}{cc|cc|c}
\alpha_1\mathbf{L}(\mathbf{du}_{\mathrm{f}},\mathbf{dv}_{\mathrm{f}}) & 0^n & 0^{2n} & 0^{2n} & 0^{2n} \\
0^n & \alpha_1\mathbf{L}(\mathbf{du}_{\mathrm{f}},\mathbf{dv}_{\mathrm{f}}) & & & \\
\hline
 & & \alpha_2\mathbf{L}(\mathbf{du}_{\mathrm{st}},\mathbf{dv}_{\mathrm{st}}) & 0^n & \\
 & & 0^n & \alpha_2\mathbf{L}(\mathbf{du}_{\mathrm{st}},\mathbf{dv}_{\mathrm{st}}) & \\
\hline
 & & & & 0^{2n}
\end{array}\right) \\
& + \beta_1 \left(\begin{array}{cc|cc|c}
 & & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{211} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{212} & 0^{2n} \\
 & & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{212} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{222} & 0^{2n} \\
\hline
\boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{211} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{212} & \boldsymbol{\Psi}'_{\mathrm{E}1}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{111} & \boldsymbol{\Psi}'_{\mathrm{E}1}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{112} & 0^{2n} \\
\boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{212} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{222} & \boldsymbol{\Psi}'_{\mathrm{E}1}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{112} & \boldsymbol{\Psi}'_{\mathrm{E}1}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{122} & 0^{2n} \\
\hline
 & & & & 0^{2n}
\end{array}\right) \\
& + \beta_2 \left(\begin{array}{cc|cc|c}
\sum\limits_{\{2,5\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m11} & \sum\limits_{\{2,5\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m12} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{211} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{212} & 0^{2n} \\
\sum\limits_{\{2,5\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m12} & \sum\limits_{\{2,5\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m22} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{212} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{222} & 0^{2n} \\
\hline
 & & \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m11} & \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m12} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{411} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{412} & 0^{2n} \\
 & & \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m12} & \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m22} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{412} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{422} & 0^{2n} \\
\hline
 & & & & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{411} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{412} & 0^{2n} \\
 & & & & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{412} & \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{422} & 0^{2n} \\
\hline
 & & & & \sum\limits_{\{3,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m11} & \sum\limits_{\{3,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m12} & 0^{2n} \\
 & & & & \sum\limits_{\{3,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m12} & \sum\limits_{\{3,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{E}}_{m22} & 0^{2n} \\
\hline
 & & & & & & 0^{2n}
\end{array}\right)
\end{aligned}
$$

$$
\mathbf{c}(\boldsymbol{p}) := \begin{pmatrix} \sum\limits_{\{1,2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{j}}_{m17} \\ \sum\limits_{\{1,2,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{j}}_{m12} \\ \hline \sum\limits_{\{2,3,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{j}}_{m37} \\ \sum\limits_{\{2,3,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{j}}_{m47} \\ \hline \sum\limits_{\{2,3,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{j}}_{m57} \\ \sum\limits_{\{2,3,4\}} \boldsymbol{\Psi}'_{\mathrm{D}m}(\mathbf{d})\,\widehat{\mathbf{j}}_{m67} \\ \hline \mathbf{0}^{2n} \end{pmatrix} - \begin{pmatrix} \alpha_1\mathbf{L}(\mathbf{du}_{\mathrm{f}}, \mathbf{dv}_{\mathrm{f}}) & 0^n & 0^{2n} & 0^{2n} \\ 0^n & \alpha_1\mathbf{L}(\mathbf{du}_{\mathrm{f}}, \mathbf{dv}_{\mathrm{f}}) & & \\ & & \alpha_2\mathbf{L}(\mathbf{du}_{\mathrm{st}}, \mathbf{dv}_{\mathrm{st}}) & 0^{2n} \\ & & 0^n & \alpha_2\mathbf{L}(\mathbf{du}_{\mathrm{st}}, \mathbf{dv}_{\mathrm{st}}) \\ & & & \alpha_3\mathbf{L}(\mathbf{du}_{\mathrm{d}}, \mathbf{dv}_{\mathrm{d}}) \quad 0^n \\ & & & 0^n \quad \alpha_3\mathbf{L}(\mathbf{du}_{\mathrm{d}}, \mathbf{dv}_{\mathrm{d}}) \end{pmatrix} \begin{pmatrix} \mathbf{u}_{\mathrm{f}} \\ \mathbf{v}_{\mathrm{f}} \\ \mathbf{u}_{\mathrm{st}} \\ \mathbf{v}_{\mathrm{st}} \\ \mathbf{u}_{\mathrm{d}} \\ \mathbf{v}_{\mathrm{d}} \\ \mathbf{u}_{\mathrm{a}} \\ \mathbf{v}_{\mathrm{a}} \end{pmatrix} + \beta_1 \begin{pmatrix} \mathbf{0}^{2n} \\ \boldsymbol{\Psi}'_{\mathrm{E}1}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{13} \\ \boldsymbol{\Psi}'_{\mathrm{E}1}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{23} \\ \mathbf{0}^{2n} \\ \mathbf{0}^{2n} \end{pmatrix} \tag{6.93}
$$

$$
+\beta_2 \begin{pmatrix} \sum\limits_{\{2,5\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{m13} \\ \sum\limits_{\{2,5\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{m23} \\ \hline \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{m13} \\ \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{m23} \\ \hline \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{m13} \\ \sum\limits_{\{2,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{m23} \\ \hline \sum\limits_{\{3,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{m13} \\ \sum\limits_{\{3,4\}} \boldsymbol{\Psi}'_{\mathrm{E}2}(\boldsymbol{p})\,\widehat{\mathbf{e}}_{m23} \end{pmatrix} + \mu \begin{pmatrix} I^n & 0^n & 0^{2n} & 0^{2n} & -I^n & 0^n \\ 0^n & I^n & & & 0^n & -I^n \\ & & 0^{2n} & 0^{2n} & 0^{2n} & \\ & & 0^{2n} & 0^{2n} & & 0^{2n} \\ & & 0^{2n} & & I^n & 0^n \\ & & & 0^{2n} & 0^n & I^n \end{pmatrix} \begin{pmatrix} \mathbf{u}_{\mathrm{f}} \\ \mathbf{v}_{\mathrm{f}} \\ \mathbf{u}_{\mathrm{st}} \\ \mathbf{v}_{\mathrm{st}} \\ \mathbf{u}_{\mathrm{d}} \\ \mathbf{v}_{\mathrm{d}} \\ \mathbf{u}_{\mathrm{a}} \\ \mathbf{v}_{\mathrm{a}} \end{pmatrix} \tag{6.94}
$$

$$
+\mu \begin{pmatrix} I^n & 0^n & 0^{2n} & 0^{2n} & -I^n & 0^n \\ 0^n & I^n & & & 0^n & -I^n \\ & & 0^{2n} & 0^{2n} & 0^{2n} & \\ & & 0^{2n} & 0^{2n} & & 0^{2n} \\ & & 0^{2n} & & I^n & 0^n \\ & & & 0^{2n} & 0^n & I^n \end{pmatrix}
$$

6.3.1.3 Solution by Coupled Point Gauß-Seidel Relaxation

Our solution will be based on a coupled point Gauß-Seidel relaxation with frozen coefficients. Hereby, we fix in a first step the entries of the matrix A and the vector $\boldsymbol{c}$ by evaluating all arguments of the diffusivity Ψ for an already known solution. By doing so, the non-linear system (6.92) will turn into a linear one. In a second step we traverse the pixels in a row-major ordering and update all flow increments simultaneously in a block-wise fashion. In essence this comes down to solving a scaled down 8×8 version of the system (6.92) in each pixel: The fixed entries of A will provide the components of the system matrix, while the entries of $\boldsymbol{c}$ will end up in the right hand side. The corresponding update instruction then reads in pixel-based notation

$$\begin{pmatrix} [\,du_{\mathrm{f}}\,]_{i,j}^{k+1} \\ [\,dv_{\mathrm{f}}\,]_{i,j}^{k+1} \\ \vdots \\ [\,du_{\mathrm{a}}\,]_{i,j}^{k+1} \\ [\,dv_{\mathrm{a}}\,]_{i,j}^{k+1} \end{pmatrix} = \begin{pmatrix} [M_{11}]_{i,j}^{k} & [M_{12}]_{i,j}^{k} & \cdots & [M_{17}]_{i,j}^{k} & [M_{18}]_{i,j}^{k} \\ [M_{12}]_{i,j}^{k} & [M_{22}]_{i,j}^{k} & \cdots & [M_{27}]_{i,j}^{k} & [M_{28}]_{i,j}^{k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ [M_{17}]_{i,j}^{k} & [M_{27}]_{i,j}^{k} & \cdots & [M_{77}]_{i,j}^{k} & [M_{78}]_{i,j}^{k} \\ [M_{18}]_{i,j}^{k} & [M_{28}]_{i,j}^{k} & \cdots & [M_{78}]_{i,j}^{k} & [M_{88}]_{i,j}^{k} \end{pmatrix}^{-1} \begin{pmatrix} [r_1]_{i,j}^{k} \\ [r_2]_{i,j}^{k} \\ \vdots \\ [r_7]_{i,j}^{k} \\ [r_8]_{i,j}^{k} \end{pmatrix}, \tag{6.95}$$

for $i = 1, \ldots, n_x$ and $j = 1, \ldots, n_y$, with matrix entries

$$\begin{aligned}
[M_{11}]_{i,j}^{k} &= \sum_{m \in \{1,2,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}^{k} \, [\widehat{J}_{m11}]_{i,j} + \alpha \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S1}}]_{\tilde{i},\tilde{j}}^{k} + [\Psi'_{\mathrm{S1}}]_{i,j}^{k} \right)}{2h^2} \\
&\quad + \beta_2 \sum_{m \in \{2,5\}} [\Psi'_{\mathrm{E2}}]_{i,j}^{k} \, [\widehat{E}_{m11}]_{i,j} + \mu, && (6.96) \\
[M_{12}]_{i,j}^{k} &= \sum_{m \in \{1,2,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}^{k} \, [\widehat{J}_{m12}]_{i,j} + \beta_2 \sum_{m \in \{2,5\}} [\Psi'_{\mathrm{E2}}]_{i,j}^{k} \, [\widehat{E}_{m12}]_{i,j}, && (6.97) \\
[M_{13}]_{i,j}^{k} &= \sum_{m \in \{1,2,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}^{k} \, [\widehat{J}_{m13}]_{i,j} + \beta_2 \, [\Psi'_{\mathrm{E2}}]_{i,j}^{k} \, [\widehat{E}_{211}]_{i,j}, && (6.98) \\
[M_{14}]_{i,j}^{k} &= \sum_{m \in \{1,2,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}^{k} \, [\widehat{J}_{m14}]_{i,j} + \beta_2 \, [\Psi'_{\mathrm{E2}}]_{i,j}^{k} \, [\widehat{E}_{212}]_{i,j}, && (6.99) \\
&\vdots \\
[M_{17}]_{i,j}^{k} &= -\,\mu, && (6.100) \\
[M_{18}]_{i,j}^{k} &= 0, && (6.101) \\
&\vdots
\end{aligned}$$

and right hand side

$$\begin{aligned}
[r_1]_{i,j}^{k} &= - \sum_{m \in \{1,2,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}^{k} \, [J_{m17}]_{i,j} + \sum_{m \in \{2,5\}} \beta_2 \, [\Psi'_{\mathrm{E2}}]_{i,j}^{k} \, [\widehat{E}_{m13}]_{i,j} \\
&\quad + \alpha_1 \sum_{l \in \{x,y\}} \sum_{(\tilde{i},\tilde{j}) \in \mathcal{N}_l^-(i,j)} \frac{\left([\Psi'_{\mathrm{S1}}]_{\tilde{i},\tilde{j}}^{k} + [\Psi'_{\mathrm{S1}}]_{i,j}^{k} \right)}{2h^2} [\,du_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}}^{k+1}
\end{aligned}$$

$$
\begin{aligned}
&+\ \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l^+(i,j)} \frac{\left([\Psi'_{\mathrm{S1}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S1}}]^k_{i,j}\right)}{2h^2} [\, du_{\mathrm{f}}\,]^k_{\tilde{i},\tilde{j}} \\
&+\ \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\left([\Psi'_{\mathrm{S1}}]^k_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S1}}]^k_{i,j}\right)}{2h^2} \left([\, u_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}} - [\, u_{\mathrm{f}}\,]_{i,j}\right) \\
&-\ \mu\,[\, u_{\mathrm{f}}\,]_{i,j} + \mu\,[\, u_{\mathrm{a}}\,]_{i,j}, \qquad (6.102)
\end{aligned}
$$

$\vdots$

where $\mathcal{N}^-(i,j)$ is the set of neighbours of pixel (i,j) in the direction of the l-axis that have already been updated and $\mathcal{N}^+(i,j)$ the set of neighbours that still need to be updated. While not all expressions have been written out in full, the remaining entries of the system matrix and the right hand side are derived in a comparable way.

6.3.2 Solving for the Fundamental Matrix

To minimise the energy $\mathcal{E}$ with respect to the fundamental matrix, we look for critical points of the Lagrangian $\mathcal{L}$ where the derivatives with respect to $\boldsymbol{f}$ and λ vanish, i.e.

$$\nabla_{\boldsymbol{f}}\,\mathcal{L}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, \boldsymbol{f}, \lambda) \;=\; \mathbf{0}, \qquad (6.103)$$

$$\frac{\partial}{\partial\lambda}\mathcal{L}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, \boldsymbol{f}, \lambda) \;=\; 0\ . \qquad (6.104)$$

For these derivatives only the two epipolar terms in $\mathcal{E}$ have to be taken into consideration, because none of the data terms and none of the smoothness terms depend on $\boldsymbol{f}$. In the subsequent derivation we will narrow the Lagrangian down to the first part of the second epipolar constraint to make the modus operandi clear. With respect to the second epipolar term $\mathcal{E}_{\mathrm{E2}}$, we can set the first part of the argument

$$\boldsymbol{d}_{2\mathrm{h}}^\top \widehat{E}_2\, \boldsymbol{d}_{2\mathrm{h}} + \boldsymbol{d}_{3\mathrm{h}}^\top \widehat{E}_3\, \boldsymbol{d}_{3\mathrm{h}} = w^2(\boldsymbol{f})\,(\boldsymbol{s}^\top \boldsymbol{f})^2\ , \qquad (6.105)$$

where $\boldsymbol{s}$ is the 9×1 constraint vector (4.3) that depends on $\boldsymbol{w}_{\mathrm{f}}$, $\boldsymbol{w}_{\mathrm{st}}$, $\boldsymbol{w}_{\mathrm{d}}$, $\boldsymbol{w}_{\mathrm{a}}$ and their increments, and the factor $w^2(\boldsymbol{f})$ is the squared weight defined in Eq. (6.58). If we further stick to the definition of Ψ'_{E2} given in Eq. (6.73), the above derivatives for the first part of the second epipolar term give rise to the eigenvalue problem

$$\left(\int_\Omega \Psi'_{\mathrm{E2}} \cdot w^2(\boldsymbol{f})\, \boldsymbol{s}\boldsymbol{s}^\top \,\mathrm{d}x\,\mathrm{d}y - \lambda I\right)\boldsymbol{f} =: (M-\lambda I)\,\boldsymbol{f} \;=\; \mathbf{0}^9 \qquad (6.106)$$

$$1 - \|\boldsymbol{f}\|^2 \;=\; 0\ . \qquad (6.107)$$

The system matrix M is a symmetric positive definite matrix with entries

$$m_{i,j} = \int_\Omega \Psi'_{\mathrm{E2}} \cdot w^2(\boldsymbol{f})\, s_i s_j \,\mathrm{d}x\,\mathrm{d}y\ , \qquad (6.108)$$

for $1 \le i,j \le 9$ and s_i being the i-th component of $\boldsymbol{s}$. In much the same way, the first epipolar term and the second part of the second epipolar term involving the tensors $\widehat{E}_4$

and $\widehat{E}_5$ will both give rise to a similar eigenvalue problem. To find a common fundamental matrix, the three system matrices have to be combined in a final set of eigenvalue equations, which can be solved in the familiar total least squares setting. Since the use of the epipolar distance requires a reasonable estimate for the fundamental matrix, we bootstrap the estimation by means of the standard 8-point algorithm with data normalisation.

As we did before in the context of the optical flow based computation of the fundamental matrix, we exclude points from the estimation process that are warped outside the image. Since we now have additional information about occluded regions within the image, we can do the same for points that have an occlusion score of 0. The occlusion scores might even be added explicitly to the epipolar terms in our global energy.

Relation to the Method F1. As already pointed out before and now made explicit via the above derivation, our minimisation with respect to the fundamental matrix is equivalent to solving the eigenvalue problem (4.47) of the feature based technique F1. The difference between the two approaches lies in the use of the robust weighing of the epipolar distance: Outliers are down weighed via the statistical tri-weight function in F1, while the regularised L_1 penaliser is applied in our method. There is, however, a more fundamental difference between F1 and our approach. Contrary to F1, our method exploits the dense set of image correspondences provided by the displacement fields between the stereo pairs. Moreover, it exploits a feedback that couples the correspondence search to the epipolar geometry computation and, as a consequence, it benefits from the same advantages as the joint optical flow and fundamental matrix estimation of Chapter 5.

6.4 Experiments

In this section we evaluate the performance of our scene flow method on synthetic stereo sequences with ground truth and on real-world stereo sequences. For both types of evaluation we use image data that is available online, as well as material that we have generated and recorded ourselves. To assess the quality of the scene flow estimation with respect to ground truth, we compute the *root mean square error* (RMSE) of the motion component of the scene flow $(u_\mathrm{f}, v_\mathrm{f}, u_\mathrm{d}, v_\mathrm{d})$, of the left optical flow $(u_\mathrm{f}, v_\mathrm{f})$ and of the first stereo flow $(u_\mathrm{st}, v_\mathrm{st})$. The RMSE between an estimated flow field $\boldsymbol{w}_\mathrm{e} = (u_\mathrm{e}, v_\mathrm{e})$ and a ground truth flow field $\boldsymbol{w}_\mathrm{g} = (u_\mathrm{g}, v_\mathrm{g})$ is defined as

$$\mathrm{RMSE}\left(\boldsymbol{w}_\mathrm{e}, \boldsymbol{w}_\mathrm{g}\right) = \left(\frac{1}{|\Omega|} \int\limits_\Omega \left(\left(u_\mathrm{e} - u_\mathrm{g}\right)^2 + \left(v_\mathrm{e} - v_\mathrm{g}\right)^2 \right) \mathrm{d}\boldsymbol{x} \right)^{1/2} , \tag{6.109}$$

with $|\Omega| = \int_\Omega d\boldsymbol{x}$ the size of the image domain. The RMSE is extended accordingly if $\boldsymbol{w}_\mathrm{e}$ and $\boldsymbol{w}_\mathrm{g}$ have more than 2 components. In addition we will compute the *absolute angular error* AbAE of the left optical flow, which is used as a quality measure in [HD07] and [WRV+08]. It is defined as

$$\mathrm{AbAE}\left(\boldsymbol{w}_\mathrm{e}, \boldsymbol{w}_\mathrm{g}\right) = \frac{1}{|\Omega|} \int\limits_\Omega \arctan\left(\frac{u_\mathrm{e} v_\mathrm{g} - u_\mathrm{g} v_\mathrm{e}}{u_\mathrm{e} u_\mathrm{g} + v_\mathrm{e} v_\mathrm{g}} \right) \mathrm{d}\boldsymbol{x} \ . \tag{6.110}$$

Tab. 6.1: Evaluation of different scene flow methods on the rectified sphere sequence.

Method	RMSE			AbAE
	$(u_\mathrm{f}, v_\mathrm{f}, u_\mathrm{d}, v_\mathrm{d})$	$(u_\mathrm{f}, v_\mathrm{f})$	$(u_\mathrm{st}, v_\mathrm{st})$	$(u_\mathrm{f}, v_\mathrm{f})$
Our method initialised with [FH06]	**1.76**	**0.63**	3.8	1.17
Our method	1.78	**0.63**	5.5	**1.16**
Wedel *et al.* [WRV+08] with ground truth	2.40	0.65	–	1.40
Wedel *et al.* [WRV+08] (87%)	2.45	0.66	**2.9**	1.50
Huguet and Devernay [HD07]	2.51	0.69	3.8	1.75
Wedel *et al.* [WRV+08] (100%)	2.55	0.77	10.9	2.76

The AbAE should not be confused with the average angular error AAE of Sec. 5.2, as it does not take into account the temporal flow component and solely measures the angle between the estimated and ground truth flow in the image plane. The reason for adopting this quality measure instead of the AAE, is that it enables us to compare our results with those of previously published methods. As a quality measure for the fundamental matrix we use anew the error d_F, which has been introduced in Sec. 4.3 and describes the deviation of the estimated epipolar geometry from ground truth in pixel units. All results have been obtained for five iteration steps of our alternating minimisation.

6.4.1 Synthetic Sequences

Rectified Stereo Sequence. In a first experiment we consider the synthetic sphere sequence of Huguet and Devernay [HD07] [2], which is composed of four 512×512 images of a textured sphere with rotating hemispheres. The left images at time t and $t+1$ and the scene set-up are shown in Fig. 6.5. Despite the fact that this sequence is rectified, and thus constitutes the special stereo case with vanishing vertical components of the stereo flow, it is a good benchmark for comparison against existing techniques.

Additionally, this sequence requires the estimation of large stereo displacements, which poses a challenge to variational methods in general. In this context we follow the idea of [HD07] and initialise $(u_\mathrm{st}, v_\mathrm{st})$ with a dedicated method for large displacements. As a standard initialisation, we use a variant of the recent optical flow technique of Brox *et al.* [BBM09] with constraint normalisation and SIFT feature matches as a prior. For consistency, we additionally include results for initialisation with the belief propagation algorithm of Felzenszwalb and Huttenlocher [FH06], as used by Huguet and Devernay. We wish to emphasise that the latter technique is only applicable for rectified stereo sequences, while our standard initialisation can be applied in all stereo settings. In practice we down sample the initialised stereo flow to a fixed intermediate resolution level in the coarse-to-fine pyramid, where we also compute the left and the right optical flow. Both flows are determined by only keeping those terms in our energy that depend on the left optical flow and solving the resulting optical flow problem for the left and right image sequences separately. At the same resolution level, the difference flow $(u_\mathrm{d}, v_\mathrm{d})$ is initialised

2. available at `http://devernay.free.fr/vision/varsceneflow/`

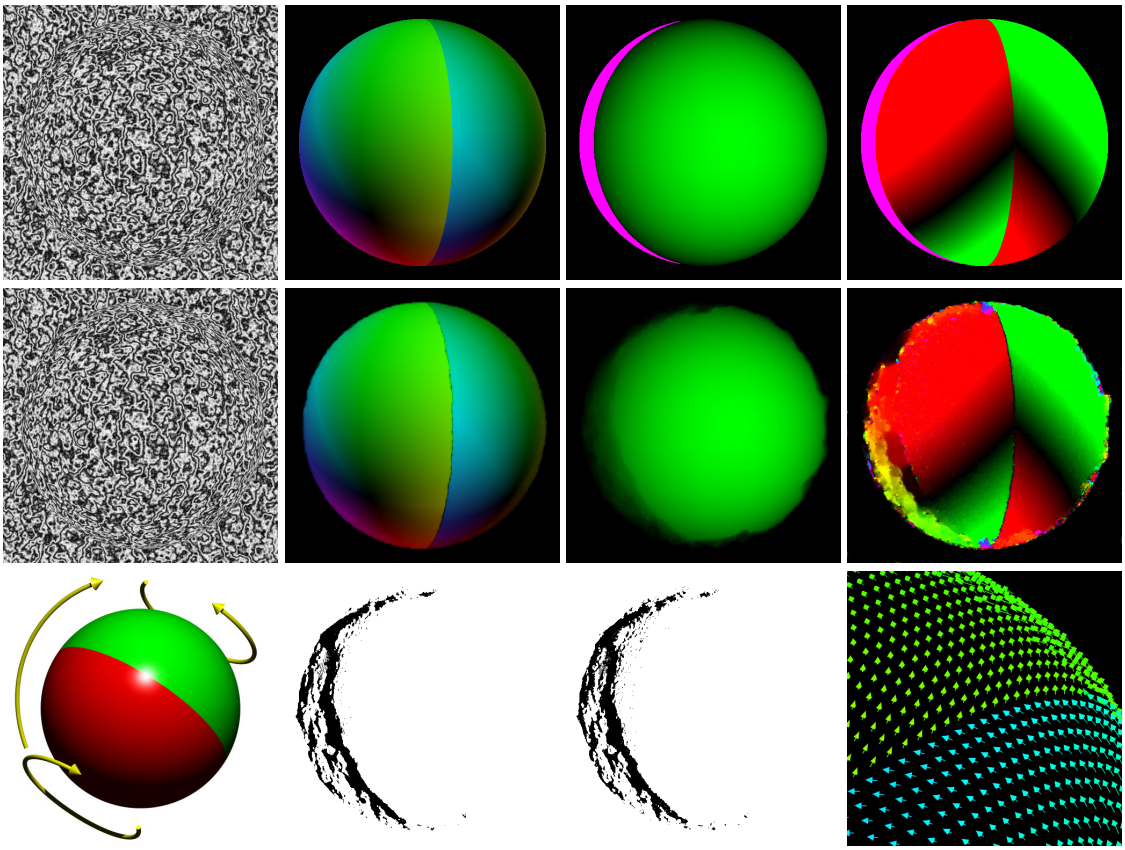

Fig. 6.5: Results for the ortho-parallel sphere sequence. **Top Row: (a)** Left frame at first time step. **(b) + (c) + (d)** Ground truth of left optical flow, first stereo flow and flow change. Colour encodes the direction and brightness the magnitude. Occlusions are coloured pink. **Middle Row: (e)** Left frame at second time step. **(f) + (g) + (h)** Estimated left optical flow, first stereo flow and flow change. **Bottom Row: (i)** The motion of the sphere (taken from [HD07]). **(j) + (k)** Estimated occlusion scores $o_{1\mathrm{r}}$ and $o_{2\mathrm{r}}$. **(l)** Estimated scene flow (run time $\approx$ 420 s, settings: $\alpha_1 = 2.0$, $\alpha_2 = 1.5$, $\alpha_3 = 0.3$, $\beta_1 = \beta_2 = 0.1$, $\gamma = 0.1$, $\mu = 1.0$).

by warping the right optical flow on the left image using $(u_{\mathrm{st}}, v_{\mathrm{st}})$ and subtracting the result from the left optical flow. Setting out from the initial value for $(u_{\mathrm{f}}, v_{\mathrm{f}}, u_{\mathrm{st}}, v_{\mathrm{st}}, u_{\mathrm{d}}, v_{\mathrm{d}})$, we now start the joint computation of all flow components by solving the Euler-Lagrange equations of Sec. 6.3.1.1 on all remaining pyramid levels.

Table 6.1 compares our results with those of the variational methods of Huguet and Devernay [HD07] and Wedel *et al.* [WRV+08] and lists the errors computed within the region of the sphere. With a substantial improvement in the RMSE for $(u_{\mathrm{f}}, v_{\mathrm{f}}, u_{\mathrm{d}}, v_{\mathrm{d}})$ and in the AbAE for $(u_{\mathrm{f}}, v_{\mathrm{f}})$, we consistently outperform the other approaches for the scene flow, although these methods are specifically tailored to the rectified stereo case. The lower RMSE of the method of Wedel *et al.* for $(u_{\mathrm{st}}, v_{\mathrm{st}})$ is due to the fact that it uses sparse stereo correspondences that do not provide results in occluded regions and that the error is not evaluated in these regions. The accuracy of their estimated scene flow is nevertheless significantly lower than ours. This even holds if they use *ground truth* for the stereo correspondences. The good performance of our method is also reflected in the accurate estimation of the stereo geometry: We reach a subpixel precision of $d_F = 0.019$.

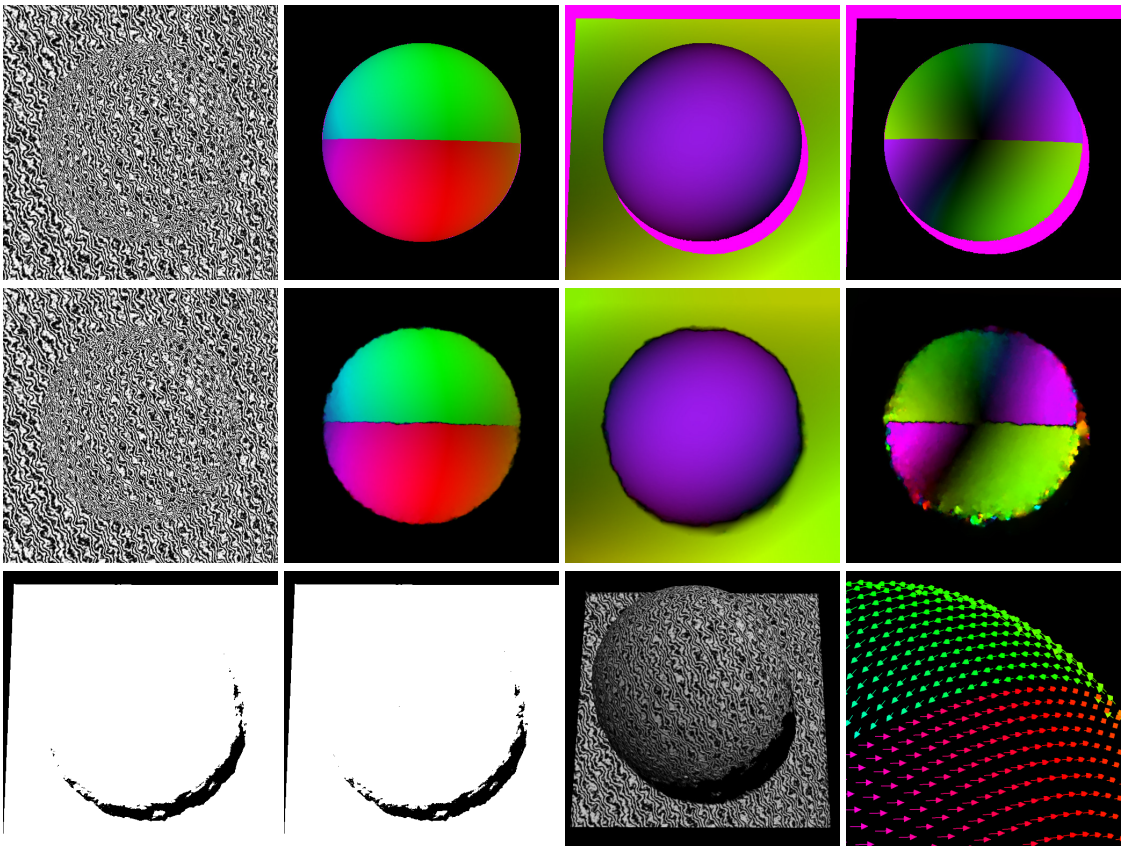

Fig. 6.6: Results for the general (converging) sphere sequence. **Top Row: (a)** Left frame at first time step. **(b) + (c) + (d)** Ground truth of left optical flow, first stereo flow and flow change. Colour encodes the direction, brightness the magnitude. Occlusions are coloured pink. **Middle Row: (e)** Left frame at second time step. **(f) + (g) + (h)** Estimated left optical flow, first stereo flow and flow change. **Bottom Row: (i) + (j)** Estimated occlusion scores $o_{1\mathrm{r}}$ and $o_{2\mathrm{r}}$. **(k)** Estimated scene reconstruction. **(l)** Estimated scene flow (run time ≈ 420 s, settings: $\alpha_1 = 1.5$, $\alpha_2 = 2.0$, $\alpha_3 = 0.8$, $\beta_1 = \beta_2 = 0.03$, $\gamma = 0.1$, $\mu = 1.0$).

The results for the scene flow components that are obtained by initialising with [BBM09] are shown in Fig. 6.5. The depicted flow fields are the estimated left optical flow, the first stereo flow and the difference flow, as well as their ground truths. For the colour coding of the flow vectors we refer to the colour circle of Fig. 4.5 (a). In addition to the flow fields we also show the estimated occlusion scores that result from the scene motion and the change in viewpoint between the cameras. Finally, the scene flow on the sphere surface is presented as a 3D vector plot in the colour code of the left optical flow. It can be observed that the cracking of the surface is captured very well due to the separate regularisation of the optical flow and the stereo flow.

Converging Stereo Sequence. In a second experiment we evaluate the performance of our method for a general stereo geometry. Due to a lack of ground truth scene flow data for general stereo settings, we generated a synthetic sequence of four 512×512 frames with ground truth ourselves [3]. The sequence is similar to the one of the previous exper-

Tab. 6.2: Evaluation of different variants of our method on the general sphere sequence.

Method		RMSE		
		$(u_\mathrm{f}, v_\mathrm{f}, u_\mathrm{d}, v_\mathrm{d})$	$(u_\mathrm{f}, v_\mathrm{f})$	$(u_\mathrm{st}, v_\mathrm{st})$
joint regularisation		0.67	0.64	2.08
joint regularisation	+ normalisation	0.63	**0.59**	1.86
separate regularisation	+ normalisation	**0.61**	**0.59**	**1.61**

iment: A textured sphere with rotating hemispheres is positioned against a plane in the background as shown in Fig. 6.6. Contrary to the rectified sphere sequence, the plane is not infinitely far and thus induces a parallax with occlusions in the background as a result. To demonstrate the benefits of the different design steps in our scene flow model we start from a variant that performs a joint regularisation of the flows as in [HD07] and does not include the constraint normalisation. We then refine the model by subsequently adding the normalisation and the separate regularisation of the flows. In all cases we initialised the stereo flow with the large displacement optical flow technique of Brox *et al.* [BBM09]. Table 6.2 lists the progressively improving results for this test. This time the errors are computed in the non-occluded regions of the whole image domain. The AbAE of the optical flow is not listed because it is not defined for the zero displacement in the background. In Fig. 6.6, the computed flow fields are shown together with the obtained occlusion scores, the 3D reconstruction and the scene flow. As one can see, the estimated displacements resemble the ground truth very well. Again, this is confirmed by a subpixel precision of $d_F = 0.021$ for the stereo geometry.

6.4.2 Real-World Sequences

In our last experiments we will compute the scene flow for several real-world stereo sequences. First we take a look back at the DinoRing and the TempleRing image sets from the Middlebury multiview data base that we have used in our stereo experiments of the previous chapters. These images have been acquired by multiple cameras that are positioned equidistantly on a spherical gantry surrounding the object [SCD+06]. Each image pair captured by two neighbouring cameras therefore forms a pair of stereo images that is related by a constant fundamental matrix, such that the set-up is equivalent to that of a fixed stereo rig encircling the object. Frames 24 and 25 of the DinoRing set thus form a stereo sequence with frames 25 and 26 as shown in the top row of Fig. 6.7. For visualisation purposes we only show a cropped square section of the original 640×480 images in the figure. The reconstruction and the scene flow computed from these four images by our method are shown in the bottom row of Fig. 6.7. As before, we have omitted the homogeneous background from the fundamental matrix estimation and only show the reconstruction within the object silhouette. For the reconstruction and the occlusion handling we used the provided internal camera parameters. In the colouring scheme of the scene flow vectors small displacements are indicated by green arrows and larger displacements by red arrows. It can be observed that the recovered 3D displacement field is in

3. available at `http://www.mia.uni-saarland.de/valgaerts/eccv10/sceneflow`

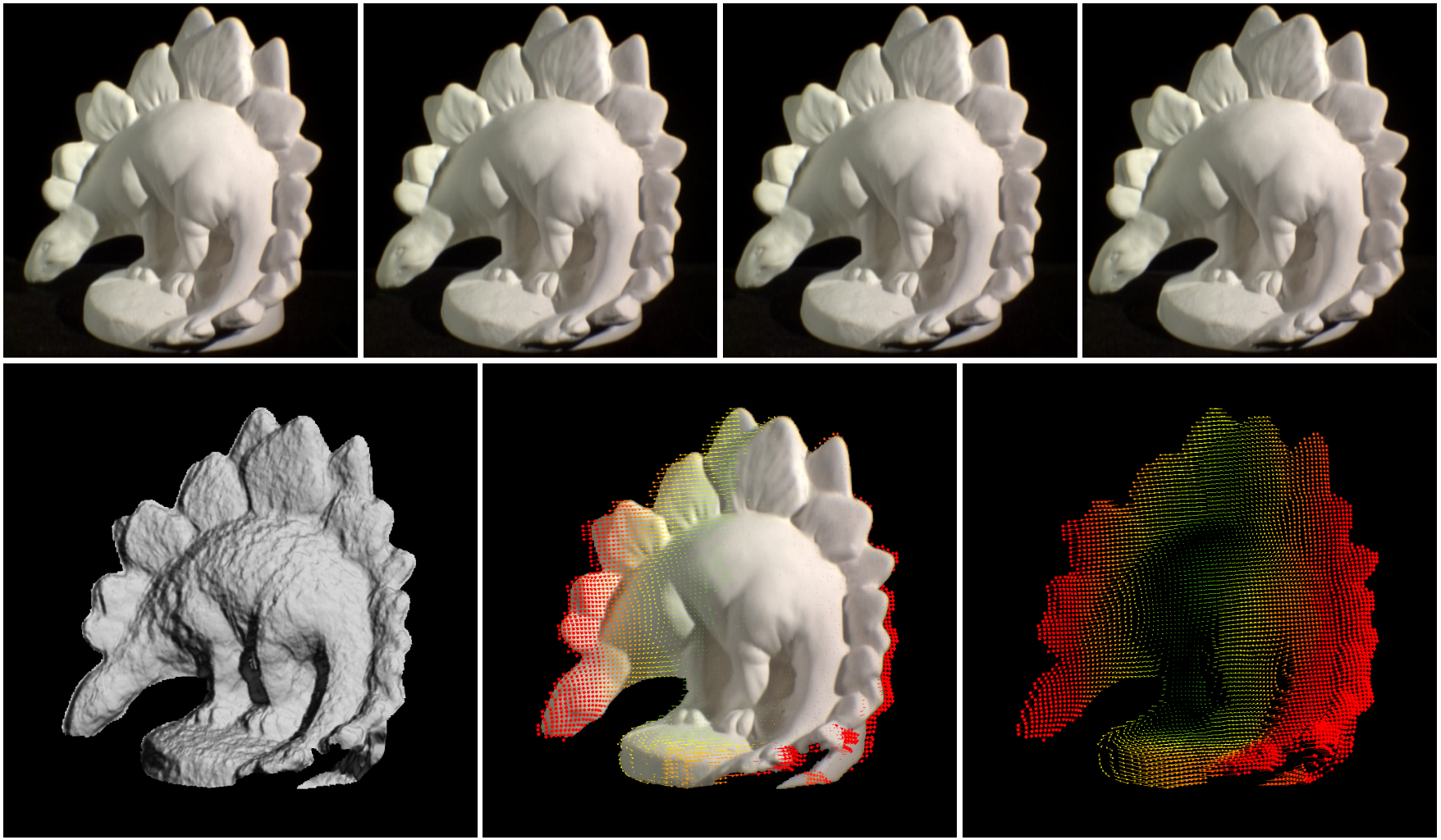

Fig. 6.7: Results for DinoRing. **Top Row: (a) + (b) + (c) + (d)** Left and right image at time t and left and right image at time $t+1$. **Bottom Row: (e)** Untextured reconstruction . **(f)** Reconstruction with the scene flow overlaid. Increasing magnitude from green over yellow to red. **(g)** Scene flow (run time: $\approx$ 490 s, settings: $\alpha_1 = 10.0$, $\alpha_2 = 10.0$, $\alpha_3 = 15.0$, $\beta_1 = \beta_2 = 5.00$, $\gamma = 20.0$, $\mu = 1.0$).

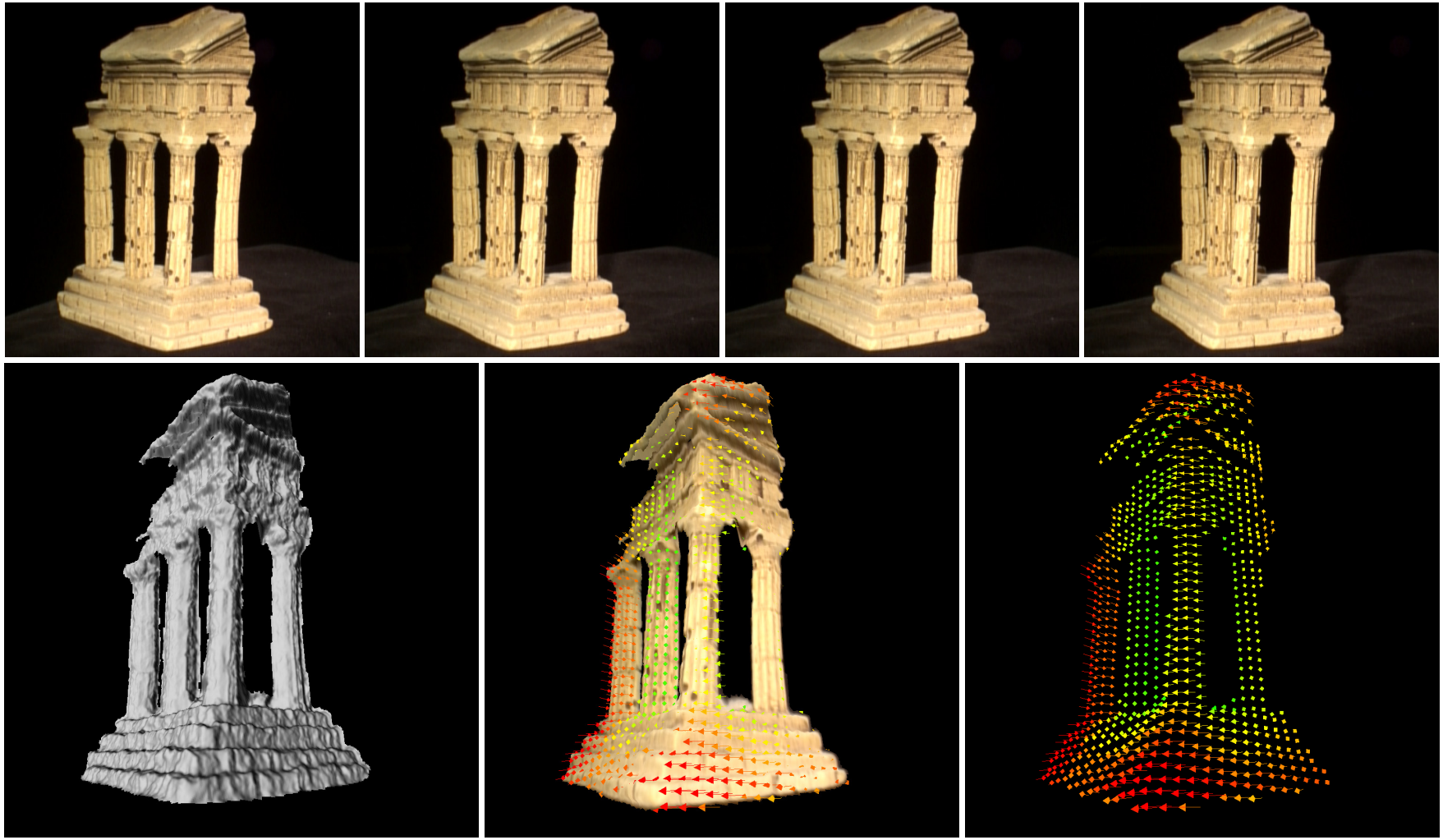

Fig. 6.8: Results for TempleRing **Top Row: (a) + (b) + (c) + (d)** Left and right image at time t and left and right image at time $t+1$. **Bottom Row: (e)** Reconstruction. **(f)** Reconstruction with the scene flow overlaid. **(g)** Scene flow (run time: $\approx$ 490 s, settings: $\alpha_1 = 10.0$, $\alpha_2 = 10.0$, $\alpha_3 = 15.0$, $\beta_1 = \beta_2 = 5.00$, $\gamma = 20.0$, $\mu = 1.0$).

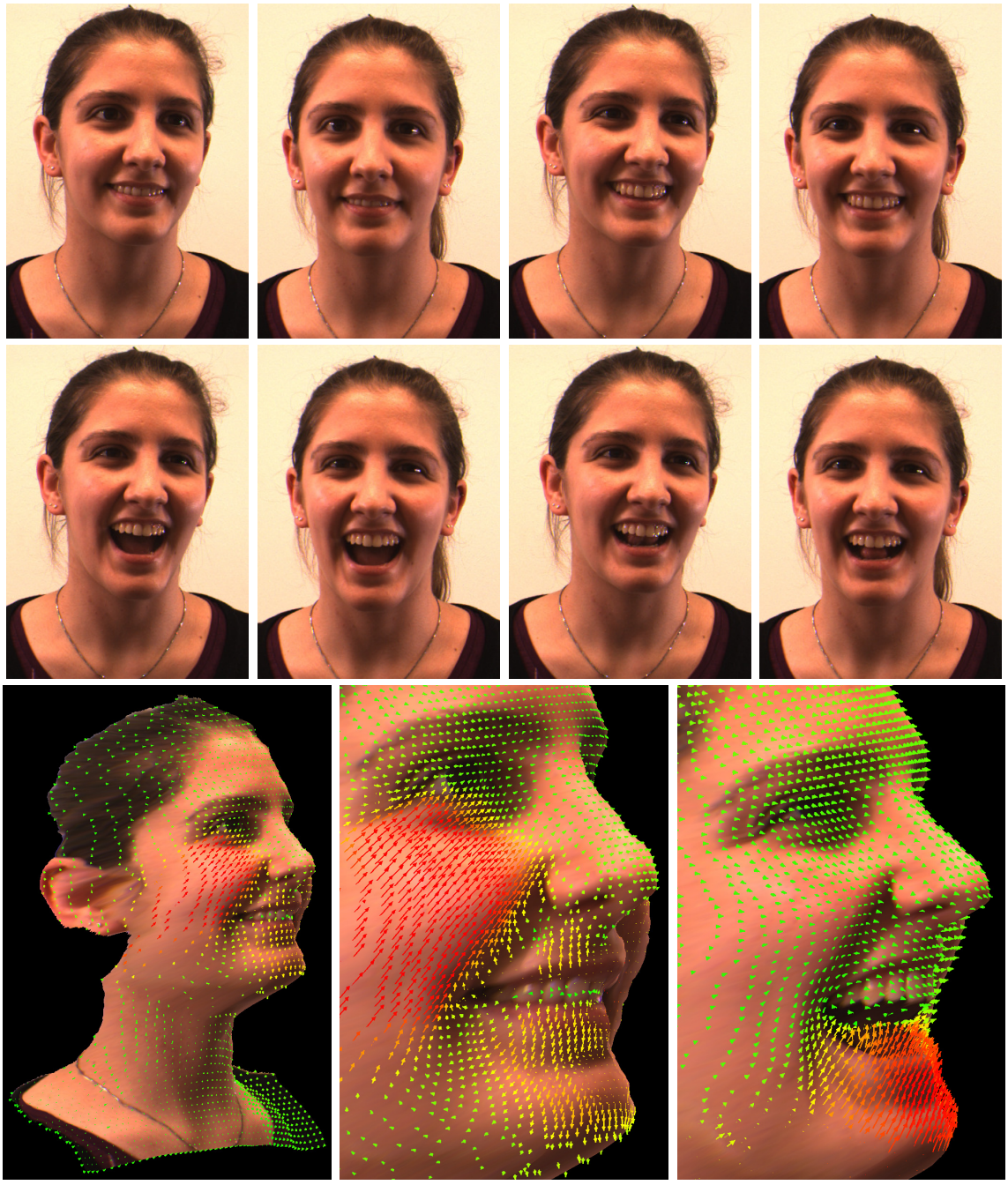

Fig. 6.9: Results for two real-world sequences. **Top Row: (a) + (b) + (c) + (d)** *Smiling*, left and right image at time t and left and right image at time $t+1$. **Middle Row: (e) + (f) + (g) + (h)** *Closing Mouth*, left and right image at time t and left and right image at time $t+1$. **Bottom Row: (i)** Reconstruction with the scene flow overlaid for *Smiling*. Increasing magnitude from green to red. **(j)** Close-up *Smiling*. **(k)** Close-up *Closing Mouth*. (run time: $\approx$ 260 s, settings: $\alpha_1 = 15.0$, $\alpha_2 = 20.0$, $\alpha_3 = 15.0$, $\beta_1 = \beta_2 = 0.5$, $\gamma = 30.0$, $\mu = 1.0$)

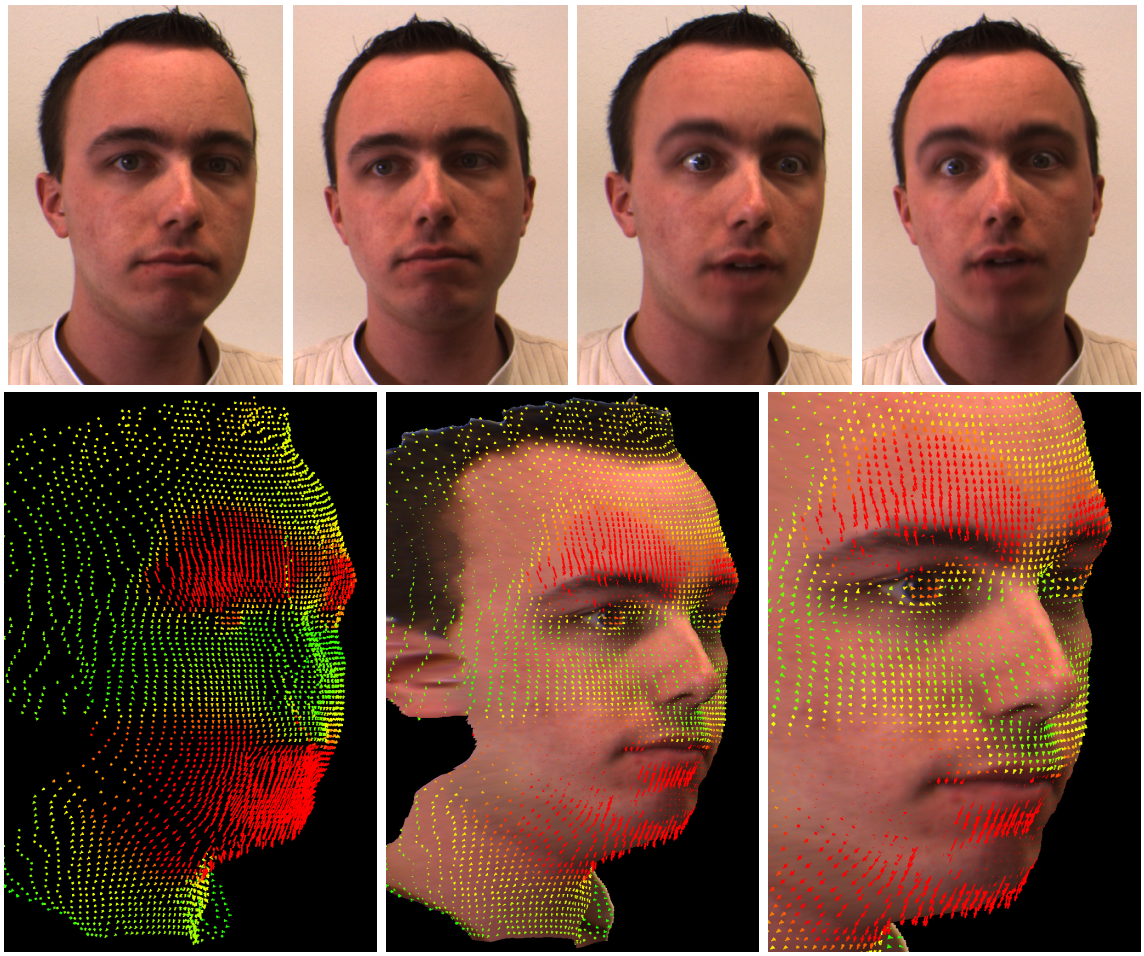

Fig. 6.10: Results for *Surprised*. **Top Row: (a) + (b) + (c) + (d)** Left and right image at time t and left and right image at time $t+1$. **Bottom Row: (e)** Scene flow. **(f)** Reconstruction with the scene flow overlaid. Increasing magnitude from green to red. **(g)** Close-up (run time: ≈ 260 s, settings: $\alpha_1 = 8.0$, $\alpha_2 = 9.1$, $\alpha_3 = 5.3$, $\beta_1 = \beta_2 = 0.5$, $\gamma = 7.0$, $\mu = 1.0$).

accordance with the motion of the stereo rig and that the dinosaur model seems to rotate around its axis relative to the camera. Similar results have been obtained for frames 13, 14 and 15 of the TempleRing data set, depicted in Fig. 6.8. For non-optimised parameters, the high accuracy of the fundamental matrix estimation is demonstrated by an error of $d_F = 0.646$ and $d_F = 0.260$ for DinoRing and TempleRing respectively.

Although the Middlebury data sets form a valid test case for scene flow estimation, the relative scene motion induced by the changing camera position is rigid and does not present the most interesting application of our method. Of greater significance are motion types that include independently moving objects and surface deformation. To test the performance of our method for more general scene motion we recorded several uncalibrated real-world sequences depicting changing expressions of different faces. Two such sequences, *Smiling* and *Closing Mouth*, are shown in the two first rows of Fig. 6.9. The resolution of all images is 340×470. As one can verify, the 3D structure and the motion of the face, shown in the bottom row of the figure, are captured well and look very realistic

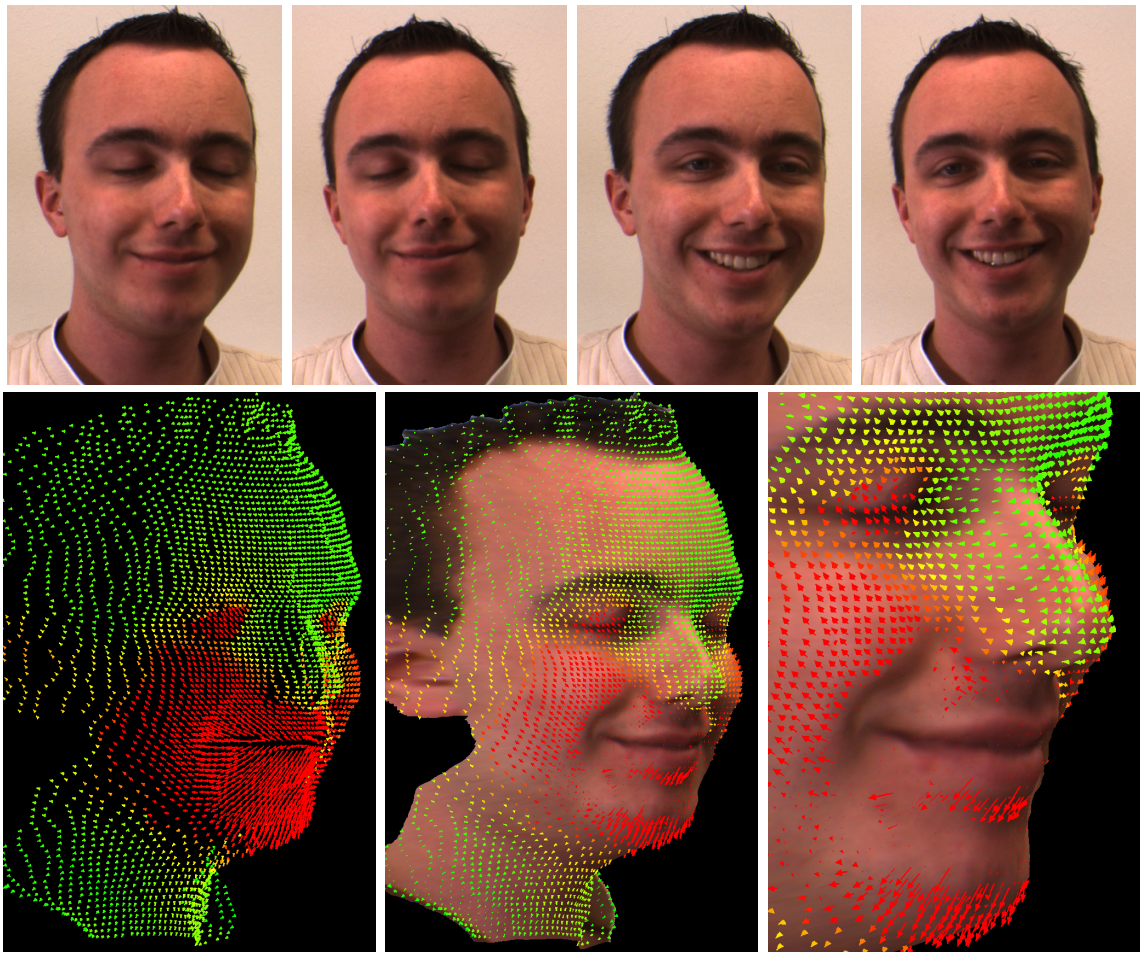

Fig. 6.11: Results for *Opening Eyes*. **Top Row: (a) + (b) + (c) + (d)** Left and right image at time t and left and right image at time $t+1$. **Bottom Row: (e)** Scene flow. **(f)** Reconstruction with scene flow overlaid. Increasing magnitude from green to red. **(g)** Close-up (run time: ≈ 260 s, settings: $\alpha_1 = 8.0$, $\alpha_2 = 9.1$, $\alpha_3 = 5.3$, $\beta_1 = \beta_2 = 0.5$, $\gamma = 7.0$, $\mu = 1.0$).

for both sequences. The colouring scheme is the same as before with small displacements indicated by green arrows and larger displacements by red arrows. We emphasise that these two results are obtained from the depicted four frames only. We do not perform a full calibration of the stereo rig but only use the focal length and an approximation of the principal point as internal camera parameters. For these experiments the homogeneous background is not excluded from the fundamental matrix estimation. Both Smiling and Closing Mouth are part of a longer stereo sequence of approximately 30 frames. By processing the complete sequence in separate chunks of four frames, an animated vector plot has been generated which can be viewed on our homepage [4]. We additionally recorded a stereo sequence of approximately 70 frames, of which two excerpts of each four images can be found in Fig. 6.10 and Fig. 6.11. It can be observed that we recover the lifting eye brows for *Surprised* and that we even capture the small motion of the eye pupils as the

4. `http://www.mia.uni-saarland.de/valgaerts/eccv10/sceneflow`

eyes change their focus. For *Opening Eyes* the movement of the eye lids is detected with equal precision. If some displacement vectors seem to be missing on the reconstructions with the scene flow overlaid, it is simply due to the fact that they point inward into the reconstructed surface and are therefore hidden out of sight. An animated vector plot for the complete sequence can be found on our website [4].

6.5 Summary

In this chapter we have presented a general approach for the dense estimation of scene flow, scene structure and stereo geometry from uncalibrated stereo sequences. Our contributions were fivefold: (i) First we generalised the classical four-frame case to arbitrary stereo setups by embedding the epipolar constraint into a joint energy functional with data and smoothness terms. (ii) We then introduced a tensor notation which allowed us to normalise the data and stereo constraints such that they become geometrically interpretable. (iii) Furthermore, we showed the equivalence of the normalised epipolar constraints to two widely-used distance measures, namely the epipolar distance and the Sampson error. (iv) In addition, we presented a separate robustification of all the flow gradients in the smoothness terms to handle scenarios where flow discontinuities do not coincide. (iv) As a final contribution, we explicitly detected occlusions due to scene motion and changes in camera viewpoint and excluded them from the estimation process.

Our evaluation has demonstrated that the proposed approach is not only more general than existing methods but also more accurate: Even without explicit knowledge about the stereo geometry, we outperform recent techniques that have been specifically designed for the rectified case. Moreover, the stereo geometry is estimated with sub-pixel precision and reconstructions for real-world data show that both the scene structure and the scene motion are determined with high quality. This clearly demonstrates the benefit of a joint approach for the computation of flow, structure and geometry.

7 Summary and Outlook

7.1 Summary

The goal of this thesis was to investigate the benefits of dense optical flow for the recovery of the stereo geometry, the 3D scene structure and the 3D scene motion from stereo image pairs and stereo image sequences. The focus hereby was on a systematic modelling in a variational framework and on the simultaneous and consistent computation of all unknowns as a joint optimisation problem. We achieved these goals in three main steps:

I. Estimation of the Fundamental Matrix from Dense Optical Flow

In Chapter 4 we introduced a *new application for optical flow*: the dense estimation of the fundamental matrix from a stereo image pair. To this end, we proposed a two-step method that first establishes a set of dense image correspondences by means of an accurate variational optical flow method and then estimates the fundamental matrix by imposing the epipolar constraint in each pixel. For the estimation of the optical flow, we based ourselves on the prototypical approach of Brox *et al.* [BBPW04], while for the estimation of the fundamental matrix we designed a robust version of the 8-point algorithm of Longuet-Higgins [Lon81]. Our approach makes successful use of *two main advantages of dense optical flow*: (i) the large amount of correspondences that is available from the filling-in effect of the smoothness term and (ii) the absence of gross outliers resulting from the combination of robust data constraints and global smoothness assumptions.

In an extensive experimental section, we presented a *systematic juxtaposition* of our dense method with widely-used sparse feature based estimation techniques. In order to represent most of the current state-of-the-art, we compared our results with *twelve different variants of feature based methods*. These were obtained by considering two feature matching algorithms (KLT [ST94] and SIFT [Low04]), three random sampling algorithms (LMedS [RL87], LORANSAC [CMO04] and DEGENSAC [CWM05]), and two different distance measures (the epipolar distance [FLP01] and the reprojection error [HZ00]). By comparing the best results for the dense and the sparse techniques, we demonstrated that modern optical flow methods can serve as novel approaches for estimating the stereo geometry with competitive quality. In our experiments, we additionally identified scenarios in which dense techniques should be preferred over sparse ones.

II. Joint Estimation of the Fundamental Matrix and the Optical Flow

In Chapter 5 we demonstrated that not only the stereo geometry can be accurately estimated from dense optical flow, but that the computation of the optical flow can at the same time benefit from knowledge about the stereo geometry. To this end we fed information

about the computed fundamental matrix back into the optical flow estimation, thereby replacing the traditionally unconstrained estimation of the optical flow by one that takes into account the stereo geometry and the rigidity of the relative scene motion. In contrast to existing stereo methods that make explicit use of a given precomputed stereo geometry, we estimated the optical flow and the fundamental matrix *simultaneously*. This resulted in an approach that is more flexible in practice and applicable for *uncalibrated* stereo images.

From a modelling point of view, we achieved this goal by incorporating the epipolar constraint as a *soft constraint* into the optical flow functional. This allowed for the optical flow and the fundamental matrix to correct each other in an alternating optimisation process, that exhibits convergence in practice. Besides yielding better optical flow results than approaches that do not estimate the epipolar geometry in the process, our experiments showed that a joint estimation of the optical flow and the fundamental matrix further improved the two-step method of Chapter 4. Since our modelling was done in a general variational framework, our strategy could be easily extended to more recent optical flow methods. This was demonstrated by replacing the prototypical approach of Brox *et al.* [BBPW04] by the more advanced method of Zimmer *et al.* [ZBW11]. Furthermore, we showed how our joint method *fuses the two steps of classical projective reconstruction*: By solving for both the camera motion and the scene structure in a single optimisation step, we obtained a dense reconstruction of the depicted scene. We concluded the chapter with a discussion of the advantages and shortcomings of our joint variational method.

III. Joint Estimation of the Geometry, the Scene Structure and the Scene Motion

In our final Chapter 6 we proposed a variational method for the computation of the *scene flow from uncalibrated stereo sequences*. We achieved this by integrating spatial and temporal information from two consecutive stereo pairs in a global energy functional that allowed us to simultaneously compute all optical flows and the fundamental matrix. By assuming that only the internal camera parameters were known, our method was able to recover the dense 3D scene structure and the dense 3D scene motion up to a scale factor.

Apart from this novel generalised model for uncalibrated scene flow, we made four additional contributions: (i) Within the multi-resolution framework required to handle large displacements, we introduced a compact notation for the linearised data constraints and epipolar constraints. This notation allowed us to normalise these constraints such that *deviations from model assumptions can be interpreted as geometrical distances*. (ii) Secondly, we showed the *equivalence of the normalised epipolar constraints to two widely-used distance measures* that had been encountered in the context of feature based methods in Chapter 4: the epipolar distance [FLP01] and the Sampson error [WHA89]. (iii) Thirdly, we proposed a regularisation strategy that penalises discontinuities in the different displacement fields *separately*. This made sense, since motion and depth continuities do not necessarily coincide. (iv) As a final contribution, we *explicitly detected occlusions* due to scene motion and changes in camera viewpoint and excluded them from the estimation process. Experiments on synthetic calibrated and uncalibrated data and on real-world sequences demonstrated that our approach is not only more general than existing methods but also more accurate: Even without explicitly knowing the stereo geometry, we *outperformed recent techniques that make explicit use of a given stereo geometry*. These results clearly demonstrate the benefit of a joint computation of flow, structure and geometry.

7.2 Future Work

This thesis does not represent the end of the road for optical flow based 3D reconstruction and 3D motion estimation. In the following we give several ideas for future research:

1. ***Efficient Numerics and Parallelisation.*** An important property of any algorithm is its run time. The implementations of the methods evaluated in this thesis were optimised for accuracy rather than for real-time performance. As already mentioned in Chapter 5, however, there is plenty of room for run time improvements in practical applications. Despite their well-known efficiency, multigrid methods still do not achieve near-real-time performance for larger images. Instead, a speed-up of several orders of magnitude should be sought in parallelisation strategies on modern graphical hardware [GZG+10, SBK10]. Since multigrid methods are in general difficult to parallelise, progress in this area mostly goes together with novel numerical schemes. One such recent scheme that lends itself to easy parallelisation is Fast Explicit Diffusion (FED) [GWB10], an explicit scheme that allows the use of varying, possibly large, time step sizes . Another type of efficient algorithms that can easily be parallelised on a GPU are so-called primal-dual approaches [ZPB07a, WTP+09]. These techniques do not minimise the energy in an Euler-Lagrange framework, but decouple the minimisation with respect to the data and smoothness term.

2. ***Extensions in Spatial Direction.*** An extension of our methods to multiple images is something that comes immediately to mind. The most straightforward of such extensions would be to use three uncalibrated stereo images instead of two. In such case we could estimate the optical flows between the central frame and the two other frames together with the so-called *trifocal tensor* [HZ00, FLP01], a $3 \times 3 \times 3$ tensor which plays a similar role in three-view geometry as the fundamental matrix in two-view geometry. Such tensor generalisations also exist for more than three views, but the chance of finding the same point correspondence in all images decreases with the number of views. An alternative way of including multiple images is therefore to decompose the problem into two-view sub-problems and to merge the resulting reconstructions [ZPB07b, BBH08]. The most challenging task for such a system would be the simultaneous calibration of all cameras with respect to a fixed reference frame. Finally, one could also think of parameterising the whole problem in 3D space, thereby bypassing the explicit computation of the optical flow. Two methods that work along this line have been proposed recently for calibrated stereo reconstruction and scene flow estimation in [SGC10] and [BMK10] respectively.

3. ***Extensions in Temporal Direction.*** Especially with respect to variational scene flow computation, one could think of including extra frames from previous and later time instances. Such additional temporal information would likely stabilise and improve the computation of all the optical flows and the fundamental matrix. A first realisation of this idea could be to apply spatio-temporal optical flow regularisers instead of the purely spatial ones used in this thesis. Although many strategies have been proposed that smooth along the temporal axis [MB87, WS01b, BBW06, ZBW11], this assumption hardly holds in the presence of larger displacements since moving objects change their location from one time instance to the other. It therefore makes more sense to penalise the flow derivatives in the direction of the flow itself. Such

a smoothness constraint would *smooth along the motion trajectories* of moving objects and involves the computation of multiple flows over time. Optical flow ideas of this type have been presented in [BA91, CM95, SS07, WTP+09]. None of these works, however, efficiently solve for all the flow fields without registering them onto the current time instance. Ideally one would like to avoid this cumbersome registration step, such that the reconstructions over time are temporally coherent. First ideas in this direction have recently been proposed by us for optical flow in [VBVZ11].

At least as challenging from a modelling point of view, is the extension of our scene flow method to temporally varying epipolar geometries. While our current model allows the estimation of different fundamental matrices at different time instances, the 3D reconstruction for internally calibrated cameras becomes more complicated. As in the case of a single fundamental matrix, there will be a scale ambiguity which now not only arises by nature of the reconstruction process but also from possible camera motion (i.e. a scaling in the baseline). Prior knowledge about the scene or the camera position might therefore be necessary to solve this ambiguity. Another interesting idea with respect to a time varying stereo system would be the assumption of temporal regularity on the epipolar geometry. Such regularity could for instance be imposed on the epipolar lines or on the fundamental matrix entries.

4. ***Hybrid methods.*** Finally, there are many possibilities of combining the methods presented in this thesis with existing techniques. Instead of juxtaposing them, one could, for instance, try to integrate dense variational methods and sparse featured based techniques to combine the best of both worlds. Such strategies would allow us to compute the optical flow in the presence of large displacements [BBM09, XJM10] and to estimate the fundamental matrix in large baseline scenarios [XS03]. Another hybrid method could combine our results with range image information that can be obtained directly from special devices, such as time-of-flight cameras [CSC+10] or structured light cameras like the Microsoft Kinect motion controller [1]. Finally, illumination and shading information could be integrated as well to refine the obtained reconstructions. This would be especially interesting for capturing detailed facial expressions from low resolution imagery; see e.g. [WWMT11].

1. `http://www.xbox.com/en-US/kinect`
   `https://github.com/OpenKinect/`

A Scene Flow Equations

Here we present the full set of Euler-Lagrange equations corresponding to energy (6.61).

$$
\begin{aligned}
0 \overset{du_\mathrm{f}}{=}\ & \Psi'_{\mathrm{D}1} \cdot \left(\widehat{J}_{111}\, du_\mathrm{f} + \widehat{J}_{112}\, dv_\mathrm{f} + \widehat{J}_{113}\, du_\mathrm{st} + \widehat{J}_{114}\, dv_\mathrm{st} + \widehat{J}_{115}\, du_\mathrm{d} + \widehat{J}_{116}\, dv_\mathrm{d} + \widehat{J}_{117}\right) \\
+\ & \Psi'_{\mathrm{D}2} \cdot \left(\widehat{J}_{211}\, du_\mathrm{f} + \widehat{J}_{212}\, dv_\mathrm{f} + \widehat{J}_{213}\, du_\mathrm{st} + \widehat{J}_{214}\, dv_\mathrm{st} + \widehat{J}_{215}\, du_\mathrm{d} + \widehat{J}_{216}\, dv_\mathrm{d} + \widehat{J}_{217}\right) \\
+\ & \Psi'_{\mathrm{D}4} \cdot \left(\widehat{J}_{411}\, du_\mathrm{f} + \widehat{J}_{412}\, dv_\mathrm{f} + \widehat{J}_{413}\, du_\mathrm{st} + \widehat{J}_{414}\, dv_\mathrm{st} + \widehat{J}_{415}\, du_\mathrm{d} + \widehat{J}_{416}\, dv_\mathrm{d} + \widehat{J}_{417}\right) \\
-\ & \alpha_1 \operatorname{div}\Big(\Psi'_{\mathrm{S}1} \cdot \nabla(u_\mathrm{f} + du_\mathrm{f})\Big) \\
+\ & \beta_2\, \Psi'_{\mathrm{E}2} \cdot \left(\widehat{E}_{211}\,(du_\mathrm{f} + du_\mathrm{st} + du_\mathrm{d}) + \widehat{E}_{212}\,(dv_\mathrm{f} + dv_\mathrm{st} + dv_\mathrm{d}) + \widehat{E}_{213}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E}2} \cdot \left(\widehat{E}_{511}\, du_\mathrm{f} + \widehat{E}_{512}\, dv_\mathrm{f} + \widehat{E}_{513}\right) \\
+\ & \mu\ \ (u_\mathrm{f} + du_\mathrm{f} - u_\mathrm{a} - du_\mathrm{a}), \qquad \text{(A.1)} \\
0 \overset{dv_\mathrm{f}}{=}\ & \Psi'_{\mathrm{D}1} \cdot \left(\widehat{J}_{112}\, du_\mathrm{f} + \widehat{J}_{122}\, dv_\mathrm{f} + \widehat{J}_{123}\, du_\mathrm{st} + \widehat{J}_{124}\, dv_\mathrm{st} + \widehat{J}_{125}\, du_\mathrm{d} + \widehat{J}_{126}\, dv_\mathrm{d} + \widehat{J}_{127}\right) \\
+\ & \Psi'_{\mathrm{D}2} \cdot \left(\widehat{J}_{212}\, du_\mathrm{f} + \widehat{J}_{222}\, dv_\mathrm{f} + \widehat{J}_{223}\, du_\mathrm{st} + \widehat{J}_{224}\, dv_\mathrm{st} + \widehat{J}_{225}\, du_\mathrm{d} + \widehat{J}_{226}\, dv_\mathrm{d} + \widehat{J}_{227}\right) \\
+\ & \Psi'_{\mathrm{D}4} \cdot \left(\widehat{J}_{412}\, du_\mathrm{f} + \widehat{J}_{422}\, dv_\mathrm{f} + \widehat{J}_{423}\, du_\mathrm{st} + \widehat{J}_{424}\, dv_\mathrm{st} + \widehat{J}_{425}\, du_\mathrm{d} + \widehat{J}_{426}\, dv_\mathrm{d} + \widehat{J}_{427}\right) \\
-\ & \alpha_1 \operatorname{div}\Big(\Psi'_{\mathrm{S}1} \cdot \nabla(v_\mathrm{f} + dv_\mathrm{f})\Big) \\
+\ & \beta_2\, \Psi'_{\mathrm{E}2} \cdot \left(\widehat{E}_{212}\,(du_\mathrm{f} + du_\mathrm{st} + du_\mathrm{d}) + \widehat{E}_{222}\,(dv_\mathrm{f} + dv_\mathrm{st} + dv_\mathrm{d}) + \widehat{E}_{223}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E}2} \cdot \left(\widehat{E}_{512}\, du_\mathrm{f} + \widehat{E}_{522}\, dv_\mathrm{f} + \widehat{E}_{523}\right) \\
+\ & \mu\ \ (v_\mathrm{f} + dv_\mathrm{f} - v_\mathrm{a} - dv_\mathrm{a}), \qquad \text{(A.2)} \\
0 \overset{du_\mathrm{st}}{=}\ & \Psi'_{\mathrm{D}2} \cdot \left(\widehat{J}_{213}\, du_\mathrm{f} + \widehat{J}_{223}\, dv_\mathrm{f} + \widehat{J}_{233}\, du_\mathrm{st} + \widehat{J}_{234}\, dv_\mathrm{st} + \widehat{J}_{235}\, du_\mathrm{d} + \widehat{J}_{236}\, dv_\mathrm{d} + \widehat{J}_{237}\right) \\
+\ & \Psi'_{\mathrm{D}3} \cdot \left(\widehat{J}_{313}\, du_\mathrm{f} + \widehat{J}_{323}\, dv_\mathrm{f} + \widehat{J}_{333}\, du_\mathrm{st} + \widehat{J}_{334}\, dv_\mathrm{st} + \widehat{J}_{335}\, du_\mathrm{d} + \widehat{J}_{336}\, dv_\mathrm{d} + \widehat{J}_{337}\right) \\
+\ & \Psi'_{\mathrm{D}4} \cdot \left(\widehat{J}_{413}\, du_\mathrm{f} + \widehat{J}_{423}\, dv_\mathrm{f} + \widehat{J}_{433}\, du_\mathrm{st} + \widehat{J}_{434}\, dv_\mathrm{st} + \widehat{J}_{435}\, du_\mathrm{d} + \widehat{J}_{436}\, dv_\mathrm{d} + \widehat{J}_{437}\right) \\
-\ & \alpha_2 \operatorname{div}\Big(\Psi'_{\mathrm{S}2} \cdot \nabla(u_\mathrm{st} + du_\mathrm{st})\Big) \\
+\ & \beta_1\, \Psi'_{\mathrm{E}1} \cdot \left(\widehat{E}_{111}\, du_\mathrm{st} + \widehat{E}_{112}\, dv_\mathrm{st} + \widehat{E}_{113}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E}2} \cdot \left(\widehat{E}_{211}\,(du_\mathrm{f} + du_\mathrm{st} + du_\mathrm{d}) + \widehat{E}_{212}\,(dv_\mathrm{f} + dv_\mathrm{st} + dv_\mathrm{d}) + \widehat{E}_{213}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E}2} \cdot \left(\widehat{E}_{411}\,(du_\mathrm{a} + du_\mathrm{st} + du_\mathrm{d}) + \widehat{E}_{412}\,(dv_\mathrm{a} + dv_\mathrm{st} + dv_\mathrm{d}) + \widehat{E}_{513}\right), \qquad \text{(A.3)}
\end{aligned}
$$

$$
\begin{aligned}
0 \overset{dv_{\mathrm{st}}}{=}\ & \Psi'_{\mathrm{D2}} \cdot \left(\widehat{J}_{214}\, du_{\mathrm{f}} + \widehat{J}_{224}\, dv_{\mathrm{f}} + \widehat{J}_{234}\, du_{\mathrm{st}} + \widehat{J}_{244}\, dv_{\mathrm{st}} + \widehat{J}_{245}\, du_{\mathrm{d}} + \widehat{J}_{246}\, dv_{\mathrm{d}} + \widehat{J}_{247}\right) \\
+\ & \Psi'_{\mathrm{D3}} \cdot \left(\widehat{J}_{314}\, du_{\mathrm{f}} + \widehat{J}_{324}\, dv_{\mathrm{f}} + \widehat{J}_{334}\, du_{\mathrm{st}} + \widehat{J}_{344}\, dv_{\mathrm{st}} + \widehat{J}_{345}\, du_{\mathrm{d}} + \widehat{J}_{346}\, dv_{\mathrm{d}} + \widehat{J}_{347}\right) \\
+\ & \Psi'_{\mathrm{D4}} \cdot \left(\widehat{J}_{414}\, du_{\mathrm{f}} + \widehat{J}_{424}\, dv_{\mathrm{f}} + \widehat{J}_{434}\, du_{\mathrm{st}} + \widehat{J}_{444}\, dv_{\mathrm{st}} + \widehat{J}_{445}\, du_{\mathrm{d}} + \widehat{J}_{446}\, dv_{\mathrm{d}} + \widehat{J}_{447}\right) \\
-\ & \alpha_2 \operatorname{div}\Big(\Psi'_{\mathrm{S2}} \cdot \nabla(v_{\mathrm{st}} + dv_{\mathrm{st}})\Big) \\
+\ & \beta_1\, \Psi'_{\mathrm{E1}} \cdot \left(\widehat{E}_{112}\, du_{\mathrm{st}} + \widehat{E}_{122}\, dv_{\mathrm{st}} + \widehat{E}_{123}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{212}\,(du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{222}\,(dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{223}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{412}\,(du_{\mathrm{a}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{422}\,(dv_{\mathrm{a}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{423}\right), \qquad \text{(A.4)} \\
0 \overset{du_{\mathrm{d}}}{=}\ & \Psi'_{\mathrm{D2}} \cdot \left(\widehat{J}_{215}\, du_{\mathrm{f}} + \widehat{J}_{225}\, dv_{\mathrm{f}} + \widehat{J}_{235}\, du_{\mathrm{st}} + \widehat{J}_{245}\, dv_{\mathrm{st}} + \widehat{J}_{255}\, du_{\mathrm{d}} + \widehat{J}_{256}\, dv_{\mathrm{d}} + \widehat{J}_{257}\right) \\
+\ & \Psi'_{\mathrm{D4}} \cdot \left(\widehat{J}_{415}\, du_{\mathrm{f}} + \widehat{J}_{425}\, dv_{\mathrm{f}} + \widehat{J}_{435}\, du_{\mathrm{st}} + \widehat{J}_{445}\, dv_{\mathrm{st}} + \widehat{J}_{455}\, du_{\mathrm{d}} + \widehat{J}_{456}\, dv_{\mathrm{d}} + \widehat{J}_{457}\right) \\
-\ & \alpha_3 \operatorname{div}\Big(\Psi'_{\mathrm{S3}} \cdot \nabla(u_{\mathrm{d}} + du_{\mathrm{d}})\Big) \\
+\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{211}\,(du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{212}\,(dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{213}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{411}\,(du_{\mathrm{a}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{412}\,(dv_{\mathrm{a}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{413}\right), \qquad \text{(A.5)} \\
0 \overset{dv_{\mathrm{d}}}{=}\ & \Psi'_{\mathrm{D2}} \cdot \left(\widehat{J}_{216}\, du_{\mathrm{f}} + \widehat{J}_{226}\, dv_{\mathrm{f}} + \widehat{J}_{236}\, du_{\mathrm{st}} + \widehat{J}_{246}\, dv_{\mathrm{st}} + \widehat{J}_{256}\, du_{\mathrm{d}} + \widehat{J}_{266}\, dv_{\mathrm{d}} + \widehat{J}_{267}\right) \\
+\ & \Psi'_{\mathrm{D4}} \cdot \left(\widehat{J}_{416}\, du_{\mathrm{f}} + \widehat{J}_{426}\, dv_{\mathrm{f}} + \widehat{J}_{436}\, du_{\mathrm{st}} + \widehat{J}_{446}\, dv_{\mathrm{st}} + \widehat{J}_{456}\, du_{\mathrm{d}} + \widehat{J}_{466}\, dv_{\mathrm{d}} + \widehat{J}_{467}\right) \\
-\ & \alpha_3 \operatorname{div}\Big(\Psi'_{\mathrm{S3}} \cdot \nabla(v_{\mathrm{d}} + dv_{\mathrm{d}})\Big) \\
+\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{212}\,(du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{222}\,(dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{223}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{412}\,(du_{\mathrm{a}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{422}\,(dv_{\mathrm{a}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{423}\right), \qquad \text{(A.6)} \\
0 \overset{du_{\mathrm{a}}}{=}\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{311}\, du_{\mathrm{a}} + \widehat{E}_{312}\, dv_{\mathrm{a}} + \widehat{E}_{313}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{411}\,(du_{\mathrm{a}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{412}\,(dv_{\mathrm{a}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{413}\right) \\
-\ & \mu\ \ (u_{\mathrm{f}} + du_{\mathrm{f}} - u_{\mathrm{a}} - du_{\mathrm{a}}), \qquad \text{(A.7)} \\
0 \overset{dv_{\mathrm{a}}}{=}\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{312}\, du_{\mathrm{a}} + \widehat{E}_{322}\, dv_{\mathrm{a}} + \widehat{E}_{323}\right) \\
+\ & \beta_2\, \Psi'_{\mathrm{E2}} \cdot \left(\widehat{E}_{412}\,(du_{\mathrm{a}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \widehat{E}_{422}\,(dv_{\mathrm{a}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) + \widehat{E}_{423}\right) \\
-\ & \mu\ \ (v_{\mathrm{f}} + dv_{\mathrm{f}} - v_{\mathrm{a}} - dv_{\mathrm{a}})\ . \qquad \text{(A.8)}
\end{aligned}
$$

Here we present the discrete Euler-Lagrange equations corresponding to energy (6.61).

$$
\begin{aligned}
0 \overset{[\,du_{\mathrm{f}}\,]_{i,j}}{=} & \sum_{m\in\{1,2,4\}} \Big([\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m11}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m12}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m13}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m14}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m15}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m16}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j}\Big) \\
& - \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}1}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}1}]_{i,j}\big)}{2} \frac{\big([\,du_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}} - [\,du_{\mathrm{f}}\,]_{i,j}\big)}{h^2} \\
& + \sum_{m\in\{2,5\}} \Big(\beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m11}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j}\Big) \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{211}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{211}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \\
& + \mu\,[\,du_{\mathrm{f}}\,]_{i,j} - \mu\,[\,du_{\mathrm{a}}\,]_{i,j} \\
& + \sum_{m\in\{1,2,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}\,[J_{m17}]_{i,j} + \sum_{m\in\{2,5\}} \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m13}]_{i,j} \\
& - \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}1}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}1}]_{i,j}\big)}{2} \frac{\big([\,u_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}} - [\,u_{\mathrm{f}}\,]_{i,j}\big)}{h^2} \\
& + \mu\,[\,u_{\mathrm{f}}\,]_{i,j} - \mu\,[\,u_{\mathrm{a}}\,]_{i,j}, \qquad \text{(A.9)}
\end{aligned}
$$

$$
\begin{aligned}
0 \overset{[\,dv_{\mathrm{f}}\,]_{i,j}}{=} & \sum_{m\in\{1,2,4\}} \Big([\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m12}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m22}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m23}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m24}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m25}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m26}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j}\Big) \\
& - \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}1}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}1}]_{i,j}\big)}{2} \frac{\big([\,dv_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}} - [\,dv_{\mathrm{f}}\,]_{i,j}\big)}{h^2} \\
& + \sum_{m\in\{2,5\}} \Big(\beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m22}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j}\Big) \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{222}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{222}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \\
& + \mu\,[\,dv_{\mathrm{f}}\,]_{i,j} - \mu\,[\,dv_{\mathrm{a}}\,]_{i,j} \\
& + \sum_{m\in\{1,2,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}\,[J_{m27}]_{i,j} + \sum_{m\in\{2,5\}} \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m23}]_{i,j} \\
& - \alpha_1 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}1}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}1}]_{i,j}\big)}{2} \frac{\big([\,v_{\mathrm{f}}\,]_{\tilde{i},\tilde{j}} - [\,v_{\mathrm{f}}\,]_{i,j}\big)}{h^2} \\
& + \mu\,[\,v_{\mathrm{f}}\,]_{i,j} - \mu\,[\,v_{\mathrm{a}}\,]_{i,j} \qquad \text{(A.10)}
\end{aligned}
$$

$$
\begin{aligned}
0 \overset{[du_{\mathrm{st}}]_{i,j}}{=} & \sum_{m\in\{2,3,4\}} \Big([\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m13}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m23}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m33}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m34}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m35}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m36}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \Big) \\
& - \alpha_2 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}2}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}2}]_{i,j}\big)}{2} \frac{\big([\,du_{\mathrm{st}}\,]_{\tilde{i},\tilde{j}} - [\,du_{\mathrm{st}}\,]_{i,j}\big)}{h^2} \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{211}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& + \beta_1\,[\Psi'_{\mathrm{E}1}]_{i,j}\,[\widehat{E}_{111}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_1\,[\Psi'_{\mathrm{E}1}]_{i,j}\,[\widehat{E}_{112}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& + \sum_{m\in\{2,4\}} \Big(\beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m11}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m11}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \Big) \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{411}]_{i,j}\,[\,du_{\mathrm{a}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{412}]_{i,j}\,[\,dv_{\mathrm{a}}\,]_{i,j} \\
& + \sum_{m\in\{2,3,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}\,[J_{m37}]_{i,j} + \beta_1\,[\Psi'_{\mathrm{E}1}]_{i,j}\,[\widehat{E}_{113}]_{i,j} + \sum_{m\in\{2,4\}} \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m13}]_{i,j} \\
& - \alpha_2 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}2}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}2}]_{i,j}\big)}{2} \frac{\big([\,u_{\mathrm{st}}\,]_{\tilde{i},\tilde{j}} - [\,u_{\mathrm{st}}\,]_{i,j}\big)}{h^2}, \qquad \text{(A.11)}
\end{aligned}
$$

$$
\begin{aligned}
0 \overset{[dv_{\mathrm{st}}]_{i,j}}{=} & \sum_{m\in\{2,3,4\}} \Big([\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m14}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m24}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m34}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m44}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m45}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m46}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \Big) \\
& - \alpha_2 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}2}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}2}]_{i,j}\big)}{2} \frac{\big([\,dv_{\mathrm{st}}\,]_{\tilde{i},\tilde{j}} - [\,dv_{\mathrm{st}}\,]_{i,j}\big)}{h^2} \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{222}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& + \beta_1\,[\Psi'_{\mathrm{E}1}]_{i,j}\,[\widehat{E}_{112}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_1\,[\Psi'_{\mathrm{E}1}]_{i,j}\,[\widehat{E}_{122}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& + \sum_{m\in\{2,4\}} \Big(\beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m22}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m22}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \Big) \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{412}]_{i,j}\,[\,du_{\mathrm{a}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{422}]_{i,j}\,[\,dv_{\mathrm{a}}\,]_{i,j} \\
& + \sum_{m\in\{2,3,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}\,[J_{m47}]_{i,j} + \beta_1\,[\Psi'_{\mathrm{E}1}]_{i,j}\,[\widehat{E}_{123}]_{i,j} + \sum_{m\in\{2,4\}} \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m23}]_{i,j} \\
& - \alpha_2 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}2}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}2}]_{i,j}\big)}{2} \frac{\big([\,v_{\mathrm{st}}\,]_{\tilde{i},\tilde{j}} - [\,v_{\mathrm{st}}\,]_{i,j}\big)}{h^2}, \qquad \text{(A.12)}
\end{aligned}
$$

$$
\begin{aligned}
0 \overset{[\,du_{\mathrm{d}}\,]_{i,j}}{=} & \sum_{m\in\{2,4\}} \Big([\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m15}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m25}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m35}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m45}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m55}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m56}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \Big) \\
& - \alpha_3 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}3}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}3}]_{i,j}\big)}{2} \frac{\big([\,du_{\mathrm{d}}\,]_{\tilde{i},\tilde{j}} - [\,du_{\mathrm{d}}\,]_{i,j}\big)}{h^2} \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{211}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& + \sum_{m\in\{2,4\}} \Big(\beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m11}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m11}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \Big) \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{411}]_{i,j}\,[\,du_{\mathrm{a}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{412}]_{i,j}\,[\,dv_{\mathrm{a}}\,]_{i,j} \\
& + \sum_{m\in\{2,3,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}\,[J_{m57}]_{i,j} + \sum_{m\in\{2,4\}} \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m13}]_{i,j} \\
& - \alpha_3 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}3}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}3}]_{i,j}\big)}{2} \frac{\big([\,u_{\mathrm{d}}\,]_{\tilde{i},\tilde{j}} - [\,u_{\mathrm{d}}\,]_{i,j}\big)}{h^2}, \qquad \text{(A.13)}
\end{aligned}
$$

$$
\begin{aligned}
0 \overset{[\,dv_{\mathrm{d}}\,]_{i,j}}{=} & \sum_{m\in\{2,4\}} \Big([\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m16}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m26}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m36}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m46}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m56}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + [\Psi'_{\mathrm{D}m}]_{i,j}\,[\widehat{J}_{m66}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \Big) \\
& - \alpha_3 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}3}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}3}]_{i,j}\big)}{2} \frac{\big([\,dv_{\mathrm{d}}\,]_{\tilde{i},\tilde{j}} - [\,dv_{\mathrm{d}}\,]_{i,j}\big)}{h^2} \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{212}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{222}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j} \\
& + \sum_{m\in\{2,4\}} \Big(\beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m22}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
& \qquad + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m22}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \Big) \\
& + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{412}]_{i,j}\,[\,du_{\mathrm{a}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{422}]_{i,j}\,[\,dv_{\mathrm{a}}\,]_{i,j} \\
& + \sum_{m\in\{2,3,4\}} [\Psi'_{\mathrm{D}m}]_{i,j}\,[J_{m67}]_{i,j} + \sum_{m\in\{2,4\}} \beta_2\,[\Psi'_{\mathrm{E}2}]_{i,j}\,[\widehat{E}_{m23}]_{i,j} \\
& - \alpha_3 \sum_{l\in\{x,y\}} \sum_{(\tilde{i},\tilde{j})\in\mathcal{N}_l(i,j)} \frac{\big([\Psi'_{\mathrm{S}3}]_{\tilde{i},\tilde{j}} + [\Psi'_{\mathrm{S}3}]_{i,j}\big)}{2} \frac{\big([\,v_{\mathrm{d}}\,]_{\tilde{i},\tilde{j}} - [\,v_{\mathrm{d}}\,]_{i,j}\big)}{h^2}, \qquad \text{(A.14)}
\end{aligned}
$$

$$
\begin{aligned}
0 \overset{[\,du_{\mathrm{a}}\,]_{i,j}}{=} \quad & \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{411}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{412}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
+ \quad & \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{411}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{412}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \\
+ \quad & \sum_{m\in\{3,4\}} \Big(\beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{m11}]_{i,j}\,[\,du_{\mathrm{a}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,dv_{\mathrm{a}}\,]_{i,j}\Big) \\
- \quad & \mu\,[\,du_{\mathrm{f}}\,]_{i,j} + \mu\,[\,du_{\mathrm{a}}\,]_{i,j} \\
+ \quad & \sum_{m\in\{3,4\}} \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{m13}]_{i,j} \\
- \quad & \mu\,[\,u_{\mathrm{f}}\,]_{i,j} - \mu\,[\,u_{\mathrm{a}}\,]_{i,j}, \qquad \text{(A.15)}
\end{aligned}
$$

$$
\begin{aligned}
0 \overset{[\,dv_{\mathrm{a}}\,]_{i,j}}{=} \quad & \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{412}]_{i,j}\,[\,du_{\mathrm{st}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{422}]_{i,j}\,[\,dv_{\mathrm{st}}\,]_{i,j} \\
+ \quad & \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{412}]_{i,j}\,[\,du_{\mathrm{d}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{422}]_{i,j}\,[\,dv_{\mathrm{d}}\,]_{i,j} \\
+ \quad & \sum_{m\in\{3,4\}} \Big(\beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{m12}]_{i,j}\,[\,du_{\mathrm{f}}\,]_{i,j} + \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{m22}]_{i,j}\,[\,dv_{\mathrm{f}}\,]_{i,j}\Big) \\
- \quad & \mu\,[\,dv_{\mathrm{f}}\,]_{i,j} + \mu\,[\,dv_{\mathrm{a}}\,]_{i,j} \\
+ \quad & \sum_{m\in\{3,4\}} \beta_2\,[\Psi'_{\mathrm{E2}}]_{i,j}\,[\widehat{E}_{m23}]_{i,j} \\
- \quad & \mu\,[\,v_{\mathrm{f}}\,]_{i,j} - \mu\,[\,v_{\mathrm{a}}\,]_{i,j}\ . \qquad \text{(A.16)}
\end{aligned}
$$

B Notation

σ	standard deviation for presmoothing
K_σ	gaussian kernel with standard deviation σ
σ_r	robust standard deviation
σ_i	singular values
μ_i	Eigenvalues
λ	Lagrange multiplier
$\mathcal{L}$	Lagrangian
ϵ	small regularisation parameter
η	downsampling factor in the interval $(0, 1)$
h	grid cell size
d	geometrical distance
$\boldsymbol{a}$	a vector
$\boldsymbol{a}^n$	vector of size $n \times 1$
a_i	i-th component of vector $\boldsymbol{a}$
$\|\boldsymbol{a}\|$	L_2 norm of vector $\boldsymbol{a}$
$\lvert\boldsymbol{a}\rvert$	magnitude of vector a (sometimes used instead of the notation $\|\boldsymbol{a}\|$)
$\mathrm{diag}(\boldsymbol{a})$	diagonal matrix with entries of vector $\boldsymbol{a}$ as main diagonal
$\hat{\boldsymbol{a}}$	normalised version of vector $\boldsymbol{a}$
$\boldsymbol{p}$	parameter vector of unknowns
$\boldsymbol{b}$	right hand side
$\boldsymbol{d}$	vector of flow increments
$\boldsymbol{v}_i$	eigenvectors of an $n \times n$ matrix

A	a matrix
$A(\boldsymbol{p})$	nonlinear operator - system matrix
P	camera projection matrix
F	fundamental matrix
E	essential matrix – epipolar tensor
J	motion tensor – scene flow tensor
(A, B)	$n \times m$ matrix composed of a $n \times k$ matrix A and a $n \times (m-k)$ matrix B
A^n	squared matrix of size $n \times n$
A_{ij}	ij-th entry of a matrix A
$\det(A)$	determinant of matrix A
$\|A\|_{\text{Frob}}$	Frobenius norm of matrix A
A^+	pseudo-inverse of matrix A
$\hat{A}$	normalised version of matrix A
g	grey value image
$(g_1, ..., g_m)^\top$	multi-channel image
x, y	spatial coordinates in an image or image sequence
t	temporal coordinate in an image sequence
$\boldsymbol{x}$	coordinate of the vector $\boldsymbol{x} = (x, y)^\top$
$\boldsymbol{x}_{\text{h}}$	homogeneous coordinate of the vector $\boldsymbol{x} = (x, y)^\top$, i.e. $(x, y, 1)^\top$
$[\boldsymbol{x}_{\text{h}}]_\times$	skew-symmetric matrix whose left and right null-space are $\boldsymbol{x}_{\text{h}}$
$(\boldsymbol{x}, \boldsymbol{x}')$	corresponding point pair in the left and right image
$(\boldsymbol{l}, \boldsymbol{l}')$	corresponding pair of epipolar lines in the left and right image
Ω	rectangular image domain
∂_x	abbreviation for $\frac{\partial}{\partial x}$
a_x	abbreviation for $\frac{\partial a}{\partial x}$

a_{xy}	abbreviation for $\frac{\partial^2 a}{\partial x \partial y}$
∇a	spatial gradient of a, i.e., $(\partial_x a, \partial_y a)^\top$
$\mathrm{div}(\boldsymbol{a})$	divergence of $\boldsymbol{a}$, i.e. $\partial_x a_1 + \partial_y a_2$
u, v	optical flow components in x and y direction
$\boldsymbol{w} = (u, v)^\top$	optical flow vector
$\mathcal{E}$	energy (integral expression)
$\mathcal{E}_\mathrm{D}$	data term (expression under the integral sign)
$\mathcal{E}_\mathrm{S}$	smoothness term (expression under the integral sign)
$\mathcal{E}_\mathrm{E}$	epipolar term (expression under the integral sign)
α	weight of the smoothness term
β	weight of the epipolar term
γ	weight of the gradient constancy assumption
$\Psi(s^2)$	nonquadratic penaliser
$\Psi'(s^2)$	diffusivity function
n_l	number of pixels in direction of dimension l
$[\,a\,]_{i,j}$	approximation/discrete version of expression a in pixel (i, j)
$\mathbf{L}$	discrete differential operator
$\mathcal{N}(i, j)$	set of neighbours of pixel (i, j)
$\lvert\mathcal{N}(i, j)\rvert$	number of neighbours of pixel (i, j)
$\mathcal{N}_l(i, j)$	set of neighbours of pixel (i, j) in direction of dimension l
$\mathcal{N}_l^-(i, j)$	set of neighbours of pixel (i, j) in direction of dimension l that have still to be processed
$\mathcal{N}_l^+(i, j)$	set of neighbours of pixel (i, j) in direction of dimension l that have already been processed

C Own Publications

Journal Papers

1. L. Valgaerts, A. Bruhn, M. Mainberger, and J. Weickert. Dense versus sparse approaches for estimating the fundamental matrix. *International Journal of Computer Vision*, 2011. Available at Springer Online. `http://dx.doi.org/10.1007/s11263-011-0466-7`

Conference Papers

2. S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer. Modeling temporal coherence for optical flow. In *Proc. 13th International Conference on Computer Vision*, Barcelona, November 2011. IEEE Computer Society Press. Accepted for publication.

3. L. Valgaerts, A. Bruhn, H. Zimmer, J. Weickert, C. Stoll, and C. Theobalt. Joint estimation of motion, structure and geometry from stereo sequences. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision – ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, pages 568–581. Springer, Berlin, 2010.

4. Oliver Vogel, Levi Valgaerts, Michael Breuß, and Joachim Weickert. Making shape from shading work for real-world images. In J. Denzler and G. Notni, editors, *Pattern Recognition*, Lecture Notes in Computer Science, pages 191–200, Jena, Germany, June 2009. Springer, Berlin.

5. H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H.-P. Seidel. Complementary optic flow. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition – EMMCVPR*, volume 5681 of *Lecture Notes in Computer Science*, pages 207–220. Springer, Berlin, 2009.

6. H. Zimmer, A. Bruhn, L. Valgaerts, M. Breuß, J. Weickert, B. Rosenhahn, and H.-P. Seidel. PDE-based anisotropic disparity-driven stereo vision. In O. Deussen, D. Keim, and D. Saupe, editors, *Proceedings of Vision, Modeling, and Visualization 2008*, pages 263–272, Konstanz, Germany, October 2008. Akademische Verlagsgesellschaft Aka.

7. L. Valgaerts, A. Bruhn, and J. Weickert. A variational model for the joint recovery of the fundamental matrix and the optical flow. In G. Rigoll, editor, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 314–324. Springer, Berlin, June 2008.

Technical Reports

8. L. Valgaerts, A. Bruhn, M. Mainberger, and J. Weickert. Dense versus sparse approaches for estimating the fundamental matrix. Technical Report 263, Department of Mathematics, Saarland University, Germany, 2010.

Bibliography

[ADPS02] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sánchez. Symmetrical dense optical flow estimation with occlusions detection. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 721–736. Springer, Berlin, 2002.

[ADSW02] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert. Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach. *Journal of Visual Communication and Image Representation*, 13(1/2):3–21, 2002.

[AELS99] L. Alvarez, J. Esclarín, M. Lefébure, and J. Sánchez. A PDE model for computing the optical flow. In *Proc. XVI Congreso de Ecuaciones Diferenciales y Aplicaciones*, pages 1349–1356, Las Palmas de Gran Canaria, Spain, September 1999.

[ALK07] T. Amiaz, E. Lubetzky, and N. Kiryati. Coarse to over-fine optical flow estimation. *Pattern Recognition*, 40(9):2496–2503, 2007.

[ASS+09] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building Rome in a day. In *Proc. Twelfth International Conference on Computer Vision*, Kyoto, October 2009. IEEE Computer Society Press.

[AWS99] L. Alvarez, J. Weickert, and J. Sánchez. A scale-space approach to nonlocal optical flow calculations. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 235–246. Springer, Berlin, 1999.

[BA91] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *Proc. 1991 IEEE Conference on Computer Vision and Pattern Recognition*, pages 292–302, Maui, HI, June 1991. IEEE Computer Society Press.

[BA96] M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, January 1996.

[BAHH92] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In F. Hodnett, editor, *Proc. Sixth European Conference on Mathematics in Industry*, pages 237–252. Teubner, Stuttgart, 1992.

[BAS07] R. Ben-Ari and N. Sochen. Variational stereo vision with sharp discontinuities and occlusion handling. In *Proc. Eleventh International Conference on Computer Vision*, Rio de Janeiro, October 2007. IEEE Computer Society Press.

[BBH08] D. Bradley, T. Boubekeur, and W. Heidrich. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In *Proc. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, June 2008. IEEE Computer Society Press.

[BBK05] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1):5–30, 2005.

[BBM09] T. Brox, C. Bregler, and J. Malik. Large displacement optical flow. In *Proc. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 41–48, Miami, FL, June 2009. IEEE Computer Society Press.

[BBPP10] L. Ballan, G. J. Brostow, J. Puwein, and M. Pollefeys. Unstructured video-based rendering: interactive exploration of casually captured videos. *ACM Transactions on Graphics*, 29(4), 2010.

[BBPW04] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optic flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, 2004.

[BBW06] T. Brox, A. Bruhn, and J. Weickert. Variational motion segmentation with level sets. In H. Bischof, A. Leonardis, and A. Pinz, editors, *Computer Vision – ECCV 2006*, volume 3951 of *Lecture Notes in Computer Science*, pages 471–483. Springer, Berlin, 2006.

[BCB97] M. J. Brooks, W. Chojnacki, and L. Baumela. Determining the ego-motion of an uncalibrated camera from instantaneous optical flow. *Journal of the Optical Society of America A*, 14:2670–2677, 1997.

[BETV08] H. Bay, A. Ess, T. Tuytelaars, and L. J. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[BFB94] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, February 1994.

[BHM00] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*. SIAM, Philadelphia, second edition, 2000.

[BHPS10] D. Bradley, W. Heidrich, T. Popa, and A. Sheffer. High resolution passive facial performance capture. *ACM Transactions on Graphics*, 29(4), 2010.

[BM10] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):500–513, August 2010.

[BMK10] T. Basha, Y. Moses, and N. Kiryati. Multi-view scene flow estimation: A view centered variational approach. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1506–1513, San Francisco, CA, June 2010. IEEE Computer Society Press.

[BPT88] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, August 1988.

[Bra77] A. Brandt. Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation*, 31(138):333–390, April 1977.

[BRN+09] D. Bitton, G. Rosman, T. Nir, A. M. Bruckstein, A. Feuer, and R. Kimmel. Over-parameterized optical flow using a stereoscopic constraint. Technical Report CIS-2009-18, Computer Science Department, Technion - Israel Institute of Technology, Israel, November 2009.

[BRS+07] S. Baker, S. Roth, D. Scharstein, M. J. Black, J. P. Lewis, and R. Szeliski. A database and evaluation methodology for optical flow. In *Proc. 2007 IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007. IEEE Computer Society Press.

[Bru06] A. Bruhn. *Variational Optic Flow Computation – Accurate Modelling and Efficient Numerics*. PhD thesis, Department of Mathematics and Comptuer Science, Saarland University, Germany, July 2006.

[BSL+09] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. Technical Report MSR-TR-2009-179, Microsoft Research, Redmond, WA, December 2009.

[BW05] A. Bruhn and J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proc. Tenth International Conference on Computer Vision*, volume 1, pages 749–755, Beijing, China, June 2005. IEEE Computer Society Press.

[BWKS05] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. Discontinuity-preserving computation of variational optic flow in real-time. In R. Kimmel, N. Sochen, and J. Weickert, editors, *Scale-Space and PDE Methods in Computer Vision*, volume 3459 of *Lecture Notes in Computer Science*, pages 279–290. Springer, Berlin, 2005.

[BWKS06] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision*, 70(3):257–277, December 2006.

[BWS05] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.

[Can86] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[CGZZ10] Q. Cai, D. Gallup, C. Zhang, and Z. Zang. 3D deformable face tracking with a commodity depth camera. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision – ECCV 2010*, volume 6313 of *Lecture Notes in Computer Science*, pages 229–242. Springer, Berlin, 2010.

[CK02] R. L. Carceroni and K. N. Kutulakos. Multi-view scene capture by surfel sampling: From video streams to non-rigid 3D motion, shape and reflectance. *International Journal of Computer Vision*, 49(2-3):175–214, September 2002.

[CLMC92] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 32:1895–1909, 1992.

[CM95] K. Chaudhury and R. Mehrotra. A trajectory-based computational model for optical flow estimation. 11(5):733–741, 1995.

[CM99] T. F. Chan and P. Mulet. On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM Journal on Numerical Analysis*, 36(2):354–367, 1999.

[CMK03] O. Chum, J. Matas, and J. Kittler. Locally optimized RANSAC. In J. van Leeuwen, G. Goos, and J. Hartmanis, editors, *Pattern Recognition*, volume 2781 of *Lecture Notes in Computer Science*, pages 236–243. Springer, Berlin, September 2003.

[CMO04] O. Chum, J. Matas, and S. Obdrzalek. Enhancing RANSAC by generalized model optimization. In K.-S. Hong and Z. Zhang, editors, *Proc. Sixth Asian Conference on Computer Vision*, volume 2 of *Lecture Notes in Computer Science*, pages 812–817, January 2004.

[Coo67] S. A. Coons. Surfaces for computer aided design of space forms. Technical Report MIT/LCS/TR-41, Massachusetts Institute of Technology, Cambridge, MA, June 1967.

[CPMK09] J. Courchay, J.-P. Pons, P. Monasse, and R. Keriven. Dense and accurate spatio-temporal multi-view stereovision. In H. Zha, R. Taniguchi, and S. Maybank, editors, *Proc. Ninth Asian Conference on Computer Vision*, volume 5995 of *Lecture Notes in Computer Science*, pages 11–22, China, September 2009.

[CSC+10] Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt. 3D shape scanning with a time-of-flight camera. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1173–1180, San Francisco, CA, June 2010. IEEE Computer Society Press.

[CV01] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, February 2001.

[CWM05] O. Chum, T. Werner, and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 772–779, San Diego, CA, June 2005. IEEE Computer Society Press.

[dAST+08] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun. Performance capture from sparse multi-view video. *ACM Transactions on Graphics*, 27(3), 2008.

[EBY99] G. Evans, J. Blackledge, and P. Yardley. *Numerical Methods for Partial Differential Equations*. Springer, Berlin, 1999.

[Els61] L. E. Elsgolc. *Calculus of Variations*. Pergamon, Oxford, 1961.

[Fau92] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In Giulio Sandini, editor, *Computer Vision – ECCV 1992*, volume 588 of *Lecture Notes in Computer Science*, pages 563–578. Springer, Berlin, 1992.

[Fau93] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, MA, 1993.

[FB81] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–385, 1981.

[FG87] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, Switzerland, June 1987.

[FH06] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 40(1):41–54, October 2006.

[Fit01] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 125–132, Kauai, HI, June 2001. IEEE Computer Society Press.

[FKN73] S. Fučik, A. Kratochvil, and J. Nečas. Kačanov–Galerkin method. *Commentationes Mathematicae Universitatis Carolinae*, 14(4):651–659, 1973.

[FLP01] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, Cambridge, MA, 2001.

[FP06] J.-M. Frahm and M. Pollefeys. RANSAC for (quasi-) degenerate data (QDEGSAC). In *Proc. 2006 IEEEComputer Society Conference on Computer Vision and Pattern Recognition*, pages 453–460, New York, NY, June 2006. IEEE Computer Society Press.

[FP08] Y. Furukawa and J. Ponce. Dense 3D motion capture from synchronized video streams. In *Proc. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, June 2008. IEEE Computer Society Press.

[FSHW04] C. Frohn-Schauf, S. Henn, and K. Witsch. Nonlinear multigrid methods for total variation denosing. *Computating and Visualization in Science*, 7(3–4):199–206, 2004.

[GF00] J. Gomes and O. Faugeras. Reconciling distance functions and level sets. *Journal of Visual Communication and Image Representation*, 11(2):209–223, 2000.

[GV89] G. H. Golub and C. M. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, MD, 1989.

[GWB10] S. Grewenig, J. Weickert, and A. Bruhn. From box filtering to fast explicit diffusion. In M. Goesele, S. Roth, A. Kuijper, B. Schiele, and K. Schindler, editors, *Pattern Recognition*, volume 6376 of *Lecture Notes in Computer Science*, pages 533–542. Springer, Berlin, 2010.

[GWW+05] I. Galić, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H.-P. Seidel. Towards PDE-based image compression. In N. Paragios, O. Faugeras, T. Chan, and C. Schnörr, editors, *Variational, Geometric and Level-Set Methods in Computer Vision*, volume 3752 of *Lecture Notes in Computer Science*, pages 37–48. Springer, Berlin, 2005.

[GZG+10] P. Gwosdek, H. Zimmer, S. Grewenig, A. Bruhn, and J. Weickert. A highly efficient GPU implementation for variational optic flow based on the Euler-Lagrange framework. In *Proc. 2010 ECCV Workshop on Computer Vision with GPUs*, Heraklion, Greece, September 2010.

[Hac85] W. Hackbusch. *Multigrid Methods and Applications*. Springer, New York, 1985.

[Han91] K. J. Hanna. Direct multi-resolution estimation of ego-motion and structure from motion. In *Proc. Workshop on Visual Motion*, pages 156–162. IEEE Computer Society Press, October 1991.

[Har92] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In Giulio Sandini, editor, *Computer Vision – ECCV 1992*, volume 588 of *Lecture Notes in Computer Science*, pages 579–587. Springer, Berlin, 1992.

[Har97] R. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, 1997.

[Har99] R. Hartley. Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2):115–127, 1999.

[HD07] F. Huguet and F. Devernay. A variational method for scene flow estimation from stereo sequences. In *Proc. Eleventh International Conference on Computer Vision*, Rio de Janeiro, October 2007. IEEE Computer Society Press.

[HG93] R. Hartley and R. Gupta. Computing matched-epipolar projection. In *Proc. 1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 549–555, New York, NY, June 1993. IEEE Computer Society Press.

[HGC92] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. 1992 IEEE International Conference on Image Processing*, pages 761–764, Champaign, IL, June 1992. IEEE Computer Society Press.

[HJ94] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1994.

[HRRS86] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. MIT Press, Cambridge, MA, 1986.

[HS81] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[HS88] C. G. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conference*, pages 147–152, Manchester, England, August 1988.

[Hub81] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[HZ00] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[IA99] M. Irani and P. Anandan. About direct methods. In Bill Triggs, Andrew Zisserman, and Richard Szeliski, editors, *Workshop on Vision Algorithms*, volume 1883, pages 267–277, Corfu, Greece, September 1999.

[IM06] M. Isard and J. MacCormick. Dense motion and disparity estimation via loopy belief propagation. In P. J. Narayanan, Shree K. Nayar, and Heung-Yeung Shum, editors, *Proc. Seventh Asian Conference on Computer Vision*, volume 3852 of *Lecture Notes in Computer Science*, pages 32–41, January 2006.

[KMK05] Y. H. Kim, A. M. Martinez, and A. C. Kak. Robust motion estimation under varying illumination. *Image and Vision Computing*, 23(4):365–375, April 2005.

[KNPS68] J. Kačur, J. Nečas, J. Polák, and J. Souček. Convergence of a method for solving the magnetostatic field in nonlinear media. *Aplikace Matematiky*, 13:456–465, 1968.

[KPT+07] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, April 2007.

[KSK06] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *Proc. 18th International Conference on Pattern Recognition, Part III*, volume 3, pages 15–18, Hong Kong, China, August 2006.

[KSO+00] K Kanatani, Y. Shimizu, N. Ohta, M. J. Brooks, W. Chojnacki, and A. van den Hengel. Fundamental matrix from optical flow: optimal computation and reliability evaluation. *Journal of Electronic Imaging*, 9:194–202, April 2000.

[KZ02] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002, Part III*, volume 2352 of *Lecture Notes in Computer Science*, pages 82–96. Springer, Berlin, 2002.

[Lev44] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *The Quarterly of Applied Mathematics*, 2:164–168, July 1944.

[LF96] Q.-T. Luong and O. D. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17(1):43–75, January 1996.

[Lin94] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer, Boston, 1994.

[LK81] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, Canada, August 1981.

[Lon81] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, September 1981.

[Low99] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. Seventh International Conference on Computer Vision*, pages 1150–1157, Corfu, Greece, September 1999. IEEE Computer Society Press.

[Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[LSY06] C. Lei, J. Selzer, and Y.-H. Yang. Region-tree based stereo using dynamic programming optimization. In *Proc. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2378–2385, Washington, DC, June 2006. IEEE Computer Society Press.

[LV96] Q.-T. Luong and T. Viéville. Canonical representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, 64(2):193–229, 1996.

[LV98] S.-H. Lai and B. C. Vemuri. Reliable and efficient computation of optical flow. *International Journal of Computer Vision*, 29(2):87–105, October 1998.

[LW08] A. W.-C. Liew and S. Wang. *Visual speech recognition: lip segmentation and mapping*. IGI Publishing, Hershey, PA, 2008.

[LZ99] C. T. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *Proc. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1125–1131, Ft. Collins, CO, June 1999. IEEE Computer Society Press.

[Mah36] P. C. Mahalanobis. On the generalised distance in statistics. In *Proceedings of the National Institute of Sciences of India*, volume 2, pages 49–55, April 1936.

[Mar63] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11:431–441, 1963.

[MB87] D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2):220–228, March 1987.

[MBW07] Y. Mileva, A. Bruhn, and J. Weickert. Illumination-robust variational optical flow with photometric invariants. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 152–162, Heidelberg, Germany, 2007. Springer, Berlin.

[MCT09] E. Murphy-Chutorian and M. M. Trivedi. Head pose estimation in computer vision: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4):607–626, February 2009.

[MG80] A. R. Mitchell and D. F. Griffiths. *The Finite Difference Method in Partial Differential Equations*. Wiley, Chichester, 1980.

[MH80] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London, Series B*, 207:187–217, 1980.

[MHK06] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2–3):90–126, 2006.

[MP98a] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5):703–719, May 1998.

[MP98b] E. Mémin and P. Pérez. A multigrid approach for hierarchical motion estimation. In *Proc. 6th International Conference on Computer Vision*, pages 933–938, Bombay, India, January 1998. IEEE Computer Society Press.

[MP02] E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision*, 46(2):129–155, 2002.

[MS05] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005.

[MS06] D. Bo Min and K. Sohn. Edge-preserving simultaneous joint motion-disparity estimation. In *Proc. 18th International Conference on Pattern Recognition*, pages 74–77, Hong Kong, August 2006.

[Nag83] H.-H. Nagel. Constraints for the estimation of displacement vector fields from image sequences. In *Proc. Eighth International Joint Conference on Artificial Intelligence*, volume 2, pages 945–951, Karlsruhe, West Germany, August 1983.

[NBK08] T. Nir, A. M. Bruckstein, and R. Kimmel. Over-parameterized variational optical flow. *International Journal of Computer Vision*, 76(2):205–216, 2008.

[NE86] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.

[OK95] N. Ohta and K. Kanatani. Optimal structure from motion algorithm for optical flow. *IEICE Transactions on Information and Systems*, E78-D(12):1559–1566, December 1995.

[PAT96] I. Patras, N. Alvertos, and G. Tziritas. Joint disparity and motion field estimation in steroscopic image sequences. In *Proc. 13th International Conference on Pattern Recognition*, volume 1, pages 359–362, Vienna, Austria, August 1996.

[PBB+06] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, April 2006.

[Pen56] R. Penrose. On best approximate solutions of linear matrix equations. *Proceedings of the Cambridge Philosophical Society*, 52:17–19, 1956.

[PKF07] J.-P. Pons, R. Keriven, and O. D. Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision*, 72(2):179–193, April 2007.

[PKV99] M. Pollefeys, R. Koch, and L. Van Gool. A simple and efficient rectification method for general motion. In *Proc. Seventh International Conference on Computer Vision*, volume 1, pages 496–501, Corfu, Greece, September 1999. IEEE Computer Society Press.

[PM87] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. In *Proc. IEEE Computer Society Workshop on Computer Vision*, pages 16–22, Miami Beach, FL, November 1987. IEEE Computer Society Press.

[PM90] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[PPV94] M. Proesmans, E. Pauwels, and L. Van Gool. Coupled geometry-driven diffusion equations for low-level vision. In B. M. ter Haar Romeny, editor, *Geometry-Driven Diffusion in Computer Vision*, volume 1 of *Computational Imaging and Vision*, pages 191–228. Kluwer, Dordrecht, 1994.

[PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, second edition, 1992.

[PVPO94] M. Proesmans, L. Van Gool, E. Pauwels, and A. Oosterlinck. Determination of optical flow and its discontinuities using non-linear diffusion. In J.-O. Eklundh, editor, *Computer Vision – ECCV '94*, volume 801 of *Lecture Notes in Computer Science*, pages 295–304. Springer, Berlin, 1994.

[RB05] S. Roth and M. Black. On the spatial statistices of optical flow. In *Proc. Tenth International Conference on Computer Vision*, volume 1, pages 42–49, Beijing, China, June 2005. IEEE Computer Society Press.

[RD96] L. Robert and R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In B. Buxton and R. Cipolla, editors, *Computer Vision – ECCV '96*, volume 1064 of *Lecture Notes in Computer Science*, pages 439–451. Springer, Berlin, 1996.

[RFP08] R. Raguram, J. M. Frahm, and M. Pollefeys. A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part II*, volume 5303 of *Lecture Notes in Computer Science*, pages 500–513. Springer, Berlin, 2008.

[RL87] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. Wiley, New York, 1987.

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[SAH91] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proc. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 310–315, Maui, HI, June 1991. IEEE Computer Society Press.

[Sam82] P. D. Sampson. Fitting conic sections to 'very scattered' data: An iterative refinement of the Bookstein algorithm. *Computer Graphics and Image Processing*, 18(1):97–108, January 1982.

[SB02] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer, New York, third edition, 2002.

[SBK10] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by GPU-accelerated large displacement optical flow. In *Computer Vision – ECCV 2010*, volume 6311 of *Lecture Notes in Computer Science*, pages 438–451. Springer, Berlin, 2010.

[SBW05] N. Slesareva, A. Bruhn, and J. Weickert. Optic flow goes stereo: a variational approach for estimating discontinuity-preserving dense dispartiy maps. In W. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 33–40. Springer, Berlin, 2005.

[SCD+06] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. 2006 IEEE Conference on Computer Vision and Pattern Recognition*, pages I: 519–528, New York, NY, June 2006. IEEE Computer Society Press.

[Sch88] H. R. Schwarz. *Numerische Mathematik*. Teubner, Stuttgart, 1988.

[Sch93] C. Schnörr. On functionals with greyvalue-controlled smoothness terms for determining optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:1074–1079, 1993.

[Sch94] C. Schnörr. Segmentation of visual motion by minimizing convex non-quadratic functionals. In *Proc. Twelfth International Conference on Pattern Recognition*, volume A, pages 661–663, Jerusalem, Israel, October 1994. IEEE Computer Society Press.

[SFS04] D. Schlesinger, B. Flach, and A. Shekhovtsov. A higher order MRF-model for stereo-reconstruction. In C. E. Rasmussen, H. H. Bülthoff, M. A. Giese, and B. Schölkopf, editors, *Pattern Recognition*, volume 3175 of *Lecture Notes in Computer Science*, pages 440–446. Springer, Berlin, 2004.

[SFV04] C. Strecha, R. Fransens, and L. Van Gool. A probabilistic approach to large displacement optical flow and occlusion detection. In D. Comaniciu, K. Kanatani, R. Mester, and D. Suter, editors, *Statistical Methods in Video Processing*, volume 3247 of *Lecture Notes in Computer Science*, pages 71–82, Berlin, 2004. Springer.

[SG07] J. Saragih and R. Goecke. Monocular and stereo methods for AAM learning from video. In *Proc. 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, June 2007. IEEE Computer Society Press.

[SGC10] J. Stühmer, S. Gumhold, and D. Cremers. Real-time dense geometry from a handheld camera. In M. Goesele, S. Roth, A. Kuijper, B. Schiele, and K. Schindler, editors, *Pattern Recognition*, volume 6376 of *Lecture Notes in Computer Science*, pages 11–20. Springer, Berlin, September 2010.

[SH89] D. Shulman and J. Hervé. Regularization of discontinuous flow fields. In *Proc. Workshop on Visual Motion*, pages 81–90, Irvine, CA, March 1989. IEEE Computer Society Press.

[Sha93] J. Shah. A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo. In *Proc. 1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 34–40, New York, NY, June 1993. IEEE Computer Society Press.

[SHS07] Y. Sheikh, A. Hakeem, and M. Shah. On the direct estimation of the fundamental matrix. In *Proc. 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, June 2007. IEEE Computer Society Press.

[SPC09a] F. Steinbrücker, T. Pock, and D. Cremers. Advanced data terms for variational optic flow estimation. In M.A. Magnor, B. Rosenhahn, and H. Theisel, editors, *Proceedings of the Vision, Modeling, and Visualization Workshop (VMV)*, pages 155–164. DNB, November 2009.

[SPC09b] F. Steinbrücker, T. Pock, and D. Cremers. Large displacement optical flow computation without warping. In *Proc. Twelfth International Conference on Computer Vision*, Kyoto, October 2009. IEEE Computer Society Press.

[SRBW11] C. Schmaltz, B. Rosenhahn, T. Brox, and J. Weickert. Region based pose tracking with occlusions using 3D models. *Machine Vision and Applications*, 2011. To appear.

[SRLB08] D. Sun, S. Roth, J. P. Lewis, and M. J. Black. Learning optical flow. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part III*, volume 5304 of *Lecture Notes in Computer Science*, pages 83–97. Springer, Berlin, 2008.

[SS02] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002.

[SS07] A. Salgado and J. Sánchez. Temporal constraints in large optical flow estimation. In R. Moreno-Díaz, F. Pichler, and A. Quesada-Arencibia, editors, *Computer Aided Systems Theory – EUROCAST 2007*, volume 4739 of *Lecture Notes in Computer Science*, pages 709–716. Springer, Berlin, 2007.

[SSS06] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics*, 25(3):835–846, 2006.

[ST94] J. Shi and C. Tomasi. Good features to track. In *Proc. 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 593–600, Seattle, WA, June 1994. IEEE Computer Society Press.

[Ste99] C. V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, 1999.

[STV03] C. Strecha, T. Tuytelaars, and L. Van Gool. Dense matching of multiple wide-baseline views. In *Proc. Ninth International Conference on Computer Vision*, volume 2, pages 1194–1201. IEEE Computer Society Press, October 2003.

[SvHV+08] C. Strecha, W. von Hansen, L. J. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proc. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, June 2008. IEEE Computer Society Press.

[TH84] R. Tsai and T. Huang. Uniqueness and estimation of thee-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1):13–26, 1984.

[TK91] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical Report CMU-CS-91-132, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, June 1991.

[TM97] P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.

[TP84] O. Tretiak and L. Pastor. Velocity estimation from image sequences with second order differential operators. In *Proc. Seventh International Conference on Pattern Recognition*, pages 16–19, Montreal, Canada, July 1984.

[TV98] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, Englewood Cliffs, 1998.

[TZ99] P. H. S. Torr and A. Zisserman. Feature based methods for structure and motion estimation. In Bill Triggs, Andrew Zisserman, and Richard Szeliski, editors, *Workshop on Vision Algorithms*, volume 1883, pages 278–294, Corfu, Greece, September 1999.

[TZM95] P. H. S. Torr, A. Zisserman, and S. J. Maybank. Robust detection of degenerate configurations for the fundamental matrix. In *Proc. Fifth International Conference on Computer Vision*, pages 1037–1042, Cambridge, MA, June 1995. IEEE Computer Society Press.

[UWSI10] C. Unger, E. Wahl, P. Sturm, and S. Ilic. Probalistic disparity fusion for real-time motion-stereo. In *Proc. Tenth Asian Conference on Computer Vision*, Lecture Notes in Computer Science, New Zealand, November 2010.

[Val07] L. Valgaerts. Combining variational and feature-based methods for motion estimation. Master's thesis, Department of Informatics, Technical University Munich, Department of Mathematics, Saarland University, Germany, 2007.

[VBMW11] L. Valgaerts, A. Bruhn, M. Mainberger, and J. Weickert. Dense versus sparse approaches for estimating the fundamental matrix. *International Journal of Computer Vision*, June 2011. Available online at http://dx.doi.org/10.1007/s11263-011-0466-7.

[VBR+05] S. Vedula, S. Baker, P. Rander, R. T. Collins, and T. Kanade. Three-dimensional scene flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):475–480, March 2005.

[VBSK00] S. Vedula, S. Baker, S. M. Seitz, and T. Kanade. Shape and motion carving in 6d. In *Proc. 2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 592–598, Hilton Head, SC, June 2000. IEEE Computer Society Press.

[VBVZ11] S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer. Modeling temporal coherence for optical flow. In *Proc. 13th International Conference on Computer Vision*, Barcelona, November 2011. IEEE Computer Society Press. Accepted for publication.

[VBW08] L. Valgaerts, A. Bruhn, and J. Weickert. A variational model for the joint recovery of the fundamental matrix and the optical flow. In G. Rigoll, editor, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 314–324. Springer, Berlin, June 2008.

[VBZ+10] L. Valgaerts, A. Bruhn, H. Zimmer, J. Weickert, C. Stoll, and C. Theobalt. Joint estimation of motion, structure and geometry from stereo sequences. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision – ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, pages 568–581. Springer, Berlin, 2010.

[VF95] T. Viéville and O. Faugeras. Motion analysis with a camera with unknown, and possibly varying intrinsic parameters. In *Proc. Fifth International Conference on Computer Vision*, pages 750–756, Cambridge, MA, June 1995. IEEE Computer Society Press.

[Vog02] C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002.

[vV91] S. van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*. SIAM, Philadelphia, 1991.

[WAH93] J. Weng, N. Ahuja, and T. Huang. Optimal motion and structure estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):864–884, 1993.

[WB94] H. Wang and M. Brady. A practical solution to corner detection. In *Proc. 1994 IEEE International Conference on Image Processing*, volume 1, pages 919–923, Austin, TX, November 1994. IEEE Computer Society Press.

[WCPB09] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure- and motion-adaptive regularization for high accuracy optic flow. In *Proc. Twelfth International Conference on Computer Vision*, Kyoto, October 2009. IEEE Computer Society Press.

[Wei94a] J. Weickert. Abschlußbericht zum Projekt "Nichtlineare Diffusionsfilter". In *Abschlußbericht und Bericht über die wissenschaftliche Tätigkeit Januar 1992 – Dezember 1993*, pages 191–209. Center for Applied Mathematics, Darmstadt and Kaiserslautern, 1994.

[Wei94b] J. Weickert. Scale-space properties of nonlinear diffusion filtering with a diffusion tensor. Technical Report 110, Laboratory of Technomathematics, University of Kaiserslautern, Germany, October 1994.

[Wei96] J. Weickert. *Anisotropic Diffusion in Image Processing*. PhD thesis, Department of Mathematics, University of Kaiserslautern, Germany, January 1996. Revised and extended version published by Teubner, Stuttgart, Germany, 1998.

[Wei00] J. Weickert. Design of nonlinear diffusion filters. In B. Jähne and H. Haußecker, editors, *Computer Vision and Applications*, pages 439–458. Academic Press, San Diego, 2000.

[Wes92] P. Wesseling. *An Introduction to Multigrid Methods*. Wiley, Chichester, 1992.

[WHA89] J. Weng, T. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):451–476, 1989.

[WHS+01] J. Weickert, J. Heers, C. Schnörr, K. J. Zuiderveld, O. Scherzer, and H. S. Stiehl. Fast parallel algorithms for a broad class of nonlinear variational diffusion approaches. *Real-Time Imaging*, 7(1):31–45, February 2001.

[Wit83] A. P. Witkin. Scale-space filtering. In *Proc. Eighth International Joint Conference on Artificial Intelligence*, volume 2, pages 945–951, Karlsruhe, West Germany, August 1983.

[WMR+09] A. Wedel, A. Meißner, C. Rabe, U. Franke, and D. Cremers. Detection and segmentation of independently moving objects from dense scene flow. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition – EMMCVPR*, volume 5681 of *Lecture Notes in Computer Science*, pages 14–27. Springer, Berlin, 2009.

[WND99] M. Woo, J. Neider, and T. Davis. *OpenGL Programming Guide*. Addison–Wesley Longman, Amsterdam, third edition, 1999.

[WPB+08] A. Wedel, T. Pock, J. Braun, U. Franke, and D. Cremers. Duality TV-$L1$ flow with fundamental matrix prior. In *Proc. Image and Vision Computing New Zealand*, Auckland, New Zealand, November 2008. IEEE Computer Society Press.

[WPB10] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2464–2471, San Francisco, CA, June 2010. IEEE Computer Society Press.

[WRV+08] A. Wedel, C. Rabe, T. Vaudrey, T. Brox, U. Franke, and D. Cremers. Efficient dense scene flow from sparse or dense stereo data. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008*, volume 5302 of *Lecture Notes in Computer Science*, pages 739–751. Springer, Berlin, 2008.

[WS01a] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, December 2001.

[WS01b] J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14(3):245–255, May 2001.

[WTP+09] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic Huber-L^1 optical flow. In *Proc. 20th British Machine Vision Conference*, London, UK, September 2009. British Machine Vision Association.

[WVM+08] A. Wedel, T. Vaudrey, A. Meissner, C. Rabe, T. Brox, U. Franke, and D. Cremers. An evaluation approach for scene flow with decoupled motion and position. In D. Cremers, B. Rosenhahn, A. L. Yuille, and F. R. Schmidt, editors, *Statistical and Geometrical Approaches to Visual Motion Analysis*, volume 5604 of *Lecture Notes in Computer Science*, pages 46–69, Berlin, September 2008. Springer.

[WWMT11] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *Proc. 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, June 2011. IEEE Computer Society Press. To appear.

[XJM10] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1293–1300, San Francisco, CA, June 2010. IEEE Computer Society Press.

[XS03] J. Xiao and M. Shah. Two-frame wide baseline matching. In *Proc. Ninth International Conference on Computer Vision*, pages 603–609. IEEE Computer Society Press, October 2003.

[XZ96] G. Xu and Z. Zhang. *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*, volume 6 of *Computational Imaging and Vision*. Kluwer, Dordrecht, 1996.

[You71] D. M. Young. *Iterative Solution of Large Linear Systems*. Academic Press, New York, 1971.

[ZBV+08] H. Zimmer, A. Bruhn, L. Valgaerts, M. Breuß, J. Weickert, B. Rosenhahn, and H.-P. Seidel. PDE-based anisotropic disparity-driven stereo vision. In O. Deussen, D. Keim, and D. Saupe, editors, *Proceedings of Vision, Modeling, and Visualization 2008*, pages 263–272, Konstanz, Germany, October 2008. Akademische Verlagsgesellschaft Aka.

[ZBW+09] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H.-P. Seidel. Complementary optic flow. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition – EMMCVPR*, volume 5681 of *Lecture Notes in Computer Science*, pages 207–220. Springer, Berlin, 2009.

[ZBW11] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, April 2011.

[ZDFL95] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78:87–119, 1995.

[ZGS+09] B. Zeisl, P. F. Georgel, F. Schweiger, E. Steinbach, and N. Navab. Estimation of location uncertainty for scale invariant feature points. In *Proc. 2009 British Machine Vision Conference*, London, England, September 2009.

[Zha98] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–195, 1998.

[ZK01] Ye Zhang and Chandra Kambhamettu. On 3D scene flow and structure estimation. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 778–785, Kauai, HI, December 2001. IEEE Computer Society Press.

[ZPB07a] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L^1 optical flow. In F.A. Hamprecht and B. Jähne C. Schnörr, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 214–223, Berlin, 2007. Springer.

[ZPB07b] C. Zach, T. Pock, and H. Bischof. A globally optimal algorithm for robust TV-L^1 range image integration. In *Proc. Ninth International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007. IEEE Computer Society Press.