Mathematical Image Analysis Group
Faculties of Mathematics and Computer Science
Saarland University

Variational Image Fusion

A Dissertation Submitted Towards the Degree
Doctor of Engineering (Dr.-Ing.)
of the Faculties of Mathematics and Computer Science
of Saarland University

submitted by
David Hafner

Saarbrücken, 2017

Day of Colloquium

05.02.2018

Dean of Faculty

Prof. Dr. Frank-Olaf Schreyer

Chair of Committee

Prof. Dr. Hans-Peter Seidel

Reviewers

Prof. Dr. Joachim Weickert
Prof. Dr. Coloma Ballester

Academic Assistant

Dr. Pascal Peter

Short Abstract

The main goal of this work is the fusion of multiple images to a single composite that offers more information than the individual input images. We approach those fusion tasks within a variational framework. First, we present iterative schemes that are well-suited for such variational problems and related tasks. They lead to efficient algorithms that are simple to implement and well-parallelisable. Next, we design a general fusion technique that aims for an image with optimal local contrast. This is the key for a versatile method that performs well in many application areas such as multispectral imaging, decolourisation, and exposure fusion. To handle motion within an exposure set, we present the following two-step approach: First, we introduce the complete rank transform to design an optic flow approach that is robust against severe illumination changes. Second, we eliminate remaining misalignments by means of brightness transfer functions that relate the brightness values between frames. Additional knowledge about the exposure set enables us to propose the first fully coupled method that jointly computes an aligned high dynamic range image and dense displacement fields. Finally, we present a technique that infers depth information from differently focused images. In this context, we additionally introduce a novel second order regulariser that adapts to the image structure in an anisotropic way.

Kurzzusammenfassung

Das Hauptziel dieser Arbeit ist die Fusion mehrerer Bilder zu einem Einzelbild, das mehr Informationen bietet als die einzelnen Eingangsbilder. Wir verwirklichen diese Fusionsaufgaben in einem variationellen Rahmen. Zunächst präsentieren wir iterative Schemata, die sich gut für solche variationellen Probleme und verwandte Aufgaben eignen. Danach entwerfen wir eine Fusionstechnik, die ein Bild mit optimalem lokalen Kontrast anstrebt. Dies ist der Schlüssel für eine vielseitige Methode, die gute Ergebnisse für zahlreiche Anwendungsbereiche wie Multispektralaufnahmen, Bildentfärbung oder Belichtungsreihenfusion liefert. Um Bewegungen in einer Belichtungsreihe zu handhaben, präsentieren wir folgenden Zweischrittansatz: Zuerst stellen wir die komplette Rangtransformation vor, um eine optische Flussmethode zu entwerfen, die robust gegenüber starken Beleuchtungsänderungen ist. Dann eliminieren wir verbleibende Registrierungsfehler mit der Helligkeitstransferfunktion, welche die Helligkeitswerte zwischen Bildern in Beziehung setzt. Zusätzliches Wissen über die Belichtungsreihe ermöglicht uns, die erste vollständig gekoppelte Methode vorzustellen, die gemeinsam ein registriertes Hochkontrastbild sowie dichte Bewegungsfelder berechnet. Final präsentieren wir eine Technik, die von unterschiedlich fokussierten Bildern Tiefeninformation ableitet. In diesem Kontext stellen wir zusätzlich einen neuen Regularisierer zweiter Ordnung vor, der sich der Bildstruktur anisotrop anpasst.

Abstract

The main goal of this work is the fusion of multiple images of the same scene to a single composite that provides more information than the individual inputs. While some applications are mainly interested in a nice looking output image, others aim for additional information such as high dynamic range or depth. In this work, we approach those fusion tasks within a variational framework that allows for a transparent modelling with a solid mathematical background.

As a first contribution, we propose novel iterative solvers for such variational techniques. In fact, they are applicable to various diffusion processes, elliptic problems, and constrained convex optimisation. In this context, basic iterative solvers such as explicit schemes, Richardson iterations, and projected gradient descent methods are simple to implement and well-suited for parallel computing. However, their efficiency suffers from severe step size restrictions. As remedy we introduce a simple and highly efficient acceleration strategy, which leads to our so-called *Fast Semi-Iterative* (FSI) schemes. They extrapolate a basic solver iteration with the previous iterate. To derive suitable extrapolation parameters, we establish a recursion relation that connects box filtering with an explicit scheme for 1D homogeneous diffusion. Our schemes avoid the main drawbacks of recent Fast Explicit Diffusion and Fast Jacobi techniques, and have an interesting connection to Polyak's heavy ball method in optimisation. Our experiments demonstrate benefits of our FSI schemes in the context of image processing as well as convex and strongly convex optimisation. In fact, we apply those schemes as efficient solvers for our fusion applications throughout this work.

Next, we present a general variational approach for image fusion that combines multiple images of the same scene to a single composite that is well-exposed and provides optimal saturation and local contrast. To this end, we design our model assumptions directly on the fusion result. In particular, we formulate the output image as a convex combination of the input and incorporate concepts from perceptually inspired contrast enhancement such as a local and nonlinear response. This output-driven approach is the key to our versatile image fusion method. We demonstrate the performance of our fusion scheme with several applications such as exposure fusion, multispectral imaging, and decolourisation. For all application domains, we conduct thorough validations that illustrate improvements w.r.t. state-of-the-art approaches that are specifically tailored to the individual tasks.

Additionally, we deal with camera and object motion during the acquisition of exposure series. Such movement might cause severe ghosting artefacts in the fusion result since pixels are merged that do not belong to the same object. As remedy we propose the following two-step approach: First, we model a variational optic flow approach that is robust w.r.t. severe illumination changes which are characteristic for exposure series. To this end, we introduce the *complete rank transform* (CRT) that stores for every pixel the intensity order in its neighbourhood. In this way, we preserve rich local information while providing an invariance to monotonically increasing intensity rescalings. We illustrate the robustness and accuracy of our method experimentally.

Next, we present an intuitive ghost removal approach to tackle remaining misalignments. In fact, we do not only identify remaining registration errors, but also compensate for them in an appropriate way. To this end, we base our method on the concept of a *brightness transfer function* (BTF) that relates brightness values between different frames. We compute those BTFs with image histograms that provide a robustness to small scene motion. Moreover, in this way the proposed approach is independent of the exposure times and the camera response function, and hence well-suited for the general exposure fusion setting. The processed images can directly serve as input of standard fusion techniques. We demonstrate this with experiments.

Provided additional knowledge about the exposure set, in particular the exposure times, we propose an alternative to the discussed two-step approach. In fact, we present the first technique that simultaneously computes an aligned *high dynamic range* (HDR) composite as well as dense displacement maps. In this way, we cannot only cope with dynamic scenes but even accurately represent the underlying scene and camera motion. The proposed joint optimisation has beneficial effects, such as an intrinsic ghost removal and an HDR-coupled smoothing. Our experiments show that both, the HDR images and the optic flow fields benefit substantially from those features and the induced mutual feedback.

Also in the context of images taken with varying focal settings, we exploit additional knowledge about the input data. In particular, we consider the relative distance of the focal planes. They specify regions where objects are captured sharply. In this way, we are able to infer depth information about the scene in addition to a fused all-in-focus image. To this end, we present a variational depth from focus technique that first identifies in-focus regions by means of specific sharpness criteria to create initial depth maps. Next, we smooth those depth maps with an advanced anisotropic diffusion process combined with robust data fidelity terms. Experiments with synthetic and real-world data demonstrate benefits of our approach compared to competing methods.

On top of that, we introduce a novel anisotropic second order regularisation strategy that explicitly accounts for the piecewise affine nature of real-world depth maps. To this end, we make use of two important concepts and link them in a profitable way: On the one hand, anisotropic regularisation is a well-established technique that has improved numerous computer vision approaches by direction-dependent smoothing. On the other hand, recent applications have uncovered the importance of second order regularisation. We build a bridge between both worlds, and propose a novel second order regulariser that allows to steer the unknown function and its slope in a direction-dependent way. To this end, we start with an isotropic second order coupling model, and systematically incorporate anisotropic concepts from first order approaches. We demonstrate benefits of the resulting approach with experiments, and apply it to improve our depth from focus method even further.

Zusammenfassung

Das grundlegende Ziel dieser Arbeit ist die Fusion mehrerer Bilder derselben Szene zu einem einzelnen Gesamtbild, das mehr Informationen bietet als die individuellen Eingangsbilder. Während einige Anwendungen hauptsächlich an einem schönen Ausgangsbild interessiert sind, zielen andere auf zusätzliche Informationen wie Hochkontrast oder Tiefe. Wir gehen diese Fusionsaufgaben in einem variationellen Rahmen an, der eine transparente Modellierung mit solider mathematischer Grundlage erlaubt.

Als ersten Beitrag stellen wir neue iterative Lösungsverfahren für solche Variationstechniken vor. Tatsächlich sind diese anwendbar für verschiedene Diffusionsprozesse, elliptische Probleme und konvexe Optimierung mit Nebenbedingungen. In diesem Kontext sind iterative Basislösungsverfahren wie explizite Verfahren, Richardsoniterationen und projizierte Gradientenabstiegsverfahren einfach zu implementieren und gut geeignet für paralleles Rechnen. Jedoch leidet deren Effizienz unter starken Schrittweitenbeschränkungen. Als Abhilfe stellen wir eine einfache und höchst effiziente Beschleunigungsstrategie vor, welche zu unseren sogenannten schnellen semi-iterativen (FSI) Verfahren führt. Diese extrapolieren eine Iteration des Basislösungsverfahrens mit der vorherigen Iterierten. Um geeignete Extrapolationsparameter herzuleiten, etablieren wir eine Rekursionsrelation, die Boxfilter mit einem expliziten Verfahren für 1D homogene Diffusion verbindet. Unsere Verfahren vermeiden die Hauptnachteile der modernen schnellen expliziten Diffusions- und schnellen Jakobi-Verfahren und haben eine spannende Verbindung zu Polyaks schwerer Kugelmethode in der Optimierung. Unsere Experimente zeigen Vorteile unserer FSI-Verfahren im Bildverarbeitungskontext sowie für konvexe und stark konvexe Optimierung. Zudem nutzen wir diese Verfahren als effiziente Lösungsmethoden für unsere Fusionsanwendungen.

Als nächstes präsentieren wir einen generellen Variationsansatz zur Bildfusion, der mehrere Bilder derselben Szene zu einem einzelnen Gesamtbild kombiniert, welches optimale Belichtung, Sättigung und lokalen Kontrast bietet. Hierzu gestalten wir unsere Modellannahmen direkt hinsichtlich des Fusionsergebnisses. Im Besonderen formulieren wir das Ausgangsbild als eine konvexe Kombination des Eingangs und integrieren Konzepte von wahrnehmungsinspirierter Kontrastverbesserung wie ein lokales und nichtlineares Ansprechverhalten. Dieser ausgangsgetriebene Ansatz ist der Schlüssel für unsere vielseitige Bildfusionsmethode. Wir demonstrieren die Leistungsfähigkeit unseres Fusionsschemas mit verschiedenen Anwendungen wie Belichtungsreihenfusion, Multispektralaufnahmen und Bildentfärbung. Wir führen für alle Anwendungsgebiete sorgfältige Evaluationen durch, welche Verbesserungen gegenüber neuartigen Verfahren veranschaulichen, die speziell auf die individuellen Aufgaben zugeschnitten sind.

Zusätzlich behandeln wir Kamera- und Objektbewegungen während einer Belichtungsreihenaufnahme. Solche Bewegungen können zu starken Geisterartefakten im Fusionsergebnis führen, weil Pixel gemischt werden, die nicht zum selben Objekt gehören. Daher schlagen wir folgenden Zweischrittansatz vor: Zunächst modellieren wir eine variationelle optische Flussmethode, die robust ist gegenüber starken, für Belichtungsreihen charakteristischen Beleuchtungsänderungen. Hierzu führen wir die komplette Rangtransformation (CRT) ein, die für jedes Pixel die Helligkeitsordnung

in seiner Nachbarschaft speichert. Damit bewahren wir reichhaltige lokale Information, während wir eine Invarianz gegenüber monoton steigenden Helligkeitsänderungen bieten. Experimente zeigen die Robustheit und Genauigkeit unserer Methode.

Als nächstes präsentieren wir eine intuitive Geisterentfernungsmethode, um verbleibende Registrierungsfehler zu handhaben. Tatsächlich identifizieren wir nicht nur Registrierungsfehler, sondern kompensieren sie sachgemäß. Hierzu gründen wir unsere Methode auf dem Konzept der Helligkeitstransferfunktion (BTF), welche die Helligkeitswerte zwischen verschiedenen Bildern in Verbindung setzt. Wir berechnen diese BTFs mit Bildhistogrammen, die eine Robustheit gegenüber kleinen Szenenbewegungen aufweisen. Des Weiteren ist die vorgeschlagene Methode dadurch unabhängig von den Belichtungszeiten und der Kameraantwortkurve und damit gut geeignet für die generelle Belichtungsreihenfusion. Wir verdeutlichen dies mit Experimenten.

Ausgestattet mit zusätzlichem Wissen über die Belichtungsreihe, insbesondere über die Belichtungszeiten, schlagen wir eine Alternative zum obigen Zweischrittansatz vor. Tatsächlich präsentieren wir die erste Technik, die gleichzeitig ein registriertes Hochkontrastbild (HDR) sowie dichte Bewegungsfelder berechnet. Dadurch können wir nicht nur dynamische Szenen behandeln, sondern die zugrundeliegende Szenen- und Kamerabewegung präzise beschreiben. Die gemeinsame Optimierung hat vorteilhafte Effekte wie eine intrinsische Geisterentfernung und eine HDR-gekoppelte Glättung. Unsere Experimente zeigen, dass sowohl die HDR Bilder als auch die Flussfelder substanziell von diesen Eigenschaften und der beidseitigen Rückkopplung profitieren.

Auch im Kontext von Bildern mit verschiedenen fokalen Einstellungen machen wir uns zusätzliches Wissen über die Eingangsdaten zu Nutzen. Im Besonderen betrachten wir die relative Distanz der Fokalebenen. Diese beschreiben Regionen, in denen Objekte scharf aufgenommen werden. Wir können somit Tiefeninformation zusätzlich zu dem fusionierten scharfen Bild ableiten. Hierzu präsentieren wir eine variationelle Tiefe-von-Fokus Technik, die zuerst fokussierte Regionen mit Hilfe von speziellen Schärfemaßen identifiziert, um initiale Tiefenkarten zu erstellen. Danach glätten wir diese Tiefenkarten mit einem hochentwickelten anisotropen Diffusionsprozess, kombiniert mit robusten Datentreuetermen. Experimente mit synthetischen und realen Daten demonstrieren Vorteile unserer Methode im Vergleich zu Konkurrenzverfahren.

Darüber hinaus stellen wir eine neuartige anisotrope Regularisierungstechnik zweiter Ordnung vor, die explizit für die stückweise affine Gestalt realer Tiefenkarten Rechnung trägt. Hierzu machen wir Gebrauch von zwei wichtigen Konzepten und verbinden sie gewinnbringend: Einerseits ist anisotrope Regularisierung eine fest etablierte Technik, die zahlreiche Bildanalysemethoden durch richtungsabhängige Glättung verbessert hat. Andererseits haben neuste Anwendungen die Wichtigkeit von Regularisierung zweiter Ordnung aufgedeckt. Wir bauen eine Brücke zwischen diesen beiden Welten und schlagen einen neuen Regularisierer zweiter Ordnung vor, der es erlaubt, die unbekannte Funktion und ihre Steigung in einer richtungsabhängigen Art und Weise zu lenken. Hierzu beginnen wir mit einem isotropen Kopplungsmodell zweiter Ordnung und integrieren systematisch anisotrope Konzepte von Methoden erster Ordnung. Wir demonstrieren die Vorteile des resultierenden Verfahrens mit Experimenten und wenden es an, um unsere Tiefe-von-Fokus Methode noch weiter zu verbessern.

Acknowledgements

At this point, I want to express my grateful thanks to all the people that supported me in doing this work. Without them, this thesis would certainly not have been possible.

First of all, I would like to express my gratitude to Prof. Joachim Weickert for giving me the opportunity to join his *Mathematical Image Analysis* (MIA) group and for his supervision. In this context, I also want to appreciate funding by the *Deutsche Forschungsgemeinschaft* (DFG) through a Gottfried Wilhelm Leibniz prize for Joachim Weickert and by the Graduate School of Computer Science at Saarland University.

Next, I want to gratefully thank Prof. Coloma Ballester for being the second reviewer of this thesis.

Certainly, my thanks also go to my collaborators Oliver Demetz, Christopher Schroers, Madina Boshtayeva, Prof. Martin Reißel, Peter Ochs, and Sven Grewenig for their support in the individual projects. Here, I also want to mention the master student Maria Luschkova since parts of her thesis contributed to this work as well. Surely, I also want to thank all other colleagues from the MIA group for providing a pleasant working atmosphere and for many fruitful discussions: Matthias Augustin, Leif Bergerhoff, Marcelo Cárdenas, Laurent Hoeltgen, Sebastian Hoffmann, Michael Krause, Markus Mainberger, Sabine Müller, Nico Persch, Pascal Peter, Simon Setzer, Sarah Andris, Martin Schmidt, Markus Schneider, and Yan Zhang. It was great fun for me to be part of this group and to enjoy the strong team spirit.

Moreover, I am very thankful to the secretary of the MIA group, Ellen Wintringer, for her uncomplicated and friendly help with organisational issues. In this regard, I also would like to acknowledge the support provided by the current and former system administrators Markus Hargarter, Heinz-Ulrich Griehl, and Peter Franke.

Last but not least, I want to thank my family for the continuous and unconditional support throughout my whole studies. Thank you very much!

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Scope and Contributions . . . 8
- 1.3 Thesis Outline . . . 11

2 Foundations **13**
- 2.1 Introduction . . . 13
- 2.2 Images as Functions . . . 14
  - 2.2.1 From Images to Functions . . . 14
  - 2.2.2 Colour Spaces . . . 14
- 2.3 Calculus of Variations . . . 17
  - 2.3.1 Variational Framework . . . 17
  - 2.3.2 Euler-Lagrange Equations . . . 17
  - 2.3.3 Numerical Solution . . . 17
- 2.4 Projections onto Closed Convex Sets . . . 18
- 2.5 Cholesky Decomposition . . . 19

3 Fast Semi-Iterative Schemes **21**
- 3.1 Introduction . . . 21
- 3.2 Related Work . . . 23
- 3.3 How to Benefit from Box Filtering . . . 24
  - 3.3.1 Explicit Scheme for 1D Linear Diffusion . . . 24
  - 3.3.2 Box Filtering via Iterative Explicit Diffusion . . . 24
  - 3.3.3 Accelerating the Explicit Scheme for 1D Linear Diffusion . . . 26
- 3.4 FSI Schemes for Diffusion Evolutions . . . 27
- 3.5 FSI Schemes for Elliptic Problems . . . 28
- 3.6 FSI Schemes for Constrained Optimisation . . . 29
- 3.7 Experiments . . . 31
  - 3.7.1 Inpainting with Edge-Enhancing Anisotropic Diffusion . . . 31
  - 3.7.2 Total Variation Regularisation . . . 32
  - 3.7.3 Performance on Nesterov's Worst Case Problems . . . 34
- 3.8 Limitations and Discussion . . . 35
- 3.9 Summary and Conclusions . . . 36
- 3.A Mathematical Analysis . . . 38
  - 3.A.1 Connection to Chebyshev Polynomials . . . 38
  - 3.A.2 Connection to FED and Fast Jacobi . . . 40
  - 3.A.3 Adaptation to Strongly Convex Problems . . . 41

4 Variational Image Fusion **45**
4.1 Introduction . . . 45
4.2 Related Work . . . 47
4.2.1 Multispectral Fusion . . . 47
4.2.2 Decolourisation . . . 47
4.2.3 Exposure Fusion . . . 48
4.2.4 Variational Contrast Enhancement . . . 49
4.3 Variational Model . . . 50
4.3.1 Energy Formulation . . . 50
4.3.2 Minimisation . . . 53
4.3.3 Model and Parameter Analysis . . . 55
4.4 Experiments . . . 58
4.4.1 Multispectral Fusion . . . 58
4.4.2 Decolourisation . . . 62
4.4.3 Exposure Fusion . . . 65
4.4.4 Further Image Fusion Applications . . . 70
4.5 Limitations and Discussion . . . 72
4.6 Summary and Conclusions . . . 73

5 Deghosting and HDR Imaging **75**
5.1 Introduction . . . 75
5.2 Related Work . . . 76
5.2.1 Image Matching . . . 77
5.2.2 Ghost Removal . . . 80
5.3 Alignment with Complete Rank Transform . . . 81
5.3.1 Relation to Previous Work . . . 83
5.3.2 Morphologically Invariant Descriptors . . . 84
5.3.3 Variational Model . . . 87
5.3.4 Minimisation . . . 89
5.3.5 Experiments . . . 91
5.3.6 Limitations and Discussion . . . 98
5.4 Ghost Removal with Brightness Transfer Function . . . 103
5.4.1 Relation to Previous Work . . . 103
5.4.2 Bidirectional BTF-Based Ghost Removal . . . 104
5.4.3 Modifications . . . 106
5.4.4 Experiments . . . 108
5.4.5 Limitations and Discussion . . . 109
5.5 Simultaneous HDR and Optic Flow Computation . . . 113
5.5.1 Relation to Previous Work . . . 114
5.5.2 Variational Model . . . 115
5.5.3 Minimisation . . . 119
5.5.4 Experiments . . . 119
5.5.5 Limitations and Discussion . . . 122

5.6 Summary and Conclusions . . . 126
5.6.1 Alignment with Complete Rank Transform . . . 126
5.6.2 Ghost Removal with Brightness Transfer Function . . . 126
5.6.3 Simultaneous HDR and Optic Flow Computation . . . 127
5.A Finite Differences for Anisotropic Diffusion . . . 129

6 Depth from Focus with Anisotropic Second Order Regularisation 131
6.1 Introduction . . . 131
6.2 Variational Depth from Focus . . . 132
6.2.1 Related Work . . . 133
6.2.2 Depth from Focus Framework . . . 135
6.2.3 Initial Depth Map Computation . . . 135
6.2.4 Depth Map Regularisation . . . 138
6.2.5 Image Fusion . . . 141
6.2.6 Experiments . . . 141
6.2.7 Limitations and Discussion . . . 146
6.3 Anisotropic Second Order Coupling Model . . . 157
6.3.1 Related Work . . . 158
6.3.2 Isotropic Coupling Model . . . 161
6.3.3 Extracting Directional Information . . . 163
6.3.4 Anisotropic Modification of Coupling Term . . . 163
6.3.5 Anisotropic Modification of Smoothness Term . . . 164
6.3.6 Anisotropic Coupling Model . . . 165
6.3.7 Minimisation . . . 166
6.3.8 Evaluation . . . 166
6.3.9 Application to Depth from Focus . . . 168
6.3.10 Limitations and Discussion . . . 170
6.4 Summary and Conclusions . . . 171
6.4.1 Variational Depth from Focus . . . 171
6.4.2 Anisotropic Second Order Coupling Model . . . 171

7 Conclusions and Outlook 173
7.1 Summary and Conclusions . . . 173
7.2 Outlook . . . 176

A Notation 179

B Abbreviations 185

C Own Publications 187

Bibliography 189

Chapter 1

Introduction

"All for one, and one for all."

Alexandre Dumas

Maurice Leloir (1894)

Contents

1.1 Motivation **1**

1.2 Scope and Contributions **8**

1.3 Thesis Outline **11**

1.1 Motivation

Pursuing his famous land- and seascape photography, the French photographer Gustave Le Gray (1820–1884) faced the following problem: Due to the physical limitations of the camera hardware and photographic material at that time, it was impossible to create a photograph that includes at the same time, the details of a bright sky and a dark sea. As remedy, he recorded two negatives with different exposure times such that each of them captured the details of the sky and the sea individually. Afterwards, he produced a single composite image by carefully combining suitable parts of both negatives in the printing process. In this way, the resulting image was well-exposed everywhere. In Figure 1.1(a), we depict an example image that was created by Le Gray in such a way already in 1857.

Another pioneer of this photographic technique was the English photographer Henry Peach Robinson (1830–1901). An early example of his work can be seen in Figure 1.1(b). Also here, the dynamic range of the captured scene exceeded the dynamic range of the camera hardware at that time. Thus, at least two differently exposed

(a) The Great Wave, Sète (1857) (b) Fading Away (1858)

Figure 1.1: Early examples of combination printing. (a) *The Great Wave, Sète* (1857) by Gustave Le Gray composed of two negatives [Le 57]. (b) *Fading Away* (1858) by Henry Peach Robinson composed of five negatives [Rob58].

negatives were required to capture the interior of the room and the sky within the window area in an appropriate way. However, to ensure an optimal focus of every part of the scene, Robinson made use of even more negatives. In total, he composed the depicted image, entitled *Fading Away*, by means of five negatives. Interestingly, Henry Peach Robinson was not only a photographer but also an author. Already in 1860, Robinson discussed this innovative multi-negative approach in an article of the British journal of photography [Rob60]. Nine years later, he even published an entire handbook for photographers that contains a chapter about this technique that he named *combination printing* [Rob69]. Historians still use this terminology to refer to such a photographic procedure that combines several negatives to a single composite image; see e.g. [Dav99, Aub02, SL15]. Besides Le Gray and Robinson, further famous pioneers of combination printing at that time have been William Lake Price (1810–1896) and Oscar Gustave Rejlander (1813–1875) [Han05].

Certainly, such problems were severe in the beginnings of photography. However, despite the enormous technical improvements of camera hardware from that time till today, similar problems still persist. In certain situations, a single photograph is still not enough to capture all the details of a scene. Indeed, there might be no camera setting such that the photograph offers the desired quality. As a common remedy, several images are taken while varying the settings of the camera. This can often be easily realised by so-called *burst* or *bracketing* modes of standard consumer cameras or also smartphones. Subsequently, in a postprocessing step, the captured set of images is combined to a single composite image that provides the desired quality. Obviously, the early combination printing techniques in the nineteenth century required a tedious manual selection and masking process of experts during the printing of the negatives. Nowadays, we aim at a fully automatic composition provided by intelligent image processing software. This digital process of combining several images to a single composite is generally referred to as *image fusion*. In the meantime, it has

(a) short exposure (b) middle exposure (c) long exposure (d) fused

Figure 1.2: Example of exposure fusion. *From left to right*: (a–c) Differently exposed input images [Jof07]. (d) Our fused composite image.

developed to an important area of research with various publications; see e.g. the text books [Sta08, Mit10, Zhe11] and the references therein.

In fact, the main goal of this work is to contribute to this image fusion research field in several aspects. To this end, let us first take a deeper look at some modern image fusion applications, and point out essential problems.

Exposure Fusion. As already mentioned, with standard consumer cameras is it not always possible to capture all the bright and dark details of a real-world scene with a single acquisition. Clearly, there exists specifically tailored camera hardware that is suited for this task; see e.g. [TKTS11, MRK+13] and references therein. However, such cameras are still on a rather prototype level or too expensive, and thus not available for average consumers. Nevertheless, taking several images with a standard camera while changing the exposure settings also allows to produce an overall well-exposed image in a postprocessing step. We illustrate this in Figure 1.2. Here, Figure 1.2(a–c) depict differently exposed input images of the same scene. None of those offers the desired quality. However, the fusion of such an exposure stack, which is referred to as *exposure fusion* [MKV09], results in an overall well-exposed composite image. In this regard, Figure 1.2(d) shows the result of our fusion approach proposed in this work.

Multispectral Fusion. Besides changing the exposure settings of the camera, it is also possible to use different filters or even different imaging modalities to acquire an input stack that contains different spectral ranges. In our example image set in Figure 1.3(a–d), we depict four images where each of them captures different details of Anton Raphael Mengs' copy of Raphael's Sistine Madonna from the Bergen Museum of Art [Cos12]. Fusing those images yields a composite that offers more details than any of the individual photographs; cf. Figure 1.3(e).

Another prominent example for multispectral image fusion is the combination of a visible spectrum and a near-infrared image; cf. Figure 1.4(a,b). While neither the visible spectrum image nor the near-infrared image captures all details of the scene, the fused composite in Figure 1.4(c) condenses the information of both inputs in an adequate way; see for example the mountains in the background. Such an enhancement of standard images with the near-infrared spectrum is for instance beneficial in the context of night vision; cf. e.g. [WGF+97, MB10].

(a) visible (b) raking light (c) infrared (d) ultraviolet (e) fused

Figure 1.3: Example of multispectral fusion. *From left to right*: (a) Visible spectrum, (b) raking light, (c) infrared, and (d) ultraviolet fluorescence image [Cos12]. (e) Our fused composite image.

(a) visible spectrum (b) near-infrared (c) fused

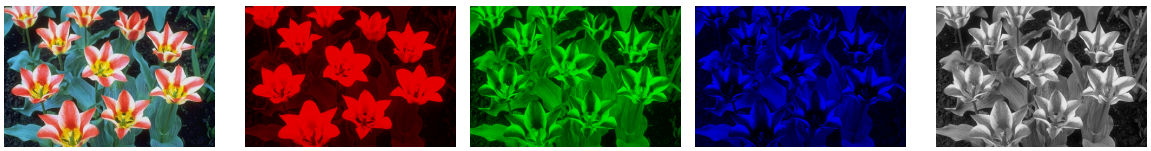
Figure 1.4: Example of visible spectrum and near-infrared fusion. *From left to right*: (a) Visible spectrum and (b) near-infrared image [BS11]. (c) Our fused composite image.

Focus Fusion. Especially in macro photography and optical microscopy, a typical problem is the limited depth of field of common cameras and microscopes, respectively. Due to this, objects only appear sharp at a certain distance range to the imaging device. Hence, it is often not possible to capture a single image that is sharp everywhere. Also in this context, a common remedy is to take several photographs while varying the focal settings. In this regard, *focus fusion* describes the task of combining the acquired focal stack to an all-in-focus composite that is desirably sharp in every image region. We illustrate this in Figure 1.5, where Figure 1.5(a–c) depict three of thirteen images of a focal stack capturing an insect. Here, the in-focus region moves away from the camera from frame to frame. Fusing this focal stack results in a composite image that provides the desired sharpness; cf. Figure 1.5(d).

Decolourisation. Although not immediately obvious, also decolourisation can be approached by image fusion. In general, decolourisation describes the conversion of a colour image to its greyscale representation. This is for instance important for black and white printing of coloured data, or for displaying such data on monochrome devices such as electronic book readers. The main challenge of decolourisation is to

(a) near focus (b) middle focus (c) far focus (d) fused

Figure 1.5: Example of focus fusion. *From left to right*: (a–c) Input images with different focal settings [ADA+04]. (d) Our fused composite image.



(a) input (b) R (c) G (d) B (e) fused

Figure 1.6: Example of decolourisation. *From left to right*: (a) Input colour image [Čad08]. (b–d) RGB colour channels. (e) Our fused greyscale composite.

preserve as much information as possible. In this work, we approach decolourisation by first decomposing the input image into its colour channels, and subsequently fusing them to a convincing greyscale counterpart; see Figure 1.6. While none of the colour channels (cf. Figure 1.6(b–d)) is able to represent the input colour image in Figure 1.6(a) appropriately, the fusion result in Figure 1.6(e) preserves the most important features. Obviously, due to the drastic intensity range restriction by a colour to greyscale transformation, it is hardly possible to preserve all details of the input image. Hence, to achieve convincing results the applied fusion technique is of great importance.

Deghosting. Up to now we have considered the case of perfectly aligned input images, where no camera or scene movement takes place during the acquisition of the image stack. However, in a practical setting, we cannot always assume such data sets. In fact, especially in the context of exposure series, such movement is often unavoidable. Unfortunately, already small misalignments might lead to severe artefacts in the fused composite image, which are often referred to as *ghosts*. We illustrate this in Figure 1.7. Here, a naïve direct fusion of the unaligned input images provides unsatisfactory results; cf. Figure 1.7(d). Hence, a registration of the input images is often inevitable. Obviously, due to the inherent drastic illumination changes between different exposures, this is a difficult problem which requires specifically tailored modifications and extensions of established alignment strategies. Moreover, in many cases an additional so-called *ghost removal* step is required to identify and to eliminate remaining misalignments.

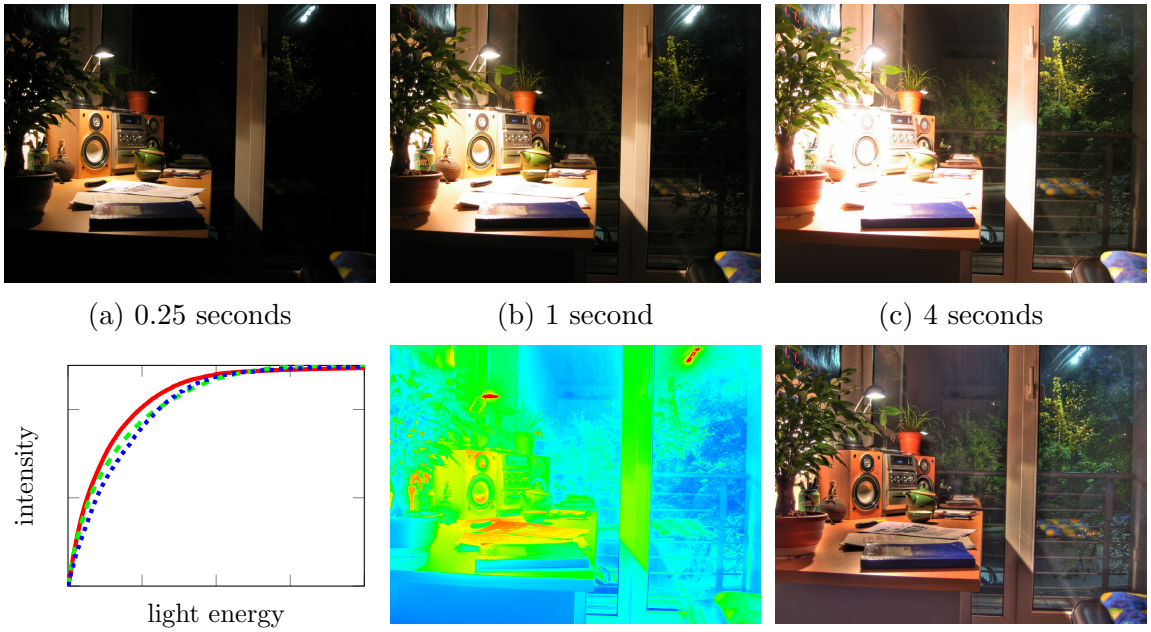
(a) short exposure (b) middle exposure (c) long exposure (d) fused

Figure 1.7: Illustration of ghosting artefacts. *From left to right*: (a–c) Unaligned input images with different exposure settings. (d) Naïve direct fusion causes undesirable ghosting artefacts due to camera and object motion during the acquisition.

High Dynamic Range Imaging. In many situations, additional information about the image stack is available. For instance, in the context of exposure bracketing, many images contain the applied camera settings in the file headers. In fact, this knowledge about the imaging process can be exploited to compute an *high dynamic range* (HDR) representation of the captured scene. We illustrate such an HDR reconstruction with differently exposed images in Figure 1.8. Here, Figure 1.8(a–c) depict three input images of an exposure set with known exposure times. In Figure 1.8(d), we plot the camera response functions for the RGB colour channels. In particular, such a response function describes the mapping of the incoming light energy (amount of photons) to the actual intensity values. In Figure 1.8(e,f), we show the recovered HDR irradiance values.[1] Compared to the input *low dynamic range* (LDR) images, the computed HDR irradiance image offers a much larger dynamic range. Since standard monitors and printers do not support such a large dynamic range, we present a logarithmic false colour representation of the HDR luminance values in Figure 1.8(e). Additionally, we compress the acquired high dynamic range again with a so-called *tone mapping* technique [FLW02] to produce the overall well-exposed LDR image in Figure 1.8(f). This tone mapped result contains bright and dark details from the entire input stack. In fact, this illustrates the high amount of information contained in the HDR image.

With the computed HDR data, it is for instance possible to produce LDR images with desired exposure times and response functions in a postprocessing step. Furthermore, such HDR values are beneficial for several image-based rendering applications; see e.g. [DM97, RHD+10]. Certainly, similar to exposure fusion (cf. Figure 1.7), also in this HDR context camera shakes and moving objects pose a big challenge for HDR reconstruction methods. In fact, an appropriate handling of such movement during the acquisition of the image set is of great importance and hence, has been an active field of research in recent years.

[1] The term *irradiance* describes how much light the image sensor receives per time (incident power). In this work, we consider HDR imaging as the computation of irradiance values incident on the imaging sensor. The actual *scene radiances* emitted/reflected by the scene objects are proportional to those irradiances. The exact relation depends on the imaging system and may be spatially varying; see e.g. [DM97] for further details.

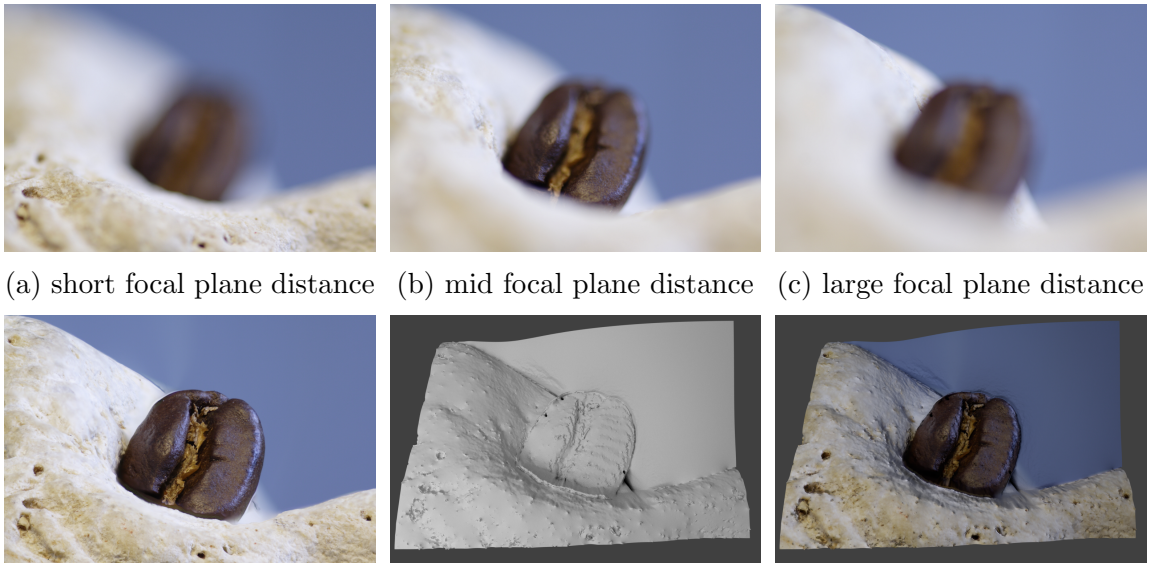


(a) 0.25 seconds (b) 1 second (c) 4 seconds

(d) camera response function (e) false colour irradiances (f) tone mapped irradiances

Figure 1.8: Illustration of high dynamic range imaging. *Top*: (a–c) Three of 13 differently exposed input images with given exposure times [Max05b]. *Bottom*: (d) Camera response function for red, green, and blue colour channel. (e) HDR irradiance values in logarithmic false colour representation from blue over green to red. (f) Tone mapped HDR irradiance values with tone mapping operator of Fattal et at. [FLW02].

Depth from Focus. Also in the context of focal stacks, knowledge about the imaging process is highly valuable. For instance, if the (relative) distance is known for which objects appear in focus for a particular image, it is possible to reconstruct 3D depth information besides the entirely sharp composite image. Such an approach is called *depth from focus* [DW88], or also *shape from focus* [NN94]. With the computed depth map and all-in-focus image, it is for instance possible to create images with desired focus settings in the postproduction; see e.g. [JBL12]. An example for depth from focus can be found in Figure 1.9. Here, Figure 1.9(a–c) depict three of the 91 input images of the focal stack, where the focal plane distance increases from frame to frame. In Figure 1.9(d,e), we show our computed all-in-focus image and depth map. Additionally, in Figure 1.9(f) we present a 3D rendering of the captured scene with the computed all-in-focus image as texture.



(a) short focal plane distance (b) mid focal plane distance (c) large focal plane distance

(d) all-in-focus image (e) rendered depth map (f) textured rendering

Figure 1.9: Example of depth from focus. *Top*: (a–c) Three of 91 input images with increasing focal plane distance [BO10]. *Bottom*: (d) Computed all-in-focus image, (e) computed depth map, and (f) rendering with all-in-focus image as texture. Here, we render our 3D models with the open source software *Blender* [Ble16].

1.2 Scope and Contributions

In this thesis, we approach those image fusion tasks within a variational framework. More specifically, we regard the input images as discrete samples of continuous 2D functions, and compute the unknowns by minimising suitable energy functionals. This not only allows a transparent modelling with an established mathematical background and well-founded optimisation techniques, but also leads to high accuracy and state-of-the-art results. Additionally, such a variational approach provides intuitive ways to couple individual tasks in joint models that highly benefit from the induced mutual feedback. In this regard, we prefer simple and comprehensible models with solid mathematical foundations to sophisticated approaches that are explicitly tuned for specific benchmark results. Hence, our main focus is not to reach unconditionally the ultimate quality, but to propose rather general concepts with strong mathematical justifications. In this way, we intend to uncover the main ingredients and ideas that are most responsible for qualitatively good results. Nevertheless, we hope that the presented concepts stimulate further research, and in this way might enter practical application software. While we consider image fusion as the main motivation for our work, the presented concepts can be seen as contributions to several research fields such as efficient numerical solvers, multispectral imaging, decolourisation, exposure fusion, high dynamic range imaging, stereo and optic flow computation, ghost removal, depth from focus, and higher order regularisation. In the following, we outline the individual contributions of our work in more detail.

Fast Semi-Iterative Schemes. As a first contribution, we introduce our novel so-called *Fast Semi-Iterative* (FSI) schemes that serve as efficient solvers for our upcoming variational fusion techniques. In fact, they are well-suited for various diffusion-like processes, elliptic problems, and constrained optimisation. For those tasks, there exist basic iterative approaches that are simple to implement and well-parallelisable; such as explicit schemes, Richardson iterations, or projected gradient descent methods. However, the efficiency of such basic solvers suffers from severe step size restrictions. As a remedy, we propose a simple and highly efficient acceleration technique: We extrapolate a basic solver iteration with the previous iterate. To derive suitable extrapolation weights, we establish a recursion relation that relates box filtering with an explicit scheme for 1D homogeneous diffusion. Our resulting FSI schemes overcome the main drawbacks of recent *Fast Explicit Diffusion* (FED) and Fast Jacobi techniques. Moreover, we uncover an interesting connection to the heavy ball method in optimisation. Our experiments show benefits of our FSI schemes for image processing applications and Nesterov's worst case problems in the context of convex and strongly convex optimisation.

Variational Image Fusion. Indeed, there exist many approaches that are tailored to the individual image fusion tasks that we discussed in the previous section. Hence, the natural question arises if those methods can be substituted by a single generic fusion technique. In fact, we observe that many fusion applications aim at a composite image with a high local contrast. Based on this finding, we propose a general variational image fusion method that works well in many application domains. In particular, we combine different images of the same scene to a single composite that is well-exposed and offers optimal saturation and local contrast. Previous research approaches this task by first precomputing application-specific weights based on the input, and then combining those weights with the images to the final composite in a second decoupled step. In contrast, we design our model assumptions directly on the fusion result. To this end, we formulate the output image as a convex combination of the input and incorporate concepts from perceptually inspired contrast enhancement techniques such as a local and nonlinear response. This output-driven approach is the key to the versatility of our general image fusion model. In this regard, we demonstrate the performance of our fusion scheme with several applications such as exposure fusion, multispectral imaging, and decolourisation. We conduct thorough validations that illustrate the improvements compared to state-of-the-art approaches that are specifically tailored to the individual tasks.

Alignment with Complete Rank Transform. Image fusion methods rely on the assumption that the input images are perfectly aligned. Otherwise, pixels that do not belong to the same object are merged in the fusion process. As discussed, this may lead to severe artefacts in the resulting composite. However, since camera shakes and moving objects during the image acquisition are omnipresent in a practical setting, we usually cannot assume perfectly aligned input images. Hence, a registration of

the image sets is often inevitable. To this end, we base on well-established variational alignment strategies and adapt them to the difficult exposure series setting. More specifically, we propose a novel patch-based descriptor that is invariant under monotonically increasing brightness changes, which are characteristic brightness changes for exposure series. We derive our novel *complete rank transform* (CRT) from successful descriptors proposed by Zabih and Woodfill in 1994, called *rank* and *census* [ZW94]. However, in contrast to rank and census, our CRT feature provides richer local information without leaving this class of invariance. We experimentally demonstrate quality improvements to related state-of-the-art techniques in the context of illumination-robust optic flow.

Ghost Removal with Brightness Transfer Function. Due to violations of the model assumptions such as occlusions or highly saturated image regions, we have to cope with remaining inaccuracies in the alignment results. As outlined above, even small misalignments may lead to severe ghosting artefacts in the fusion results. Hence, an additional postprocessing step is required to eliminate remaining misalignments. For this task, we present an intuitive *ghost removal* approach that identifies registration errors and compensates for them in an appropriate way. To this end, we base on the concept of a *brightness transfer function* (BTF) that relates the brightness values between the differently exposed input images. Following Grossberg and Nayar [GN03], we compute the BTFs by means of image histograms. In this way, the proposed approach is independent of the knowledge of the exposure times and the camera response function, and hence well-suited for the general exposure fusion setting. On top of that, we design our ghost removal technique in such a way that the processed images can directly serve as input of standard fusion techniques without modification.

Simultaneous HDR and Optic Flow Computation. As mentioned above, further knowledge about the image acquisition process such as the exposure times or the camera response function allows to infer additional information about the captured scene. In this regard, we discuss approaches to reconstruct physically plausible HDR irradiance values from a set of images taken with varying exposure times. Also here, we consider the case of unaligned input images, and demonstrate that this additional prior knowledge does not only allow an HDR reconstruction, but also supports the alignment process in a beneficial way. More specifically, we present the first fully coupled approach that simultaneously computes an aligned HDR composite as well as accurate dense displacement maps (optic flow fields). In this way, we cannot only cope with dynamic scenes but even precisely represent the underlying scene motion. The proposed joint optimisation has advantageous effects, such as an intrinsic ghost removal and an HDR-coupled smoothing. Both the HDR images and the optic flow fields benefit substantially from those features and the induced mutual feedback. We demonstrate this with synthetic and real-world experiments.

Variational Depth from Focus. Focus fusion is the task of combining a set of differently focused images to a single composite that is entirely sharp. Also in this context, we make beneficial use of further information. In fact, if for each image the (relative) distance is known at which objects appear in focus, it is possible to infer 3D depth information of the captured scene. The crucial point of such a *depth from focus* approach is the decision about the in-focus areas. To this end, we consider several sharpness criteria to construct initial depth maps. Next, we present a variational method that introduces novel regularisation strategies onto such depth maps. In particular, we regularise the depth maps with a sophisticated anisotropic technique combined with robust data fidelity terms. Experiments with synthetic and real-world data demonstrate that our new model provides significantly better quality than related state-of-the-art approaches.

Anisotropic Second Order Coupling Model. On top of that, we account for the piecewise affine shape of real-world depth maps. To this end, we propose a novel anisotropic second order regularisation technique that combines two important concepts in a beneficial way: On the one hand, anisotropic regularisation is a successful and well-established strategy that enables direction-dependent smoothing. On the other hand, recent computer vision applications show the need for second order smoothness in the context of depth map computation. In this work, we combine the benefits of both worlds, and introduce a second order regulariser that allows to penalise the unknown function and its slope in a direction-dependent way. In particular, we build on an isotropic coupling model and systematically incorporate anisotropic ideas from first order approaches. Moreover, we embed the resulting regulariser into our variational depth from focus approach, and experimentally demonstrate the benefits of the new model.

1.3 Thesis Outline

The order of the discussed contributions in the previous paragraphs reveals already a coarse outline of the presented work. The complete structure of this thesis is as follows: After a discussion of the foundations that build the basis of our work in Chapter 2, we present our novel iterative schemes in Chapter 3. In Chapter 4, we explain our general variational image fusion method that performs well for many fusion applications. Chapter 5 covers our two-step approach for the alignment of exposure series that consists of an optic flow-based registration and a ghost removal method. Provided further knowledge about the exposure series, we additionally present our joint model for HDR and optic flow computation. In the subsequent Chapter 6, we propose our depth from focus approach and introduce our novel anisotropic second order regularisation technique. We conclude this thesis with a summary and outlook in Chapter 7. In the Appendix A and B, we give a list of the notation and abbreviations applied throughout this thesis. Moreover, in Appendix C we itemise our individual publications.

Chapter 2

Foundations

"You have to learn the rules of the game.
And then you have to play better
than anyone else."

Albert Einstein

pixabay.com

Contents

2.1 Introduction **13**

2.2 Images as Functions **14**

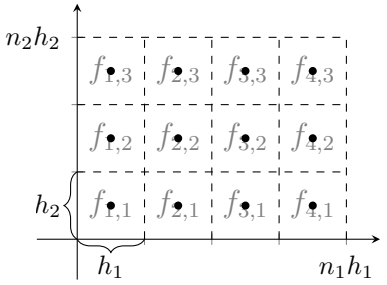
2.3 Calculus of Variations **17**

2.4 Projections onto Closed Convex Sets **18**

2.5 Cholesky Decomposition **19**

2.1 Introduction

This chapter builds the basis of our work. First, we describe how we interpret digital images as continuous functions in Section 2.2. In this context, we also introduce different colour spaces. Next, we present our general variational framework and explain its minimisation by means of the Euler-Lagrange equations in Section 2.3. In Section 2.4, we discuss additional side-constraints that restrict the unknowns to some convex sets. Moreover, we consider the direct solution of equation systems by means of a Cholesky decomposition in Section 2.5.



(a) sampling

$$(f_{1,3}, \ldots, f_{4,3}, f_{1,2}, \ldots, f_{4,2}, f_{1,1}, \ldots, f_{4,1})^\mathsf{T} \in \mathbb{R}^{n_1 \cdot n_2}$$

(b) vector notation

Figure 2.1: Illustration of 2D function sampling with $n_1 = 4$ and $n_2 = 3$, and its representation in vector notation, where we concatenate the image rows on top of each other.

2.2 Images as Functions

2.2.1 From Images to Functions

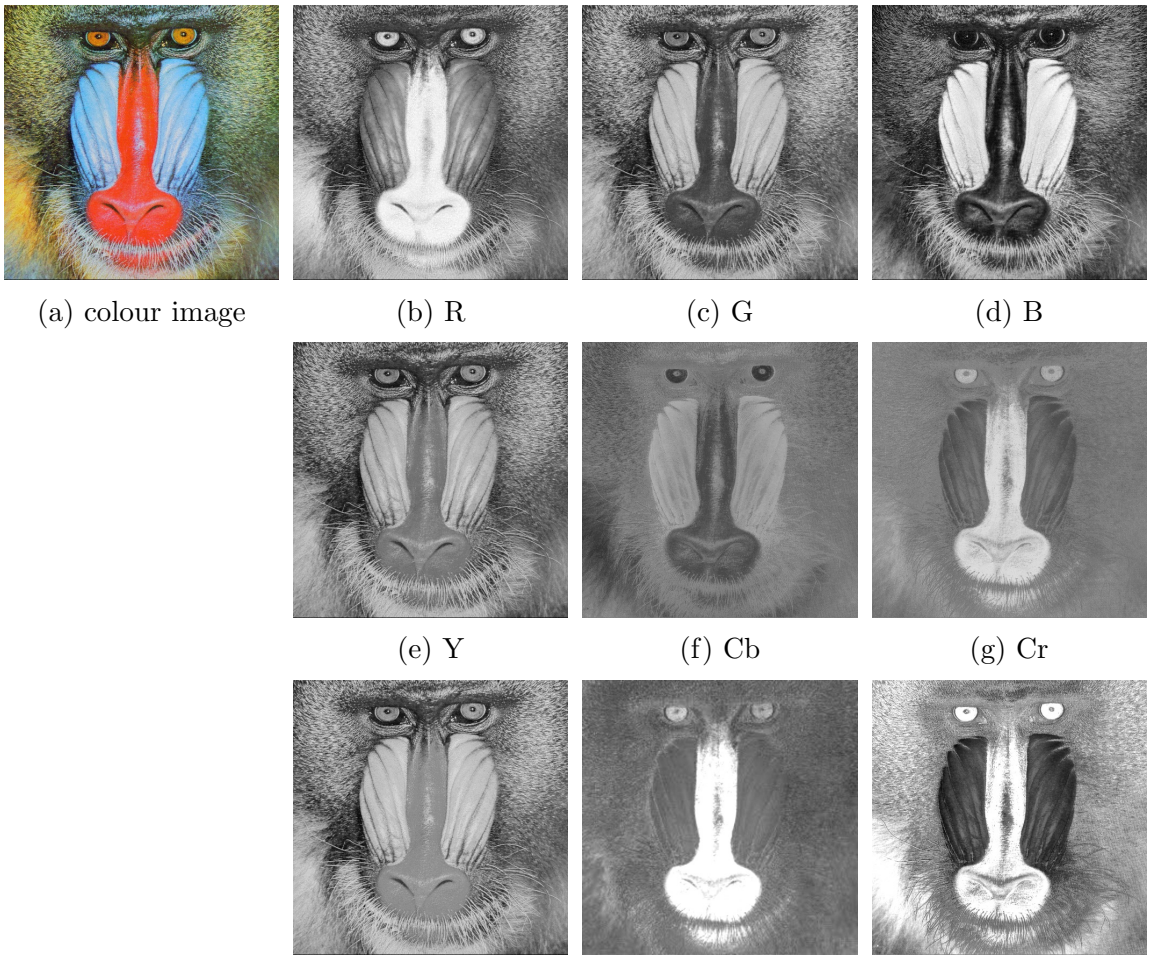
Generally, a digital greyscale image is a two-dimensional array where each cell represents a single picture element (*pixel*). The associated cell value specifies the brightness of the corresponding pixel. We interpret those pixels as discrete sampling points of a continuous 2D function $f\colon \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^2$ describes the rectangular image domain. Let h_1 and h_2 define the horizontal and vertical grid sizes, and let n_1 and n_2 denote the corresponding number of pixels in horizontal and vertical direction. Then, the value of each pixel (i, j) with $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$ is obtained by sampling the 2D function f at $\big((i - \frac{1}{2}) \cdot h_1, (j - \frac{1}{2}) \cdot h_2\big)$; cf. Figure 2.1(a).

2.2.2 Colour Spaces

The extension of the discussed concept from greyscale to colour images is straightforward. Instead of considering scalar-valued functions $f\colon \Omega \to \mathbb{R}$, we regard vector-valued functions $\boldsymbol{f}\colon \Omega \to \mathbb{R}^3$, where each pixel contains a 3D vector that specifies its colour. In this work, we consider three different colour spaces illustrated in Figure 2.2. For a broad discussion of various colour spaces and related concepts, we refer the interested reader e.g. to the extensive text book of Wyszecki and Stiles [WS00].

RGB

The most common colour space is the RGB colour space where R represents the red, G the green, and B the blue colour channel. This colour space is motivated by the trichromatic human visual system that contains three different kind of cones. Each of them is most sensitive to the light wavelength that corresponds to the mentioned colours red, green, and blue; see e.g. [Fai13]. If not explicitly mentioned otherwise, we apply this RGB colour space in our work.



(a) colour image (b) R (c) G (d) B

(e) Y (f) Cb (g) Cr

(h) L (i) a (j) b

Figure 2.2: Illustration of colour spaces. *In reading order*: (a) Colour test image *Baboon*. (b–d) Decomposition in RGB colour space. (e–g) Decomposition in YCbCr colour space. (h–j) Decomposition in CIE-Lab colour space. For visibility reasons, the CIE-Lab channels are shifted and scaled appropriately.

YCbCr

The RGB colour space separates a colour into the primary colours red, green, and blue. However, for some applications a different colour decomposition might be beneficial. To this end, the YCbCr colour space is frequently applied. Here, the luma channel Y describes the luminance of the colour, whereas the chroma channels Cb and Cr describe the blue-yellow and red-cyan chromatic components, respectively. For RGB and YCbCr values in the range $[0, 1]$, the transformations between the two colour spaces are given by (cf. [Pra01])

$$\begin{pmatrix} \mathrm{Y} \\ \mathrm{Cb} \\ \mathrm{Cr} \end{pmatrix} = \begin{pmatrix} 0.0 \\ 0.5 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.299000 & 0.587000 & 0.144000 \\ -0.168736 & -0.331264 & 0.500000 \\ 0.500000 & -0.418668 & -0.081312 \end{pmatrix} \begin{pmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{pmatrix} \tag{2.1}$$

and

$$\begin{pmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1.000000 & 0.000926 & 1.401687 \\ 1.000000 & -0.343695 & -0.714169 \\ 1.000000 & 1.772160 & 0.000990 \end{pmatrix} \left(\begin{pmatrix} \mathrm{Y} \\ \mathrm{Cb} \\ \mathrm{Cr} \end{pmatrix} - \begin{pmatrix} 0.0 \\ 0.5 \\ 0.5 \end{pmatrix} \right) . \tag{2.2}$$

CIE-Lab

The CIE-Lab colour space is designed in such a way that the Euclidean distance between two colours reflects the colour difference perceived by a human observer; see e.g. [Fai13]. Since the RGB colour space is device-dependant while the CIE-Lab space is not, a direct transformation from RGB values to CIE-Lab values is not possible. Hence, in a fist step, the RGB values need to be transformed to a device-independent representation. Following the sRGB standard (IEC 61966-2-1:1999), the conversion from RGB to the device-independent CIE-XYZ colour space is given by

$$\bar{\mathrm{C}} = \begin{cases} \frac{\mathrm{C}}{12.92} & \text{if } \mathrm{C} \leq 0.04045, \\ \left(\frac{\mathrm{C}+0.055}{1.055}\right)^{2.4} & \text{else} \end{cases} \tag{2.3}$$

with $\mathrm{C} \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$, followed by

$$\begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{pmatrix} = \begin{pmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{pmatrix} \begin{pmatrix} \bar{\mathrm{R}} \\ \bar{\mathrm{G}} \\ \bar{\mathrm{B}} \end{pmatrix} . \tag{2.4}$$

Then, the nonlinear transformation from CIE-XYZ to CIE-Lab reads (cf. [Pra01, GW07])

$$\begin{pmatrix} \mathrm{L} \\ \mathrm{a} \\ \mathrm{b} \end{pmatrix} = \begin{pmatrix} 116 \cdot \Theta(\mathrm{Y}/\mathrm{Y_n}) - 16 \\ 500 \cdot (\Theta(\mathrm{X}/\mathrm{Y_n}) - \Theta(\mathrm{Y}/\mathrm{Y_n})) \\ 200 \cdot (\Theta(\mathrm{Y}/\mathrm{Y_n}) - \Theta(\mathrm{Z}/\mathrm{Z_n})) \end{pmatrix} , \tag{2.5}$$

where

$$\begin{pmatrix} \mathrm{X_n} \\ \mathrm{Y_n} \\ \mathrm{Z_n} \end{pmatrix} = \begin{pmatrix} 0.950456 \\ 1.000000 \\ 1.089058 \end{pmatrix} \tag{2.6}$$

is the reference white point under the illuminant D65, and

$$\Theta(z) \;=\; \begin{cases} 7.787 \cdot z + 0.1379 & \text{if } z \leq 0.008856, \\ \sqrt[3]{z} & \text{else.} \end{cases} \tag{2.7}$$

2.3 Calculus of Variations

2.3.1 Variational Framework

In this section, we explain the general variational framework that forms the basis of our models. Since we work with two-dimensional images, we restrict ourselves to functions of $\boldsymbol{x} = (x_1, x_2)^\mathsf{T} \in \Omega \subset \mathbb{R}^2$. Moreover, we consider vector-valued functions $\boldsymbol{u} = (u_1, \ldots, u_{n_u})^\mathsf{T}$, where the number of unknown functions n_u and the meaning of u_i depend on the actual application. Our general energy functional is given by

$$E(\boldsymbol{u}) \;=\; \int_\Omega \mathcal{F}(u_1, \ldots, u_{n_u}, \boldsymbol{\nabla} u_1, \ldots, \boldsymbol{\nabla} u_{n_u}) \,\mathrm{d}\boldsymbol{x}\,, \tag{2.8}$$

where $\boldsymbol{\nabla} \coloneqq (\partial_{x_1}, \partial_{x_2})^\mathsf{T}$ denotes the spatial gradient operator. The design of the integrand $\mathcal{F}(u_1, \ldots, u_{n_u}, \boldsymbol{\nabla} u_1, \ldots, \boldsymbol{\nabla} u_{n_u})$ is the crucial point of all our variational methods. It depends on the considered problem and the associated model assumptions. Generally, we are interested in the desired solution $\boldsymbol{u}$ that is computed as a minimiser of the energy functional in (2.8).

2.3.2 Euler-Lagrange Equations

According to the calculus of variations (see e.g. [GF00]), the necessary conditions for a minimiser of the energy in (2.8) are given by the so-called *Euler-Lagrange equations*

$$\mathcal{F}_{u_i} - \partial_{x_1} \mathcal{F}_{\partial_{x_1} u_i} - \partial_{x_2} \mathcal{F}_{\partial_{x_2} u_i} \;=\; 0 \qquad (i = 1, \ldots, n_u)\,. \tag{2.9}$$

With $\boldsymbol{n}$ as the outer normal vector on the image boundary $\partial\Omega$, the corresponding boundary conditions read

$$\boldsymbol{n}^\mathsf{T} \begin{pmatrix} \mathcal{F}_{\partial_{x_1} u_i} \\ \mathcal{F}_{\partial_{x_2} u_i} \end{pmatrix} \;=\; 0 \qquad (i = 1, \ldots, n_u)\,. \tag{2.10}$$

2.3.3 Numerical Solution

Generally, we discretise the Euler-Lagrange equations on a rectangular grid with uniform horizontal and vertical grid sizes h_1 and h_2, respectively. The resulting discrete Euler-Lagrange equations lead to (non)linear systems of equations in the form of

$$\boldsymbol{B}(\boldsymbol{u})\, \boldsymbol{u} \;=\; \boldsymbol{d}(\boldsymbol{u})\,, \tag{2.11}$$

where the matrix- and vector-valued functions $\boldsymbol{B}\colon \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N$ and $\boldsymbol{d}\colon \mathbb{R}^N \to \mathbb{R}^N$ might depend in a nonlinear way on $\boldsymbol{u} \in \mathbb{R}^N$. Please note that we consider here in the discrete setting a vector notation for the unknown $\boldsymbol{u}$. In fact, there is an unique one-to-one mapping between the 2D pixel notation in (2.9) and the single index notation in (2.11). More specifically, this is achieved by stacking the rows of each channel u_i, $i = 1, \ldots, n_u$, on top of each other to form the vector $\boldsymbol{u}$; cf. Figure 2.1(b). With n_1 and n_2 denoting the number of pixels in horizontal and vertical direction, the number of equations N is given by $N = n_u \cdot n_1 \cdot n_2$. We present efficient schemes for the solution of such large equation systems in Chapter 3.

2.4 Projections onto Closed Convex Sets

For some applications it is beneficial to constrain the solution of the minimisation problem to a specific set. In this regard, we deal with closed convex sets $\mathcal{C}$: A set $\mathcal{C}$ is said to be convex if for all $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{C}$ and $\epsilon \in [0, 1]$ the point $(1 - \epsilon)\boldsymbol{u} + \epsilon\boldsymbol{v}$ also belongs to the set $\mathcal{C}$ (cf. [BV04]).

To realise such constrains, we project the current solution after each solver step onto the permitted convex set. In this regard, the Euclidean projection of a vector $\boldsymbol{u} \in \mathbb{R}^n$ onto a closed convex set $\mathcal{C} \subset \mathbb{R}^n$ is given by minimising (see e.g. [PB14])

$$E(\check{\boldsymbol{u}}) \;=\; |\check{\boldsymbol{u}} - \boldsymbol{u}|^2 \qquad \text{s.t.} \quad \check{\boldsymbol{u}} \in \mathcal{C}\,, \tag{2.12}$$

where $|\cdot|$ denotes the Euclidean norm. Next, we give three examples of convex sets and their corresponding projections that we consider in this work; cf. Figure 2.3. Further details about those projections onto convex sets can for instance be found in [PB14].

Box. First, we consider the convex set $\mathcal{C}_{\text{box}} = \{\boldsymbol{u} \mid \boldsymbol{a} \leq \boldsymbol{u} \leq \boldsymbol{b}\}$, where the relation $\leq$ is meant component-wise, and in this sense $\boldsymbol{a} \leq \boldsymbol{b}$. The analytic projection onto this set is for each component u_i of the vector $\boldsymbol{u}$ given by

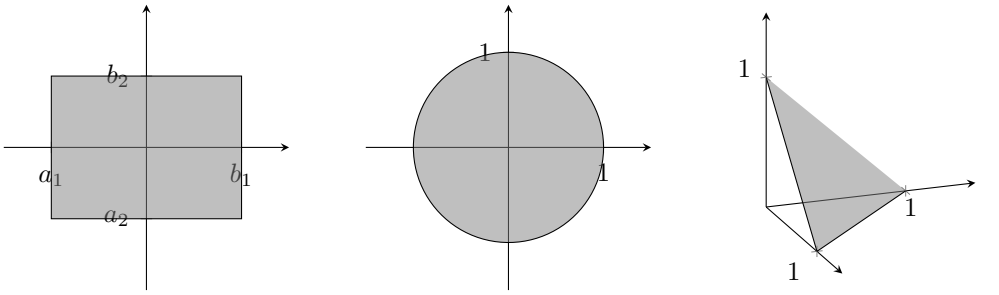
$$\mathcal{P}_{\mathcal{C}_{\text{box}}}(u_i) \;=\; \begin{cases} a_i & \text{if } u_i \leq a_i\,, \\ u_i & \text{if } a_i < u_i < b_i\,, \\ b_i & \text{else.} \end{cases} \tag{2.13}$$

Unit Ball. A unit ball is defined by the set $\mathcal{C}_{\text{ball}} = \{|\boldsymbol{u}| \leq 1\}$. The corresponding projection of $\boldsymbol{u}$ onto $\mathcal{C}_{\text{ball}}$ reads

$$\mathcal{P}_{\mathcal{C}_{\text{ball}}}(\boldsymbol{u}) \;=\; \frac{\boldsymbol{u}}{\max\{1, |\boldsymbol{u}|\}}\,. \tag{2.14}$$

Simplex. The unit simplex is defined by the convex set $\mathcal{C}_{\text{simp}} = \{\boldsymbol{u} \mid \boldsymbol{u} \geq \boldsymbol{0} \text{ and } \boldsymbol{u}^{\mathsf{T}}\boldsymbol{1} = 1\}$, where $\boldsymbol{1} \in \mathbb{R}^n$ is a vector of ones. The analytic projection onto this set is for each component u_i given by

$$\mathcal{P}_{\mathcal{C}_{\text{simp}}}(u_i) \;=\; \max\{0, u_i - \theta\}\,, \tag{2.15}$$



(a) box in $\mathbb{R}^2$ (b) unit ball in $\mathbb{R}^2$ (c) simplex in $\mathbb{R}^3$

Figure 2.3: Illustration of convex sets.

Algorithm 2.1: Projection onto simplex in $\mathbb{R}^n$ [SS06].

input : vector $\boldsymbol{u} \in \mathbb{R}^n$
output: projected vector $\check{\boldsymbol{u}} \in \mathcal{C}_{\text{simp}}$

$\boldsymbol{v} = \text{sort}(\boldsymbol{u})$ such that $v_1 \geq \ldots \geq v_n$

$q = \max\left\{ j \in \{1, \ldots, n\} \;\middle|\; v_j - \frac{1}{j}\big(\sum_{i=1}^{j} v_i - 1\big) > 0 \right\}$

$\theta = \frac{1}{q}\big(\sum_{i=1}^{q} v_i - 1\big)$

for $i = 1, \ldots, n$ **do**

  $\check{u}_i = \max\{0, u_i - \theta\}$

where θ is chosen such that the projected vector $\check{\boldsymbol{u}}$ satisfies $\check{\boldsymbol{u}}^\mathsf{T}\mathbf{1} = 1$. An efficient way for computing such a projection onto a simplex in $\mathbb{R}^n$ is given in Algorithm 2.1 [SS06].

2.5 Cholesky Decomposition

The discussed Euler-Lagrange equations lead to large (non)linear systems in the form of (2.11). Efficient iterative solvers for such large systems are presented in Chapter 3. However, in this thesis, we are also concerned with small linear systems where direct solution methods are applicable. In this context, we deal with small symmetric positive definite matrices $\boldsymbol{B} \in \mathbb{R}^{M \times M}$, where $M \ll N$. Accordingly, our goal is to compute the unknown solution $\boldsymbol{u} \in \mathbb{R}^M$ of the small linear system

$$\boldsymbol{B}\,\boldsymbol{u} \;=\; \boldsymbol{d} \tag{2.16}$$

with known right hand side $\boldsymbol{d} \in \mathbb{R}^M$. A well-suited method for such problems is the Cholesky decomposition (see e.g. [GV96]). Here, the goal is to decompose the system matrix $\boldsymbol{B}$ such that

$$\boldsymbol{B} \;=\; \boldsymbol{C}\boldsymbol{C}^\mathsf{T}, \tag{2.17}$$

where $\boldsymbol{C}$ is a lower triangular matrix with real and positive diagonal entries. More specifically, we compute the entries c_{ij} of $\boldsymbol{C}$ by means of the entries b_{ij} of $\boldsymbol{B}$ as follows

(see e.g. [PTVF07]):

$$c_{ij} \;=\; \begin{cases} \sqrt{b_{ii} - \sum_{q=1}^{i-1} c_{iq}^2} & \text{if } i = j\,, \\ \frac{1}{c_{jj}} \left(b_{ij} - \sum_{q=1}^{j-1} c_{iq} c_{jq} \right) & \text{if } i > j\,, \\ 0 & \text{else.} \end{cases} \tag{2.18}$$

With the help of this unique decomposition $\boldsymbol{B} = \boldsymbol{C}\boldsymbol{C}^{\mathsf{T}}$ the solution of the system $\boldsymbol{B}\boldsymbol{u} = \boldsymbol{d}$ is straightforward: First, we replace $\boldsymbol{C}^{\mathsf{T}}\boldsymbol{u}$ by the vector $\boldsymbol{v}$ and solve

$$\boldsymbol{C}\boldsymbol{v} \;=\; \boldsymbol{d} \tag{2.19}$$

with forward substitution. Next, backward substitution allows to solve

$$\boldsymbol{C}^{\mathsf{T}}\boldsymbol{u} \;=\; \boldsymbol{v} \tag{2.20}$$

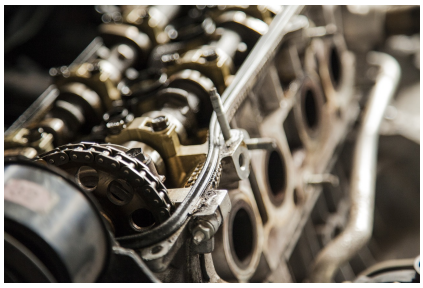
which finally yields $\boldsymbol{u}$. Note that such forward and backward substitutions can be realised efficiently.

Chapter 3

Fast Semi-Iterative Schemes

"FSI engines achieve higher performance and better dynamics than conventional engines, with better efficiency."

AUDI AG



pixabay.com

Contents

3.1 Introduction **21**
3.2 Related Work **23**
3.3 How to Benefit from Box Filtering **24**
3.4 FSI Schemes for Diffusion Evolutions **27**
3.5 FSI Schemes for Elliptic Problems **28**
3.6 FSI Schemes for Constrained Optimisation **29**
3.7 Experiments **31**
3.8 Limitations and Discussion **35**
3.9 Summary and Conclusions **36**
3.A Mathematical Analysis **38**

Main parts of this chapter base on our work published in [HOW+16].

3.1 Introduction

In this chapter we propose efficient numerical solvers for three problem classes in image processing and computer vision: *(i)* diffusion evolutions, *(ii)* variational models leading to elliptic *partial differential equations* (PDEs), and *(iii)* constrained convex optimisation problems.

Diffusion processes have applications e.g. as linear or nonlinear scale-spaces [Iij62, PM90, Wei98]. In the space-discrete case, they lead to dynamical systems defined by

$$\partial_t \boldsymbol{u} \;=\; \boldsymbol{A}(\boldsymbol{u})\,\boldsymbol{u}\,, \tag{3.1}$$

where the vector $\boldsymbol{u} \in \mathbb{R}^N$ contains the grey values in N different pixel locations, $t \in (0, \infty)$ denotes the diffusion time, and $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ is a symmetric negative semi-definite matrix that may depend in a nonlinear way on the evolving image $\boldsymbol{u}$. This abstract model holds in any dimension and includes isotropic as well as anisotropic diffusion models with differential operators of second or higher order.

Elliptic problems frequently arise as steady states of diffusion evolutions or as Euler-Lagrange equations of variational models [BPT88]. Space-discrete formulations lead to systems of equations in the form of

$$\boldsymbol{B}(\boldsymbol{u})\,\boldsymbol{u} \;=\; \boldsymbol{d}(\boldsymbol{u})\,, \tag{3.2}$$

where $\boldsymbol{B} \in \mathbb{R}^{N \times N}$ is symmetric positive definite, and $\boldsymbol{d} \in \mathbb{R}^N$ is the known right hand side. In case of nonlinear evolutions or nonquadratic variational models, the system matrix $\boldsymbol{B}$ and the vector $\boldsymbol{d}$ may depend on the evolving image $\boldsymbol{u}$.

Constrained convex optimisation problems appear e.g. in dual formulations of certain nonsmooth minimisation tasks such as *total variation* (TV) regularisation [ROF92, AV94]. A general framework can be cast as

$$\text{minimise } F(\boldsymbol{u}) \qquad \text{s.t. } \boldsymbol{u} \in \mathcal{C}\,, \tag{3.3}$$

where $F\colon \mathbb{R}^N \to \mathbb{R}$ is a smooth convex function, and $\mathcal{C}$ denotes a convex set that models the constraint. Such constrained optimisation methods are flexible modelling tools that have a broad range of applications.

For all three problem classes there exist basic iterative schemes, namely *(i)* explicit (Euler forward) finite difference schemes, *(ii)* Richardson iterations, and *(iii)* projected gradient descent methods. These schemes are easy to implement and well-suited for parallel architectures such as *graphics processor units* (GPUs). Unfortunately, severe restrictions of the time step sizes or the relaxation parameters render such algorithms rather inefficient. Hence, it would be highly desirable to find acceleration strategies that improve the efficiency of those basic schemes while preserving their advantages.

Main Contributions. We propose an acceleration strategy that consists of a semi-iterative approach in the sense of Varga [Var57]. It computes the new iterate $\boldsymbol{u}^{k+1}$ by applying the basic iterative scheme to $\boldsymbol{u}^k$ and extrapolating the result by means of $\boldsymbol{u}^{k-1}$. Here, the extrapolation step is responsible for a substantial acceleration. We call our techniques *Fast Semi-Iterative* (FSI) schemes. In contrast to classical semi-iterative approaches from the numerical literature, we obtain different extrapolation parameters that can be derived in an intuitive way from box filter recursions.

A first appearance of this recursion relation can be found in the Ph.D. thesis of Grewenig [Gre13], where box filters are shown to give a good compromise between efficiency and numerical stability (cf. also [WGSB16]). Additionally, we present several extensions of our techniques such as an adaptation to constrained optimisation and to strongly convex problems. On top of that, we uncover interesting connections of our FSI schemes to well-performing iterative procedures such as *Fast Explicit Diffusion* (FED) [WGSB16] or Polyak's heavy ball method [Pol64].

Chapter Outline. After a discussion of related work in Section 3.2, we review 1D linear diffusion and expose its relation to box filtering in Section 3.3. Subsequently, we transfer this concept to iterative schemes and present our novel FSI techniques for diffusion evolutions (Section 3.4), elliptic problems (Section 3.5), and constrained convex optimisation (Section 3.6). Our experiments in Section 3.7 illustrate the benefits of our algorithms. We discuss possible limitations in Section 3.8, and conclude this chapter with a summary and outlook in Section 3.9. Additionally, we present details of our mathematical analysis in the Appendix 3.A of this chapter.

3.2 Related Work

Our schemes are closely related to nonstationary iterative schemes, where the algorithmic parameters vary from iteration to iteration. In this context, already in 1911, Richardson discussed possible benefits of varying relaxation parameters in his iterative scheme [Ric11]. Later, based on Chebyshev polynomials of the first kind, cyclic parameter choices were proposed that allow substantial speed-ups; see e.g. [Lan52, Sho53, You54]. Inherently, the Richardson method in [Ric11] is closely related to gradient descent schemes, and thus to the solution of parabolic PDEs. In this context, similar ideas have been proposed by Yaun'Chzhao-Din [Yua58] and Saul'yev [Sau64]. They are known under the name super time stepping [GS78, AAG96]. Recently, motivated by box filter factorisations, Weickert et al. [WGSB16] proposed cyclically varying parameters that substantially improve the damping properties of the resulting schemes. Additionally, the authors introduced a Jacobi-like scheme for elliptic problems. Setzer et al. [SSM13] built on the work of [WGSB16] and provided an extension to projection methods with application to nonsmooth optimisation. Furthermore, we can relate such cyclic Richardson approaches to so-called semi-iterative procedures that rely on Chebyshev recursion formulas [Var57, GV61]; see e.g. [You89] for a historical overview. In this regard, we additionally discuss relations to closely related *Runge-Kutta-Chebyshev* methods for parabolic problems [vS80, VHS90]. For the interested reader, we refer to [vdH96] for a survey of Runge-Kutta methods for partial differential equations. On top of that, we are able to derive a connection of our semi-iterative schemes to Polyak's heavy ball method [Pol64], where Ochs et al. [OBP15] recently proposed an extension that includes proximal mappings. Similarly, our technique relates to the so-called momentum method that is frequently applied in machine learning approaches; see e.g. [RHW86, Qia99, SMDH13]. The term *momentum* stems

from an analogy to physics, in particular to the movement of a Newtonian particle through a viscous medium in a conservative force field; see e.g. [Qia99].

With the knowledge gained in the upcoming sections, we will be able to discuss the connections of our FSI schemes to the mentioned algorithms in more detail.

3.3 How to Benefit from Box Filtering

3.3.1 Explicit Scheme for 1D Linear Diffusion

As starting point of our work we consider linear diffusion of a 1D signal $u(x,t)$:

$$\partial_t u \;=\; \partial_{xx} u\,. \tag{3.4}$$

With grid size h and time step size τ, an explicit scheme for (3.4) is given by

$$u_i^{k+1} \;=\; (I + \tau\, L)\, u_i^k\,, \tag{3.5}$$

where

$$L \;\coloneqq\; \frac{1}{h^2} \cdot \boxed{1}\boxed{-2}\boxed{1} \tag{3.6}$$

denotes the standard discretisation of the 1D Laplacian in stencil notation,

$$I \;\coloneqq\; \boxed{1} \tag{3.7}$$

the identity operator, and u_i^k approximates u in pixel i at time level k. For stability reasons, all stencil weights should be nonnegative. This implies that the time step size must satisfy $\tau \leq h^2/2$. Obviously, this restriction makes such an explicit scheme inefficient: With n explicit diffusion steps, we can only reach a stopping time of $\mathcal{O}(n)$.

3.3.2 Box Filtering via Iterative Explicit Diffusion

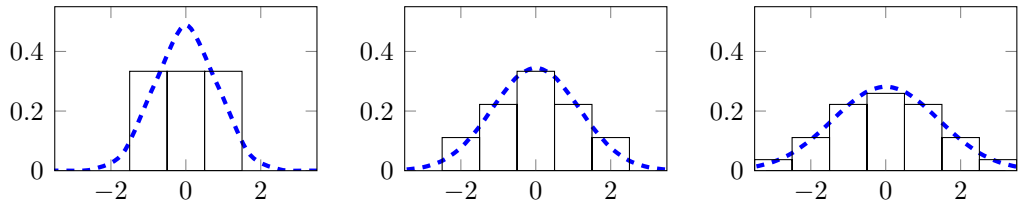
In order to apply explicit schemes more efficiently, let us make a didactic excursion to box filters. A box filter B_{2n+1} of length $(2n+1)h$ is defined by

$$B_{2n+1} \;\coloneqq\; \frac{1}{2n+1} \cdot \underbrace{\boxed{1}\boxed{\cdots}\boxed{1}}_{(2n+1)h} \tag{3.8}$$

such that

$$[B_{2n+1}\, u]_i \;=\; \frac{1}{2n+1} \sum_{j=-n}^{n} u_{i+j}\,. \tag{3.9}$$

It is well-known [Hel77] that linear diffusion with stopping time T is equivalent to a convolution with a Gaussian of variance $\sigma^2 = 2T$. Moreover, the central limit theorem



(a) $m = 1$ (b) $m = 2$ (c) $m = 3$

Figure 3.1: Iterated box filtering. An m-fold iteration of a box filter B_{2n+1} with variance σ_n^2 (here: $n = 1$, black) approximates Gaussian convolution with variance $m\,\sigma_n^2$ (blue).

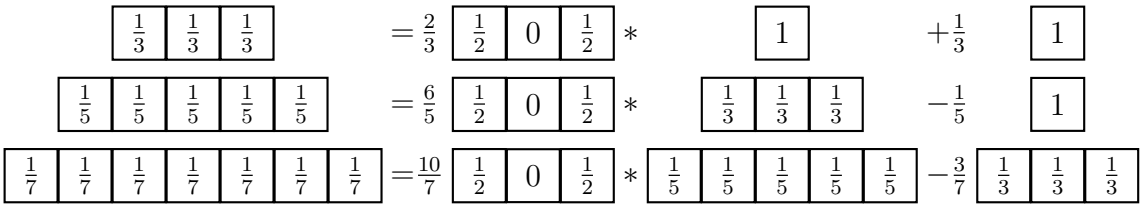


Figure 3.2: Illustration of the connection between box filtering and an explicit scheme for 1D homogeneous diffusion ($n = 3$). *From top to bottom*: $k = 0, 1, 2$. This illustration is inspired by Prof. Joachim Weickert.

tells us that iterated box filtering approximates Gaussian convolution. Indeed, an m-fold iteration of a box filter B_{2n+1} with variance σ_n^2 approximates a Gaussian with variance $m\,\sigma_n^2$; cf. Figure 3.1. The variance σ_n^2 of B_{2n+1} is given by (cf. also [Wel86])

$$\sigma_n^2 \;=\; \frac{1}{2n+1}\sum_{j=-n}^{n}(jh-0)^2 \;=\; \frac{2h^2}{2n+1}\,\frac{n(n+1)(2n+1)}{6} \;=\; \frac{n(n+1)}{3}h^2\,. \qquad (3.10)$$

This implies that a single application of a box filter B_{2n+1} approximates linear diffusion with stopping time $T_{B_{2n+1}} = \sigma_n^2/2 = \frac{n(n+1)}{3}\frac{h^2}{2}$. Note that this stopping time is $\mathcal{O}(n^2)$. Hence, if we were able to implement B_{2n+1} by means of n explicit linear diffusion steps, we could accelerate the explicit scheme from $\mathcal{O}(n)$ to $\mathcal{O}(n^2)$. To this end, let us consider the following theorem (cf. [Gre13, Theorem 4.1]):

Theorem 1 (Connection of Box Filters and Explicit Diffusion)**.** *A box filter B_{2n+1} of length $(2n+1)h$ can be constructed iteratively by n explicit linear diffusion steps:*

$$B_{2k+3} \;=\; \alpha_k \cdot (I + \tau L)\, B_{2k+1} \;+\; (1-\alpha_k)\cdot B_{2k-1} \qquad (k = 0, \ldots, n-1)$$

with $\tau := h^2/2$, $\alpha_k := (4k+2)/(2k+3)$, and $B_{-1} := I$.

We illustrate this connection for $n = 3$ in Figure 3.2. Note that for $k = 0$ we have $B_3 = I + \frac{h^2}{3}L$, which is a single diffusion step with time step size $\frac{2}{3}\tau$.

Proof. We proof Theorem 1 by induction. Let us fist consider $k=0$ and $k=1$:

$$[B_3\,u]_i \;=\; \frac{2}{3}(I+\tau L)\,u_i + \frac{1}{3}u_i \;= \frac{1}{3}\,(u_{i-1}+u_{i+1}) + \frac{1}{3}u_i \;=\; \frac{1}{3}\sum_{j=-1}^{1} u_{i+j}\,,$$

$$[B_5\,u]_i \;=\; \frac{6}{5}(I+\tau L)\,[B_3\,u]_i - \frac{1}{5}u_i \;=\; \frac{1}{5}\Big(\sum_{j=-2}^{0}u_{i+j} + \sum_{j=0}^{2}u_{i+j} - u_i\Big) \;=\; \frac{1}{5}\sum_{j=-2}^{2}u_{i+j}\,.$$

Then, the inductive step for $k \geq 2$ yields

$$\begin{aligned}[B_{2k+3}\,u]_i \;&=\; \frac{4k+2}{2k+3}\cdot(I+\tau L)\,[B_{2k+1}\,u]_i \;-\; \frac{2k-1}{2k+3}\cdot[B_{2k-1}\,u]_i \\ &=\; \frac{1}{2k+3}\Big(\sum_{j=-k-1}^{k-1}u_{i+j} + \sum_{j=-k+1}^{k+1}u_{i+j} - \sum_{j=-k+1}^{k-1}u_{i+j}\Big) \;=\; \frac{1}{2k+3}\sum_{j=-(k+1)}^{k+1}u_{i+j}\,.\end{aligned}$$

□

3.3.3 Accelerating the Explicit Scheme for 1D Linear Diffusion

To apply Theorem 1 for accelerating the explicit diffusion scheme (3.5), let us first rewrite it in matrix-vector notation:

$$\boldsymbol{u}^{k+1} \;=\; (\boldsymbol{I}+\tau\,\boldsymbol{L})\,\boldsymbol{u}^k\,, \tag{3.11}$$

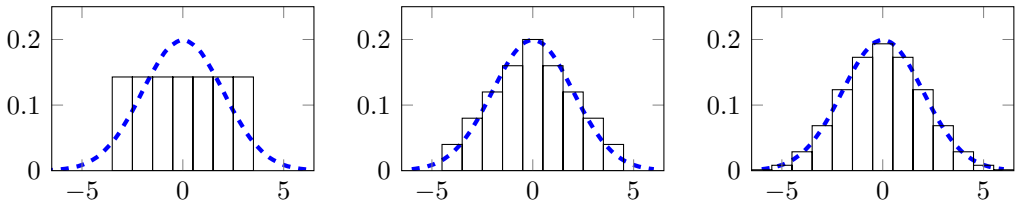
where the vector $\boldsymbol{u}\in\mathbb{R}^N$ contains the discrete entries of u, $\boldsymbol{I}\in\mathbb{R}^{N\times N}$ is the identity matrix, and the symmetric negative semi-definite matrix $\boldsymbol{L}\in\mathbb{R}^{N\times N}$ implements the Laplacian. The box filter relation in Theorem 1 suggests the following scheme to accelerate the explicit diffusion scheme (3.11) such that $\boldsymbol{u}^k$ corresponds to an application of a box filter B_{2k+1}:

$$\boldsymbol{u}^{k+1} \;=\; \alpha_k\cdot(\boldsymbol{I}+\tau\,\boldsymbol{L})\,\boldsymbol{u}^k \;+\; (1-\alpha_k)\cdot\boldsymbol{u}^{k-1} \tag{3.12}$$

with $\tau = h^2/2$, $\alpha_k = (4k+2)/(2k+3)$, and $\boldsymbol{u}^{-1} \coloneqq \boldsymbol{u}^0$. As we have seen, n iterations of this scheme implement a box filter B_{2n+1} of length $(2n+1)h$. However, a single box filter might be a poor approximation for the actual linear diffusion process that is equivalent to Gaussian convolution. To improve the approximation quality, we should iterate the box filter; cf. Figure 3.3. Hence, we propose a *cyclic* application of (3.12), where the m-th cycle with cycle length n is given by

$$\boxed{\begin{array}{c}\boldsymbol{u}^{m,k+1} \;=\; \alpha_k\cdot\big(\boldsymbol{I}+\tau\,\boldsymbol{L}\big)\,\boldsymbol{u}^{m,k} \;+\; (1-\alpha_k)\cdot\boldsymbol{u}^{m,k-1} \\ \text{with}\;\; \boldsymbol{u}^{m,-1} \coloneqq \boldsymbol{u}^{m,0}\;\; \text{and}\;\; \alpha_k = (4k+2)/(2k+3)\;\; \text{for}\;\; k=0,\dots,n-1.\end{array}} \tag{3.13}$$

For the next cycle, we set $\boldsymbol{u}^{m+1,0} \coloneqq \boldsymbol{u}^{m,n}$. Here, the number of cycles is responsible for the accuracy, while the cycle length n accounts for the $\mathcal{O}(n^2)$ efficiency; cf. also Figure 3.3.



(a) $m = 1$ and $n = 3$ (b) $m = 2$ and $n = 2$ (c) $m = 6$ and $n = 1$

Figure 3.3: Different approximations of Gaussian convolution with variance $\sigma^2 = 4$ (blue) by means of m-fold applications of a box filter B_{2n+1} with length $2n + 1$ (black). We see that the approximation quality improves with the number of box filter iterations. Moreover, this figure illustrates the tradeoff between accuracy and efficiency: While the approximation in (a) requires only three solver iterations, the more accurate approximation in (c) requires six iterations. A good tradeoff in this example is given by (b) with four solver iterations.

3.4 FSI Schemes for Diffusion Evolutions

Our discussion so far was for didactic reasons only, since linear diffusion can be implemented directly as an efficient box filter without explicit iterations. However, it suggests how we could generalise these ideas to arbitrary isotropic or anisotropic nonlinear diffusion processes (3.1) that have explicit schemes of type

$$\boldsymbol{u}^{k+1} \;=\; \big(\boldsymbol{I} + \tau\, \boldsymbol{A}(\boldsymbol{u}^k)\big)\, \boldsymbol{u}^k\,, \tag{3.14}$$

where $\boldsymbol{A}$ is symmetric and negative semi-definite. Such schemes have the time step size restriction $0 \leq \tau < 2/\varrho(\boldsymbol{A}(\boldsymbol{u}^k))$, where ϱ denotes the spectral radius. Obviously, there is a strong similarity of (3.14) to the linear diffusion scheme in (3.11). Hence, it appears to be natural to formulate, in analogy to (3.13), the following *Fast Semi-Iterative* (FSI) scheme with a symmetric and negative semi-definite $\boldsymbol{A}$:

$$\boxed{\begin{aligned} \boldsymbol{u}^{m,k+1} \;&=\; \alpha_k \cdot \big(\boldsymbol{I} + \tau\, \boldsymbol{A}(\boldsymbol{u}^{m,k})\big)\, \boldsymbol{u}^{m,k} \;+\; (1 - \alpha_k) \cdot \boldsymbol{u}^{m,k-1} \\ \text{with}\;\; \boldsymbol{u}^{m,-1} &:= \boldsymbol{u}^{m,0} \;\text{ and }\; \alpha_k = (4k+2)/(2k+3) \;\text{ for }\; k = 0, \dots, n-1. \end{aligned}} \tag{3.15}$$

This scheme describes the m-th cycle with length n of our FSI algorithm. We repeat it several times to reach a specific stopping time.

Connection to Fast Explicit Diffusion (FED)

As explained in Section 3.3, we build on the connection between box filtering and explicit schemes. This was inspired by the cyclic FED approach of Weickert et al. [WGSB16] which demonstrates the benefits of a reliance on box filters. More specifically, Weickert et al. [WGSB16] exploit a factorisation of a box filter into several explicit diffusion steps to construct highly efficient algorithms. In case of linear problems, one can even show that FED and FSI schemes yield identical results after each cycle; cf. Section 3.A.2. However, in case of nonlinear problems, the discussed

semi-iterative structure of our FSI schemes seems to be beneficial since it shows a higher robustness to nonlinear updates within one cycle. We demonstrate this by means of experiments in Section 3.7. Moreover, FED schemes are highly sensitive to numerical inaccuracies. Due to this, FED approaches require sophisticated rearrangements of possibly unstable time steps to avoid the explosion of rounding errors; see e.g. [GS78, CR96, WGSB16] and references therein. We eliminate this drawback by our FSI technique and its semi-iterative structure.

Connection to Runge-Kutta-Chebyshev Method

Additionally, we can relate our FSI schemes to another class of numerical methods for the solution of parabolic PDEs, so-called *Runge-Kutta-Chebyshev* methods; see e.g. [vS80, VHS90, vdH96] and references therein. Here, the algorithmic parameters of the multi-step Runge-Kutta scheme are determined by means of Chebyshev polynomials of the first kind. In comparison to that, we are able to show a connection of our schemes to Chebyshev polynomials of the second kind; cf. Section 3.A.1. Interestingly, this underlying difference seems to be the main reason for better damping properties: While for instance the method of van der Houwen and Sommeijer [vS80] requires an additional damping parameter, our FSI schemes can be used without any additional damping; cf. also [Gre13].

3.5 FSI Schemes for Elliptic Problems

So far, we have considered diffusion-like processes that correspond to parabolic PDEs. Next, we explain how to transfer this concept to discretised elliptic PDEs, or more generally to the solution of equation systems in the form of (3.2). In the linear case, formulating (3.2) as $\boldsymbol{u} = \boldsymbol{u} - \omega(\boldsymbol{B}\boldsymbol{u} - \boldsymbol{d})$ gives rise to the Richardson scheme [Ric11]

$$\boldsymbol{u}^{k+1} \;=\; (\boldsymbol{I} - \omega\,\boldsymbol{B})\,\boldsymbol{u}^k + \omega\,\boldsymbol{d} \tag{3.16}$$

with $0 \leq \omega < 2/\varrho(\boldsymbol{B})$. Considering the error vector $\boldsymbol{e}^k \coloneqq \boldsymbol{u}^k - \boldsymbol{u}^*$ between the current estimate $\boldsymbol{u}^k$ and the unknown exact solution $\boldsymbol{u}^*$ yields

$$\boldsymbol{e}^{k+1} \;=\; \boldsymbol{u}^{k+1} - \boldsymbol{u}^* \;=\; \boldsymbol{u}^k - \boldsymbol{u}^* - \omega(\boldsymbol{B}\,\boldsymbol{u}^k - \boldsymbol{d} - \boldsymbol{B}\,\boldsymbol{u}^* + \boldsymbol{d}) \;=\; (\boldsymbol{I} - \omega\,\boldsymbol{B})\,\boldsymbol{e}^k\,, \tag{3.17}$$

where we used that $\boldsymbol{B}\boldsymbol{u}^* = \boldsymbol{d}$. We observe a strong similarity of (3.17) to the explicit diffusion scheme in (3.11). Hence, we propose the following FSI scheme for elliptic problems:

$$\boxed{\begin{aligned} \boldsymbol{u}^{m,k+1} \;&=\; \alpha_k \cdot \big((\boldsymbol{I} - \omega\,\boldsymbol{B})\,\boldsymbol{u}^{m,k} + \omega\,\boldsymbol{d}\big) \;+\; (1 - \alpha_k) \cdot \boldsymbol{u}^{m,k-1} \\ \text{with}\;\; \boldsymbol{u}^{m,-1} &\coloneqq \boldsymbol{u}^{m,0} \;\text{ and }\; \alpha_k = (4k+2)/(2k+3) \;\text{ for }\; k = 0, \ldots, n-1. \end{aligned}} \tag{3.18}$$

To extend (3.18) to nonlinear systems of equations in the form of (3.2), we replace $\boldsymbol{B}$ and $\boldsymbol{d}$ by their nonlinear counterparts $\boldsymbol{B}(\boldsymbol{u}^{m,k})$ and $\boldsymbol{d}(\boldsymbol{u}^{m,k})$. Moreover, we choose

$0 \leq \omega < 2/\mathscr{L}$, where $\mathscr{L} > 0$ denotes the Lipschitz constant of the smooth strictly maximal monotone map $\boldsymbol{G}(\boldsymbol{u}) \coloneqq \boldsymbol{B}(\boldsymbol{u})\,\boldsymbol{u} - \boldsymbol{d}(\boldsymbol{u})$ with

$$(\boldsymbol{u}-\boldsymbol{v})^{\mathsf{T}}\,(\boldsymbol{G}(\boldsymbol{u})-\boldsymbol{G}(\boldsymbol{v})) \;>\; 0 \qquad \forall \boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^N \quad \text{with} \quad \boldsymbol{u}\neq\boldsymbol{v} \tag{3.19}$$

and

$$|\boldsymbol{G}(\boldsymbol{u})-\boldsymbol{G}(\boldsymbol{v})| \;\leq\; \mathscr{L}\cdot|\boldsymbol{u}-\boldsymbol{v}| \qquad \forall \boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^N\,. \tag{3.20}$$

Preconditioning

The discussed FSI method yields fast convergence for problems where the coefficients of the equation system have a similar value of magnitude. However, in case of strongly differing coefficients a preconditioning or, in other words, a different splitting of the system matrix $\boldsymbol{B}$, turns out to be highly beneficial. As an example, we consider the Jacobi overrelaxation splitting $\boldsymbol{B} = \frac{1}{\omega}\boldsymbol{P} + (\boldsymbol{B} - \frac{1}{\omega}\boldsymbol{P})$, where $\boldsymbol{P}$ denotes a positive definite diagonal matrix. This leads to

$$\boldsymbol{u}^{m,k+1} \;=\; \alpha_k\cdot\left((\boldsymbol{I}-\omega\,\boldsymbol{P}^{-1}\boldsymbol{B})\,\boldsymbol{u}^{m,k}+\omega\,\boldsymbol{P}^{-1}\boldsymbol{d}\right) \;+\; (1-\alpha_k)\cdot\boldsymbol{u}^{m,k-1}\,. \tag{3.21}$$

Assuming a symmetric positive definite matrix $\boldsymbol{B}$, we require a suitable ω such that the eigenvalues of $\boldsymbol{I}-\omega\,\boldsymbol{P}^{-1}\boldsymbol{B}$ lie in $(-1,1]$.

Connection to Fast Jacobi

Similar to the connection to FED in the parabolic case, we can show a relation of our FSI scheme to the recent Fast Jacobi solver of Weickert et al. [WGSB16] for elliptic problems; cf. Section 3.A.2. As discussed before, we eliminate the drawback of rearranging the relaxation parameters and we can observe a larger robustness to intermediate nonlinear updates in our experiments.

3.6 FSI Schemes for Constrained Optimisation

Often, the elliptic problem from the previous section can be interpreted as the minimality condition of a suitable optimisation problem. In fact, the gradient descent scheme to compute a minimiser of the optimisation problem (3.3) without side-constraints is given by

$$\boldsymbol{u}^{k+1} \;=\; \boldsymbol{u}^{k} - \omega\,\boldsymbol{\nabla}F(\boldsymbol{u}^{k}) \tag{3.22}$$

with $0 \leq \omega < 2/\mathscr{L}$, where $\mathscr{L} > 0$ is the Lipschitz constant of $\boldsymbol{\nabla}F$. Again, the structural similarity to (3.11) suggests the following FSI iteration:

$$\boldsymbol{u}^{m,k+1} \;=\; \alpha_k\cdot\left(\boldsymbol{u}^{m,k}-\omega\,\boldsymbol{\nabla}F(\boldsymbol{u}^{m,k})\right) \;+\; (1-\alpha_k)\cdot\boldsymbol{u}^{m,k-1}\,. \tag{3.23}$$

Adaptation to Constrained Problems

Several optimisation problems constrain the solution $\boldsymbol{u}$ to some closed convex set $\mathcal{C}$. To apply our FSI scheme also to such constrained optimisation problems, we adapt it in the following way:

$$\boxed{\begin{aligned} \boldsymbol{u}^{m,k+1} \;&=\; \mathcal{P}_{\mathcal{C}}\Big(\alpha_k \cdot \big(\boldsymbol{u}^{m,k} - \omega\, \boldsymbol{\nabla} F(\boldsymbol{u}^{m,k})\big) + (1-\alpha_k)\cdot \boldsymbol{u}^{m,k-1}\Big) \\ \text{with } \boldsymbol{u}^{m,-1} &:= \boldsymbol{u}^{m,0} \text{ and } \alpha_k = (4k+2)/(2k+3) \text{ for } k = 0,\dots,n-1. \end{aligned}} \tag{3.24}$$

Here, $\mathcal{P}_{\mathcal{C}}$ denotes the orthogonal projection operator that projects onto the closed convex set $\mathcal{C}$; cf. also Section 2.4.

Adaptation to Strongly Convex Problems

So far, we have considered the case where the Lipschitz constant $\mathscr{L}$ is assumed to be known. However, strongly convex problems additionally provide information about the strong convexity parameter ℓ. To make use of this additional knowledge, we propose the following recursive parameter choice to accelerate our iterative scheme:

$$\alpha_k \;=\; \frac{1}{1 - \frac{\alpha_{k-1}}{4}\cdot\left(\frac{\mathscr{L}-\ell}{\mathscr{L}+\ell}\right)^2} \qquad (k = 1,\dots,n-1)\,, \tag{3.25}$$

where $\alpha_0 = 2(\mathscr{L}+\ell)/(3\mathscr{L}+\ell)$ and $\omega = 2/(\mathscr{L}+\ell)$. Since our derivation of the parameter choice in (3.25) is quite technical, we skip it at the current point and refer the interested reader to Section 3.A.3. Note that for $\ell = 0$ these parameters come down to (3.24).

We present a full description of our FSI method in Algorithm 3.1. Also in the elliptic case (Section 3.5) we can apply such a parameter choice. In this case, ℓ and $\mathscr{L}$ are related to the smallest and largest eigenvalues of the positive definite system matrix $\boldsymbol{B}$.

Connection to Heavy Ball Method

Let us now derive a close relation of the proposed FSI scheme to Polyak's *heavy ball* (HB) method [Pol64] where one iteration is given by

$$\boldsymbol{u}^{k+1} \;=\; \boldsymbol{u}^k \;-\; \alpha_{\mathrm{HB}}\cdot\boldsymbol{\nabla} F(\boldsymbol{u}^k) \;+\; \beta_{\mathrm{HB}}\cdot(\boldsymbol{u}^k - \boldsymbol{u}^{k-1})\,. \tag{3.26}$$

The first part of this iterative scheme can be seen as a gradient descent step, while the second part represents an inertial term. It relates the current iterate $\boldsymbol{u}^k$ to the old time step $\boldsymbol{u}^{k-1}$. This allows significant speed-ups. A similar concept is known under the name *momentum* which has shown its usefulness in many machine learning approaches; see e.g. [RHW86, Qia99, SMDH13].

Interestingly, we can connect (3.26) to our FSI approach by applying *cyclically varying* parameters

$$\alpha_{\mathrm{HB}k} \;=\; \omega\alpha_k \qquad \text{and} \qquad \beta_{\mathrm{HB}k} \;=\; \alpha_k - 1 \tag{3.27}$$

Algorithm 3.1: FSI scheme with projection operator $\mathcal{P}_{\mathcal{C}}$ and $0 \leq \ell < \mathscr{L}$.

input : $\boldsymbol{\nabla} F$, $\mathcal{P}_{\mathcal{C}}$, $\mathscr{L}$, ℓ, n, $\boldsymbol{u}^0$

$\omega = 2/(\mathscr{L} + \ell)$

$\alpha_0 = 2(\mathscr{L} + \ell)/(3\mathscr{L} + \ell)$

for $k = 1, \ldots, n-1$ **do**

  $\alpha_k = 1 \Big/ \Big(1 - \frac{\alpha_{k-1}}{4} \cdot \big(\frac{\mathscr{L}-\ell}{\mathscr{L}+\ell}\big)^2\Big)$

$\boldsymbol{u}^{0,-1} = \boldsymbol{u}^{0,0} = \boldsymbol{u}^0$

for $m = 0, 1, \ldots$ **do**

  for $k = 0, 1, \ldots, n-1$ **do**

    $\boldsymbol{u}^{m,k+1} = \mathcal{P}_{\mathcal{C}}\Big(\alpha_k \left(\boldsymbol{u}^{m,k} - \omega\, \boldsymbol{\nabla} F(\boldsymbol{u}^{m,k})\right) + (1 - \alpha_k)\, \boldsymbol{u}^{m,k-1}\Big)$

  $\boldsymbol{u}^{m+1,-1} = \boldsymbol{u}^{m+1,0} = \boldsymbol{u}^{m,n}$

with α_k given by (3.23) and combined with a restart $\boldsymbol{u}^{m,-1} := \boldsymbol{u}^{m,0}$ after each cycle. We illustrate benefits of our approach by means of experiments in the upcoming section.

3.7 Experiments

Our experimental evaluation consists of three main parts. In Section 3.7.1 and 3.7.2, we illustrate the performance of our FSI schemes on image processing applications. Additionally, we compare our methods to related iterative schemes by means of Nesterov's worst case problems in Section 3.7.3. As error measure, we apply the *mean squared error* (MSE) that is given by

$$\mathrm{MSE}(u, u_{\mathrm{ref}}) \;=\; \frac{1}{|\Omega|} \int_{\Omega} \big(u(\boldsymbol{x}) - u_{\mathrm{ref}}(\boldsymbol{x})\big)^2 \,\mathrm{d}\boldsymbol{x}\,, \tag{3.28}$$

where u denotes the computed and u_{ref} the reference solution. Furthermore, $\Omega \subset \mathbb{R}^2$ denotes the rectangular image domain, and $|\Omega|$ its cardinality.

3.7.1 Inpainting with Edge-Enhancing Anisotropic Diffusion

Our first experiment is concerned with image inpainting. Here, the task is to recover missing information by an interpolation of known data points. In the extreme case, where only very sparse data points are given (cf. Figure 3.4(a)), such an inpainting approach can be exploited for image compression; see e.g. [GWW+08]. In this regard, successful approaches rely on edge-enhancing anisotropic diffusion [Wei96], and consider the steady state solution of the following PDE for image inpainting:

$$\partial_t u \;=\; (1 - w(\boldsymbol{x})) \cdot \mathrm{div}\left(\boldsymbol{D}(\boldsymbol{\nabla} u_\sigma)\, \boldsymbol{\nabla} u\right) \;-\; w(\boldsymbol{x}) \cdot (u - f)\,, \tag{3.29}$$

where $w\colon \Omega \to \{0,1\}$ is a binary mask that indicates known data. Furthermore, $\boldsymbol{D}(\boldsymbol{\nabla} u_\sigma)$ is the so-called *diffusion tensor* that we determine by

$$\boldsymbol{D}(\boldsymbol{\nabla} u_\sigma) \;\coloneqq\; \mu_1 \cdot \boldsymbol{r}_1 \boldsymbol{r}_1^\mathsf{T} \;+\; \mu_2 \cdot \boldsymbol{r}_2 \boldsymbol{r}_2^\mathsf{T}\,, \tag{3.30}$$

where the first eigenvector $\boldsymbol{r}_1 \parallel \boldsymbol{\nabla} u_\sigma$ points across image edges, and the second one $\boldsymbol{r}_2 \perp \boldsymbol{\nabla} u_\sigma$ along them. Here, $u_\sigma \coloneqq G_\sigma * u$ denotes convolution of u with a Gaussian G_σ of standard deviation σ. Since we want to perform full diffusion along edges but reduced smoothing across them, we set $\mu_2 = 1$ and determine μ_1 by means of the Charbonnier diffusivity [CBAB94] with a contrast parameter $\lambda > 0$, i.e.

$$\mu_1 \;=\; \frac{1}{\sqrt{1 + |\boldsymbol{\nabla} u_\sigma|^2 / \lambda^2}}\,. \tag{3.31}$$

We discretise Equation (3.29) with the finite difference approach presented by Weickert et al. [WWW13], where we apply the parameters $\alpha_{\text{stencil}} = 0.4$ and $\gamma_{\text{stencil}} = 1.0$; cf. also Section 5.A.

Figure 3.4 depicts the sparse input data as well as our inpainting results. First, the provided speed-up by our FSI technique compared to the baseline *explicit diffusion* (ED) scheme is obvious. Moreover, we perform nonlinear updates after *each* iteration step. As discussed in [WGSB16], FED schemes are not robust to such intermediate updates within one cycle. Experimentally, our FSI schemes show this desired robustness. Hence, we offer more flexibility in this regard which further allows for a better performance.[1]

3.7.2 Total Variation Regularisation

In this experiment we deal with image regularisation, where the goal is to recover an undisturbed version of a noisy input image; cf. Figure 3.5. To this end, we consider the following energy functional [ROF92, AV94]:

$$E(u) \;=\; \frac{1}{2} \int_\Omega (u - f)^2 \,\mathrm{d}\boldsymbol{x} \;+\; \alpha \cdot R_{\mathrm{TV}}(u)\,, \tag{3.32}$$

where $f\colon \Omega \to \mathbb{R}$ represents a noisy input image, and the parameter $\alpha > 0$ steers the amount of smoothness. The regularisation term[2]

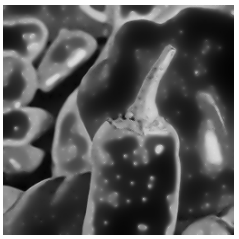
$$R_{\mathrm{TV}}(u) \;\coloneqq\; \sup_{\|\|\boldsymbol{p}\|\|_\infty \le 1} \int_\Omega u \operatorname{div} \boldsymbol{p} \,\mathrm{d}\boldsymbol{x} \tag{3.33}$$

[1] When optimising the actual algorithmic runtime, one should also take into account the time needed for a single iteration step and a nonlinear update. In this regard, there is always a tradeoff between few iterations with many nonlinear updates on the one hand, and more iterations with less iterations on the other hand.

[2] Note that the regularisation term in (3.33) is for differential functions u equivalent to $\int_\Omega |\boldsymbol{\nabla} u| \,\mathrm{d}\boldsymbol{x}$. This can be shown by means of the divergence (Gauss-Ostrogradsky) theorem.

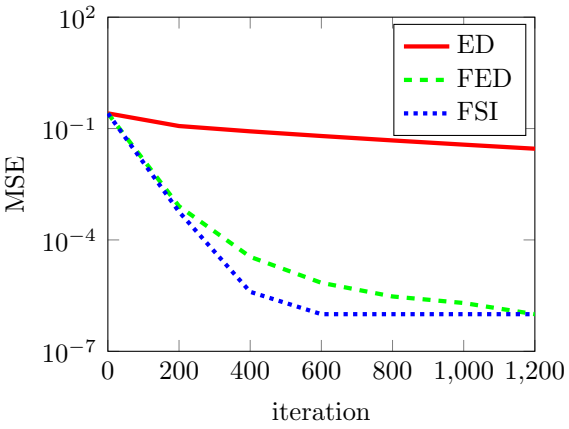
(a) input

(b) reference



(c) ED

(d) FSI



(e) mean squared error

Figure 3.4: Inpainting with edge-enhancing anisotropic diffusion, initialised with a black image. (a) Input data where only 5% of the pixels are known. (b) Converged inpainting result of a standard *explicit diffusion* (ED) scheme that serves as reference. (c) Result of ED after 600 iterations. (d) Result of FSI after three cycles with length $n = 200$. (e) *Mean squared error* (MSE) between current estimates and reference solution. The grey value range is $[0, 1]$.

penalises the *total variation* (TV) of u, where $\boldsymbol{p}\colon \Omega \to \mathbb{R}^2$ denotes a smooth vector field with compact support on Ω and

$$\|\|\boldsymbol{p}\|\|_\infty \coloneqq \sup_{\boldsymbol{x}\in\Omega} |\boldsymbol{p}(\boldsymbol{x})| \,. \tag{3.34}$$

The minimality condition of (3.32) w.r.t. u reads

$$u \;=\; f - \alpha \operatorname{div} \boldsymbol{p} \,. \tag{3.35}$$

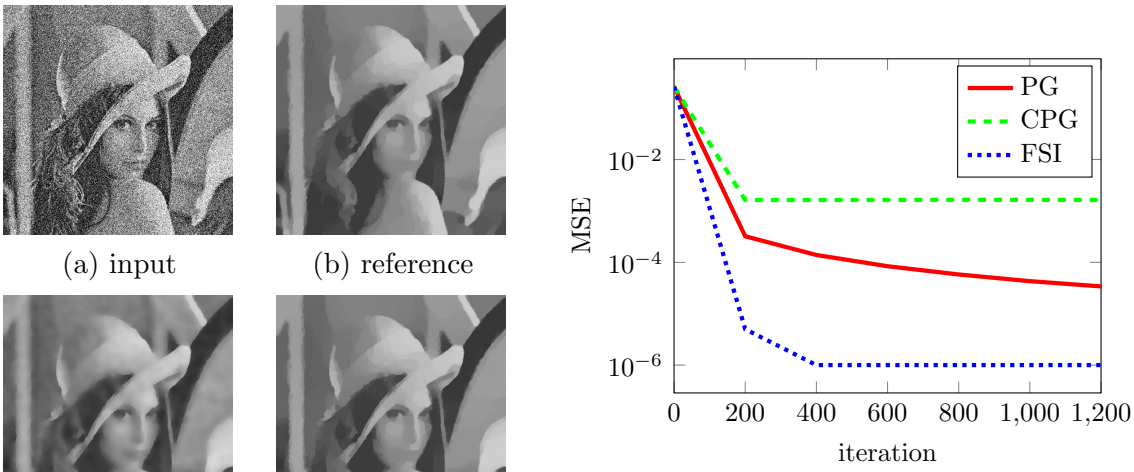
Next, we replace in (3.32) u by (3.35) which leads to the dual problem (cf. [Cha04])

$$\sup_{\|\|\boldsymbol{p}\|\|_\infty \le 1} \int_\Omega \left(\frac{1}{2} (\alpha \operatorname{div} \boldsymbol{p})^2 + \alpha f \operatorname{div} \boldsymbol{p} - (\alpha \operatorname{div} \boldsymbol{p})^2 \right) \mathrm{d}\boldsymbol{x} \tag{3.36}$$

$$= \sup_{\|\|\boldsymbol{p}\|\|_\infty \le 1} -\frac{1}{2} \int_\Omega \left((\alpha \operatorname{div} \boldsymbol{p})^2 - 2\alpha f \operatorname{div} \boldsymbol{p} \right) \mathrm{d}\boldsymbol{x} \,. \tag{3.37}$$

We discretise (3.37) with finite differences, and solve the resulting constrained optimisation problem for $\boldsymbol{p}$ with the proposed FSI scheme in Algorithm 3.1. Here, we account for the constraint $\|\|\boldsymbol{p}\|\|_\infty \le 1$ by pointwise projections onto the unit ball (cf. also Section 2.4):

$$\mathcal{P}_C\big(\boldsymbol{p}(\boldsymbol{x})\big) \;=\; \frac{\boldsymbol{p}(\boldsymbol{x})}{\max\{1, |\boldsymbol{p}(\boldsymbol{x})|\}} \,. \tag{3.38}$$



(a) input (b) reference

(c) PG (d) FSI

(e) mean squared error

Figure 3.5: *Total variation* (TV) regularisation. (a) Input with white Gaussian noise of standard deviation $\sigma = 0.25$. The grey value range is $[0, 1]$. (b) Converged reference result of standard *projected gradient* (PG) approach. (c) Result of PG after 200 iterations. (d) Result of FSI after one cycle with length $n = 200$. (e) *Mean squared error* (MSE) between current estimates and reference solution.

Finally, we determine the solution u by means of (3.35).

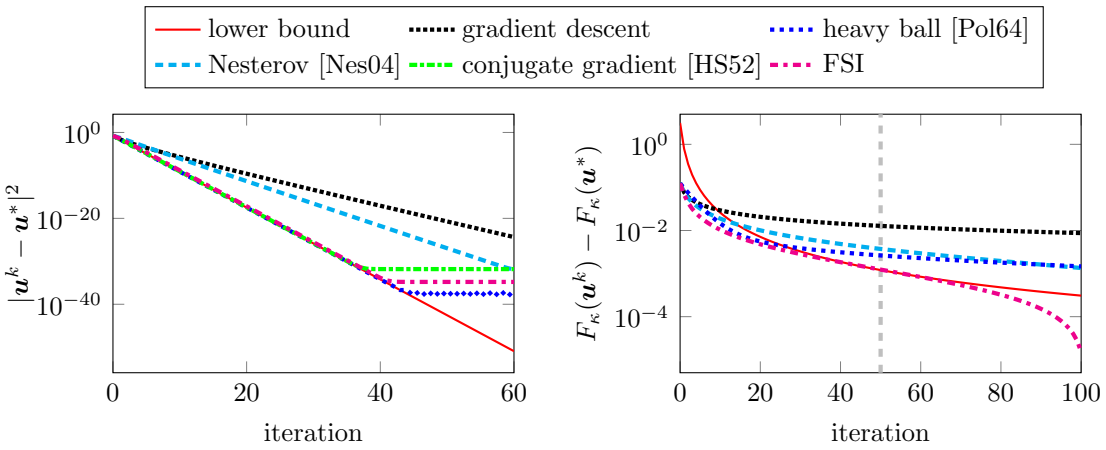
Figure 3.5 shows the noisy input image and our smoothed results. Also here, the provided acceleration of the baseline *projected gradient* (PG) scheme by our FSI approach is obvious. Moreover, the comparison to the cyclic FED-like projection method by Setzer et al. [SSM13] (CPG) demonstrates that our FSI scheme is more robust under the applied projections within each iteration, while the standard CPG approach without backtracking does not converge properly for such large cycle lengths.

3.7.3 Performance on Nesterov's Worst Case Problems

In our last experiment, we evaluate the performance of our FSI techniques w.r.t. related iterative schemes on Nesterov's worst case problems for smooth convex and strongly convex optimisation [Nes04]. These are quadratic minimisation problems which are difficult for any algorithm that can solve all instances of the respective class of problems. On the one hand, the *strongly convex* problem is given by minimising

$$F_\varkappa(\boldsymbol{u}) \;=\; \frac{\varkappa - 1}{8}\left(u_1^2 + \sum_{i=1}^{N}(u_i - u_{i+1})^2 - 2u_1\right) + \frac{1}{2}\,|\boldsymbol{u}|^2 \tag{3.39}$$

with some positive parameter $\varkappa$ and $\boldsymbol{u} \in \mathbb{R}^N$. In [Nes04, Theorem 2.1.13], Nesterov provides a lower bound for all iterations and predicts a linear convergence rate for



(a) strongly convex optimisation (b) convex optimisation

Figure 3.6: Numerical comparisons by means of worst case functions of Nesterov [Nes04]. (a) Strongly convex optimisation. (b) Convex optimisation. Here, the lower bound only holds for $k = 50$ (dashed grey line). In both cases, the problem dimension is $N = 10^5$ and $\boldsymbol{u}^*$ represents a known exact solution.

this problem class. On the other hand, minimising

$$F_\kappa(\boldsymbol{u}) \;=\; \frac{\mathscr{L}}{8}\left(u_1^2 + \sum_{i=1}^{\kappa-1}(u_i - u_{i+1})^2 + u_\kappa^2\right) \;-\; \frac{\mathscr{L}}{4}\,u_1 \tag{3.40}$$

constitutes a *convex* optimisation problem with some positive parameters κ and $\mathscr{L}$. The lower bound for such smooth convex problems derived in [Nes04, Theorem 2.1.7] is only valid for one specific iteration count $k \in \mathbb{N}$ (here: $k = 50$ since $\kappa = 101$). Nothing is said about the error before and after the k-th iteration.

Figure 3.6 plots for both problems the resulting error curves of different solvers. While our FSI scheme offers state-of-the-art performance for the strongly convex problem (Figure 3.6(a)), it even outperforms competing methods in the case of the convex optimisation problem (Figure 3.6(b)). In particular, this illustrates the benefits of our nonstationary cyclic parameter choice compared to Polyak's heavy ball method.

3.8 Limitations and Discussion

The experimental evaluation in the previous section illustrated benefits of our approach. As a mainly practical drawback we have to mention that, in comparison to a standard gradient descent scheme, our FSI schemes require the storage of the older iterate $\boldsymbol{u}^{k-1}$. This might be problematic for applications that are limited in the available resources. However, exactly this usage of the older iterate is essential for an algorithmic speed-up; cf. also Nesterov's method [Nes04] and the heavy ball method by Polyak [Pol64]. In this regard, it is worth to note that also well-established succes-

sive overrelaxation and conjugate gradient methods can be interpreted as techniques that make efficient use of the older iterate; see e.g. [HY04, Var09].

Referring to the conjugate gradient method [HS52], another drawback of our technique is the required knowledge of the maximal spectral radius or the Lipschitz constant, and the appropriate choice of τ or ω, respectively. Indeed, we require a precise approximation to provide high efficiency. Conjugate gradient methods circumvent this by a clever usage of inner products. On the other hand, such inner products are unfavourable for parallel computing. In fact, our FSI method refrains from such inner products and thus, is well-suited for highly parallel architectures such as GPUs.

Last but not least, let us comment on a stability or convergence proof of the presented iterative schemes. We are only able to prove global convergence in case of linear problems or quadratic optimisation without constraints; cf. Section 3.A. Further convergence results are not given in this thesis, but would certainly be highly interesting to investigate in future work. Nevertheless, we could observe an asymptotic convergence of our algorithms in all conducted experiments, i.e. our algorithms converged to the desired result in any norm with sufficiently many iterations. The theoretical proof of such a convergence is a goal of future work. It may also allow us to state precise convergence rates for different optimisation problems. Moreover, we could observe that our method is not a descent method; i.e. the error in the Euclidean norm might not decrease for every intermediate iteration. However, such statements might be nevertheless possible for the results after a complete cycle.

A good starting point for future work might be the work of Golub and Kannan [GK86], who analyse a related *stationary* semi-iterative method for the solution of nonlinear equation systems. Additionally, our FSI schemes can be related to the so-called (k, ℓ)-step methods of Gutknecht [Gut89]. Hence, also the mathematical investigations in this context [GNV86, Gut89, Gut15] might be helpful for future research. Moreover, also the recent analysis of heavy ball methods for convex optimisation by Ghadimi et al. [GFJ15] seems to be a promising starting point for future work.

3.9 Summary and Conclusions

In this chapter, we have presented our novel *Fast Semi-Iterative* (FSI) schemes that offer efficient solutions for diffusion evolutions, elliptic problems, and constrained optimisation. The proposed schemes are simple to implement and well-suited for parallel implementations. Hence, they are applicable for a wide range of image processing and computer vision tasks. More specifically, we investigated the relation between box filtering and explicit schemes for 1D homogeneous diffusion. It is fascinating to see how such a rather simple concept generalises to flexible and highly efficient algorithms. Compared to FED-like approaches, our techniques do not require sophisticated rearrangements to avoid numerical instabilities and experimentally provide a significantly higher robustness to nonlinear updates and projections within each iteration. Moreover, our experiments demonstrate benefits w.r.t. related iterative procedures.

As mentioned in Section 3.8, a goal of future work is an extended theoretical analysis of the presented approaches which may allow us to state a general convergence theory. Furthermore, generalisations of our schemes, e.g. to general nonexpansive proximal operators and nonsymmetric system matrices, would be highly interesting to investigate in future work.

3.A Mathematical Analysis

In this appendix, we show a connection of our FSI schemes to Chebyshev polynomials; see e.g. [AS72, MH03]. First, this allows to show stability of our schemes for linear processes in the Euclidean norm after each cycle. Next, we can establish a connection to related iterative schemes. Moreover, it opens the door to a beneficial adaptation of our FSI scheme to strongly convex problems or, in the elliptic case, to a known lower bound of the eigenvalues of the system matrix.

3.A.1 Connection to Chebyshev Polynomials

To investigate the connection of our schemes to Chebyshev polynomials, let us first consider a linear process with the following explicit scheme:

$$\boldsymbol{u}^{k+1} \;=\; \big(\boldsymbol{I} + \tau\,\boldsymbol{A}\big)\,\boldsymbol{u}^{k} \tag{3.41}$$

with a symmetric negative semi-definite matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$. Next, let us proof the following statement:

Theorem 2 (Connection to Chebyshev Polynomials)**.** *The result of one FSI step for linear problems in the form of*

$$\boldsymbol{u}^{k+1} \;=\; \alpha_k \cdot \big(\boldsymbol{I} + \tau\,\boldsymbol{A}\big)\,\boldsymbol{u}^{k} \;+\; (1-\alpha_k)\cdot \boldsymbol{u}^{k-1} \tag{3.42}$$

with $\boldsymbol{u}^{-1} \coloneqq \boldsymbol{u}^0$ *and* $\alpha_k = (4k+2)/(2k+3)$ *for* $k = 0,\dots,n-1$ *can be expressed as*

$$\boldsymbol{u}^{k+1} \;=\; \frac{1}{2k+3}\cdot U_{2k+2}\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\,\boldsymbol{u}^{0}\,, \tag{3.43}$$

where U_{2k+2} *denotes an even Chebyshev polynomial of the second kind.*

Proof. An even Chebyshev polynomial of the second kind is defined by (cf. [AS72, MH03] and Figure 3.7)

$$U_0(z) \;=\; 1 \qquad \text{and} \qquad U_2(z) \;=\; 4z^2 - 1\,, \tag{3.44}$$

and the recurrence relation for $k > 0$

$$U_{2k+2}(z) \;=\; 2\cdot T_2(z)\cdot U_{2k}(z) - U_{2k-2}(z)\,, \tag{3.45}$$

where

$$T_2(z) \;=\; 2z^2 - 1 \tag{3.46}$$

denotes a Chebyshev polynomial of the first kind. Similar to Theorem 1, we proof Theorem 2 by induction. For $k = 0$, we have

$$\begin{aligned} \boldsymbol{u}^{1} \;&=\; \frac{2}{3}\cdot\big(\boldsymbol{I} + \tau\,\boldsymbol{A}\big)\,\boldsymbol{u}^{0} \;+\; \frac{1}{3}\cdot\boldsymbol{u}^{0} \;=\; \frac{1}{3}\cdot\big(3\boldsymbol{I} + 2\tau\boldsymbol{A}\big)\,\boldsymbol{u}^{0} \\ &\overset{(3.44)}{=}\; \frac{1}{3}\cdot U_2\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\,\boldsymbol{u}^{0}\,. \end{aligned} \tag{3.47}$$

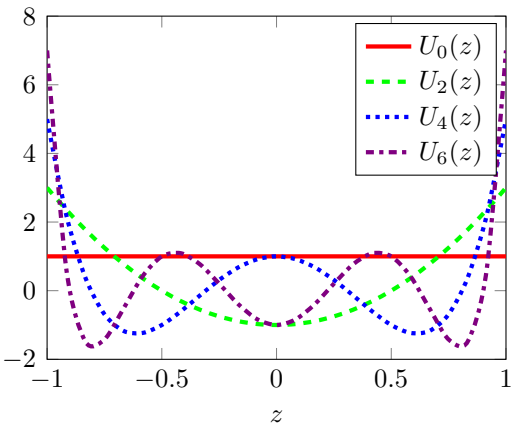


Figure 3.7: Illustration of even Chebyshev polynomials of the second kind U_{2k} for $k \in \mathbb{N}$.

Similarly, $k = 1$ yields

$$\begin{aligned}
\boldsymbol{u}^2 &= \frac{6}{5} \cdot \left(\boldsymbol{I} + \tau\, \boldsymbol{A}\right) \boldsymbol{u}^1 - \frac{1}{5} \cdot \boldsymbol{u}^0 \\
&= \frac{1}{5} \cdot \left(6 \cdot \left(\boldsymbol{I} + \tau\, \boldsymbol{A}\right) \boldsymbol{u}^1 - \boldsymbol{I}\, \boldsymbol{u}^0\right) \\
&\overset{(3.47)}{=} \frac{1}{5} \cdot \left(2 \cdot \left(\boldsymbol{I} + \tau\, \boldsymbol{A}\right) \cdot U_2\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big) - \boldsymbol{I}\right) \boldsymbol{u}^0 \\
&\overset{(3.44)}{=} \frac{1}{5} \cdot \left(2 \cdot \left(\boldsymbol{I} + \tau\, \boldsymbol{A}\right) \cdot U_2\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big) - U_0\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\right) \boldsymbol{u}^0 \\
&\overset{(3.46)}{=} \frac{1}{5} \cdot \left(2 \cdot T_2\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big) \cdot U_2\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big) - U_0\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\right) \boldsymbol{u}^0 \\
&\overset{(3.45)}{=} \frac{1}{5} \cdot U_4\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\, \boldsymbol{u}^0 \,.
\end{aligned} \tag{3.48}$$

Assuming that (3.47) holds for all $k \geq 0$ and (3.48) for all $k \geq 1$, we get for the inductive step ($k \geq 2$)

$$\begin{aligned}
\boldsymbol{u}^{k+1} &= \frac{4k+2}{2k+3} \cdot \left(\boldsymbol{I} + \tau\, \boldsymbol{A}\right) \boldsymbol{u}^k - \frac{2k-1}{2k+3} \cdot \boldsymbol{u}^{k-1} \\
&\overset{(3.46)}{=} \frac{1}{2k+3} \cdot \left((4k+2) \cdot T_2\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big) \cdot \boldsymbol{u}^k - (2k-1) \cdot \boldsymbol{u}^{k-1}\right) \\
&\overset{(3.47)}{=} \frac{1}{2k+3} \cdot \left((4k+2) \cdot T_2\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big) \cdot \boldsymbol{u}^k - U_{2k-2}\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\, \boldsymbol{u}^0\right) \\
&\overset{(3.48)}{=} \frac{1}{2k+3} \cdot \left(2 \cdot T_2\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big) \cdot U_{2k}\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big) - U_{2k-2}\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\right) \boldsymbol{u}^0 \\
&\overset{(3.45)}{=} \frac{1}{2k+3} \cdot U_{2k+2}\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\, \boldsymbol{u}^0 \,.
\end{aligned} \tag{3.49}$$

□

As an important consequence of Theorem 2, we can formulate the result of one

complete FSI cycle with cycle length n for linear processes as

$$\boldsymbol{u}^n \;=\; \frac{1}{2n+1} \cdot U_{2n}\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\, \boldsymbol{u}^0 \,. \tag{3.50}$$

In Theorem 2, we have considered our FSI scheme for linear diffusion evolutions. However, in the same way one can show that similar results also hold for linear elliptic problems and quadratic optimisation without constraints, where $F(\boldsymbol{u}) = \frac{1}{2}\boldsymbol{u}^\mathsf{T}\boldsymbol{B}\boldsymbol{u} - \boldsymbol{u}^\mathsf{T}\boldsymbol{d}$. In this context, we consider the error vector $\boldsymbol{e}^k = \boldsymbol{u}^k - \boldsymbol{u}^*$ that yields

$$\boldsymbol{e}^{k+1} \;=\; \alpha_k \cdot \big(\boldsymbol{I} - \omega\,\boldsymbol{B}\big)\,\boldsymbol{e}^k \;+\; (1-\alpha_k)\cdot \boldsymbol{e}^{k-1} \,. \tag{3.51}$$

Accordingly, we obtain

$$\boldsymbol{e}^n \;=\; \frac{1}{2n+1} \cdot U_{2n}\Big(\sqrt{\boldsymbol{I} - \tfrac{\omega}{2}\boldsymbol{B}}\Big)\, \boldsymbol{e}^0 \,. \tag{3.52}$$

Since we require $0 \le \tau < 2/\varrho(\boldsymbol{A})$, $0 \le \omega < 2/\varrho(\boldsymbol{B})$, and $0 \le \omega < 2/\mathscr{L}$, respectively (cf. Section 3.4, 3.5, and 3.6), our schemes for linear processes are stable in the Euclidean norm; i.e. the Euclidean norm is nonincreasing after each cycle:

$$|\boldsymbol{u}^n| = \left|\frac{1}{2n+1} U_{2n}\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\boldsymbol{u}^0\right| \le \left|\frac{1}{2n+1} U_{2n}\Big(\sqrt{\boldsymbol{I} + \tfrac{\tau}{2}\boldsymbol{A}}\Big)\right| \big|\boldsymbol{u}^0\big| \le \big|\boldsymbol{u}^0\big| \tag{3.53}$$

and

$$|\boldsymbol{e}^n| = \left|\frac{1}{2n+1} U_{2n}\Big(\sqrt{\boldsymbol{I} - \tfrac{\omega}{2}\boldsymbol{B}}\Big)\boldsymbol{e}^0\right| \le \left|\frac{1}{2n+1} U_{2n}\Big(\sqrt{\boldsymbol{I} - \tfrac{\omega}{2}\boldsymbol{B}}\Big)\right| \big|\boldsymbol{e}^0\big| \le \big|\boldsymbol{e}^0\big| \,, \tag{3.54}$$

where we used the property (cf. also Figure 3.7)

$$-1 \;<\; \frac{U_{2n}(z)}{2n+1} \;\le\; 1 \qquad \text{for} \quad z \in (0,1] \,. \tag{3.55}$$

Please note at this point that a diagonal preconditioning with $\boldsymbol{P}^{-1}$ (cf. Section 3.5) requires an appropriate adaptation of ω such that the eigenvalues of $\boldsymbol{I} - \frac{\omega}{2}\,\boldsymbol{P}^{-1}\boldsymbol{B}$ lie in $(0,1]$.

3.A.2 Connection to FED and Fast Jacobi

The discussed representation in terms of Chebyshev polynomials helps to show a close connection to FED schemes [WGSB16], where one cycle with length n reads

$$\boldsymbol{u}^{k+1} \;=\; \big(\boldsymbol{I} + \tau_k\,\boldsymbol{A}\big)\,\boldsymbol{u}^k \qquad (k = 0, \dots, n-1) \tag{3.56}$$

with varying time step sizes

$$\tau_k \;=\; \frac{\tau}{2\,\cos^2\left(\pi\frac{2k+1}{4n+2}\right)} \,. \tag{3.57}$$

Interestingly, $1/\tau_k$ are given as the zeros of the polynomial $U_{2n}\left(\sqrt{1+\frac{\tau}{2}z}\right)$ (cf. [Gre13]). With (3.56) the result of one FED cycle is given by

$$\boldsymbol{u}^n \;=\; \prod_{k=0}^{n-1}(\boldsymbol{I}+\tau_k\boldsymbol{A})\,\boldsymbol{u}^0\,. \tag{3.58}$$

Using Vieta's formula and the property $U_{2n}(1) = 2n+1$ [AS72], we can rewrite the matrix product in (3.58) as

$$\prod_{k=0}^{n-1}(\boldsymbol{I}+\tau_k\boldsymbol{A}) \;=\; \prod_{k=0}^{n-1}\tau_k\cdot\prod_{k=0}^{n-1}(\tfrac{1}{\tau_k}\boldsymbol{I}+\boldsymbol{A}) \;=\; \frac{1}{2n+1}\cdot U_{2n}\Big(\sqrt{\boldsymbol{I}+\tfrac{\tau}{2}\boldsymbol{A}}\Big)\,. \tag{3.59}$$

Comparing (3.50) and (3.59) uncovers the equivalence of the results of one FED and one FSI cycle in case of linear evolutions. In a straightforward way, the same equivalence holds for Fast Jacobi and FSI cycles in case of linear equation systems.

3.A.3 Adaptation to Strongly Convex Problems

On top of that, the representation of our FSI schemes in terms of Chebyshev polynomials (3.50) allows for a natural adaptation of our algorithms to strongly convex problems or, in the elliptic case, to the eigenvalues of the positive definite system matrix $\boldsymbol{B}$. More specifically, we assume a larger bound $\mathscr{L} > 0$ of the largest eigenvalue of the system matrix and a lower bound $\ell \geq 0$ of the smallest eigenvalue.

To derive our recursive parameter choice in (3.25), let us first revisit the general case. With Theorem 2, the error vector (3.17) after one cycle can be written as

$$\boldsymbol{e}^n \;=\; \frac{1}{2n+1}\cdot U_{2n}\Big(\sqrt{\boldsymbol{I}-\tfrac{\omega}{2}\boldsymbol{B}}\Big)\,\boldsymbol{e}^0\,, \tag{3.60}$$

where the parameter ω in (3.60) is chosen such that the eigenvalues of the matrix $\boldsymbol{I}-\frac{\omega}{2}\boldsymbol{B}$ are mapped to the interval $(0,1]$; cf. Equation (3.54). Now, we are given the lower and upper bounds ℓ and $\mathscr{L}$ of the eigenvalues of $\boldsymbol{B}$. Beneficially, this allows for a tighter mapping to the interval $(0,1]$ in the form of

$$\frac{\mathscr{L}\cdot\boldsymbol{I}-\boldsymbol{B}}{\mathscr{L}-\ell}\,. \tag{3.61}$$

Moreover, the factor $\frac{1}{2n+1}$ in (3.60) provides consistency of the iterative scheme: If $\boldsymbol{B}$ would be identical to the zero matrix $\boldsymbol{O} \in \mathbb{R}^{N\times N}$, all iterates should remain unchanged, i.e. $\boldsymbol{e}^k = \boldsymbol{e}^0\ \forall k$. This implies the normalisation factor

$$\frac{1}{U_{2n}\left(\sqrt{1-\frac{\omega}{2}0}\right)} \;=\; \frac{1}{U_{2n}(1)} \;=\; \frac{1}{2n+1} \tag{3.62}$$

such that

$$\frac{1}{2n+1}\cdot U_{2n}\Big(\sqrt{\boldsymbol{I}-\tfrac{\omega}{2}\boldsymbol{O}}\Big) \;=\; \boldsymbol{I}\,. \tag{3.63}$$

Analogously, we determine the normalisation factor for the mapping in (3.61) by

$$\frac{1}{U_{2n}\left(\sqrt{\frac{\mathscr{L}\cdot 1-0}{\mathscr{L}-\ell}}\right)} \;=\; \frac{1}{U_{2n}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)}\,, \tag{3.64}$$

which finally leads to the matrix polynomial

$$\frac{1}{U_{2n}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)} \cdot U_{2n}\left(\sqrt{\frac{\mathscr{L}\,\boldsymbol{I}-\boldsymbol{B}}{\mathscr{L}-\ell}}\right)\,. \tag{3.65}$$

Consequently (cf. Equation (3.43)), the error at $k+1$ is given by

$$\boldsymbol{e}^{k+1} \;=\; \frac{1}{U_{2k+2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)} \cdot U_{2k+2}\left(\sqrt{\frac{\mathscr{L}\,\boldsymbol{I}-\boldsymbol{B}}{\mathscr{L}-\ell}}\right)\,\boldsymbol{e}^{0}\,. \tag{3.66}$$

Applying the recurrence relation (cf. Equation (3.45))

$$U_{2k+2}\left(\sqrt{\frac{\mathscr{L}\,\boldsymbol{I}-\boldsymbol{B}}{\mathscr{L}-\ell}}\right) = 2\cdot T_2\left(\sqrt{\frac{\mathscr{L}\,\boldsymbol{I}-\boldsymbol{B}}{\mathscr{L}-\ell}}\right)\cdot U_{2k}\left(\sqrt{\frac{\mathscr{L}\,\boldsymbol{I}-\boldsymbol{B}}{\mathscr{L}-\ell}}\right) - U_{2k-2}\left(\sqrt{\frac{\mathscr{L}\,\boldsymbol{I}-\boldsymbol{B}}{\mathscr{L}-\ell}}\right) \tag{3.67}$$

leads to

$$\boldsymbol{e}^{k+1} \;=\; \frac{2\cdot T_2\left(\sqrt{\frac{\mathscr{L}\,\boldsymbol{I}-\boldsymbol{B}}{\mathscr{L}-\ell}}\right)\cdot U_{2k}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)}{U_{2k+2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)}\cdot \boldsymbol{e}^{k} - \frac{U_{2k-2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)}{U_{2k+2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)}\cdot \boldsymbol{e}^{k-1}\,. \tag{3.68}$$

With

$$T_2\left(\sqrt{\frac{\mathscr{L}\,\boldsymbol{I}-\boldsymbol{B}}{\mathscr{L}-\ell}}\right) \;=\; \frac{(\mathscr{L}+\ell)\,\boldsymbol{I}-2\boldsymbol{B}}{\mathscr{L}-\ell} \;=\; \frac{\mathscr{L}+\ell}{\mathscr{L}-\ell}\cdot\left(\boldsymbol{I}-\frac{2}{\mathscr{L}+\ell}\boldsymbol{B}\right) \tag{3.69}$$

and

$$\begin{aligned} U_{2k-2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right) \;&=\; 2\cdot T_2\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)\cdot U_{2k}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right) - U_{2k+2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right) && (3.70)\\ &=\; 2\cdot\frac{(\mathscr{L}+\ell)}{\mathscr{L}-\ell}\cdot U_{2k}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right) - U_{2k+2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right) && (3.71)\end{aligned}$$

we obtain

$$\begin{aligned} \boldsymbol{e}^{k+1} \;&=\; \alpha_k\cdot(\boldsymbol{I}-\omega\,\boldsymbol{B})\,\boldsymbol{e}^{k} \;+\; (1-\alpha_k)\cdot\boldsymbol{e}^{k-1} && \mid +\boldsymbol{u}^{*} && (3.72)\\ \boldsymbol{u}^{k+1} \;&=\; \alpha_k\cdot\big((\boldsymbol{I}-\omega\,\boldsymbol{B})\,\boldsymbol{u}^{k}+\omega\,\boldsymbol{d}\big) \;+\; (1-\alpha_k)\cdot\boldsymbol{u}^{k-1} && && (3.73)\end{aligned}$$

where $\omega = \frac{2}{\mathscr{L}+\ell}$ and

$$\alpha_k \;=\; \frac{2(\mathscr{L}+\ell)}{\mathscr{L}-\ell}\cdot\frac{U_{2k}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)}{U_{2k+2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)}\,. \tag{3.74}$$

To determine those parameters efficiently, we make once again use of the recurrence relation in (3.45). For $k = 0$, we have

$$\alpha_0 \;=\; \frac{2(\mathscr{L}+\ell)}{\mathscr{L}-\ell} \cdot \frac{1}{4\frac{\mathscr{L}}{\mathscr{L}-\ell} - 1} \;=\; \frac{2(\mathscr{L}+\ell)}{3\mathscr{L}+\ell}\,. \tag{3.75}$$

For $k \geq 1$, we consider $\frac{1}{\alpha_k}$ and exchange U_{2k+2} by the right hand side of (3.45), i.e.

$$\frac{1}{\alpha_k} \;=\; \frac{\mathscr{L}-\ell}{2(\mathscr{L}+\ell)} \cdot \left(2 \cdot T_2\left(\sqrt{\tfrac{\mathscr{L}}{\mathscr{L}-\ell}}\right) - \frac{U_{2k-2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)}{U_{2k}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)} \right) \tag{3.76}$$

$$\overset{(3.46)}{=} \frac{\mathscr{L}-\ell}{2(\mathscr{L}+\ell)} \cdot \left(\frac{2(\mathscr{L}+\ell)}{\mathscr{L}-\ell} - \frac{U_{2k-2}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)}{U_{2k}\left(\sqrt{\frac{\mathscr{L}}{\mathscr{L}-\ell}}\right)} \right) \tag{3.77}$$

$$\overset{(3.74)}{=} 1 \;-\; \frac{\alpha_{k-1}}{4} \cdot \left(\frac{\mathscr{L}-\ell}{\mathscr{L}+\ell}\right)^2 . \tag{3.78}$$

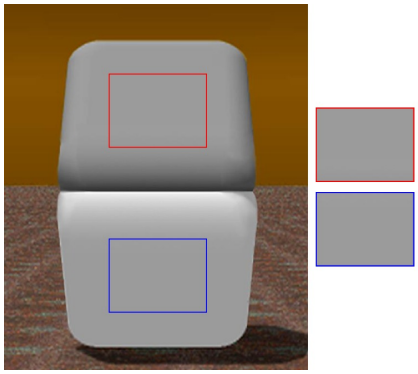
This finally leads to our efficient recursive parameter choice in (3.25).

Chapter 4

Variational Image Fusion

"Contrast is what makes photography interesting."

Conrad L. Hall



Beau Lotto [Lot02]

Contents

4.1	**Introduction**	**45**
4.2	**Related Work**	**47**
4.3	**Variational Model**	**50**
4.4	**Experiments**	**58**
4.5	**Limitations and Discussion**	**72**
4.6	**Summary and Conclusions**	**73**

Main parts of this chapter base on our work published in [HW15, HW16].

4.1 Introduction

As outlined in Chapter 1, the fusion of multiple images is an important component in many visual computing applications. In particular, it is an essential tool when several photographs or sensors are required to capture all important structures of a scene. In this context, image fusion approaches aim at condensing the most important information from the acquired image stack to a single composite that is richer in details than any of the input images.

In fact, there exist many fusion approaches that are tailored to one specific application. Most of them pursue the following two-step pipeline: Based on application-specific quality measures, they determine weights for each of the input images in a first

step. Then, in a second step, those weights are combined with the input images to form the output image. Contrary to such a decoupled two-stage approach, we propose a conceptually different idea. Instead of precomputing weights based on the input images, we present a variational method that directly aims for an optimal output image with high quality. This has several advantages: First, all our model assumptions and parameters have an intuitive meaning and directly influence the fusion result in the desired way. Moreover, our approach produces images with visual phenomena such as a Cornsweet illusion that optimise the contrast locally.[1] This is hardly achievable with standard two-step fusion methods. Most importantly, this output-driven idea is the key concept to refrain from an application-specific weight precomputation and, in this way, to provide a general image fusion framework that performs well in many fusion applications.

Main Contributions. While the input data differ for each image fusion application, most applications aim at a similar goal: an output image with important perceptual qualities such as being well-exposed or offering a high local contrast. Based on this observation, we introduce a general variational technique for image fusion. We achieve this by refraining from an application-specific precomputation of weights based on the input images. Instead, we formulate the fusion result as a convex combination of the input, and directly opt for an output that is optimal w.r.t. our energy functional. Here, we base our model assumptions on important perceptually inspired image enhancement concepts that account for the local contrast adaptation of the human visual system. The minimisation of our energy yields fusion results that capture the most important information from the input and feature a desirable visual quality.

We demonstrate the general applicability of our technique with several fusion applications. For our three main application domains, i.e. multispectral fusion, exposure fusion, and decolourisation, we conduct thorough evaluations on public image data sets and compare to previous work in the individual research areas. Last but not least, we present an efficient algorithmic solution by exploiting ideas of [BCPR07] and applying our FSI scheme from Chapter 3. In combination with a parallel implementation on the graphics card, this leads to desirably fast runtimes of our fusion approach.

Chapter Outline. We start with a discussion of related work in Section 4.2. Afterwards, we present our general variational image fusion model in Section 4.3. Its minimisation yields the desired composite image. Here, we also analyse our model and investigate its parameters. The experiments in Section 4.4 demonstrate the versatility and performance of our technique for different image fusion applications. After

[1] An illustrative example of a Cornsweet illusion is given in the teaser image of this chapter. Although the red and blue patches have exactly the same brightness value, we perceive the red patch slightly darker than the blue one. This is achieved by the local intensity variation between the lower and upper part of the object. For further details about the Cornsweet illusion and related concepts, we refer the interested reader to [Cor70, SE84, PSL99] and references therein.

discussing possible limitations of our method in Section 4.5, we conclude this chapter with a summary and outlook in Section 4.6.

4.2 Related Work

To explain how our model relates to previous work, let us first give a survey of our three main application domains (Section 4.2.1–4.2.3). Furthermore, since we base our model assumptions on variational contrast enhancement concepts, we additionally review related work in this field (Section 4.2.4).

4.2.1 Multispectral Fusion

In multispectral imaging, multiple sensors or filters are used to acquire a set of images. They all capture a different spectral range of the scene, e.g. the visible and the near-infrared band (cf. for instance [FS08, BS11]). The fusion of those images results in an image that offers more details than any of the input photographs; see e.g. our exemplary results in Figure 1.3 and Figure 1.4. Based on brightness and saturation of the visible spectrum image, Zhang et al. [ZSM08] precompute a weight map to identify image regions that should be improved. Then, contrast and texture is transferred from the near-infrared band to the visible spectrum image. Lau et al. [LHM11] and Eynard et al. [EKB14] interpret the near-infrared as a fourth colour channel, and regard multispectral image fusion as a colour transformation task. However, this has the following drawback: In some parts of the image the visible spectrum is more reliable than the near-infrared, and vice versa. Consequently, it is not optimal to assume the same importance of all spectral ranges in every image region. Hence, contrary to [LHM11] and [EKB14], we interpret the near-infrared information as an additional *spatial* lightness information. Interestingly, this is in accordance to the psychophysical study of Fredembach and Süsstrunk [FS08] in the context of near-infrared imaging. Thus, we explicitly model a spatially varying importance of the different spectral ranges. We demonstrate the resulting benefits of our approach by means of experiments in Section 4.4.1.

4.2.2 Decolourisation

Although not immediately obvious, also decolourisation can be approached by image fusion. In general, decolourisation describes the conversion of a colour image to its greyscale representation, while preserving as much information as possible. For a detailed review and evaluation of decolourisation approaches before 2009, we refer to the article of Čadík [Čad08]. Here, we mainly focus on more recent decolourisation methods.

A straightforward idea is to compute the greyscale image as the Y channel of the CIE-XYZ colour space; cf. Section 2.2.2. However, such a simple global mapping is not always sufficient to preserve the colour contrast. Hence, Smith et al. [SLTM08] first

apply a global mapping that is based on the Helmholtz-Kohlrausch effect. Afterwards, they locally add details with an unsharp masking technique. Gooch et al. [GOTG05] propose a fully local decolourisation method by forcing the differences between pixels in the greyscale image to resemble them from the colour image. Lu et al. [LXJ12] relax this constraint with a weak colour ordering, i.e. the sign of the greyscale differences is not assumed to be identical to the sign of the colour counterparts. Also the structure preserving technique of Eynard et al. [EKB14] is applicable for decolourisation. Here, they aim at a greyscale image whose graph Laplacian resembles the one of the colour image. Another successful idea to approach the decolourisation task is to fuse the colour channels of the input image; see e.g. [AAHB11]. Generally, none of the colour channels or a simple global combination of them is sufficient to represent the original colour image adequately. However, a spatially varying channel fusion can provide satisfying results. In this work, we also approach decolourisation by image fusion; cf. Figure 1.6. However, contrary to [AAHB11], we do not precompute decolourisation-specific fusion weights based on the input. Rather, we directly aim at an optimal output greyscale image. This not only affords the general applicability of our model to various fusion applications, but also realises visual phenomena such as a Cornsweet illusion that optimise the contrast locally (cf. [Cor70, SE84, PSL99]). As we will see, this is an important feature to visually preserve the contrast of the input image if one considers the drastic intensity range restriction by a colour to greyscale transformation.

4.2.3 Exposure Fusion

Classical *high dynamic range* (HDR) methods combine several *low dynamic range* (LDR) images to one HDR image with the help of the exposure times and the camera response function; see e.g. [MP95, DM97, MN99]. However, displaying those HDR results on standard monitors or printing them requires to compress the high dynamic range again. This process is called *tone mapping*; see e.g. [RHD+10] for a survey and [ČWNA08] for a discussion and evaluation of various tone mapping operators. Since tone mapping is not the focus of our work, we restrict our discussion to tone mapping operators that are most related to our fusion technique.

The tone mapping operator of Fattal et al. [FLW02] works in the gradient domain, and accounts for the local contrast adaptation of the visual system by attenuating large gradients, and maintaining or even enhancing smaller ones. Comparably, Durand and Dorsey [DD02] decompose the HDR image into a base and a detail layer by means of bilateral filtering. Then, they compress the base while keeping the details. Reinhard et al. [RSSF02] apply first a global transform, and locally increase the contrast afterwards. Also Mantiuk et al. [MMS06] show and discuss the importance of the contrast adaptation of the human visual system with regard to tone mapping. Most related to our work is the two-stage tone mapping operator of Ferradans et al. [FBPC11] that applies a variational contrast enhancement in the second stage; cf. also Section 4.2.4.

If the ultimate goal is anyway a displayable and well-exposed LDR image, there is an interesting alternative to the described two-step procedure of HDR imaging and tone mapping, namely *exposure fusion* [MKV09]. Here, the idea is to skip the HDR image generation by a direct fusion of the differently exposed LDR images to an overall well-exposed composite. Due to the visual similarity to a tone-mapped HDR image, the result of exposure fusion is also sometimes referred to as *pseudo*-HDR image. Such an exposure fusion approach has several advantages: First, there is no need to know the exposure times or the camera response function. It is even possible to include images that do not follow the HDR imaging model, such as images captured with different camera settings, e.g. with and without photoflash. Second, this one-step approach allows a direct tuning of the final results without the detour via an intermediate HDR image. Obviously, exposure fusion is related to tone mapping. However, the different types of input data ask for different requirements and model assumptions. In the meantime, exposure fusion has even developed to an own research area with various publications that we review next.

Most existing exposure fusion methods pursue the following processing pipeline: In a first step, based on exposure fusion-specific quality measures, weighting maps are computed for each of the input images. Such quality measures are for instance the magnitude of the Laplacian [Bog00, MKV09], the entropy [Gos05, HP10], or the colour saturation [MKV09, SCSB11, SKB14]. A similar idea, e.g. applied by Raman and Chaudhuri [RC09] or by Singh et al. [SKB14], is to decompose the input images into base and detail layers. Then, the amount of detail is considered as measure to determine the input image weights. In a second step, these weighting maps are combined with the input images to form the final composite image. Here, the fusion strategies vary from region-based blending [Gos05] and pixelwise weighted averaging [RC09, HP10, SCSB11, SCB13, SKB14] to gradient domain fusion [CH04, STC$^+$12] and pyramid-based techniques [BK93, Bog00, MKV09]. Different to those two-step approaches, Raman and Chaudhuri [RC07] propose a variational method to directly compute the fused composite. However, they require a smoothness constraint of the final image which is prone to over-smoothed, blurry results. Hence, a more suitable idea by Kotwal and Chaudhuri [KC11] is to formulate the output image as a weighted average of the input. Then, they design an energy on this composite. In this work, we follow a similar idea. However, we base our model assumptions on perceptually inspired contrast enhancement concepts. This allows e.g. for optimising the local contrast in the sense of the Cornsweet effect and for producing images with vivid colours.

4.2.4 Variational Contrast Enhancement

The discussions above show that there are many approaches that are specifically tailored to the individual fusion tasks. However, all presented applications share a similar goal: The fusion of several images to one composite that offers *optimal local contrast*. In fact, we make use of exactly this observation to present a general variational fusion approach. To this end, we profit from important findings in histogram modification

and contrast enhancement that we review next.

Based on the seminal work of Sapiro and Caselles [SC97] on histogram modification with differential equations, Bertalmío et al. [BCPR07] introduce a variational approach to locally increase the contrast of an image. In this context, Palma-Amestoy et al. [PPBC09] investigate several perceptually inspired energy terms. Interestingly, there are strong connections of such models to Land's retinex theory [LM71] and to visual neuroscience; see e.g. [Ber14]. Moreover, Batard and Bertalmío [BB15] discuss and connect nonlocal image regularisation (cf. [GO08]) and such variational contrast enhancement methods [BCPR07] by means of their dual formulations.

Those contrast enhancement techniques have already found several successful applications: In Section 4.2.3, we have already mentioned the two-stage tone mapping operator of Ferradans et al. [FBPC11] that applies such a variational contrast enhancement in the second stage. Further application areas are gamut mapping [ZVB14] and image dehazing [GVPB15]. More related to image fusion is the work of Bertalmío and Levine [BL13]. They propose a gradient-based variational approach to fuse a pair of images with different exposure times. Unfortunately, this method is specifically tailored to two input images and cannot be extended to multiple images in a straightforward way. Piella [Pie09] incorporates a gradient domain term in the energy of [BCPR07]. This forces the similarity to a precomputed gradient field that combines the gradients from multiple images. Here, the weights for the individual images are predetermined based on the input. In contrast, our energy minimisation that directly aims at an optimal composite is able to refrain from such an application-specific weight precomputation. Nevertheless, those existing successful applications motivate us to also base our model on such perceptually inspired contrast enhancement concepts. They clearly have demonstrated their usefulness, and their perceptual basis has been extensively discussed in various publications; see e.g. [BCPR07, PPBC09, BCP09, PC14, Ber14, PAB+17].

4.3 Variational Model

4.3.1 Energy Formulation

Our general goal is to fuse n_f input images $f_1, \dots, f_{n_f}$ to a single composite image u that condenses the most important information from the image stack. For didactic reasons, we consider greyscale images first and discuss our adaptations to colour images later. Let us start with formulating the output image as a pointwise convex combination of the input:

$$u(\boldsymbol{x}) \;=\; \sum_{i=1}^{n_f} w_i(\boldsymbol{x}) \cdot f_i(\boldsymbol{x})\,, \tag{4.1}$$

where

$$w_i(\boldsymbol{x}) \geq 0 \qquad \text{and} \qquad \sum_{i=1}^{n_f} w_i(\boldsymbol{x}) = 1\,. \tag{4.2}$$

Here, $\boldsymbol{x} = (x_1, x_2)^\mathsf{T}$ denotes the position on the rectangular image domain $\Omega \subset \mathbb{R}^2$, and $w_i \colon \Omega \to [0,1]$ is the weight map of the image f_i.

As discussed in Section 4.2, most previous research concentrates on determining weights based on application-specific quality measures that are defined on the input images. In contrast to such a weight precomputation, we directly opt for an optimal fusion result u. To this end, we propose the following energy functional:

$$\begin{aligned} E(\boldsymbol{w}) \;=\; & \frac{1}{2} \int_\Omega \Big(\Big(u(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x}) \Big)^2 + \delta \cdot \big(u(\boldsymbol{x}) - \mu\big)^2 \Big) \,\mathrm{d}\boldsymbol{x} \\ & - \frac{\gamma}{2} \int_\Omega \int_\Omega G_\sigma(\boldsymbol{x} - \boldsymbol{y}) \cdot \Psi_\lambda(u(\boldsymbol{x}) - u(\boldsymbol{y})) \,\mathrm{d}\boldsymbol{x} \,\mathrm{d}\boldsymbol{y} \\ & + \frac{\alpha}{2} \int_\Omega \sum_{i=1}^{n_f} |\boldsymbol{\nabla} w_i(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x} \end{aligned} \tag{4.3}$$

subject to the constraints (4.1) and (4.2). We will discuss the meaning and interplay of all components of this energy functional step by step in the upcoming paragraphs. At this point, we ask for the patience of the reader. The image weights $\boldsymbol{w} = (w_1, \ldots, w_{n_f})^\mathsf{T}$ that follow from this energy optimisation can be seen as a side-product of our output-driven approach. In fact, we are mainly interested in the fused image u. However, formulating this image as a convex combination of the input allows to impose a smoothness constraint on the weights and *not* on the image itself. While the latter is prone to cause over-smoothed, blurry fusion results, the former is a much more intuitive and meaningful assumption. A further important advantage of this formulation is the inherent attachment of u to the input data that prevents so-called *halo* artefacts and an unrealistic appearance of the fusion results.

As discussed in Section 4.2, the energy functional in (4.3) is inspired by successful variational histogram modification and contrast enhancement techniques; see e.g. [SC97, BCPR07, PPBC09, BCP09]. In particular, these works discuss and analyse in detail in which way the energy terms mimic important properties of the human visual system, and for instance how they relate to Land's retinex theory [LM71]. In line with this, let us now explain our model assumptions and corresponding energy terms in (4.3).

Dispersion Term

Following [BCPR07], we model in the first line of (4.3) a so-called *dispersion term*. The first part of this dispersion term forces u to resemble the attachment image $\tilde{f}$, which we choose as an average of the input images. As discussed in [PPBC09], this provides a desirable attachment to the original data and accounts for the colour constancy assumption [LM71]. The second part implements the grey world principle [Buc80, SE84, PPBC09]. In fact, it provides well-exposed images by keeping the solution close to the constant μ. Here, the influence of the second assumption can be steered with the positive parameter δ.

Contrast Term

The second term, the *contrast term*, counteracts this dispersion term since it favours images with a high local contrast more than uniform images that are close to a constant. Please note the minus sign in front of the contrast term. Intuitively speaking, this energy term favours solutions that differ much from pixel to pixel. Here, locality is introduced by the Gaussian weighting

$$G_\sigma(\boldsymbol{x}-\boldsymbol{y}) \;=\; \frac{1}{2\pi\sigma^2}\,\exp\left(\frac{-\,|\boldsymbol{x}-\boldsymbol{y}|^2}{2\sigma^2}\right) \tag{4.4}$$

with standard deviation σ. Furthermore, the function

$$\Psi_\lambda(z) \;=\; \sqrt{z^2+\lambda^2} \tag{4.5}$$

provides a nonlinear behaviour. More specifically, in accordance with [BCPR07], its sigmoid-shaped derivative $\Psi'_\lambda(z) = z/\sqrt{z^2+\lambda^2}$ that appears in the algorithmic iteration (cf. Equation 4.7) mimics the nonlinear response of the human visual system in the sense of a contrast transducer function [Wil80, McC89]. Here, the parameter λ allows to tune this nonlinearity, and $\gamma > 0$ weights the influence of the contrast term.

Regularisation Term

The third term of our energy functional is a *regularisation term* that rewards smooth weight maps. More specifically, it renders the assumption that neighbouring pixels in the fused composite should have similar weights. Here, $\boldsymbol{\nabla} \coloneqq (\partial_{x_1}, \partial_{x_2})^\mathsf{T}$ denotes the gradient operator, and $\alpha > 0$ steers the amount of smoothness.

In the context of image regularisation, more sophisticated edge-preserving smoothness terms have shown advantages compared to the proposed linear isotropic regularisation; see e.g. [Wei98] and references therein. However, as it turns out, this is not the case for our image fusion task. Here, smooth weight maps and a resulting *smooth blending* of the input images are beneficial (also at image edges). In particular, this explains why edge-preserving regularisation does not lead to better results in the considered image fusion setting. In fact, it may even lead to undesirable visual seams that deteriorate the output image.

Simplex Constraint

Our energy in (4.3) is equipped with the *simplex constraint* (4.2) on the image weights $\boldsymbol{w}$; cf. also Section 2.4. In particular, this constraint restricts the fusion result u to pointwise convex combinations of the input images. Hence, it provides a close attachment to the input. In combination with the smoothness constraint, this prevents an unnatural high amount of contrast and, in this way, undesirable artefacts such as colour shifts or halos in the output image.

Colour Image Processing

For the sake of simplicity, we have restricted ourselves to greyscale images so far. In case of colour images, we transform the input images from the RGB to the YCbCr colour space (cf. Section 2.2.2), and define the dispersion and contrast term on the luminance channel $u_{\mathrm{Y}} \colon \Omega \to [0,1]$. Additionally, to prevent colour casts, we compute *joint* weight maps for all channels.

Moreover, largely saturated colours make images to look vivid and expressive. To this end, we extend our energy (4.3) in case of colour images with the following *saturation term*:

$$-\frac{\vartheta}{2} \int_{\Omega} \Big(\big(u_{\mathrm{Cb}}(\boldsymbol{x}) - 1/2\big)^2 + \big(u_{\mathrm{Cr}}(\boldsymbol{x}) - 1/2\big)^2 \Big) \, \mathrm{d}\boldsymbol{x} \,, \tag{4.6}$$

where $u_{\mathrm{Cb}} \colon \Omega \to [0,1]$ and $u_{\mathrm{Cr}} \colon \Omega \to [0,1]$ denote the chroma channels of u. This term favours values different from grey, i.e. where both chroma channels are equal to $1/2$. Hence, this term aims for images with vivid colours. Here, the positive parameter ϑ allows to control the amount of colour saturation. Once again, the minus sign in front of this saturation term should be noted.

4.3.2 Minimisation

Gradient Descent

With iteration index k and time step size τ, the gradient descent of the energy in (4.3) with the saturation term in (4.6) is for $i = 1, \dots, n_f$ given by

$$\begin{aligned} w_i^{k+1}(\boldsymbol{x}) \; = w_i^k(\boldsymbol{x}) - \tau \Big(& f_{\mathrm{Y}i}(\boldsymbol{x}) \Big(u_{\mathrm{Y}}^k(\boldsymbol{x}) - \tilde{f}_{\mathrm{Y}}(\boldsymbol{x}) + \delta \, (u_{\mathrm{Y}}^k(\boldsymbol{x}) - \mu) \\ & - \gamma \int_{\Omega} G_\sigma(\boldsymbol{x} - \boldsymbol{y}) \cdot \Psi_\lambda'(u_{\mathrm{Y}}^k(\boldsymbol{x}) - u_{\mathrm{Y}}^k(\boldsymbol{y})) \, \mathrm{d}\boldsymbol{y} \Big) \\ & - \vartheta \Big(f_{\mathrm{Cb}i}(\boldsymbol{x}) \big(u_{\mathrm{Cb}}^k(\boldsymbol{x}) - 1/2 \big) + f_{\mathrm{Cr}i}(\boldsymbol{x}) \big(u_{\mathrm{Cr}}^k(\boldsymbol{x}) - 1/2 \big) \Big) \\ & - \alpha \, \Delta w_i^k(\boldsymbol{x}) \Big) \,, \end{aligned} \tag{4.7}$$

where $(u_{\mathrm{Y}}^k, u_{\mathrm{Cb}}^k, u_{\mathrm{Cr}}^k)^{\mathsf{T}} = \sum_{i=1}^{n_f} w_i^k(\boldsymbol{x}) \cdot (f_{\mathrm{Y}i}, f_{\mathrm{Cb}i}, f_{\mathrm{Cr}i})^{\mathsf{T}}$. We discretise Equation (4.7) with finite differences on a rectangular grid with uniform grid sizes, and approximate the integral with the rectangle method. Furthermore, we initialise w_i with $1/n_f$, and assume the images to be mirrored at the boundaries.

Algorithmic Speed-Up

Although the contrast term in (4.3) accounts for an optimal *local* contrast, we want to point out that it is *nonlocal* in its nature: For each point, we have to consider a large

nonlocal neighbourhood to determine the local contrast. The size of this neighbourhood depends on the standard deviation σ of the Gaussian weighting function, which is chosen large to provide good quality; cf. Section 4.3.3. The contrast term results in the integral term in the second line of (4.7). Hence, the computational complexity of a single iteration step is $\mathcal{O}(N^2)$, where N denotes the number of pixels. Obviously, this shows the need for efficient algorithms. To this end, we apply the following numerical approximation [BCPR07]: First, we approximate the sigmoid-shaped function $\Psi'_\lambda(z)$ with a polynomial $\sum_{i=0}^{n_a} a_i z^i$ of degree n_a (cf. Figure 4.1). Specifically, we obtain the coefficients $a_0, \ldots, a_{n_a}$ by minimising the quadratic error function[2]

$$F(a_0, \ldots, a_{n_a}) \;=\; \int_{-1}^{1} \Big(\sum_{i=0}^{n_a} a_i z^i - \Psi'_\lambda(z) \Big)^2 \,\mathrm{d}z \,, \tag{4.8}$$

where we assume an intensity range $[0, 1]$. We solve the resulting small linear system with the discussed Cholesky decomposition from Section 2.5. Obviously, since $\Psi'_\lambda(z)$ is an odd function, we only need to consider the odd parts of the polynomial, i.e. $a_i = 0$ for all even i. In our experiments, we apply a polynomial of degree $n_a = 7$ since it provides a good trade-off between approximation quality and efficiency.

Next, the determined coefficients allow us to approximate the term $\Psi'_\lambda(u(\boldsymbol{x}) - u(\boldsymbol{y}))$ as follows:

$$\begin{aligned} \Psi'_\lambda \left(u(\boldsymbol{x}) - u(\boldsymbol{y}) \right) \;&\approx\; \sum_{i=0}^{n_a} a_i \left(u(\boldsymbol{x}) - u(\boldsymbol{y}) \right)^i \\ &=\; \sum_{i=0}^{n_a} a_i \left(\sum_{j=0}^{i} \binom{i}{j} u^{i-j}(\boldsymbol{x}) \, (-1)^j \, u^j(\boldsymbol{y}) \right) \\ &=\; \sum_{j=0}^{n_a} \left(\sum_{i=j}^{n_a} (-1)^j \, a_i \binom{i}{j} u^{i-j}(\boldsymbol{x}) \right) u^j(\boldsymbol{y}) \,, \end{aligned} \tag{4.9}$$

where we applied in the second step the binomial formula with $\binom{i}{j} = \frac{i!}{j!(i-j)!}$. Moreover, the last step follows from an exchange of the summations. With (4.9), we finally approximate the integral term in the second line of (4.7) as

$$\begin{aligned} &\int_\Omega G_\sigma(\boldsymbol{x} - \boldsymbol{y}) \cdot \Psi'_\lambda(u^k_{\mathrm{Y}}(\boldsymbol{x}) - u^k_{\mathrm{Y}}(\boldsymbol{y})) \,\mathrm{d}\boldsymbol{y} \\ &\approx \sum_{j=0}^{n_a} \Bigg(\underbrace{\Big(\sum_{i=j}^{n_a} (-1)^j a_i \binom{i}{j} u^{i-j}(\boldsymbol{x}) \Big)}_{\text{(I)}} \cdot \underbrace{\int_\Omega G_\sigma(\boldsymbol{x} - \boldsymbol{y}) \, u^j(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}}_{\text{(II)}} \Bigg) . \end{aligned} \tag{4.10}$$

[2] In this context, one could also make use of Chebychev polynomials that are known to minimise the maximum norm of the approximation [MH03] or Bernstein polynomials as proposed in [PM16, PAB+17]. However, the applied approximation in the form of (4.8) was sufficient for our purposes.

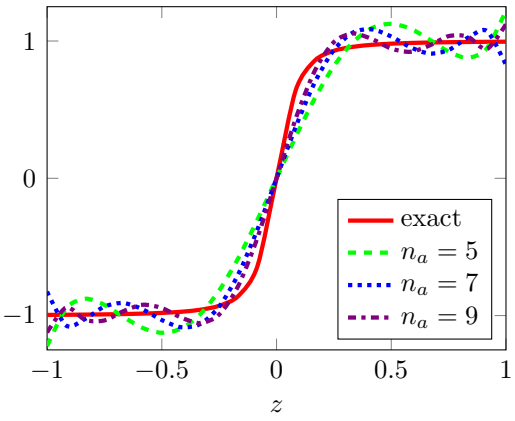


Figure 4.1: Polynomial approximations of the sigmoid-shaped function $\Psi'_\lambda(z) = \frac{z}{\sqrt{z^2+\lambda^2}}$, where n_a denotes the order of the polynomial, and $\lambda = 0.1$ in this example.

We see that we can compute (I) pointwise with powers of $u(\boldsymbol{x})$. Moreover, (II) is nothing else than a Gaussian convolution of u^j, i.e.

$$G_\sigma * u^j(\boldsymbol{x}) \;=\; \int_\Omega G_\sigma(\boldsymbol{x}-\boldsymbol{y})\, u^j(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}\,. \tag{4.11}$$

We compute those convolutions with a fast recursive algorithm [Yv95]. In this way, we can reduce the overall complexity of one iteration step from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$.

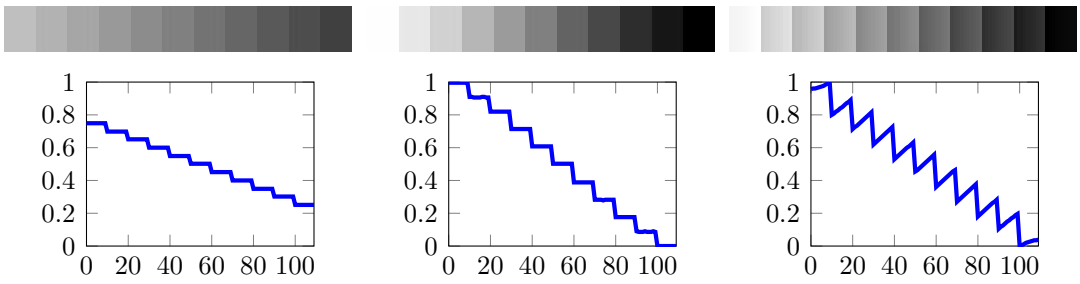
FSI Scheme

To compute the steady state solution of (4.7) with the discussed polynomial approximation, we apply our projected FSI scheme from Chapter 3 (Algorithm 3.1). Here, we realise the projection of $\boldsymbol{w}^k$ onto the simplex in $\mathbb{R}^{n_f}$ as discussed in Section 2.4 (Algorithm 2.1). The combination of both techniques, i.e. the polynomial approximation with a fast recursive realisation of Gaussian convolution [Yv95] and the application of our FSI scheme, enables a fast parallel implementation of our fusion method on the graphics card; cf. our upcoming runtime evaluation in Section 4.3.3.

4.3.3 Model and Parameter Analysis

Benefits of Local Contrast Term

Inspired by Bertalmío et al. [BCPR07], let us first consider a variational histogram equalisation in Figure 4.2. In our framework, this can be achieved by setting the smoothness parameter $\alpha = 0$ and by replacing the simplex constraint (4.2) with $0 \leq w(\boldsymbol{x})f(\boldsymbol{x}) \leq 1$. Applying a global contrast term, i.e. degrading $G_\sigma(\boldsymbol{x}-\boldsymbol{y})$ to the constant $1/|\Omega|$, yields a standard histogram equalisation (cf. Figure 4.2(b)). On the other hand, a local contrast term allows to visually increase the contrast in the sense of a Cornsweet illusion [Cor70,SE84,PSL99] (cf. Figure 4.2(c)). This experiment illustrates the general advantages of a local contrast term compared to a global one.



(a) input (b) global (c) local

Figure 4.2: Comparison of global and local contrast term for histogram equalisation. *Top*: Intensity images. *Bottom*: Corresponding horizontal scanlines, where the horizontal axes represent the pixel indices and the vertical axes the intensity values. This experiment is inspired by Bertalmío et al. [BCPR07].

(a) input [Čad08] (b) Ancuti et al. [AAHB11] (c) our result

Figure 4.3: Benefits of output-driven optimisation. This example demonstrates that our output-driven optimisation implements a local contrast adaptation that keeps the colour patches in the greyscale image better distinguishable.

Benefits of Output-Driven Optimisation

Thanks to our output-driven optimisation, our fusion technique is capable of producing fusion results with the just discussed Cornsweet illusions. This is hardly possible with standard two-stage fusion methods, since they do not take into account the quality of the final output image when precomputing weights for the input images in advance. Let us consider a decolourisation example in Figure 4.3: Ancuti et al. [AAHB11] precompute decolourisation-specific weights based on the input and fuse the images later on. In this way, it is not possible to create greyscale gradients in the individual patches; cf. Figure 4.3(b). In contrast, our output-driven optimisation produces such gradients that optimise the contrast locally (cf. Figure 4.3(c)). Furthermore, our direct optimisation w.r.t. the output image is the key concept to present a general technique for image fusion without tailoring it to the individual applications.

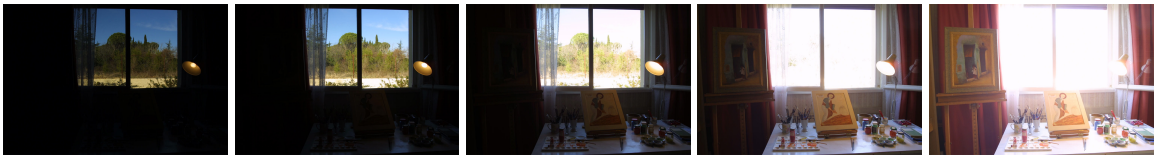


Figure 4.4: Input images of example exposure set with image sizes 1200×800 [Jof07].

$\mu = 0.1$ $\sigma = 1\%$ $\gamma = 0$ $\vartheta = 0$

$\mu = 0.9$ $\sigma = 100\%$ $\gamma = 0.5$ $\vartheta = 2$

Figure 4.5: Influence of main model parameters. Parameter σ is given in percent of the image diagonal. The corresponding input images can be found in Figure 4.4.

Model Parameters

Let us now illustrate the influence of our main model parameters on the example of fusing an exposure set. The corresponding input image set is depicted in Figure 4.4.

In the first column of Figure 4.5, we apply different values of μ. Since the dispersion term favours solutions that are close to μ, it is obvious that larger values lead to brighter results. We propose to compute μ automatically as an average of the input images. The second column in Figure 4.5 depicts composite images for different scales σ of the Gaussian G_σ in the contrast term. We observe a larger *local* contrast with decreasing σ. Obviously, there is a trade-off: A too large local contrast might be perceived as unnatural, a too small one as too flat. As a rule of thumb, we propose to set σ to 10% of the image diagonal. Similar observations apply to the contrast parameter γ in the third column of Figure 4.5. Choosing it too small yields an image with low contrast, while choosing it too large gives unrealistic looking results. In general, setting it to $1/4$ provides good quality. Last but not least, the fourth column of Figure 4.5 shows the effects of the proposed saturation term and its parameter ϑ. Generally, a larger value of ϑ leads to more saturated colours, and in this way to a more vivid appearance. Setting $\vartheta = 1$ yields a good compromise in this regard.

In all our experiments below, we apply the discussed procedure to determine μ and σ automatically. All other parameters are fixed for the individual applications. Table 4.1 shows the standard parameter setting. The corresponding final fusion result

Table 4.1: Default parameter setting of our fusion technique.

α	ϑ	γ	δ	λ	μ	σ
1	1	$1/4$	1	$1/10$	input average	10% of diagonal

Table 4.2: Runtimes on an *NVIDIA GeForce GTX 970*.

number of images	image size	runtime in seconds
3	640×480	4
3	1024×768	7
5	640×480	8
5	1024×768	18

for the exposure set from Figure 4.4 with this standard parameter choice can be found in Figure 1.2(d).

To conclude, all our model parameters have an intuitive meaning and are fixed or can be determined automatically. This allows an easy and straightforward use of our approach, also for non-experts. Obviously, this is an important feature of our fusion technique since it provides the users with intuitively comprehensible parameters to tune the output w.r.t. their personal preferences.

Runtime Experiments

Our reference implementation is written in *CUDA* and runs on an *NVIDIA GeForce GTX 970* graphics card.[3] We assume the algorithm to be converged if the root mean squared difference of two fusion results between 100 iterations is less than 10^{-4}. Table 4.2 lists the runtimes for different number of images and different image sizes. This evaluation demonstrates that the discussed algorithm from Section 4.3.2 provides fast execution times for common image set sizes.

4.4 Experiments

In the following sections, we evaluate our approach in the context of its main application domains, i.e. multispectral fusion (Section 4.4.1), decolourisation (Section 4.4.2), and exposure fusion (Section 4.4.3). Last but not least, we additionally demonstrate its potential for further image fusion applications in Section 4.4.4.

4.4.1 Multispectral Fusion

The idea of multispectral imaging is to capture different spectral ranges with two or more photographs of the same scene. Fusing them allows for producing an image that

[3] At this point, special thanks go to Sabine Müller for providing her personal computer to conduct those runtime experiments.

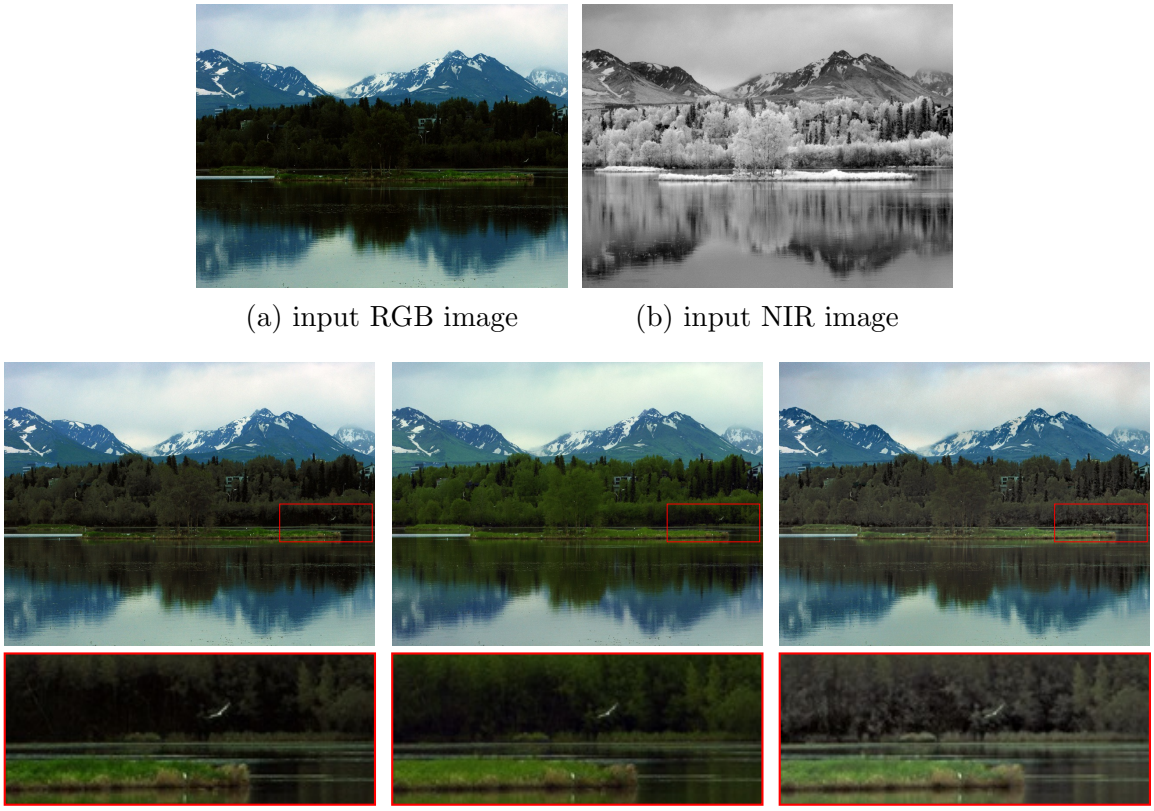
offers details that cannot be captured within the visible spectrum only. In this regard, Figure 1.4 illustrates the performance of our method for the enhancement of standard photographs (RGB) with a *near-infrared* (NIR) image. We process the images in the following way: First, we convert the RGB image to the YCbCr colour space (cf. Section 2.2.2). Next, we apply the presented technique to fuse the luminance channel Y with the single channel NIR image. Since we want to stay close to the visible spectrum image while adding details from the near-infrared range, we choose the luminance channel Y as attachment image $\tilde{f}$. Finally, we combine the fused luminance composite with the original chroma channels to form the output image.

In Figure 4.6, we compare our RGB-NIR fusion results to two state-of-the-art approaches from the literature. As mentioned in Section 4.2.1, contrary to Lau et al. [LHM11] (Figure 4.6(c)) and Eynard et al. [EKB14] (Figure 4.6(d)), we do *not* regard the near-infrared range as a fourth colour channel. Instead, in accordance to psychophysical studies [FS08], we treat it as an additional *spatial* lightness information. As our result demonstrates, this is an important model assumption. Compared to both competing approaches, our fused image (Figure 4.6(e)) offers the most details from both spectral ranges; see for instance the trees and the mountains. Besides providing better quality, our approach has an additional important advantage compared to the approach of Eynard et al.: Their method requires the user to specify the output colour of some pixels explicitly. This is pivotal for reasonable results. In contrast, our output-driven approach requires no such user interaction, and is fully automatic.

To underline those findings, we conduct a detailed comparison to the approach of Eynard et al. [EKB14] on further image sets. To this end, we use the public database provided by Brown and Süsstrunk [BS11], which contains RGB and NIR images for different real-world sceneries. Figure 4.7 depicts for three of those image sets the original RGB input images (Figure 4.7(a)), the results of Eynard et al. (Figure 4.7(b)), and our results (Figure 4.7(c)). Since the method of Eynard et al. requires user interaction by an expert, we compare to resulting images that are kindly provided by the authors. First, the experiments show that both approaches add important structures to the final images that have not been visible in the input RGB images. However, in this respect, the zooms illustrate that our results are richer in details. Second, for the results of Eynard et al. one can observe undesirable colour casts compared to the RGB images; cf. for instance the last row in Figure 4.7. This would be acceptable if the details would increase significantly in this way. However, this is not the case. Since we regard the NIR image as an additional spatial lightness information, our results do not suffer from such artefacts.

To demonstrate the overall good performance of our general fusion approach for multispectral imaging, we provide our resulting images for all 477 image sets from the mentioned database of Brown and Süsstrunk [BS11] online: `www.mia.uni-saarland.de/Research/Image_Fusion`.

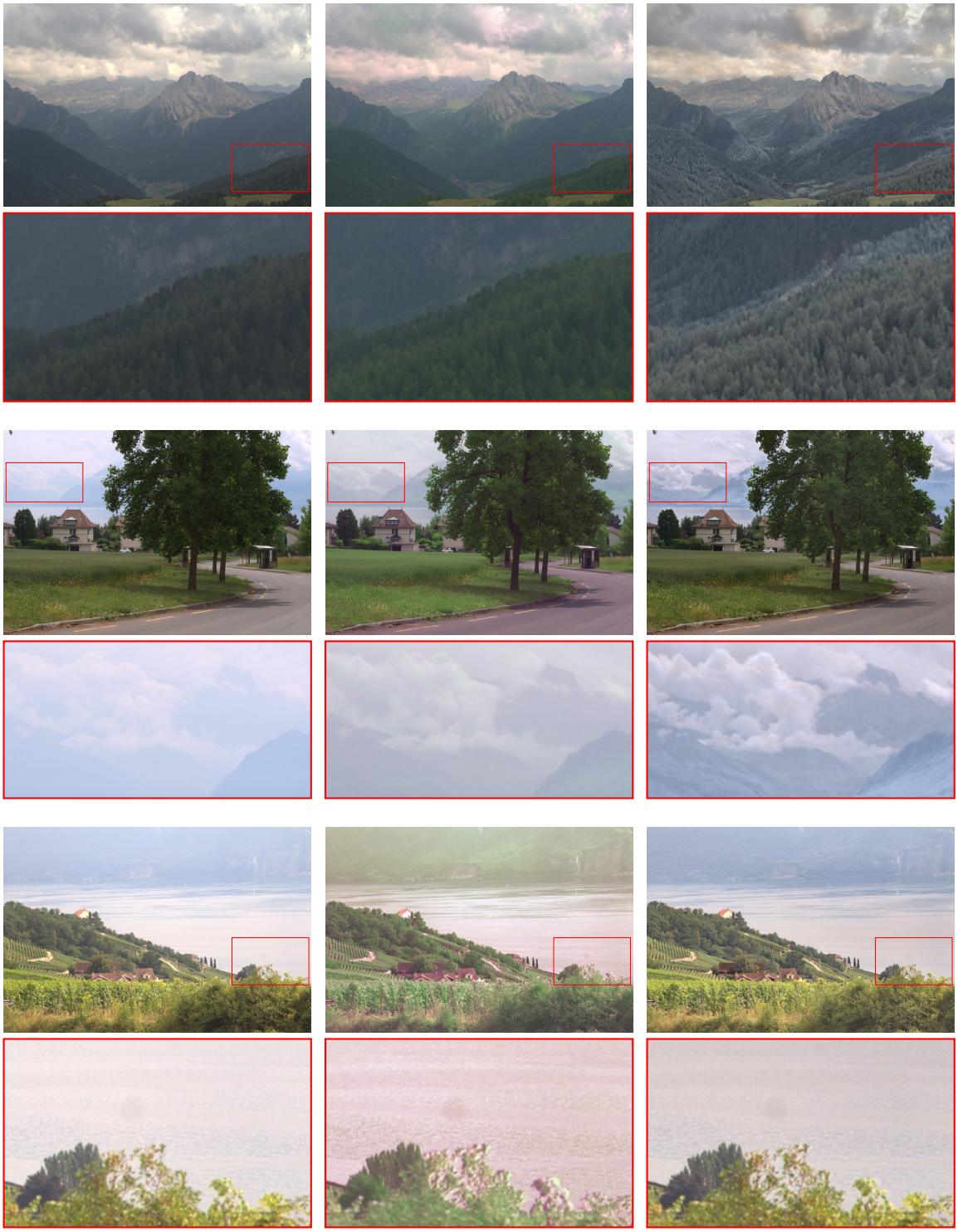
In our last multispectral experiment, we deal with more than two input images (Figure 1.3(a)–(d)). Also here, our fused image in Figure 1.3(e) demonstrates the capability of our method to incorporate structures from all spectral ranges, without leading to an unrealistic appearance.



(a) input RGB image (b) input NIR image

(c) Lau et al. [LHM11] (d) Eynard et al. [EKB14] (e) our result

Figure 4.6: Comparison of multispectral image fusion. *Top*: Input images [Nat08]. *Bottom*: Resulting images with zooms. Our fusion result features the most local details from both spectral ranges without appearing unrealistic.



(a) input RGB (b) Eynard et al. [EKB14] (c) our results

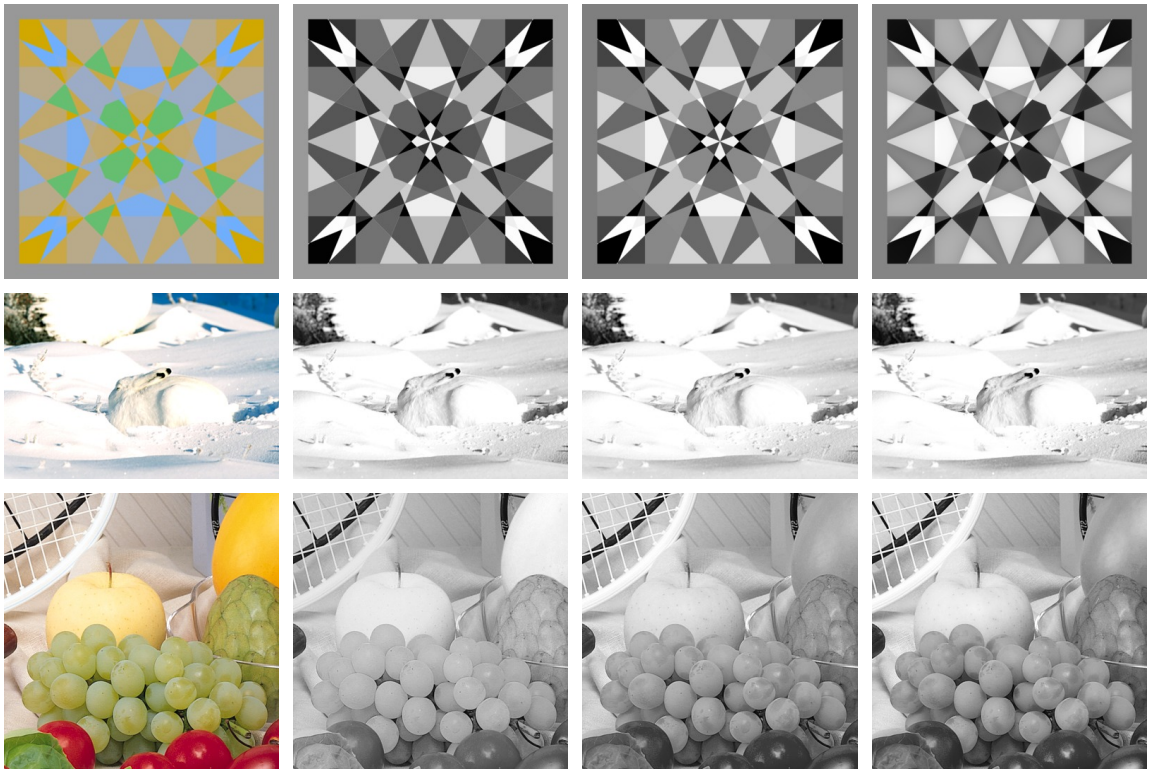
Figure 4.7: RGB and NIR fusion [BS11]. Our fused images in (c) show more details than the standard photographs in (a), and also than the resulting images of Eynard et al. [EKB14] in (b).

4.4.2 Decolourisation

We approach the decolourisation task as follows: First, we decompose the input colour image into its RGB channels, and regard them as input of our fusion algorithm. Then, we fuse all three channels to a grey-valued image that offers an optimal local contrast. Here, we use the pointwise average over all colour channels as attachment image $\tilde{f}$, and set $\delta = 0$. Figure 1.6 illustrates this procedure and depicts our corresponding decolourisation result. While a single channel is not sufficient to preserve the details from the colour image, our fused composite offers a desirable quality.

In [Čad08], Čadík presents a decolourisation benchmark that consists of 24 images, which are intended to represent various image classes. With the help of this benchmark set we compare our method to different state-of-the-art decolourisation approaches from the literature. For illustration, we depict three of those 24 benchmark images in Figure 4.8(a). The remaining images with a visual comparison to various decolourisation techniques can be found online: `www.mia.uni-saarland.de/Research/Image_Fusion`. Here, we restrict ourselves to a comparison with the best performing approaches, namely the method of Lu et al. [LXJ12] (Figure 4.8(b)) and Eynard et al. [EKB14] (Figure 4.8(c)). All decolourisation results in Figure 4.8 are of similar high quality. However, in some image regions, our results (Figure 4.8(d)) tend to preserve more details from the colour image than the other ones; cf. also our upcoming quantitative comparison.

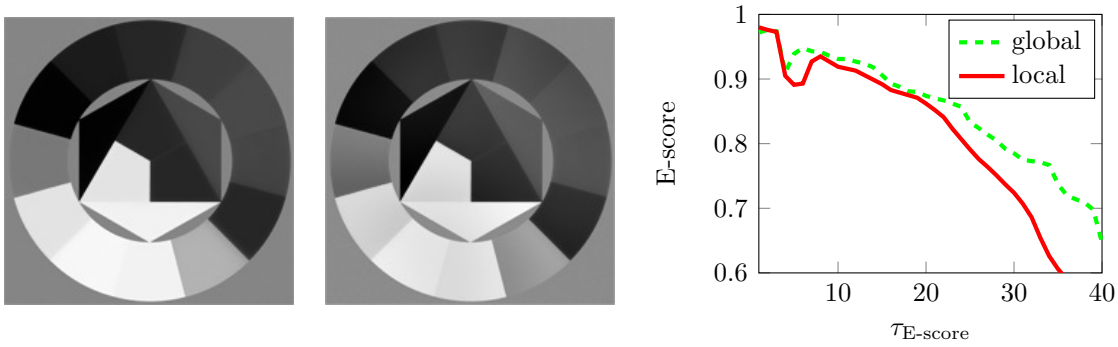
Lu et al. [LXJ14] argue that the decolourisation benchmark of Čadík [Čad08] is biased to synthetic images. Hence, they propose a new benchmark dataset that consists of 250 colour images. Additionally, they introduce the so-called *E-score* quality measure to quantitatively judge different decolourisation results. Definitely, this is an important step to provide objective quality metrics and in this way, to trigger further research in the area of decolourisation. However, as it turns out, this measure unfortunately does not take into account the local contrast adaptation of the human visual system. Rather, it even penalises effects such as a Cornsweet illusion with high errors. This is illustrated in Figure 4.9: Although Figure 4.9(b) offers a better visual quality than Figure 4.9(a), it receives a significantly worse E-score measure. Here, a higher E-score means a better quality, and $\tau_{\text{E-score}}$ defines a colour difference that is not perceivable by the human visual system; cf. [LXJ14] for further details. This leads us to apply our method with a global contrast term for the conducted E-score evaluation. The corresponding E-score graphs for the benchmarks of Čadík [Čad08] and Lu et al. [LXJ14] are depicted in Figure 4.10. Considering the benchmark of Čadík (Figure 4.10(a)), we obtain the best quality measures for almost all $\tau_{\text{E-score}}$ values. Similar observations hold for the decolourisation benchmark of Lu et al. [LXJ14] (Figure 4.10(b)). Here, we outperform the competing methods for large $\tau_{\text{E-score}}$ values. First, this shows that our general image fusion approach performs well for decolourisation. It produces greyscale images that are of similar or even superior quality than state-of-the-art approaches, which are especially tailored to the decolourisation task. Second, as Figure 4.9 illustrates, our results could be even more improved with a local contrast term. In this regard, we believe that modifications and extensions of



(a) input (b) Lu et al. (c) Eynard et al. (d) our results

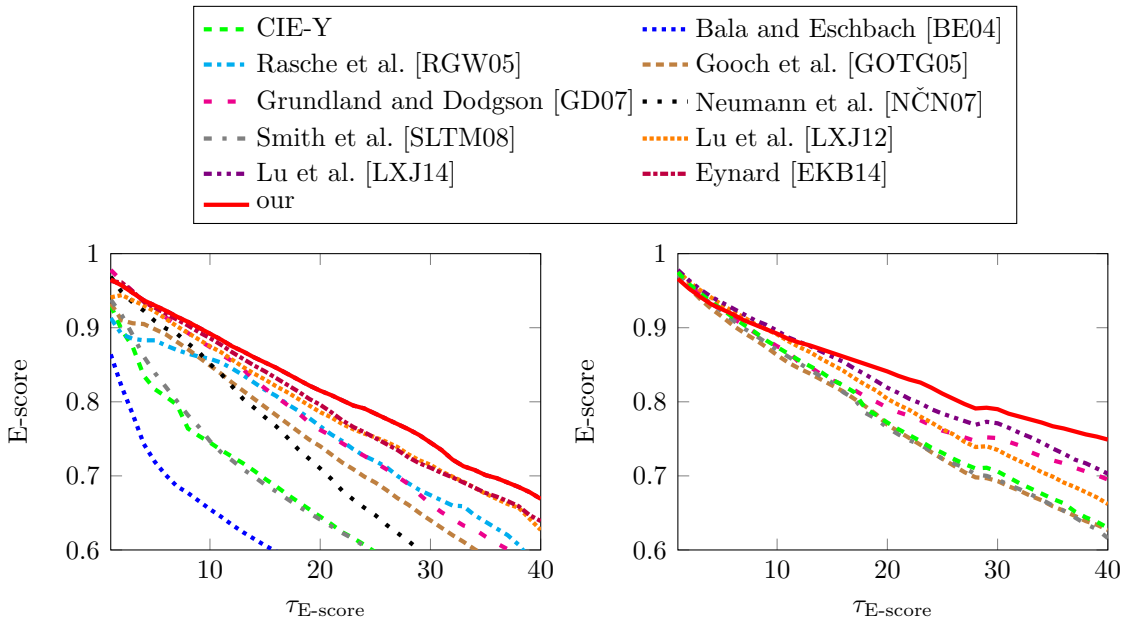
Figure 4.8: Results for the decolourisation benchmark of [Čad08]. *From left to right*: (a) Input colour images, (b) results of Lu et al. [LXJ12], (c) results of Eynard et al. [EKB14], and our results.

existing decolourisation quality measures are highly desirable in order to account for such properties of the visual system. This is an important goal for future research.



(a) global (b) local (c) E-score

Figure 4.9: Effect of (a) global and (b) local contrast term on the E-score measure of Lu et al. [LXJ14] in (c). The higher the E-score value, the better. We see that the E-score measure does unfortunately not account for the local contrast adaptation of the visual system. It even penalises it with much smaller E-score values.



(a) Čadík [Čad08] (b) Lu et al. [LXJ14]

Figure 4.10: E-score evaluation [LXJ14] for decolourisation benchmarks of (a) Čadík [Čad08] and (b) Lu et al. [LXJ14]. The higher the E-score value, the better.

4.4.3 Exposure Fusion

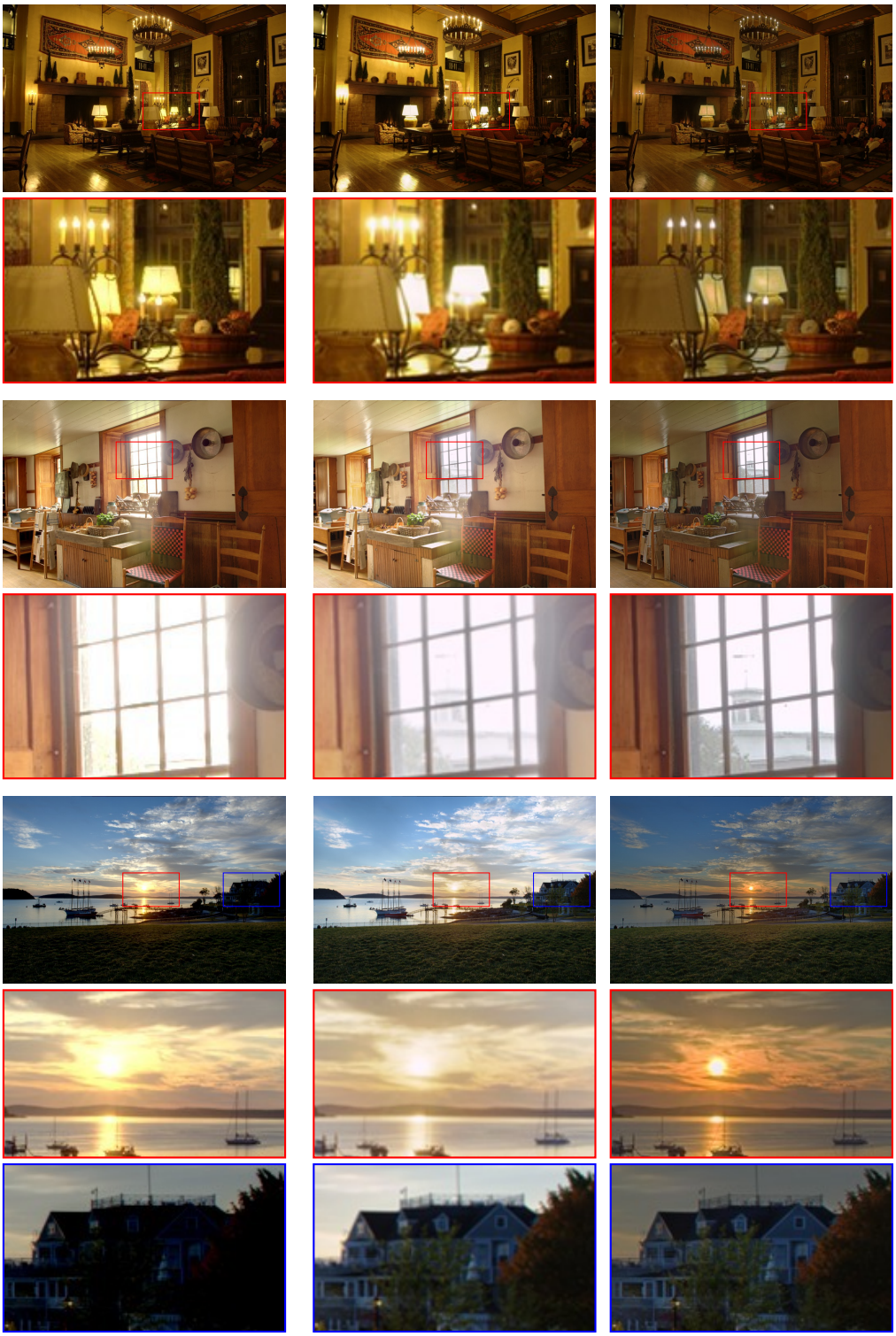
The task of exposure fusion is to combine differently exposed images to a single composite that is well-exposed and offers an optimal local contrast. We show an example result of our method for exposure fusion in Figure 1.2. Here, our fused image contains details from all input images, is well-exposed, and features a good local contrast.

To thoroughly evaluate our general image fusion method for the task of exposure fusion, we conduct an extensive comparison to state-of-the-art approaches. In particular, we first consider the popular exposure fusion method of Mertens et al. [MKV09]. It builds the basis of many exposure fusion approaches and still provides top results compared to more recent techniques. Additionally, we compare to the recent method of Singh et al. [SKB14]. This exposure fusion method provides high quality results and compares favourably to other state-of-the-art techniques; cf. [SKB14]. As mentioned in Section 4.2.3, both methods precompute weights based on the input images, and combine them later on. We will show that our idea, which directly aims for an optimal output image, is able to outperform those input-based approaches. In cooperation with the authors of [SKB14], we selected twelve representative exposure sets from the HDR photographic survey of Fairchild [Fai07]. Each set consists of nine differently exposed LDR images. For illustration, we depict in Figure 4.11 results for three of those exposure sets. Especially in the zooms in Figure 4.11, the higher amount of local contrast offered by our approach is obvious, both in dark and bright image regions. Moreover, our images do not suffer from halos or blurring effects that can be observed for the method of Singh et al. [SKB14]; cf. Figure 4.11(b,c). A comparison for the other exposure sets can be found online: `www.mia.uni-saarland.de/Research/Image_Fusion`. There, we additionally include our fusion results for all 105 images sets provided by Fairchild to demonstrate the generally good performance of our method for exposure fusion.

Unfortunately, no established quality measures exist to objectively evaluate exposure fusion results based on the given input LDR images. However, in a related context, Aydin et al. [AMMS08] introduced the so-called *dynamic range independent metric* (DRIM). This metric bases on properties of the human visual system, and can be applied to compare images with different dynamic ranges. In particular, it is especially suited to compare an HDR reference image and an LDR representative of it. Hence, it is frequently applied in the context of tone mapping. To make use of this metric for evaluating our exposure fusion results, we consider publicly available HDR images [Max05a], and create sets of LDR images with different exposure times from them. To ensure that those images are representatives of real LDR images, we apply appropriate exposure times and a common camera response function. Specifically, we compute for each HDR image five LDR images separated by one exposure value. These images serve as input for the exposure fusion techniques. Finally, we apply DRIM to compare the reference HDR image with the exposure fusion results. DRIM does not output a single quality measure, but informative distortion maps that use the following colour code: Green indicates a loss of visible contrast, blue an amplification

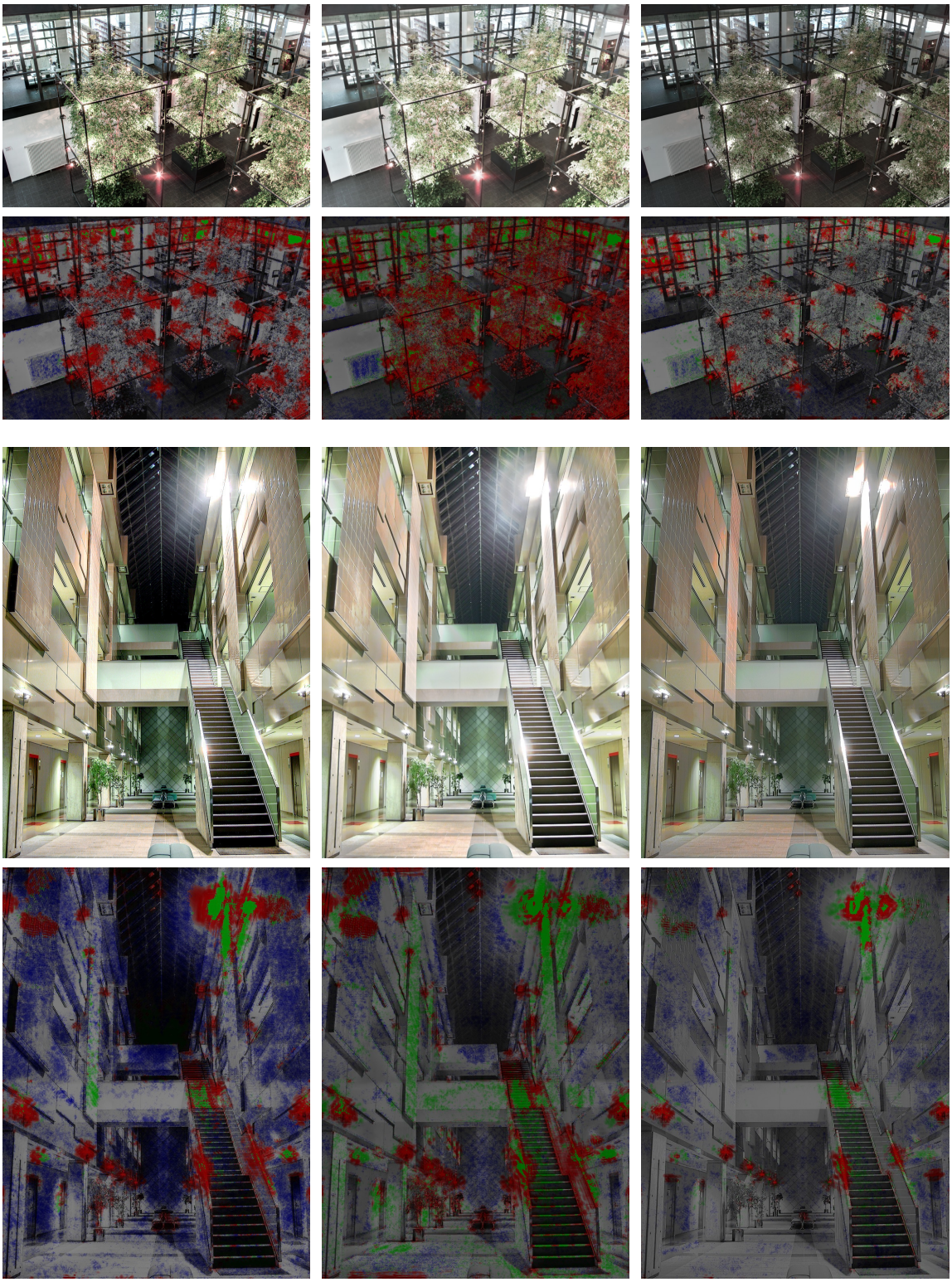
of invisible contrast, and red a reversal of visible contrast. In addition, the colour saturation is proportional to the amount of distortion; cf. [AMMS08] for further details. Figure 4.12 depicts the fusion results and the corresponding DRIM distortion maps of the method of Mertens et al. [MKV09] (Figure 4.12(a)), Singh et al. [SKB14] (Figure 4.12(b)), and our approach (Figure 4.12(c)). As the DRIM maps indicate, our results show less distortions compared to the competing approaches. This illustrates that our model assumptions, which are based on perceptually inspired concepts from variational contrast enhancement, are well-suited to create exposure fusion results with a high perceptual quality.

In our last exposure fusion experiment in Figure 4.13, we additionally compare to the method of Kotwal and Chaudhuri [KC11] since it is a related exposure fusion approach. Also here, we observe a higher local contrast provided by our approach compared to the competing methods. Furthermore, in comparison to Kotwal and Chaudhuri, our method does not suffer from a colour cast and provides a vivid colour impression. We achieve this by the proposed saturation term and the coupled handling of all colour channels. Referring to the method of Singh et al., our result does not only feature more local contrast, but is also free from an unrealistic detail enhancement; see for instance the texture at the wall in Figure 4.13(c).



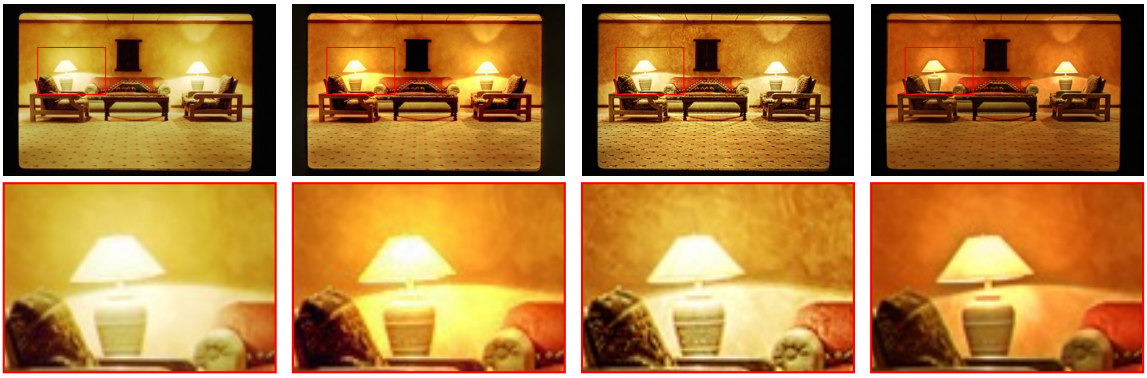
(a) Mertens et al. [MKV09] (b) Singh et al. [SKB14] (c) our results

Figure 4.11: Exposure fusion results (with zooms) for LDR image sets provided by Fairchild [Fai07]. High resolution images and results for various further exposure sets can be found online: `www.mia.uni-saarland.de/Research/Image_Fusion`.



(a) Mertens et al. [MKV09] (b) Singh et al. [SKB14] (c) our results

Figure 4.12: DRIM-based exposure fusion comparison. *Odd rows*: Resulting images. *Even rows*: Corresponding DRIM distortion maps. The more colours, the more visible distortions.



(a) [KC11] (b) [MKV09] (c) [SKB14] (d) our result

Figure 4.13: Exposure fusion results of (a) Kotwal and Chaudhuri [KC11], (b) Mertens et al. [MKV09], (c) Singh et al. [SKB14], and (d) our result. Input exposure stack is available at [Col99]. Particularly at the lamps, the better local contrast provided by our method is obvious.

4.4.4 Further Image Fusion Applications

We have seen that our output-driven approach leads to a versatile image fusion method with a broad range of applications. To further demonstrate its general applicability, let us illustrate its potential for two additional image fusion tasks.

Multilight Image Collection

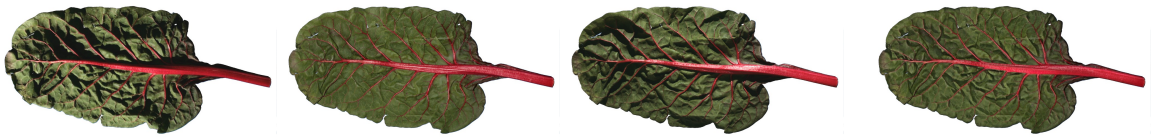
In contrast to the classical two-stage pipeline of HDR imaging and tone mapping, our approach is completely independent of the knowledge of the exposure times and the camera response function. On top of that, also images that do not follow the HDR imaging model, e.g. images captured with different camera settings or under non-constant lighting conditions, can be included easily into the input stack. We illustrate this by fusing a flash and no-flash image set in Figure 4.14. Additionally, we show an example for the fusion of images captured with changing illumination in Figure 4.15.

Focus Fusion

As discussed in Chapter 1, the goal of focus fusion is to combine a stack of images with different focal settings to a single composite that is desirably sharp everywhere. In Figure 4.16(a,b), we depict two input images from such a focal stack. The first one is sharp in the back, but blurred in the front; the second one vice versa. The computed weight maps in Figure 4.16(c,d) illustrate that our method works as expected. It fuses the images in such a way that the result has a high local contrast, and is desirably sharp; cf. Figure 4.16(e). Here, we set the attachment image $\tilde{f}$ to the average of both input images. Moreover, since the dispersion term is obviously not that important for focus fusion as for the other applications, we weight it down by increasing the weights for the contrast and regularisation term, i.e. we set $\gamma = 2.5$ and $\alpha = 10$.

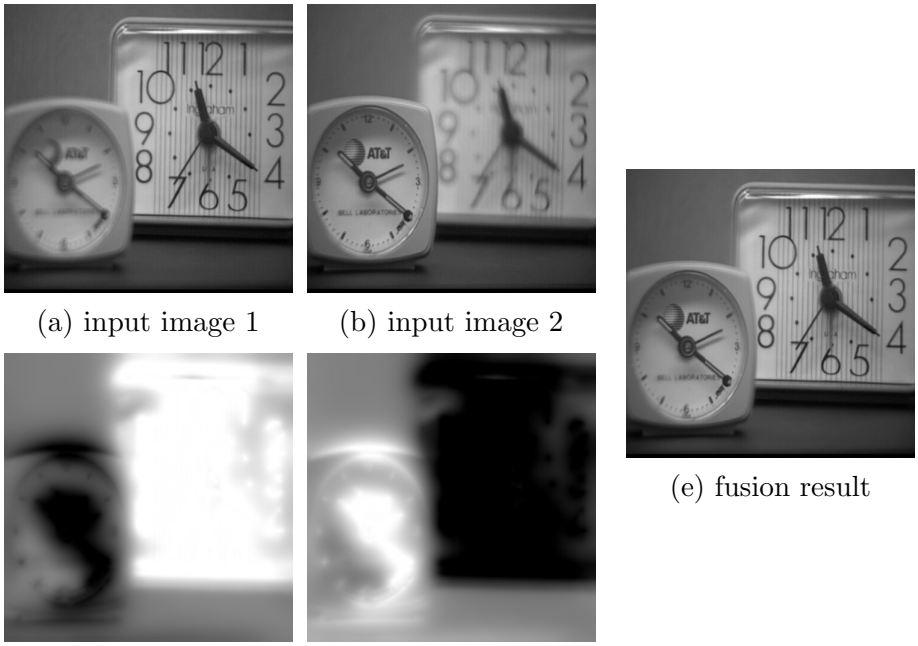
(a) ambient image (b) flash image (c) fusion result

Figure 4.14: Fusion of flash and no-flash photographs.



(a) lighting 1 (b) lighting 2 (c) lighting 3 (d) fusion result

Figure 4.15: Fusion of images captured under varying lighting conditions. (a–c) Multilight image collection [FAR07], and (d) our fused result. It contains details from all input images where the amount of details can be steered with the contrast parameters γ and σ.



(a) input image 1 (b) input image 2

(e) fusion result

(c) weight map 1 (d) weight map 2

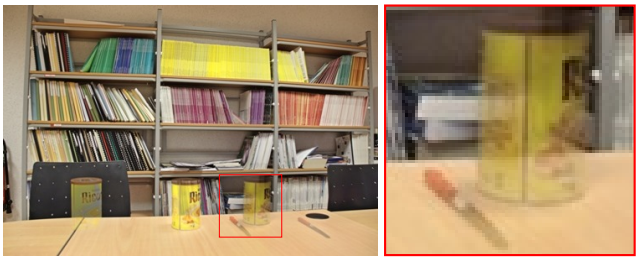
Figure 4.16: Focus fusion. (a,b) Differently focused input images [BO10]. (c,d) Our computed weight maps and (e) corresponding fusion result illustrate that our method picks meaningful in-focus regions.

4.5 Limitations and Discussion

Our model is tailored to static scenes, i.e. the input images have to be aligned properly. Obviously, this is no problem for decolourisation. However, especially exposure series may suffer from either camera shakes or object motion during the acquisition of the image set. Fusing unregistered images might result in so-called *ghosting* artefacts in the fused composite; cf. Figure 4.17. We thoroughly examine this *deghosting* task in Chapter 5. There, we additionally explain how to make use of supplementary information about the image stack. In this regard, knowledge about the image acquisition process allows us to reconstruct physically plausible HDR irradiances from an exposure series. The general fusion scheme that we presented in this chapter does not consider such additional information. On the other hand, this increases its applicability if this information is not available. Similarly, knowledge about the image set is also highly beneficial in the context of focus fusion. In particular, knowing the relative distances of the focal planes allows for inferring depth information; cf. Chapter 6.

At this point, we also want to comment on the numerical minimisation of the presented energy functional; cf. Section 4.3.2. Depending on the parameter setting our energy functional is nonconvex. Thus, the solution depends on the initialisation. In all our experiments, setting $w_i = 1/n_f$, where n_f denotes the number of input images, yields good results. Moreover, it is hard to establish theoretical limits for the allowed time step size τ of our iterative scheme. Even in the case of variational contrast enhancement of a single image only very conservative results are known in this regard; see e.g. [BCPR07, PPBC09]. Considering the fact that the resulting differential equation can be interpreted as a nonlocal variant of backward diffusion underlines the difficulty of finding theoretically justified time step size restrictions. However, in all our experiments we could easily determine an appropriate time step size τ in an empiric way such that the resulting algorithm is stable and fast. Here, we want to note that the contrast parameter γ has the greatest influence on the restriction of the time step size.

(a) short exposure (b) middle exposure (c) long exposure



(d) fused composite

Figure 4.17: Moving objects may lead to ghosting artefacts in the fused composite. (a–c) Input images from [SS12]. (d) Fused image with zoom into a ghost region.

4.6 Summary and Conclusions

In this chapter, we have presented a general variational method for image fusion. The main difference to previous research is that we intentionally refrain from precomputing application-specific weights based on the input images and combining the images with those predetermined weights later on. Instead, we model an energy functional that directly opts for an optimal composite image. This output-driven idea is the key concept why our method works that well in various application areas. We demonstrate this by means of thorough evaluations in all three main application areas, namely multispectral imaging, decolourisation, and exposure fusion. Here, we compare to state-of-the-art and well-performing methods in each field. As it turns out, our general approach produces results of similar high quality, and even outperforms competing methods for various example images. This shows the generally good performance and versatility of our technique. On top of that, all components of our variational model have a plain meaning and allow a direct manipulation of desired properties of the output. In this regard, all parameters can be fixed or determined automatically in a straightforward way. This allows an intuitive use, even by non-experts. This is an important feature to easily adapt the fusion results to the personal preferences, display and printer settings. To conclude, we believe that our technique is generally suitable for applications that require the fusion of multiple images. It provides a composite that condenses the most important information from the input images in an adequate way, and offers important properties such a local contrast adaptation.

To further improve the fusion results, it might be beneficial to adapt the weighting function in our contrast term in an image-dependant way. This would allow to steer

the contrast by means of the spatial and tonal differences in a bilateral sense; cf. for instance [AW95, TM98, CWS15]. Moreover, from a more general point of view, in this work we have considered a variational formulation for histogram *equalisation* (cf. [SC97]) to design our general image fusion method. In a similar way, it would be highly interesting to consider variational histogram *matching* to propose novel methods for further applications such as the colour transfer between two images; see e.g. [PPC11, BL13].

Chapter 5

Deghosting and HDR Imaging

"I ain't afraid of no ghost."

Ray Parker, Jr.

pixabay.com

Contents

5.1 Introduction 75
5.2 Related Work 76
5.3 Alignment with Complete Rank Transform 81
5.4 Ghost Removal with Brightness Transfer Function . . . 103
5.5 Simultaneous HDR and Optic Flow Computation 113
5.6 Summary and Conclusions 126
5.A Finite Differences for Anisotropic Diffusion 129

Main parts of this chapter base on our work published in [DHW13, HDW14, DHW15] and the master's thesis of Maria Luschkova [Lus13].

5.1 Introduction

As illustrated in Section 4.5, already small misalignments in the exposure set may lead to undesirable artefacts in the fusion result. However, in a practical setting we cannot assume perfectly aligned input images. In fact, camera shakes and moving objects during the acquisition of the exposure stack are omnipresent. Even when using a tripod, already the triggering of the camera shutter may cause significant misalignments. Additionally, completely static scenes without any moving objects such as moving clouds or waving leafs are seldom. Hence, an appropriate processing

step to handle such motion within an exposure series is often indispensable. To this end, we present two different strategies in this chapter. First, we consider the following two-step approach: In a first step, we compute displacement maps between a chosen reference image of the exposure set and all other frames. To this end, we base on successful variational optic flow approaches and adapt them to the difficult exposure series setting, where common constancy assumptions such as the brightness constancy assumption do not hold. Subsequently, the computed displacement maps are used to align all input images. In a second step, we remove remaining misalignments by a so-called *ghost removal* technique that identifies regions of the input images which are prone to produce artefacts in the fused composite image.

Alternatively, provided further information about the exposure set, we present a conceptually different approach. Since knowledge about the *camera response function* (CRF) and the exposure times allows for inferring *high dynamic range* (HDR) information, we propose to simultaneously compute an aligned HDR image and dense displacement fields for all input images. We will see that both tasks influence each other, and thus highly profit form the induced mutual feedback provided by our joint approach.

Chapter Outline. In Section 5.2, we give a broad overview of related work in exposure series alignment and ghost removal. We deepen this discussion in the individual sections, where we explicitly explain the relation to previous work. The core part of this chapter consists of three main sections: In Section 5.3, we discuss our variational framework for displacement estimation. Here, we focus on the general case of differently exposed images, and propose our *complete rank transform* (CRT) for this purpose. Subsequently, in Section 5.4 we discuss our ghost removal approach for identifying and eliminating remaining misalignments. Next, we present our simultaneous HDR and optic flow computation in Section 5.5. We conclude this chapter with a summary and outlook in Section 5.6. Additionally, we give details of our finite difference discretisation in the Appendix 5.A of this chapter.

5.2 Related Work

Previous work on deghosting can be classified into two main classes. First, we consider approaches that try to match all images of the exposure set, and in this way try to make optimal use of the available information. Accordingly, we call them *image matching* methods. Second, we discuss so-called *ghost removal* techniques that already assume an adequate alignment of the images, and aim at removing remaining misalignments. Here, the main goal is a result free of artefacts at the price of discarding valuable information. Certainly, we also cover some hybrid approaches that may fit into both classes. Recent surveys of deghosting approaches in the context of HDR imaging can be found in [SS12] and [TAEE15].

5.2.1 Image Matching

Let us first review methods that aim at aligning the input images of the exposure stack. In this context, we again differentiate two types of methods: First, we discuss approaches that are suited for pure camera motion. Subsequently, we consider techniques that additionally cope with local scene motion, i.e. moving objects.

Camera Motion

The techniques in this class assume a global motion pattern between two input images that is mainly caused by a movement of the camera. Here, the methods differ in the considered motion models. In particular, several methods model a simple global translation between the different frames [War03, KHC+06, Aky11]. Grosch [Gro06] and Jacobs et al. [JLW08] extend the method in [War03] and additionally account for global rotations. Similarly, Im et al. [ILLP11] determine an affine transformation between an image pair by minimising a quadratic error function. Alternatively, the techniques in [CH06, RMVS07, GG07, Yao11] determine a global translation and rotation by means of the Fourier transform. Moreover, several approaches relate the input images by so-called *homographies* which turn out to be well-suited for realistic camera movements; see e.g. [MMF03, Can03, EUS06, TM07, HLL+10, GKTT13, OLTK15].

Besides the considered motion models, the methods differ in the actual matching assumptions. Ward [War03] introduces the so-called *median threshold bitmaps* (MTBs) that are invariant under monotonically increasing grey value rescalings, and thus well-suited for exposure series. Similarly, Akyüz [Aky11] proposes so-called *correlation maps* that are robust w.r.t. varying exposure times. Interestingly, both approaches share similarities to our illumination-robust descriptors that we consider in Section 5.3. Further descriptor-based methods applied in such a setting are based on the *scale-invariant feature transform* (SIFT) [Low04, EUS06, TM07, HLL+10], *speeded-up robust features* (SURF) [BETV08, GKTT13], or the *contrast invariant feature transform* CIFT [GG07]. Contrary to considering descriptors that are robust against brightness changes, Mann et al. [MMF03] and Candocia [Can03] present a joint registration of the input images and photometric alignment. Here, photometric alignment or also photometric calibration describes the process of relating the intensity values between the different frames; cf. also Section 5.4.

Camera and Object Motion

The approaches discussed so far assume a *global* transformation between an input image pair. This may be suited for pure camera movement. However, any moving scene object during the acquisition violates this assumption. Hence, the following methods try to *locally* match all parts of the input images under arbitrary scene motion.

Optic Flow-Based Alignment. Bogoni [Bog00] and Kang et al. [KUWS03] first compute a global transform for the whole image, and refine the resulting motion field

locally in a second step with a variant of the local optic flow approach of Lucas and Kanade [LK81]. Sand and Teller [ST04] match sparse feature points and locally refine and interpolate the flow fields afterwards. In [MG07, MG10, CPOL12] hierarchical block matching schemes are applied to locally register all input images. Moreover, there exit several energy-based techniques for computing the optic flow fields. For instance, Hossain and Gunturk [HG11] and Bouzaraa et al. [BUC15] apply histogram matching to photometrically align the intensity values of differently exposed frames; cf. also Section 5.4. Then, the resulting constancy assumption is incorporated as data term into a variational optic flow framework. In a similar context, Bengtsson et al. [BML15] investigate a temporal smoothness of the flow fields in the form of [VBVZ11], and analyse different weightings of the data terms to handle saturated image regions. Ferradans et al. [FBPC12] propose an irradiance-based data term, and embed it into the optic flow approach of Chambolle and Pock [CP11]. Comparably, Jinno and Okuda [JO12] present a Markov random field approach that also implements an irradiance-based constancy assumption. Moreover, Gupta et al. [GIN13] advocate an exposure bracketing technique based on the Fibonacci numbers to improve the optic flow computation. Lin and Chang [LC09] propose a disparity map computation in case of differently exposed stereo pairs. In particular, they estimate the response function of the camera by means of sparse SIFT correspondences, and subsequently apply it to formulate an appropriate constancy assumption. In [SEM11], Sellent et al. explicitly tackle the problem of motion blur caused by large exposure times. Zimmer et al. [ZBW11a] align the exposure stack with a variational approach, where the gradient constancy assumption is imposed. The resulting motion fields with subpixel precision are subsequently applied to construct an HDR composite with super-resolution.

To a great extent, the above mentioned approaches are especially tailored to exposure sets. For instance, some of the discussed techniques require explicit knowledge about the exposure times or the camera response function. However, there exist various optic flow approaches that are suited for the more general alignment of differently exposed input images. Indeed, there is a long tradition of designing methods for illumination-robust optic flow computation, either by introducing robust features or by modelling the illumination changes explicitly. In the following, we review alignment approaches that fit into this category.

The gradient constancy assumption renders the optic flow model robust w.r.t. global additive brightness changes. It goes back to Nagel [Nag83], Tretiak and Pastor [TP84], and Uras et al. [UGVT88], while embeddings in a variational setting have been studied e.g. by Schnörr [Sch93] and Brox et al. [BBPW04]. Constancy assumptions on higher order derivatives have been proposed by Papenberg et al. [PBB+06]. There are further gradient-based features that have shown their usefulness for optic flow computation. Famous examples are the *scale-invariant feature transform* (SIFT) [Low04, LYT11], the *histogram of oriented gradients* (HOG) [DT05, RMG+13], or *local directional patterns* (LDP) [JKC10, MRM+14]. Also the gradient orientation is a suitable constancy assumption in case of drastic illumination changes; see e.g. Haber and Modersitzki [HM07]. Another illumination-robust feature that has found great attention

in recent years is the so-called *census transform* by Zabih and Woodfill [ZW94]. The resulting census signatures offer a strong invariance against monotonically increasing brightness changes. Interestingly, also those census features share a strong relation to the gradient constancy assumption as discussed in [HDW13] and [HDWR15]. Stein [Ste04] uses the census signatures in an efficient feature matching approach, where a hash table-based indexing scheme provides flow estimates in real-time and is well-suited for large displacements. Müller et al. [MRRF11] as well as Mohamed and Mertsching [MM12] exploit those sparse feature matches to handle large displacements and to recover image details lost in a coarse-to-fine minimisation technique. Furthermore, Müller et al. [MRR+11] embed the census transform as data term into a variational optic flow framework. Tests in real-world scenarios demonstrate the desired robustness of the resulting dense flow fields w.r.t. illumination changes. Vogel et al. [VRS13] compare different data terms and show that the census transform is well-suited for challenging lighting conditions. Puxbaum and Ambrosch [PA10] apply the modified census transform of Fröba and Ernst [FE04] to the image gradients and use it for optic flow estimation. Braux-Zin et al. [BZDB13] combine the brightness and census-based constancy assumption in a variational framework. In [RBP14], Ranftl et al. propose an extension of the census transform to increase the robustness to scale changes. Also in the context of stereo estimation, Mei et al. [MSZ+11] and Ranftl et al. [RGPB12] have illustrated the usefulness of the census transform. In Section 5.3, we will present variants of the census and the related rank transform of Zabih and Woodfill [ZW94], and embed it into our variational optic flow approach.

Moreover, Steinbrücker et al. [SPC09], Molnár and Chetverikov [MCF10], and Werlberger et al. [WPB10] achieve an invariance against multiplicative illumination changes via patch-based data terms using normalised cross correlation. In this context, Drulea and Nedevschi [DN13] introduce the so-called *correlation transform* for motion estimation. Wedel et al. [WPZ+09] perform a structure-texture decomposition of the input images by means of the image denoising method of Rudin et al. [ROF92]. Here, the textural part shows an increased robustness under shadows and shading reflections. In presence of colour imagery, robustness w.r.t. illumination changes can also be achieved by exploiting photometric invariants [GB97, vG04, MBW07] or by switching to other colour spaces as performed in [ZBW11b]. Also the mutual information [VW97] turns out to be a useful feature for registering images with different illumination; see e.g. Hermosillo et al. [HCF02] and Panin [Pan12]. In [XJM10] and [KLL13] robust data terms are incorporated in an adaptive way by switching locally between different constancy assumptions.

Instead of matching illumination-robust features, the following approaches follow a different idea to tackle brightness changes: They model them explicitly. Based on *comparagrams* [Man00], Kim and Pollefeys [KP04] as well as Dederscheck et al. [DMM12] estimate the *brightness transfer function* (BTF) that allows to compensate for global brightness changes. In order to handle also local changes, the model of Cornelius and Kanade [CK83] allows smooth additive variations from the brightness constancy assumption. Further approaches that model such additive changes

are presented in [Shu89, Muk90, CP11, ARS12]. Gennert and Negahdaripour [GN87] extend this idea and jointly optimise for spatially varying multiplicative and additive illumination changes in addition to the optic flow field. Also the methods in [Neg98, HSGL11, FDW12] apply such an affine illumination model. Hager and Belhumeur [HB96] apply a principal component analysis to set up the basis of an illumination model whose parameters are jointly computed with the motion. Similarly, Demetz et al. [DSV+14] learn a set of basis functions and jointly recover locally varying illumination parameters and a dense flow field. The methods in [BFY00, HF01] propose to estimate physically based illumination parameters to support the motion estimation. Kim and Kak [KK06] compare different local and global approaches under brightness changes and in particular improve the robustness against noise.

Patch Matching Techniques. If the actual displacement maps (optic flow fields) are of rather less importance and the main goal is a nice looking composite image, there exists an interesting alternative to the discussed optic flow-based alignment strategies. Instead of computing explicit displacement maps and subsequently aligning the input images, several approaches apply so-called *patch matching* techniques. Here, the basic idea is to choose a reference image and to find for every image region of this reference image corresponding patches in the other images of the exposure stack. In this context, Zheng et al. [ZLZ+13] propose a hybrid approach, where they search for corresponding pixels by means of the brightness transfer function if the intensity value of the reference pixel is reliable. Otherwise, a block-based template matching is performed. Also Hu et al. [HGP12] make use of an advanced patch matching algorithm to find corresponding image regions. Additionally, this method contains a ghost removal step after the registration to handle wrong correspondences. In their follow-up work [HGPS13], the authors explicitly improve the handling of saturated regions. In an HDR setting, Sen et al. [SKY+12] jointly align the input images and estimate an HDR composite image. In [KSB+13], this work is extended to an HDR video setting.

5.2.2 Ghost Removal

Assuming the images to be sufficiently aligned, the task of so-called *ghost removal* methods is to identify and eliminate remaining motion artefacts that may show up as ghosts in the final composite images; cf. Figure 4.17. In recent years, a variety of ghost removal approaches has been proposed. In particular, they differ in their underlying assumptions about the exposure set. For instance, some require knowledge of the exposure times and the camera response function, while others do not make specific assumptions about the imaging process. While the former one might be useful in an HDR imaging context, the later one has a wider area of application; cf. exposure fusion. Detailed surveys of recent ghost removal literature can be found in [SSM12] and [TAEE15]. Here, we give a broad overview and discuss the main concepts.

One idea for detecting ghosts in an exposure sequence is to predict the intensity values of an LDR image with the help of the *brightness transfer function* (BTF) that transfers the brightness values from one frame to another. If the predicted

and the measured pixel values differ more than a predefined threshold, the pixels are indicated as ghosts. While some methods assume a linear *camera response function* (CRF) [KHC+06, GGC+09, SSM12, TAEE15], others explicitly estimate the CRF/BTF that is applied for the brightness transfer. In particular, several approaches determine the BTF by means of pixels that are static along the exposure set. To detect such static pixels, Wu et al. [WXRL10] consider the image gradients while Raman [RC11] apply a variance-based measure. Alternatively, Heo et al. [HLL+10] compute the BTF based on the joint probability density function between different exposures, where they iteratively discard pixels that significantly differ from the predicted values. In contrast to such a sample-based BTF estimation, Grosch [Gro06], Li et al. [LRZ+10], and Moon et al. [MTCL12] transform the brightness values by means of histogram matching. This procedure is similar to the approach of Grossberg and Nayar [GN03] who show that a CRF estimation based on the image histograms is robust w.r.t. a certain amount of camera and scene motion. Wang et al. [WT13] perform a brightness normalisation by an additive offset and compute the prediction difference by means of the lightness channel of the CIE-Lab colour space (cf. also Section 2.2.2). Moreover, several methods make use of the observation that the pixel value should increase if the exposure time increases. Violations of this consistency check are an indication of a moving object; see e.g. [SPS09, WXRL10, ALKC11].

Another approach for ghost detection is to compare features that are robust under varying exposure times. Example features are local entropy measures [JLW08], coarsely quantised intensity values [MPC09], the median threshold bitmaps [War03, PK10], the rank of the intensity values [LPC11], zero mean cross correlation [ALKC11, SPLC13], or the gradient direction [ZC12a]. In all methods, a significant change of the considered feature indicates a ghost. Furthermore, Reinhard et al. [RHD+10] examine the variance of a pixel value over the exposure stack to identify ghosts, and Granados et al. [GKTT13] incorporate a camera noise model for ghost detection. At this point, we also want to mention that several methods impose specific types of spatial smoothness on the computed ghost maps. Examples are applications of morphological operations in [PK10, SPLC13] or the explicit modelling of a regularisation term such as in [HLL+10].

Alternative approaches for ghost removal were proposed by Granados et al. [GSL08] as well as Silk and Lang [SL12]. Instead of explicitly detecting ghosts, they aim at identifying a static background that is later used for the HDR reconstruction. Two techniques that also do not require an explicit detection of moving objects are the methods by Khan et al. [KAR06] and by Pedone and Heikkilä [PH08]. In their HDR reconstruction, they iteratively reduce the influence of pixels with a high probability of being nonstatic.

5.3 Alignment with Complete Rank Transform

As we have seen, a popular approach to render computer vision algorithms robust against illumination changes is to incorporate features that are invariant w.r.t. the

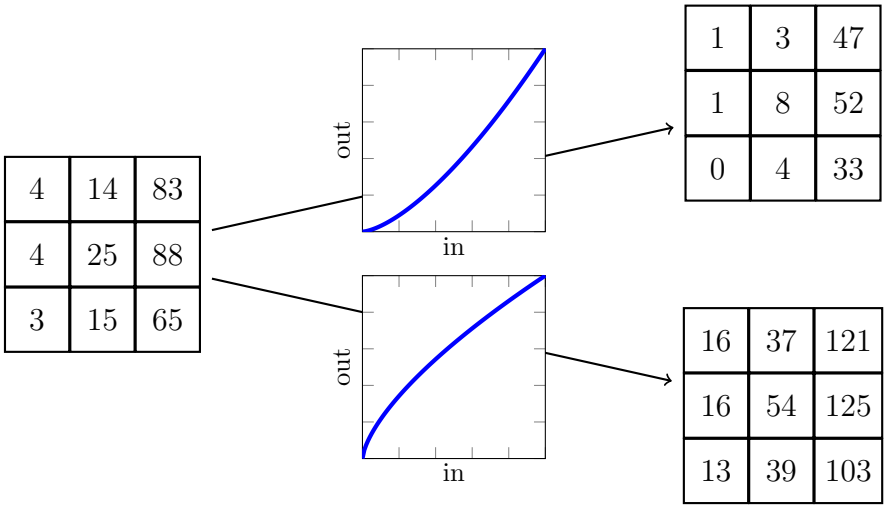


Figure 5.1: Monotonically increasing grey value rescalings. *From left to right*: Input intensity patch, monotonically increasing rescalings, resulting rescaled intensity patches. Values are rounded to the nearest integers. Note that the intensity order within the patch remains unchanged. We exploit this observation to design our illumination-robust descriptors.

occurring brightness variations. In case of exposure series we have to expect strong illumination changes. In fact, without further knowledge about the exposure set, we can only assume a general global monotonically increasing transformation of the brightness values from frame to frame; cf. Figure 5.1. Evidently, a monotonically increasing rescaling does not alter the ordering of the intensity values.[1] Hence, it seems intuitive to consider descriptors that are based on this grey value order to design illumination-robust techniques. In accordance to [AGLM93], we call such features that are invariant under monotonically increasing rescalings *morphologically invariant*. A prominent representative of such illumination-robust descriptors is the rank transform of Zabih and Woodfill [ZW94]. It computes the rank of a pixel's intensity within a local neighbourhood. Indeed, this transform is invariant against any (global) monotonically increasing intensity changes. However, it is also clear that only storing the rank of the pixel means to discard all other local information. In the same paper [ZW94], the census transform is proposed. It compares a pixel with all its neighbours and stores which one is larger. In this way (besides encoding the rank in a different form), also some spatial information is stored. However, also here a lot of information is discarded. Thus, it would be highly desirable to develop a feature that is on the one hand morphologically invariant and on the other hand discards as little information as possible.

[1] To a certain extent, quantisation noise and a clamping of the intensity values in saturated image regions might violate this assumption in practice. However, as we will see in our upcoming experiments (cf. Section 5.3.5), our proposed optic flow approach is robust against such effects and nevertheless provides good results.

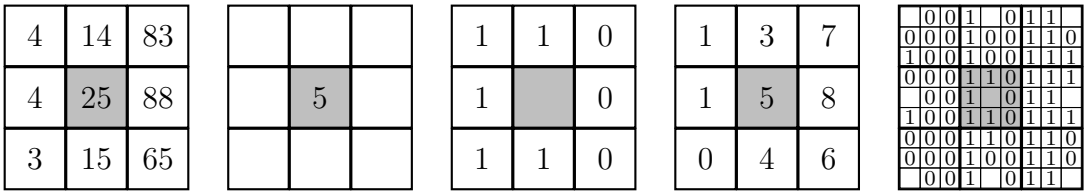
Main Contributions. In this section, we propose descriptors that exhibit the same strong morphological invariance as the rank, but carry significantly more information. In fact, our general idea is to extend the concept of the rank transform in a beneficial way: Instead of solely considering the rank of the central pixel, we compute the ranks of all pixels in its neighbourhood. In this way, we store the *complete* local intensity order. We claim that our resulting *complete rank transform* (CRT) can be used as a generally superior alternative to the census transform: It is equally computationally efficient, as parameter-free as the census transform, and leads to clearly improved results. We embed our feature into a variational energy functional for optic flow computation. In this way, we demonstrate the benefits of the proposed descriptor w.r.t. existing approaches. Moreover, we illustrate the overall performance of our technique for the alignment of exposure sets. Since we assume the exposure times and the camera response function to be unknown, our method is applicable for general alignment tasks that exhibit severe illumination changes; i.e. our method is not limited to the HDR reconstruction or exposure fusion setting.

Section Outline. We start with a discussion of related illumination-robust descriptors in Section 5.3.1. In Section 5.3.2, we explain the rank and census transform, and introduce their complete counterparts. Next, we embed them into our variational framework for optic flow computation in Section 5.3.3. We describe its numerical minimisation in Section 5.3.4. In Section 5.3.5, we evaluate our model by means of experiments, and discuss possible limitations in Section 5.3.6.

5.3.1 Relation to Previous Work

There are many other transforms that are related to our idea: Independently of Zabih and Woodfill's rank and census transform [ZW94], Pietikäinen et al. performed broad research on various kinds of so-called *local binary patterns*; see e.g. the comparative study in [OPH96] or the book [PHZA11] and references therein. However, the majority of these local binary patterns discard rather more than less available information (cf. e.g. [CKZP13]), and hence go in the opposite direction of our research.

The *binary robust independent elementary features* (BRIEF) of [CLO+12] are a variation of the census transform which performs the comparisons on arbitrary pixel pairs in the neighbourhood. Stein [Ste04] applies a ternary variant of the census transform as an efficient descriptor for sparse structure matching in driver assistance systems, and Fröba and Ernst [FE04] propose the modified census transform for face recognition. Another variation of the census transform is proposed by Ranftl et al. [RBP14], who develop a scale-robust census descriptor by sampling radial stencils with different radii that depend on the locally estimated scale. The first appearance of ordinal measures of full patches goes back to the work of Bhat and Nayar [BN98] for block matching in the context of stereo computation. Related to that, several sparse interest point descriptors building on intensity order-based ideas have been proposed: With their chained circular neighbourhoods, Chan et al. [CGKC12] make a first step towards representing neighbourhood ordinal information. The *local intensity order*



(a) input (b) RT (c) CT (d) CRT (e) CCT

Figure 5.2: Illustration of the presented intensity order transforms with a 3×3 neighbourhood patch ($K = 9$), where the reference pixel is marked in grey. *From left to right*: (a) Input intensity patch, (b) *rank transform* (RT), (c) *census transform* (CT), (d) *complete rank transform* (CRT), and (e) *complete census transform* (CCT).

pattern (LIOP) descriptor of Wang et al. [WFW11] describes the intensity order of a very large neighbourhood and is specifically tailored for sparse interest point matching. A similar idea of matching order distributions is proposed by Tang et al. [TLCT09]. Moreover, Mittal and Ramesh [MR06] combine order and intensity information to increase the robustness against noise.

In Section 5.2.1, we have already discussed several successful applications of the census transform or variants of it for optic flow and stereo computation; see e.g. [Ste04, PA10,MSZ+11,MRR+11,MRRF11,RGPB12,MM12,HDW13,VRS13,BZDB13,RBP14, HSW15]. In fact, those approaches might benefit from our illumination-robust CRT descriptor that we introduce in the following sections.

5.3.2 Morphologically Invariant Descriptors

In Figure 5.2, we give an overview of the considered morphologically invariant transformations, i.e. transforms that are invariant under any global monotonically increasing rescalings of the input signal. Formally, each transform maps a local image patch to a ζ-dimensional feature vector $\boldsymbol{s}\colon \mathbb{R}^K \to \mathbb{R}^\zeta$, where we define the image patch as the K closest neighbouring pixels w.r.t. the spatial Euclidean distance. For didactic reasons, we represent in the following the patch intensity values by a K-dimensional vector $\mathbf{f} \in \mathbb{R}^K$, where the values are ordered by increasing spatial distance from the centre. Hence, the intensity of the central pixel is assigned to the first entry f_1. In case of equal distances, we always start with the right neighbour and proceed counter-clockwise. Accordingly, for the example intensity patch in Figure 5.2(a) the neighbourhood vector is given by $\mathbf{f} = (25, 88, 14, 4, 15, 83, 4, 3, 65)^\mathsf{T}$.

Rank Transform

The *rank transform* (RT) proposed by Zabih and Woodfill [ZW94] encodes for each pixel the position of its grey value in the ranking of all grey values in the neighbourhood. In other words, it describes the number of neighbours with a smaller grey value than the central one. Formally, the rank transform maps each pixel to its scalar rank

signature $s_{\mathrm{RT}} \in \{0, \ldots, K-1\}$, and can be computed as

$$s_{\mathrm{RT}}(\mathfrak{f}) \coloneqq \sum_{i=2}^{K} \chi_{(\mathfrak{f}_i < \mathfrak{f}_1)}\,, \tag{5.1}$$

where $\chi_{(z_1<z_2)}$ denotes the indicator function

$$\chi_{(z_1<z_2)} \coloneqq \begin{cases} 1 & \text{if } z_1 < z_2\,, \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

Census Transform

In the same paper [ZW94], Zabih and Woodfill introduce the *census transform* (CT). This transform has attracted a lot of attention in recent years and can be seen as an extension of the rank transform. Besides encoding the rank, it adds a spatial component by expressing the relationship between the central pixel and each of its neighbours explicitly. Specifically, one bit of information is stored for each pixel of the neighbourhood: If the neighbour is smaller than the reference pixel the bit is 1, and 0 otherwise. In the final binary signature, all bits are concatenated. While the order of this concatenation is in general arbitrary, it has to be consistent such that each bit can be uniquely associated with one neighbour.

In mathematical terms, each image patch of size K is mapped to a binary signature vector $\boldsymbol{s}_{\mathrm{CT}} \in \{0,1\}^{K-1}$. Note that we do not need to store a bit for the central pixel itself, which explains the signature length $K-1$. Consequently, we compute the $K-1$ digits of the census signature $\boldsymbol{s}_{\mathrm{CT}}$ as follows:

$$\left(\boldsymbol{s}_{\mathrm{CT}}(\mathfrak{f})\right)_i \coloneqq \chi_{(\mathfrak{f}_i < \mathfrak{f}_1)} \qquad \forall i \in \{2, \ldots, k\}\,. \tag{5.3}$$

Hence, we compare every neighbouring pixel to the central one $\mathfrak{f}_1$. Furthermore, it is interesting to note that the sum of the digits of a census signature coincides with the rank s_{RT} of that pixel. Hence, the information of the rank transform is naturally contained by a census signature. However, as discussed, the census transform adds additional valuable spatial information.

Complete Rank Transform

Although the two signatures proposed in [ZW94] exhibit the same morphological invariance, the census transform obviously encodes by construction more information than the pure rank. However, there is still some more information that can be used without losing the desired class of invariance. To this end, let us now introduce an extension of Zabih and Woodfill's basic transform: the *complete rank transform* (CRT). We will see that the resulting signature carries much more information than its predecessors.

Given the census signature $\boldsymbol{s}_{\mathrm{CT}}$ of an image patch, we know which pixels in the patch are smaller than the *central* one. However, the relationships among *all* neighbours cannot be determined by the pure census information. For instance, if two

neighbouring pixels are both smaller than the central one, it is still unclear which of the two neighbours is smallest. Hence, to also encode this valuable information, we propose the complete rank transform. Specifically, we determine the K-dimensional CRT signature $\boldsymbol{s}_{\mathrm{CRT}} \in \{0, \dots, K-1\}^K$ by computing and storing the rank of all pixels in the patch:

$$\boldsymbol{s}_{\mathrm{CRT}}(\mathfrak{f}) \;\coloneqq\; (s^1_{\mathrm{RT}}, \dots, s^K_{\mathrm{RT}})^{\mathsf{T}}\,, \tag{5.4}$$

where

$$s^i_{\mathrm{RT}} \;\coloneqq\; \sum_{\substack{j=1 \\ j \neq i}}^{K} \chi_{(\mathfrak{f}_j < \mathfrak{f}_i)}\,. \tag{5.5}$$

With this CRT signature, the whole intensity order is represented. From the viewpoint of morphological invariance, this is the *maximal* amount of information that can be extracted without leaving this class of invariance.

The explanation of the CRT signatures in Equation (5.4) is demonstrative and intuitive to understand, but also inefficient. In fact, the complexity of the discussed computation of $\boldsymbol{s}_{\mathrm{CRT}}$ is quadratic in K. However, essentially what has to be done is to sort the intensity values of the patch. To this end, we propose to apply an efficient sorting algorithm such as *Quicksort* which leads to $\mathcal{O}(K \log K)$; see e.g. [PTVF07].

Complete Census Transform

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

Instead of storing all K ranks, it stores for each pixel of the patch whether it is smaller or larger than *any other* pixel in the patch. Thus, we obtain a signature $\boldsymbol{s}_{\mathrm{CCT}} \in \{0,1\}^{K(K-1)}$ which contains all census signatures with each of the pixels as reference:

$$\boldsymbol{s}_{\mathrm{CCT}}(\mathfrak{f}) \;\coloneqq\; (\boldsymbol{s}^{1\,\mathsf{T}}_{\mathrm{CT}}, \dots, \boldsymbol{s}^{K\,\mathsf{T}}_{\mathrm{CT}})^{\mathsf{T}}\,, \tag{5.6}$$

where, similar to (5.3), the $K-1$ digits of each individual signature $\boldsymbol{s}^i_{\mathrm{CT}}$ are given by

$$\left(\boldsymbol{s}^i_{\mathrm{CT}}\right)_j \;\coloneqq\; \chi_{(\mathfrak{f}_j < \mathfrak{f}_i)} \qquad \forall j \in \{1, \dots, K\} \setminus i\,. \tag{5.7}$$

Evidently, the original census signature from Equation (5.3) coincides with $\boldsymbol{s}^1_{\mathrm{CT}}$. Please note that the information contained in complete rank and complete census is equivalent. This can be seen from the bijection between them: On the one hand, it makes no difference if we compute the CCT signature directly from the intensity values or from the CRT signature of a patch, i.e.

$$\boldsymbol{s}_{\mathrm{CCT}}(\mathfrak{f}) \;=\; \boldsymbol{s}_{\mathrm{CCT}}\left(\boldsymbol{s}_{\mathrm{CRT}}(\mathfrak{f})\right)\,. \tag{5.8}$$

In the opposite direction, the complete rank digits are just the sums of corresponding CCT bits:

$$\left(\boldsymbol{s}_{\mathrm{CRT}}(\mathfrak{f})\right)_i \;=\; \sum_{j=1}^{K-1} \left(\boldsymbol{s}^i_{\mathrm{CT}}\right)_j\,. \tag{5.9}$$

Both complete rank and complete census signatures, do also represent tied ranks, i.e. if pixels in the patch have the same intensity value. Thus, the number of possible signatures for a patch with K pixels is given by the K-th *ordered Bell number* $\mathrm{OBN}(K)$ (also called K-th Fubini number) which is defined by [SP95]

$$\mathrm{OBN}(K) \;=\; \sum_{i=0}^{K} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} j^K \,. \tag{5.10}$$

It expresses the maximally possible number of weak orderings of a set of K elements.

Discussion

In each pixel, our complete rank signature contains the full local image intensity order. Obviously, this is much more information than the rank or census signatures carry. In particular, it is not possible to encode more local image information without leaving the class of morphologically invariant descriptors. The reason for this is that the only property that cannot be changed by a monotonic function is monotonicity, i.e. whether one pixel is larger than the other or not. Here, we want to note that our notion of *image information* does not coincide with the classical notion of information content in terms of coding length. The work on *actionable information* by Soatto [Soa09] goes more in our direction, as it only includes the information content of image (sequences) that is relevant for the task at hand. In particular, this means that so-called *nuisances* (illumination and viewpoint changes) are discarded and do not count as actionable information. Interestingly, Soatto [Soa09] models illumination changes with monotonically increasing continuous functions. In that sense, our morphologically invariant descriptors do only discount nuisances and no actionable information.

In fact, due to its much more compact representation and lower dimensionality, the CRT signatures are generally preferable to the CCT signatures. However, this alternative census-inspired perspective of our CRT feature offers an unexpected insight: As pointed out in [HSW15], each binary digit of a census signature can be regarded as the sign of the corresponding directional derivative (in a finite difference sense). Thus, from this point of view, one can conclude that the complete rank transform inherently contains rich local differential information. In this regard, dealing with derivatives of such signatures as in [PA10] actually corresponds to second order image derivative information. This fact is not obvious from just considering the rank representation and should be kept in mind.

For the sake of clarity, we summarise the discussed transforms and compare their essential properties in Table 5.1.

5.3.3 Variational Model

Our ultimate goal is to align the images of an exposure series to a selected reference frame. This comes down to the computation of displacement fields (*optic flow*) that

Table 5.1: Comparison of the proposed intensity order transforms. The number of pixels in the considered neighbourhood is given by K.

transform	range $\mathcal{D}$ of one digit	signature length ζ	spatial information	size of descriptor space
rank (RT)	$\{0, \dots, K-1\}$	1	–	K
census (CT)	$\{0, 1\}$	$K-1$	✓	2^{K-1}
complete rank (CRT)	$\{0, \dots, K-1\}$	K	✓	OBN(K)
complete census (CCT)	$\{0, 1\}$	$K(K-1)$	✓	OBN(K)

relate each position in a first (reference) frame f_1 to the corresponding position in a second frame f_2. To this end, we present in this section our general variational framework for optic flow computation. It is based on the seminal work of Horn and Schunck from 1981 [HS81], and allows a transparent and flexible modelling while being able to provide accurate state-of-the-art results as demonstrated in various optic flow benchmarks. First, we assume that the input images have been mapped by one of the introduced transforms to a vector-valued function $\boldsymbol{s}\colon \Omega \to \mathcal{D}^\zeta$, where $\Omega \subset \mathbb{R}^2$ denotes the 2D rectangular image domain and $\mathcal{D}$ describes the range of a single signature digit; cf. Table 5.1. Next, we model the constancy assumption that signatures of corresponding pixels in the first and second frame coincide. To this end, let $\boldsymbol{v} = (v_1, v_2)^\mathsf{T}\colon \Omega \to \mathbb{R}^2$ denote the unknown flow field, where v_1 represents the horizontal and v_2 the vertical flow component. With this notation, our energy functional is given by

$$E(\boldsymbol{v}) \;=\; D(\boldsymbol{v}) \;+\; \alpha \cdot R(\boldsymbol{v})\,. \tag{5.11}$$

The first term $D(\boldsymbol{v})$ is the *data term* that implements our constancy assumption. We reward small cost if a feature $\boldsymbol{s}_1$ of the first image at position $\boldsymbol{x}$ is similar to the feature $\boldsymbol{s}_2$ of the second image at the corresponding position $\boldsymbol{x} + \boldsymbol{v}(\boldsymbol{x})$. More specifically, we realise this by modelling the data term

$$D(\boldsymbol{v}) \;=\; \int_\Omega \Psi\Big(\tfrac{1}{\zeta}|\boldsymbol{s}_2(\boldsymbol{x}+\boldsymbol{v}) - \boldsymbol{s}_1(\boldsymbol{x})|^2\Big)\,\mathrm{d}\boldsymbol{x}\,, \tag{5.12}$$

where we omit the argument $\boldsymbol{x}$ of the optic flow field $\boldsymbol{v}$ for the sake of readability. Here, ζ denotes the signature length; cf. Table 5.1. In case of colour images, this signature length is tripled since we concatenate the signatures of each colour channel. Moreover, we apply the penalisation function $\Psi(z^2) = \sqrt{z^2 + \varepsilon^2}$ with the small positive parameter ε to increase the robustness w.r.t. outliers that may be caused by noise or occlusions; see e.g. [Hub81,BA91,Coh93]. In the exposure series setting, it additionally provides a robustness w.r.t. highly under- and over-saturated image regions where the imposed constancy assumption might be violated.

The second term $R(\boldsymbol{v})$ in (5.11) is the so-called *regularisation term*. It rewards smooth flow fields with small spatial variations, while the positive smoothness param-

eter α determines the amount of smoothness. In particular, we model the regularisation term

$$R(\boldsymbol{v}) \;=\; \int_\Omega \Psi\Big(|\boldsymbol{\nabla} v_1(\boldsymbol{x})|^2 + |\boldsymbol{\nabla} v_2(\boldsymbol{x})|^2\Big)\,\mathrm{d}\boldsymbol{x} \tag{5.13}$$

to provide a discontinuity preserving smoothing with the penalisation function $\Psi(z^2) = \sqrt{z^2+\varepsilon^2}$, $\varepsilon > 0$. It can be interpreted as regularised *total variation* (TV) regularisation of the optic flow $\boldsymbol{v}$; see e.g. [ROF92, AV94, PBB+06].

Discussion. Certainly, one can also think of more advanced smoothness terms which for instance adapt to the local image structure in an anisotropic way; cf. e.g. [NE86, ZBW11b] or our upcoming Chapter 6. However, as for instance pointed out by Zimmer [Zim11, p. 124], they do not consistently improve the flow computation in the exposure series setting. This can be explained by the possibly large under- and oversaturated image regions that yield unreasonable or even misleading directional cues. Thus, we refrain from such smoothness terms in our setting. Furthermore, there exist also higher order regularisation terms; cf. also our upcoming Chapter 6. For instance, second order terms are especially beneficial if the optic flow fields are assumed to be (piecewise) affine, which is for instance a well-suited assumption for traffic scenes [GLU12, MG15]. However, here we cannot assume such a prior knowledge, and hence stick to the discussed discontinuity preserving smoothness term in (5.13) that penalises first order flow derivatives.

Moreover, one can also think of modelling the alignment task of all image pairs in the exposure series within one joint energy functional. This would enable us to apply some kind of *temporal* smoothness, i.e. a smoothness of the optic flow along different acquisitions in the sense of [Nag90, WS99, VBVZ11, BML15]. However, in the general setting such a temporal smoothness is not fulfilled due to arbitrary camera shakes between the acquisitions. Thus, we refrain from such a model assumption in our general setting. Nevertheless, such a temporal smoothness might be beneficial in the HDR video setting; cf. our discussion of future work in Section 5.6.

5.3.4 Minimisation

Multiresolution Technique

Due to the data term, the energy in (5.11) is nonconvex w.r.t. the optic flow variables v_1 and v_2. As a remedy we consider, similar to [BBPW04], a fixed point iteration scheme that is embedded into a multiresolution pyramid-based approach. This leads to a linearised and convex version of the data term that allows for handling large displacements. More specifically, on each pyramid level ℓ we compute small flow increments $\boldsymbol{dv}^\ell := (dv_1^\ell, dv_2^\ell)^\mathsf{T}$. Introducing these flow increments to the data term in (5.12) yields

$$D^\ell(\boldsymbol{dv}^\ell) \;=\; \int_\Omega \Psi\Big(\tfrac{1}{\zeta}|\boldsymbol{s}_2(\boldsymbol{x}+\boldsymbol{v}^\ell+\boldsymbol{dv}^\ell) - \boldsymbol{s}_1(\boldsymbol{x})|^2\Big)\,\mathrm{d}\boldsymbol{x}\,. \tag{5.14}$$

Subsequently, first order Taylor linearisation of $\boldsymbol{s}_2(\boldsymbol{x} + \boldsymbol{v}^\ell + \boldsymbol{dv}^\ell)$ leads to

$$\boldsymbol{s}_2(\boldsymbol{x} + \boldsymbol{v}^\ell + \boldsymbol{dv}^\ell) \;\approx\; \boldsymbol{s}_2(\boldsymbol{x} + \boldsymbol{v}^\ell) + \partial_{x_1}\boldsymbol{s}_2(\boldsymbol{x} + \boldsymbol{v}^\ell)\, dv_1^\ell + \partial_{x_2}\boldsymbol{s}_2(\boldsymbol{x} + \boldsymbol{v}^\ell)\, dv_2^\ell \,. \tag{5.15}$$

Next, we interpret the difference $\boldsymbol{s}_2(\boldsymbol{x} + \boldsymbol{v}^\ell) - \boldsymbol{s}_1(\boldsymbol{x})$ as a temporal derivative and abbreviate it by $\boldsymbol{s}_t$. Consequently, the corresponding linearised version of the data term in (5.14) reads

$$\bar{D}^\ell(\boldsymbol{dv}^\ell) \;=\; \int_\Omega \Psi\Big(\tfrac{1}{\zeta}|\boldsymbol{s}_{x_1}\, dv_1^\ell + \boldsymbol{s}_{x_2}\, dv_2^\ell + \boldsymbol{s}_t|^2\Big)\,\mathrm{d}\boldsymbol{x}\,, \tag{5.16}$$

where the vector-valued derivatives

$$\boldsymbol{s}_{x_1} \;\coloneqq\; \partial_{x_1}\boldsymbol{s}_2(\boldsymbol{x} + \boldsymbol{v}^\ell) \qquad \text{and} \qquad \boldsymbol{s}_{x_2} \;\coloneqq\; \partial_{x_2}\boldsymbol{s}_2(\boldsymbol{x} + \boldsymbol{v}^\ell) \tag{5.17}$$

are calculated componentwise. After computing the flow increment $\boldsymbol{dv}^\ell$, we successively update the overall flow in the form of

$$\boldsymbol{v}^{\ell+1} \;=\; \boldsymbol{v}^\ell \,+\, \boldsymbol{dv}^\ell \,. \tag{5.18}$$

Moreover, we penalise on each level the derivatives of the *complete* flow by means of the regularisation term

$$R^\ell(\boldsymbol{dv}^\ell) \;=\; \int_\Omega \Psi\Big(|\boldsymbol{\nabla}(v_1^\ell + dv_1^\ell)|^2 + |\boldsymbol{\nabla}(v_2^\ell + dv_2^\ell)|^2\Big)\,\mathrm{d}\boldsymbol{x}\,. \tag{5.19}$$

To conclude, the final energy that we solve on each pyramid level ℓ is with (5.16) and (5.19) given by

$$E^\ell(\boldsymbol{dv}^\ell) \;=\; \bar{D}^\ell(\boldsymbol{dv}^\ell) \,+\, \alpha \cdot R^\ell(\boldsymbol{dv}^\ell)\,. \tag{5.20}$$

Euler-Lagrange Equations

Following the calculus of variations (cf. Section 2.3), a minimiser of the energy in (5.20) has to fulfil the Euler-Lagrange equations w.r.t. the flow increments dv_1^ℓ and dv_2^ℓ. For the sake of readability, let us first introduce the following abbreviations:

$$\Psi'_D \;\coloneqq\; \tfrac{1}{\zeta} \cdot \Psi'\big(\tfrac{1}{\zeta}|\boldsymbol{s}_{x_1}\, dv_1^\ell + \boldsymbol{s}_{x_2}\, dv_2^\ell + \boldsymbol{s}_t|^2\big)\,, \tag{5.21}$$

$$\Psi'_R \;\coloneqq\; \Psi'\big(|\boldsymbol{\nabla}(v_1^\ell + dv_1^\ell)|^2 + |\boldsymbol{\nabla}(v_2^\ell + dv_2^\ell)|^2\big)\,. \tag{5.22}$$

With these abbreviations, the Euler-Lagrange equations of (5.20) are given by

$$\Psi'_D \cdot \boldsymbol{s}_{x_1}^\mathsf{T}\big(\boldsymbol{s}_{x_1}\, dv_1^\ell + \boldsymbol{s}_{x_2}\, dv_2^\ell + \boldsymbol{s}_t\big) \,-\, \alpha\ \mathrm{div}\,\big(\Psi'_R \cdot \boldsymbol{\nabla}(v_1^\ell + dv_1^\ell)\big) \;=\; 0\,, \tag{5.23}$$

$$\Psi'_D \cdot \boldsymbol{s}_{x_2}^\mathsf{T}\big(\boldsymbol{s}_{x_1}\, dv_1^\ell + \boldsymbol{s}_{x_2}\, dv_2^\ell + \boldsymbol{s}_t\big) \,-\, \alpha\ \mathrm{div}\,\big(\Psi'_R \cdot \boldsymbol{\nabla}(v_2^\ell + dv_2^\ell)\big) \;=\; 0\,, \tag{5.24}$$

with the associated boundary conditions

$$\boldsymbol{n}^\mathsf{T}\boldsymbol{\nabla}(v_1^\ell + dv_1^\ell) \;=\; 0 \qquad \text{and} \qquad \boldsymbol{n}^\mathsf{T}\boldsymbol{\nabla}(v_2^\ell + dv_2^\ell) \;=\; 0\,, \tag{5.25}$$

where $\boldsymbol{n}$ denotes the outer normal vector on the boundary of Ω.

Implementation Details

All occurring spatial derivatives of the feature vector $\boldsymbol{s}$ in (5.23) and (5.24) are computed by means of the fourth order stencil $(1, -8, 0, 8, -1)/(24h_q)$, where $q \in \{1, 2\}$. Moreover, to compute the downsampled transformed images, we presmooth them with a Gaussian whose standard deviation is proportional to the current grid size $\eta^\ell \cdot h_q$, where $\eta = 0.95$ is the downsampling factor. Accordingly, we solve a sparse nonlinear system of equations on each pyramid level, where the nonlinearities are caused by the Ψ' terms in (5.21) and (5.22). To solve those systems, we apply our FSI scheme which we presented in Chapter 3. More precisely, one iteration step at the pyramid level ℓ with single index notation i, cycle index m, and iteration index k is for the flow increments ${dv_1}^\ell$ and ${dv_2}^\ell$ given by

$$
\begin{aligned}
{dv_1}_i^{\ell,m,k+1} &= (1-\alpha_k)\cdot {dv_1}_i^{\ell,m,k-1} + \alpha_k \cdot \qquad (5.26)\\
&\Bigg((1-\omega)\cdot {dv_1}_i^{\ell,m,k} \;+\; \omega \;\cdot\; \Bigg(-\Psi'_{Di}\cdot \boldsymbol{s}_{x_1 i}{}^{\mathsf{T}} (\boldsymbol{s}_{x_2 i}\cdot {dv_2}_i^{\ell,m,k} + \boldsymbol{s}_{ti}) \\
&+ \sum_{q=1}^{2} \sum_{j\in\mathcal{N}_q(i)} \alpha \cdot \frac{\Psi'_{Ri}+\Psi'_{Rj}}{2h_q} \cdot \Big(\frac{{v_1}_j^{\ell,m,k} - {v_1}_i^{\ell,m,k} + {dv_1}_j^{\ell,m,k}}{h_q} \Big) \Bigg) \\
&\Big/ \Bigg(\Psi'_{Di}\cdot \boldsymbol{s}_{x_1 i}{}^{\mathsf{T}} \boldsymbol{s}_{x_1 i} + \sum_{q=1}^{2} \sum_{j\in\mathcal{N}_q(i)} \alpha \cdot \frac{\Psi'_{Ri}+\Psi'_{Rj}}{2h_q^2} \Bigg) \Bigg), \qquad (5.27)
\end{aligned}
$$

$$
\begin{aligned}
{dv_2}_i^{\ell,m,k+1} &= (1-\alpha_k)\cdot {dv_2}_i^{\ell,m,k-1} + \alpha_k \cdot \qquad (5.28)\\
&\Bigg((1-\omega)\cdot {dv_2}_i^{\ell,m,k} \;+\; \omega \;\cdot\; \Bigg(-\Psi'_{Di}\cdot \boldsymbol{s}_{x_2 i}{}^{\mathsf{T}} (\boldsymbol{s}_{x_1 i}\cdot {dv_1}_i^{\ell,m,k} + \boldsymbol{s}_{ti}) \\
&+ \sum_{q=1}^{2} \sum_{j\in\mathcal{N}_q(i)} \alpha \cdot \frac{\Psi'_{Ri}+\Psi'_{Rj}}{2h_q} \cdot \Big(\frac{{v_2}_j^{\ell,m,k} - {v_2}_i^{\ell,m,k} + {dv_2}_j^{\ell,m,k}}{h_q} \Big) \Bigg) \\
&\Big/ \Bigg(\Psi'_{Di}\cdot \boldsymbol{s}_{x_2 i}{}^{\mathsf{T}} \boldsymbol{s}_{x_2 i} + \sum_{q=1}^{2} \sum_{j\in\mathcal{N}_q(i)} \alpha \cdot \frac{\Psi'_{Ri}+\Psi'_{Rj}}{2h_q^2} \Bigg) \Bigg), \qquad (5.29)
\end{aligned}
$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ describe the neighbouring pixels in horizontal and vertical direction, respectively. Moreover, ω represents the numerical relaxation parameter, and α_k are the cyclic parameters of our FSI scheme; cf. Chapter 3.

5.3.5 Experiments

Our experimental section is structured as follows: In the first part, we demonstrate that the invariance against any monotonically increasing rescalings is indeed fulfilled by simulating such mappings artificially. Additionally, we investigate the behaviour

under noise and the impact of the neighbourhood size K. Next, we assess the performance of our approach on public benchmark data sets. Last but not least, we apply our method for the alignment of exposure series.

Choice of Parameters. Due to the simplicity of our model, only very few parameters have to be chosen. The main free parameter of our optic flow method is the weight of the smoothness term α. We optimise α for every experiment. The parameter of the nonlinear function Ψ has less influence and we set in all our experiments $\varepsilon = 10^{-2}$.

Evaluation

Concerning synthetic perturbations, we consider the eight training image sequences of the Middlebury optic flow benchmark [BSL+11] because no severe illumination changes are present and reliable ground truth flow fields are available. To assess the accuracy of the optic flow fields, we evaluate the *average angular error* (AAE) [BFB94]:

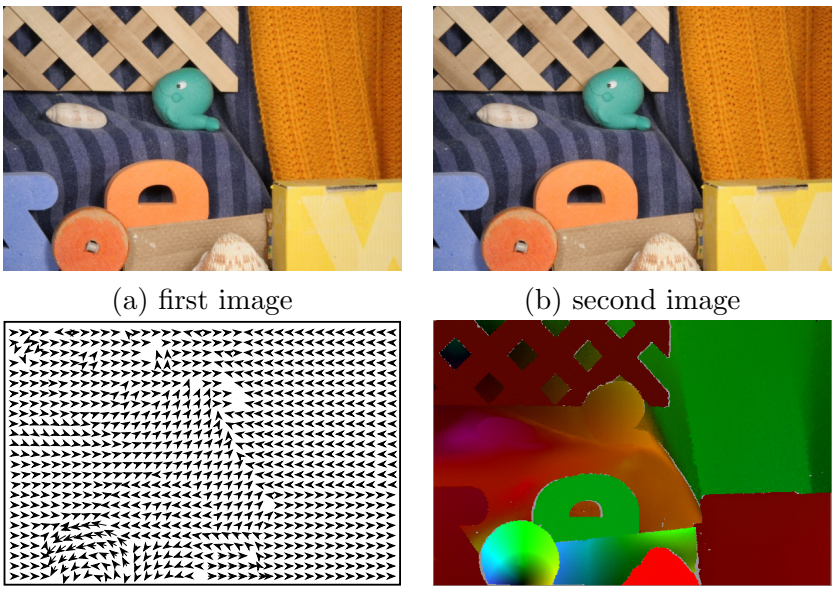
$$\mathrm{AAE}(\boldsymbol{v}, \boldsymbol{v}_{\mathrm{GT}}) \;=\; \frac{1}{|\Omega|} \int_{\Omega} \arccos\left(\frac{v_1 v_{\mathrm{GT}1} + v_2 v_{\mathrm{GT}2} + 1}{\sqrt{({v_1}^2 + {v_2}^2 + 1)(v_{\mathrm{GT}1}^2 + v_{\mathrm{GT}2}^2 + 1)}} \right) \mathrm{d}\boldsymbol{x}\,, \quad (5.30)$$

where $\boldsymbol{v} = (v_1, v_2)^{\mathsf{T}}$ is the estimated and $\boldsymbol{v}_{\mathrm{GT}} = (v_{\mathrm{GT}1}, v_{\mathrm{GT}2})^{\mathsf{T}}$ the known *ground truth* (GT) displacement field. As usual we state the AAE values in degree.

Morphological Invariance. Our first experiment examines the behaviour of the proposed method under monotonically increasing intensity changes. To this end, we consider the eight Middlebury training image sequences, where we show in Figure 5.3 the *RubberWhale* sequence as an example. In particular, in Figure 5.3(a,b) we depict the two unaligned input images while Figure 5.3(c,d) represent the ground truth displacement field (*optic flow*) between both frames as arrow and colour plot, respectively. Here, the hue specifies the direction, and the saturation determines the amount of motion; cf. Figure 5.4. Moreover, a grey colour means that no reliable flow vector is given at this position.

To synthetically generate severe illumination changes, we apply a gamma correction to the second frames, i.e. $f_2^{\tilde{\gamma}}(\boldsymbol{x})$ with $\tilde{\gamma} \in [0.2, 3]$, where we assume intensity values in the range of $[0, 1]$. The results of this experiment are depicted in Figure 5.5. To ensure a fair comparison, the regularisation parameter α has been optimised for each graph. First, the plot in Figure 5.5 indeed demonstrates the morphological invariance of the considered features. Second, we see that we outperform the rank and census transform by the proposed complete descriptors. In this regard, our CRT and CCT features show a similar performance. However, one should keep in mind that CRT is, due to its lower dimensionality (cf. Table 5.1), much more computationally efficient.

Sensitivity to Noise. In this experiment we perturb both input images of the image sequences with zero-mean Gaussian noise of varying standard deviations σ, and



(a) first image (b) second image

(c) ground truth flow (arrows) (d) ground truth flow (colour)

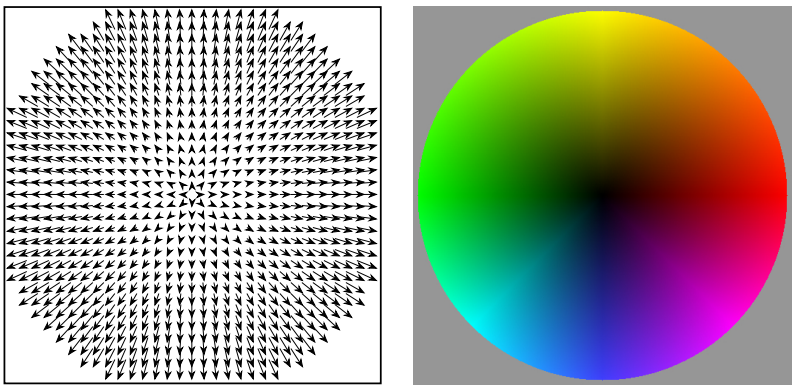
Figure 5.3: *RubberWhale* training sequence from the Middlebury optic flow benchmark [BSL+11]. *In reading order*: (a,b) Unaligned input images, (c) ground truth displacement map as arrow plot (black boundary added for visibility reasons), and (d) ground truth displacement map as colour plot; cf. Figure 5.4.

measure the resulting accuracy. The outcome of this experiment is depicted in Figure 5.6. Compared to the rank and census transform, the complete rank as well as the complete census transform loose a bit less accuracy if the contamination with noise increases. Also here, CRT and CCT provide a similar performance. Hence, as for the same reasons as before, we prefer the use of our CRT features for computational efficiency.

Neighbourhood Size. Let us investigate our CRT feature in more detail. As it turns out, the best neighbourhood size K, i.e. the one that leads to the smallest error measure, depends on the input image size. To systematically analyse this behaviour, we consider a benchmark set where the image sequences are available in different resolutions. In particular, we test different neighbourhood sizes by means of the Middlebury stereo benchmark from 2005 [HS07]. It provides six image sets with ground truth data in a low, middle, and high resolution. Since only horizontal displacements are present in the considered stereo setting, we adapt our energy functional accordingly and only optimise for the horizontal flow component v_1. Moreover, we apply the *average endpoint error* (AEE) [ON94]

$$\mathrm{AEE}(\boldsymbol{v}, \boldsymbol{v}_{\mathrm{GT}}) \;=\; \frac{1}{|\Omega|} \int_{\Omega} |\boldsymbol{v}(\boldsymbol{x}) - \boldsymbol{v}_{\mathrm{GT}}(\boldsymbol{x})| \, \mathrm{d}\boldsymbol{x} \,, \tag{5.31}$$

which is a more reasonable error measure for stereo computation. Obviously, in our setting we have $v_2(\boldsymbol{x}) = v_{\mathrm{GT}2}(\boldsymbol{x}) = 0$.



(a) arrow plot (b) colour plot

Figure 5.4: Illustration of applied optic flow colour coding. Here, the hue denotes the direction while the colour saturation represents the amount of motion. Moreover, a grey colour indicates that no reliable flow vector is given at this position.

In Figure 5.7(f), we plot the AEE for different CRT neighbourhood sizes and image resolutions. Here, we optimised the smoothness parameter α for every neighbourhood size, where we considered for each resolution the average error over all six image sequences. In Figure 5.7(a–e), we depict the corresponding shapes of the neighbourhood patches for different K. First of all, we observe that the chosen neighbourhood size does not have a drastic effect on the results. However, we can also see that the best neighbourhood size K increases with the image resolution. Hence, we recommend to adapt K to the actual image size in order to reach ultimate quality.

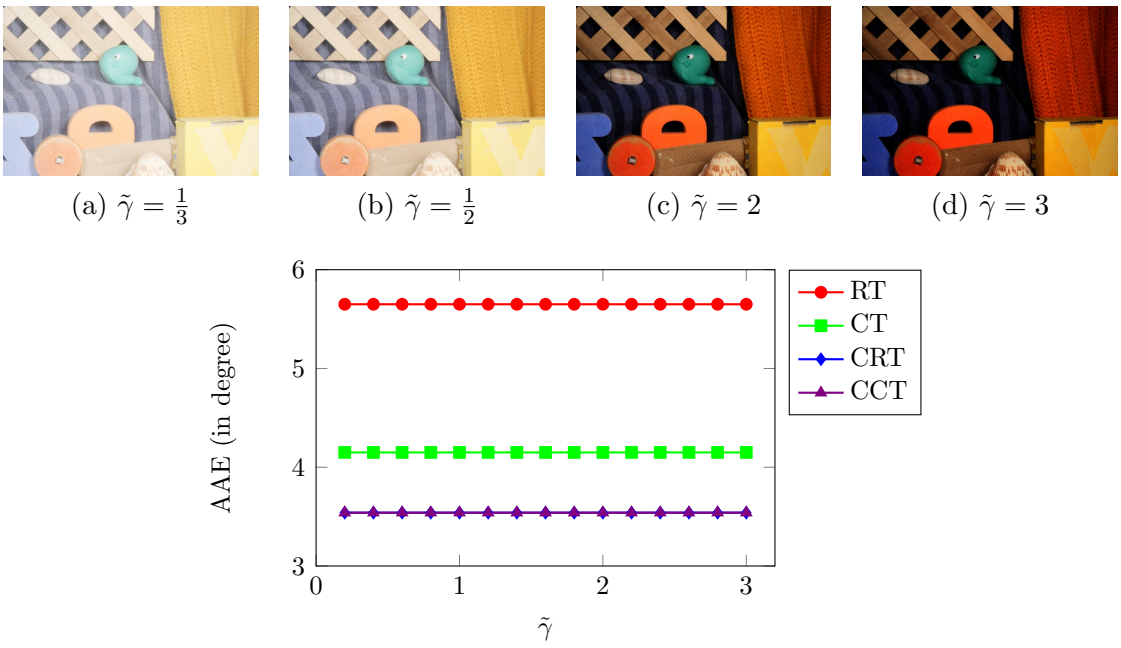
Performance on Public Benchmark Systems

Since our method is tailored to challenging illumination conditions, we also focus our evaluation on image material where such conditions are present. In that respect, the KITTI vision suite [GLU12] offers a good testbed for our needs. It provides a huge amount of image sequences captured from a driving car, along with corresponding ground truth flow fields that are acquired with a laser scanning technique. Due to the inherent small-scale imprecisions of the ground truth data acquisition process of the KITTI benchmark, the usual error measures such as the average angular error are not well suited for a quantitative evaluation. Thus, the common measure for the KITTI benchmark is the so-called *bad pixel error* (BPE) [SS02, GLU12]:

$$\mathrm{BPE}_\theta(\boldsymbol{v}, \boldsymbol{v}_{\mathrm{GT}}) \;=\; \frac{1}{|\Omega|} \int_\Omega \chi_{(\theta < |\boldsymbol{v}(\boldsymbol{x}) - \boldsymbol{v}_{\mathrm{GT}}(\boldsymbol{x})|)} \, \mathrm{d}\boldsymbol{x} \,. \tag{5.32}$$

In particular, we consider BPE_3 which expresses the percentage of estimated flow vectors that differ by more than 3 pixels form the measured ground truth solution, i.e. the percentage of pixels with an endpoint error above 3 pixels.

Let us first compare our method to related optic flow approaches from the literature. To this end, we compute flow fields for four real-world test sequences of the KITTI



(a) $\tilde{\gamma} = \frac{1}{3}$ (b) $\tilde{\gamma} = \frac{1}{2}$ (c) $\tilde{\gamma} = 2$ (d) $\tilde{\gamma} = 3$

(e) behaviour under $\tilde{\gamma}$ changes

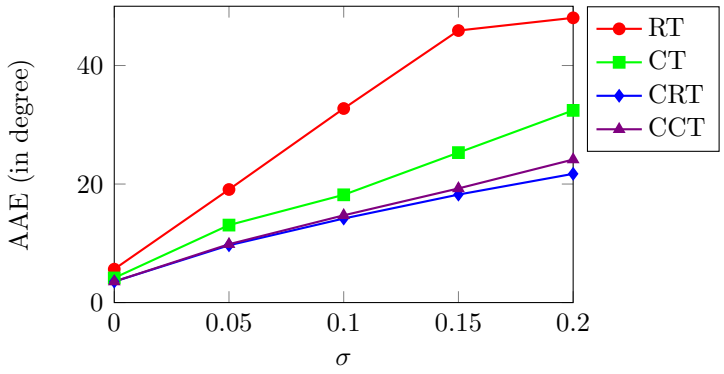
Figure 5.5: Investigation of morphological invariance. For illustration, (a–d) depict the second frame of the *RubberWhale* sequence from Figure 5.3 with different gamma corrections. In (e), we plot the resulting *average angular errors* (AAEs), where we consider for each $\tilde{\gamma}$ the mean AAE over all eight training image sequences of the Middlebury benchmark [BSL+11].

benchmark suite [GLU12], which exhibit severe illumination changes. In particular, we have chosen the same data set as selected for the *GCPR 2013 – Special Session on Robust Optical Flow* [Deu13]. Table 5.2 summarises the obtained results. As reference, the numbers for the methods of [ZBW11b] and [BW05] are taken from the website of this special session. The method of [BW05] is particularly interesting to compare, since our regularisation strategy is similar to the one applied in this paper. As one can see, our complete rank transform consistently outperforms the competing methods.

Next, we assess the error rates on the Middlebury optic flow training images which are less demanding w.r.t. illumination changes. Hence, the goal of this experiment is to test our approach under normal lighting conditions. For each signature type, the regularisation parameter α has been optimised and then kept constant over all images. Table 5.3 depicts the errors for the different image sets. Also in this setting, the rank and census transform are outperformed by their complete counterparts. Moreover, we see again that CRT and CCT provide similar results. Hence, we generally propose the use of CRT features since they are much more computationally efficient.

For the sake of completeness, we also evaluate our method on the test images of the Middlebury optic flow benchmark [BSL+11]. Since the test sequences of this benchmark exhibit almost no illumination changes or other scenarios that our highly invariant descriptors are designed for, we cannot expect top-ranking results on this bench-

(a) $\sigma = 0.05$ (b) $\sigma = 0.1$ (c) $\sigma = 0.15$ (d) $\sigma = 0.2$



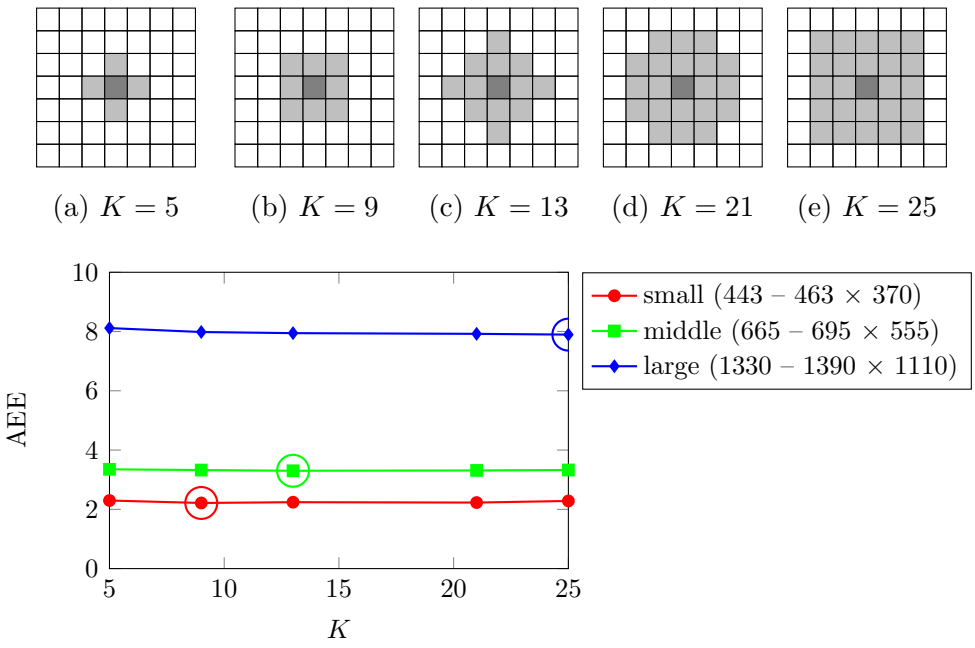
(e) behaviour under Gaussian noise

Figure 5.6: Investigation of behaviour under noise. For illustration, (a–d) depict the second frame of the *RubberWhale* sequence from Figure 5.3 perturbed by Gaussian noise with zero mean and varying standard deviation σ. In (e), we plot the resulting *average angular errors* (AAEs), where we consider for each choice of σ the mean AAE over all eight training image sequences of the Middlebury optic flow benchmark [BSL+11].

mark. Nevertheless, it turns out that our variational model can keep up with its nearest competitors (cf. [Sch16]). In fact, our method ranks between the method of Brox et al. [BBPW04] and the much more advanced method by Zimmer et al. [ZBW+09]. These results are remarkable in the sense that they prove our invariant data term to include hardly less information than the combined grey value and gradient information of [BBPW04,ZBW+09] that offers significantly less robustness w.r.t. illumination changes.

Application to Exposure Series Alignment

Last but not least, let us demonstrate the performance of our approach for the alignment of exposure series. To this end, we consider first a synthetic exposure set that allows for comparing the computed flow fields by means of the AAE. In Figure 5.8(a–c), we depict the differently exposed input images, where the ground truth flow fields from the reference image to the other images are given in Figure 5.8(d,e). In particular, we generated this scene with the open source software *Blender* [Ble16], where we computed the ground truth flow with a plug-in that we especially developed for this purpose and which makes efficient use of the *Blender*-internal rendering engine. Here, we apply a common camera response function and appropriate exposure times to ensure that the created exposure set constitutes a realistic example sequence.



(a) $K = 5$ (b) $K = 9$ (c) $K = 13$ (d) $K = 21$ (e) $K = 25$

(f) impact of CRT-neighbourhood size K

Figure 5.7: Investigation of CRT-neighbourhood size for different image resolutions. (a–e) Different neighbourhood patches depending on neighbourhood size K. Reference pixel is marked in dark grey and considered neighbouring pixels in light grey. (f) Resulting *average endpoint errors* (AEEs) for different image sizes. Here, we consider for each image size the mean AAE over all six image pairs of the Middlebury stereo benchmark set [HS07]. Best results are highlighted by the coloured circles.

In Figure 5.8(f,g), we show the computed optic flow fields using our CRT features. First of all, we observe that our approach provides desirable alignment results for such difficult input images. A comparison with the method of Zimmer et al. [ZBW11a] in terms of the AAE underlies this finding: While the approach in [ZBW11a], which was in fact also designed for such difficult exposure series settings, yields an AAE of 12.60°, we achieve an AAE of 5.17°. For both approaches, we optimised the smoothness parameter to gain the best error measures.

However, we also observe that the discussed approach fails in aligning the balloon correctly. This is also obvious in Figure 5.8(i) where we show the exposure fusion result of the aligned short and long exposure (Figure 5.8(h,j)) and the reference frame (Figure 5.8(b)). For fusing those images, we applied our general fusion technique presented in Chapter 4 with the default parameter setting. While most parts of the scene are registered precisely, the unaligned balloon produces undesirable ghosting artefacts in the fusion result; cf. in particular the zoom in Figure 5.10(a).

Similar results apply for real-world exposure sets such as the one depicted in Figure 5.9(a–c). The computed flow fields in Figure 5.9(d,e) allow for an accurate alignment for most of the image regions; cf. Figure 5.9(f,g). However, also here incorrectly matched parts such as the ears of the rabbit produce ghosts in the fusion result;

Table 5.2: Behaviour for real-world scenarios of the KITTI benchmark suite [GLU12]. Errors are given in terms of the BPE_3 measure, i.e. the percentage of pixels having a Euclidean error larger than 3. Best results are written in bold face.

KITTI image sequence:	#11	#15	#44	#74	average
Zimmer et al. [ZBW11b]	37.3	32.3	23.2	62.9	38.9
Bruhn and Weickert [BW05]	33.9	47.7	32.4	71.4	46.7
census transform	36.5	28.6	28.5	63.8	39.4
complete tank transform	**29.8**	**22.8**	**22.6**	**61.5**	**34.2**

Table 5.3: Quantitative comparison of the *rank* (RT), *census* (CT), *complete rank* (CRT) and *complete census transform* (CCT) on the Middlebury optic flow training images [BSL+11] in terms of the *average angular error* (AAE). Values are given in degree and the best results are written in bold face.

	Dimetrodon	Grove2	Grove3	Hydrangea	RubberWhale	Urban2	Urban3	Venus	average
RT	5.46	3.69	8.89	2.17	3.55	3.59	7.51	10.30	5.65
CT	4.67	2.76	6.41	1.72	2.78	3.21	4.58	7.10	4.15
CRT	3.80	2.61	6.37	1.86	2.92	**2.56**	**3.85**	4.37	3.54
CCT	**3.73**	**2.51**	**6.33**	**1.70**	**2.72**	2.77	4.30	**4.26**	**3.54**

cf. Figure 5.9(g) and Figure 5.10(b). Here, we want to mention that decreasing the smoothness weight α indeed allows for aligning the rabbit ears, but unfortunately introduces more severe errors elsewhere.

5.3.6 Limitations and Discussion

We have seen that our variational approach with the proposed CRT features provides accurate alignment results and compares favourably with related illumination-robust optic flow approaches from the literature. However, we have also observed that we cannot expect perfect alignment results in every image region. Violations of the model assumptions such as occlusions or highly under- or over-saturated image regions may lead to registration errors that further produce ghosts in the fused composite images; cf. for instance Figure 5.10. Hence, an additional postprocessing step is often unavoidable. To this end, we propose an appropriate ghost removal technique in Section 5.4.

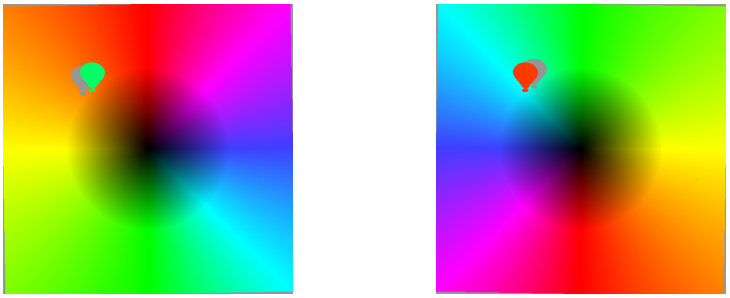
Indeed, our CRT-based approach is rather general and can be applied to match differently exposed images. However, further knowledge about the input data might

allow for more specific model assumptions, which would allow a better handling of saturated image regions or of quantisation effects introduced by the camera response function. We present such an approach in Section 5.5.

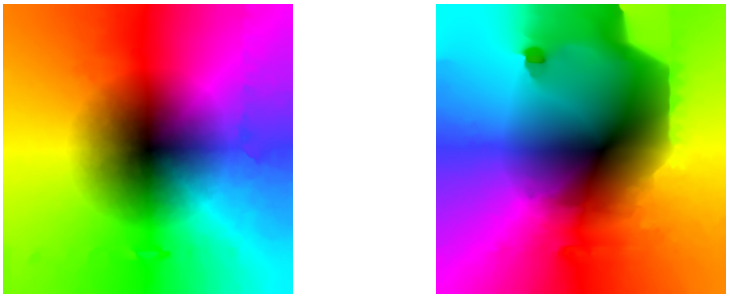
Moreover, in certain sceneries the amount and type of illumination varies spatially. However, our CRT features are only invariant to *global* monotonically increasing brightness rescalings, i.e. where every pixel undergoes the same brightness transformation. Any local illumination changes violates this assumption. At this point, we also would like to mention that our CRT feature lacks rotational invariance. To ensure such an invariance, we have to stick to the original rank transform (with a circular neighbourhood). Note that also the original census transform and the CCT lack such a rotational invariance. However, this invariance comes at a high cost of discarding a lot of valuable information. In fact, we could not confirm a benefit of such an invariance in our experiments. This might be explained by a sufficiently small rotational motion from frame to frame. Nevertheless, the design of an illumination-robust descriptor that provides a good tradeoff between the loss of information and a high robustness to specific motion patterns such as rotations or scale changes is an interesting topic for future work.

Last but not least, as for every variational pyramid-based approach large motion of small-scale objects pose a big challenge. In this regard, an application of our CRT features for sparse feature matching might be also promising. At least, it would be interesting to formulate a CRT-based variational approach that consequently incorporates sparse CRT matches in the energy formulation; cf. e.g. [BM11] for related ideas. Due to the patch-based nature of our CRT feature also an application of the fast nearest neighbour search algorithm *PatchMatch* [BSFG09] seems to be natural in this context. In particular, a combination of *PatchMatch* with a variational optic flow method constitutes a promising research topic for an adequate handling of large motion; see e.g. [Mus16].

(a) short exposure (b) middle exposure (c) long exposure



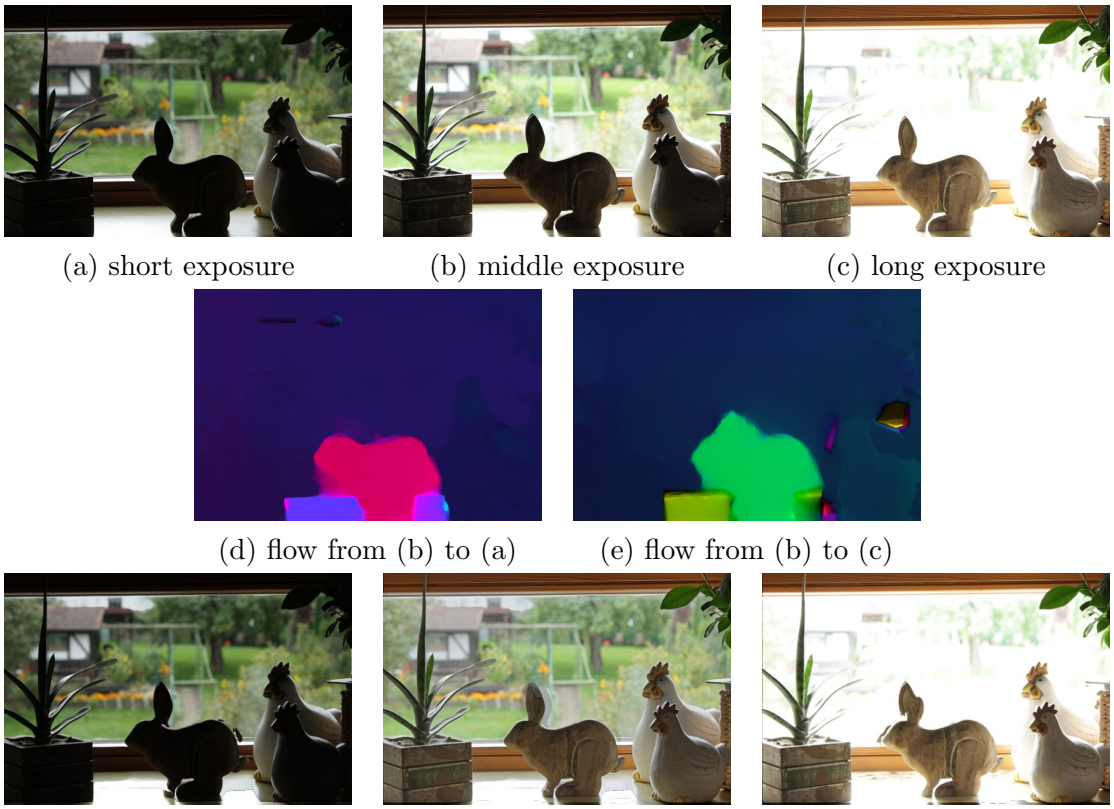
(d) ground truth flow (b) to (a) (e) ground truth flow (b) to (c)



(f) computed flow (b) to (a) (g) computed flow (b) to (c)

(h) aligned short exposure (i) fusion result (j) aligned long exposure

Figure 5.8: Exposure set *Balloon* synthetically generated with *Blender* [Ble16]. (a–c) Input images created with a common camera response function and appropriate exposure times to mimic a real-world scene. (d,e) Ground truth flow field produced with our plug-in for *Blender*. (f,g) Resulting flow fields computed by our CRT-based variational approach. (h,j) Aligned short and long exposures using the flow fields from (f) and (g). (i) Exposure fusion of (b,h,j) with the proposed method from Chapter 4. A zoom into (i) can be found in Figure 5.10(a).

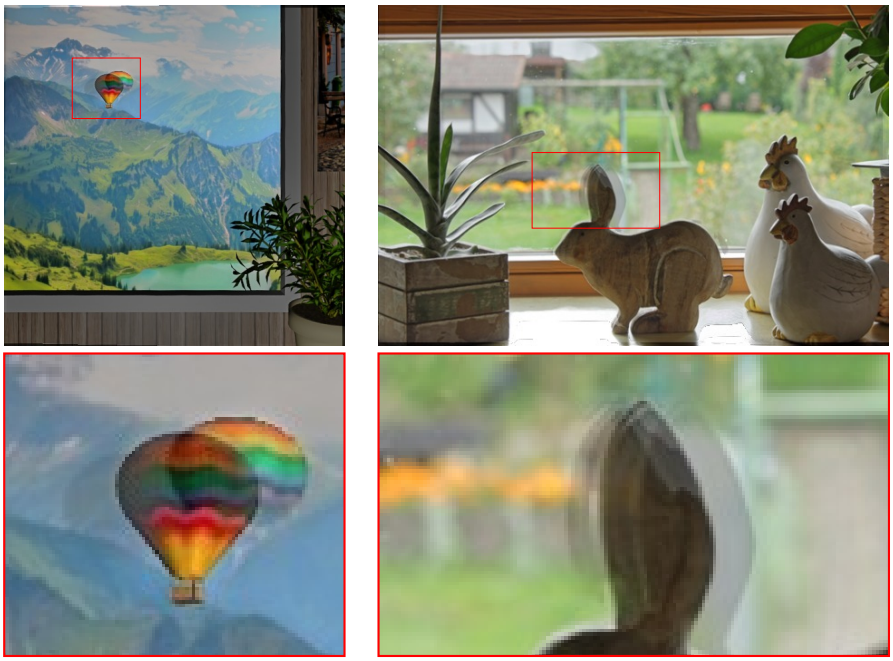


(a) short exposure (b) middle exposure (c) long exposure

(d) flow from (b) to (a) (e) flow from (b) to (c)

(f) aligned short exposure (g) fusion result (h) aligned long exposure

Figure 5.9: Real-world exposure set *Rabbit*. (a–c) Three of the five input images of the exposure set. (d,e) Computed flow fields using our CRT features. (f,h) Aligned short and long exposures using those flow fields. (g) Exposure fusion result created with the approach from Chapter 4. A zoom into (g) can be found in Figure 5.10(b).



(a) *Balloon* fusion result (b) *Rabbit* fusion result

Figure 5.10: Exposure fusion results of the aligned exposure sets from Figure 5.8 and 5.9. *Top*: Full images. *Bottom*: Zooms. While our registration strategy provides accurate alignment results for most image regions, it fails in some difficult parts. Unfortunately, this carries over to undesirable ghosting artefacts when fusing the images.

5.4 Ghost Removal with Brightness Transfer Function

In the last section of this chapter, we have described how to align differently exposed images to a chosen reference frame of the exposure set. The discussed variational approach is able to handle local scene motion, i.e. motion caused by camera shakes as well as moving objects. However, e.g. due to violations of the model assumptions such as large displacements of small objects or occlusions, we have to cope with small remaining imprecisions; cf. Figure 5.10. In such cases, an additional postprocessing step is required to identify and to eliminate remaining misalignments. This step is called *ghost removal*. In particular, we compute for each image a so-called *ghost map* that indicates if an image region is reliable, i.e. if it should be considered in the fusion process, or not.

Main Contributions. In Section 5.2.2 we have seen that, depending on their underlying assumptions about the exposure set, ghost removal approaches differ in their area of application. For instance, some require the knowledge of the exposure times and the *camera response function* (CRF), while others do not make such assumptions about the imaging process. While the former one might be useful in an HDR reconstruction context, the later one has a wider application area. Indeed, in certain set-ups the CRF and the exposure times are unknown. Moreover, in a general setting, the images in the exposure set do not necessarily have to follow the HDR observation model. This motivates us to present in this section a ghost removal algorithm that is independent of the knowledge of the CRF and the exposure times. To this end, we base our ghost removal technique on the concept of a *brightness transfer function* (BTF) that relates the brightness values between differently exposed frames. Moreover, we design our method in such a way that the processed ghost-free output images can directly serve as input for exposure fusion methods; without any specific modifications. This enables a straightforward application of our method as an intermediate processing step between alignment and image fusion.

Section Outline. We explain the relation of our ghost removal scheme to previous work in Section 5.4.1. Next, we present the core part of our technique in Section 5.4.2, and propose modifications in Section 5.4.3. In Section 5.4.4, we evaluate our approach by means of experiments, and discuss possible limitations in Section 5.4.5.

5.4.1 Relation to Previous Work

As discussed in Section 5.2.2, there exist several approaches that also make use of the BTF for ghost removal. In this work, we aim at a versatile technique and thus, do not restrict the CRF to be linear such as in [KHC+06, GGC+09, SSM12, TAEE15], but allow CRFs with arbitrary shape. Moreover, since we assume the input images to be pre-aligned (e.g. with the variational technique from Section 5.3), the histogram-based BTF estimation as proposed by Grossberg and Nayar [GN03] is well-suited for our needs. Thus, our ghost removal approach is closely related to the methods of

Grosch [Gro06], Li et al. [LRZ+10], and Moon et al. [MTCL12] that also apply such a histogram-based BTF estimation. In fact, Li et al. [LRZ+10] and our approach share strong similarities. As main differences, our method is completely independent from knowing the exposure times (no adaptive thresholding based on the exposure time), and we perform our pixel comparison in the perceptually motived CIE-Lab colour space; cf. Section 2.2.2.

5.4.2 Bidirectional BTF-Based Ghost Removal

Formally, a *brightness transfer function* (BTF) $\varphi_{i\to j}\colon [0,1] \to [0,1]$ maps the intensity values from the image f_i to the image f_j. Mann [Man00] presents a method for estimating the BTF by means of the joint probability density function (*comparagram*) between corresponding pixels in the first and in the second image. However, already small misalignments render the construction of such a comparagram difficult since it requires the computation of correspondences. Hence, we follow a different idea in this work: Grossberg and Nayar [GN03] propose to compute the BTF by means of the (cumulative) image histograms. As demonstrated in [GN03], such histograms provide a robustness w.r.t. small scene motion and thus, are well-suited for our setting.

Let us describe this method in more detail. Generally, a histogram measures the frequency of the intensity values in an image. Normalising each frequency by the total number of pixels leads to a normalised histogram that maps an intensity value $z \in [0,1]$ to a probability between 0 and 1. Moreover, a cumulative histogram $H\colon [0,1] \to [0,1]$ can be seen as a cumulative distribution function that represents the probability that an intensity value is smaller or equal to z. Hence, the value of the cumulative histogram of the maximal intensity values is always 1, i.e. $H(1) = 1$. Let H_1 and H_2 denote the cumulative histograms of the images f_1 and f_2. Eventually, we can relate both cumulative histograms as follows:

$$H_2(\varphi_{1\to 2}(z)) \;=\; H_1(z)\,, \tag{5.33}$$

where $z \in [0,1]$ represents an intensity value and $\varphi_{1\to 2}\colon [0,1] \to [0,1]$ the unknown BTF from f_1 to f_2. Accordingly, $\varphi_{1\to 2}$ is given by

$$\varphi_{1\to 2}(z) \;=\; H_2^{-1} H_1(z)\,. \tag{5.34}$$

Obviously, in the discrete setting H_2 is not necessarily a strictly monotonically increasing function. Hence, computing the inverse H_2^{-1} of the cumulative histogram H_2 is strictly speaking not possible. However, it can be approximated by so-called *histogram specification* that transforms the (cumulative) histogram of one image into the other one; cf. for instance [GW07]. More specifically, for each intensity value z_1 in f_1, we find the smallest value z_2 such that

$$H_1(z_1) \;\leq\; H_2(z_2)\,, \tag{5.35}$$

and assign the corresponding output of the BTF $\varphi_{1\to 2}$ for the considered intensity value z_1 to z_2, i.e.

$$\varphi_{1\to 2}(z_1) \;=\; z_2\,. \tag{5.36}$$

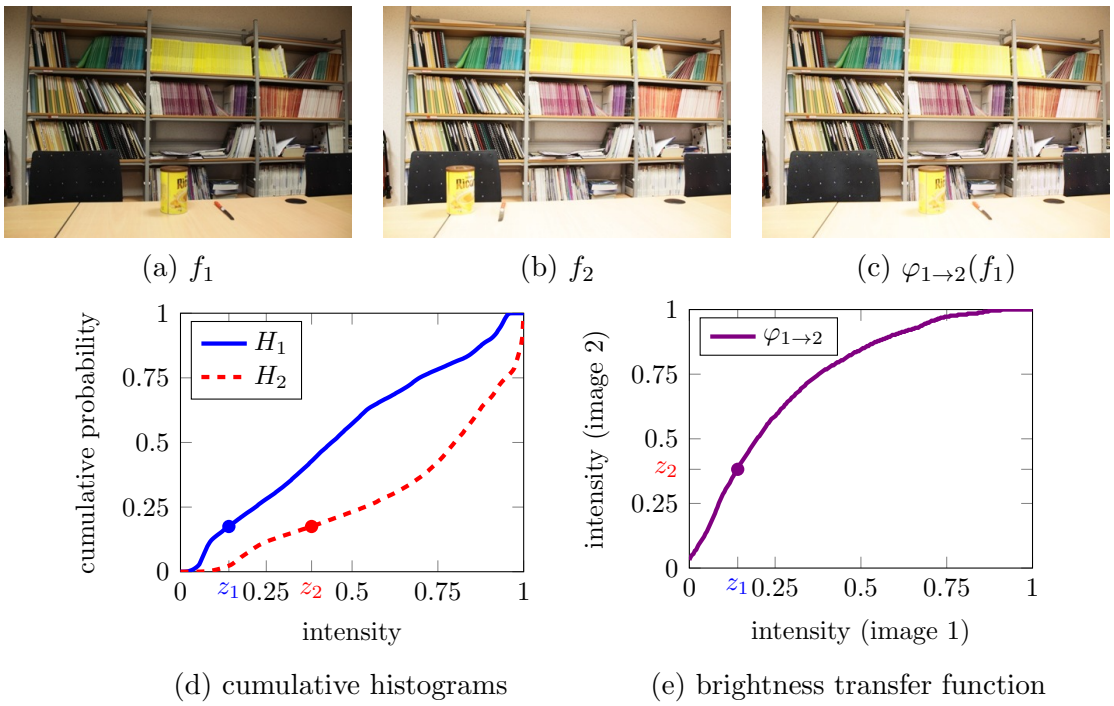


Figure 5.11: Illustration of cumulative histograms and brightness transfer function. For didactic reasons, we consider here only the green colour channel. (a) First image f_1. (b) Second image f_2. (c) Transferred image $\varphi_{1\to2}(f_1)$. (d) Cumulative histograms H_1 and H_2 of first and second image. (e) Corresponding BTF $\varphi_{1\to2}$ that maps the intensity values from f_1 to f_2. In this example, the intensity value z_1 is mapped to z_2; cf. Equation (5.35) and (5.36). In this way, the cumulative histogram of the transferred image in (c) matches H_2.

In this way, the (cumulative) histogram of the transferred image $\varphi_{1\to2}(f_1)$ approximates the (cumulative) histogram of the second image f_2. We illustrate this in Figure 5.11.

Now, having understood how to compute the BTF between an image pair, we first select a reference image f_r from the exposure set. Then, we compute the brightness transfer functions between this reference image and all other images in the set with the discussed histogram specification. Next, we apply those BTFs to transfer the intensity values of the reference to the other images, and compare the resulting values. If the difference is too large, this region is likely to produce a ghosting artefact in the fusion result and thus, should be excluded from the fusion process.

However, whenever an intensity value is under- or over-saturated in the reference image f_r, the output of the BTF does not provide reasonable information. In fact, in such cases, it is more reasonable to transfer the pixel values the other way around, i.e. from the tested image to the reference f_r, because this pixel might be less saturated. Accordingly, we apply a *bidirectional* check, and compute for each image f_i a binary

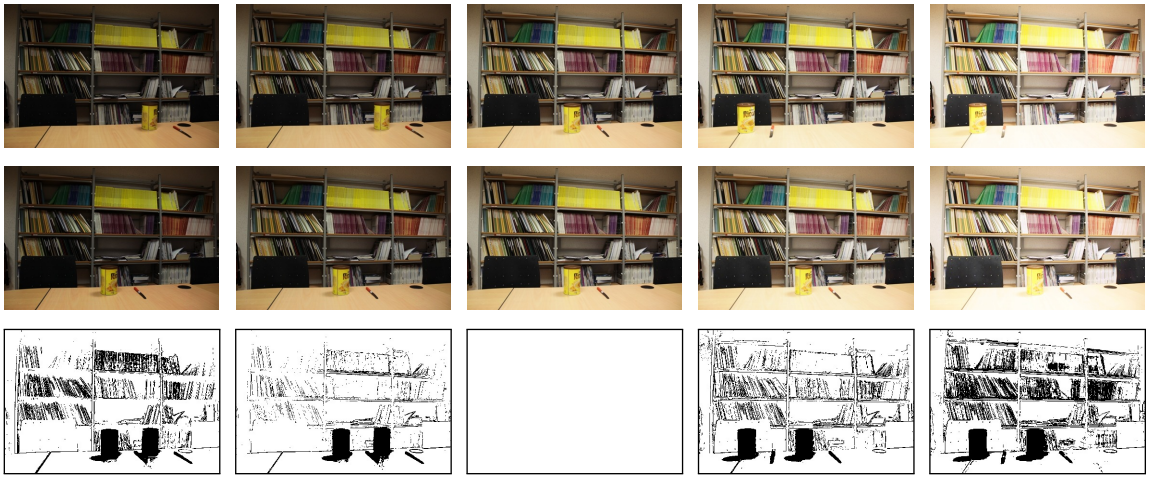


Figure 5.12: BTF-based deghosting. *Top*: Input images f_i with reference image f_r in the centre [SS12]. *Middle*: Transferred images $\varphi_{r\to i}(f_r)$ where $\varphi_{r\to i}$ is the BTF from the reference image f_r to the image f_i. *Bottom*: Computed binary ghost maps w_i, where black pixels indicate ghost-prone regions (black boundary added for visibility reasons).

ghost map $w_i \colon \Omega \to [0,1]$ as follows:

$$w_i(\boldsymbol{x}) \;=\; \begin{cases} 1 & \text{if } |b(\boldsymbol{x})| < \theta\,, \\ 0 & \text{else}\,, \end{cases} \tag{5.37}$$

where $\theta > 0$ is a user-defined threshold parameter. Here, $b(\boldsymbol{x})$ measures the distance of the transferred intensity value and the observed one in a bidirectional way, i.e.

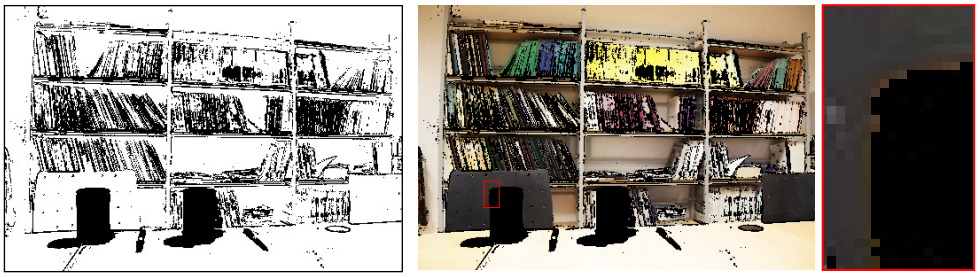
$$b(\boldsymbol{x}) \;=\; \begin{cases} \varphi_{r\to i}\big(f_r(\boldsymbol{x})\big) - f_i(\boldsymbol{x}) & \text{if } |f_r(\boldsymbol{x}) - \tfrac{1}{2}| < |f_i(\boldsymbol{x}) - \tfrac{1}{2}|\,, \\ \varphi_{i\to r}\big(f_i(\boldsymbol{x})\big) - f_r(\boldsymbol{x}) & \text{else,} \end{cases} \tag{5.38}$$

such that we always transfer the brightness value that is closer to the middle grey value $\frac{1}{2}$, i.e. least saturated. In case of colour images, we consider the Euclidean distance in the CIE-Lab colour space (cf. Section 2.2.2), and use the lightness channel L for the bidirectional decision in (5.38).

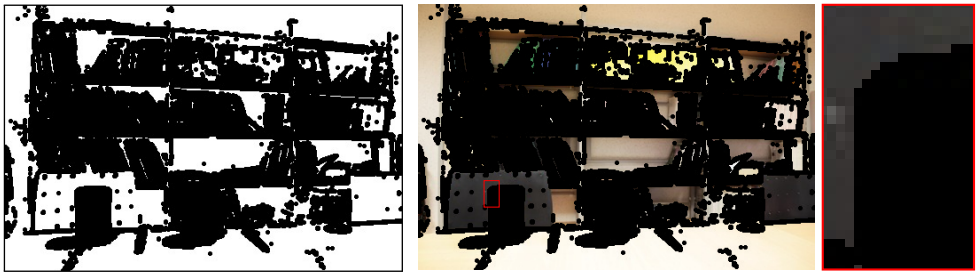
Figure 5.12 depicts an example exposure set, the transferred images $\varphi_{r\to i}(f_r)$, and the corresponding ghost maps w_i. Obviously, the ghost map w_r of the reference image f_r is by construction 1 (white) everywhere.

5.4.3 Modifications

The depicted ghost maps demonstrate that our technique provides a suitable ghost removal. However, in the following we propose two modifications that improve our deghosting results even further.



(a) origninal ghost map (b) input image overlaid with (a)



(c) eroded ghost map (d) input image overlaid with (c)

Figure 5.13: Erosion of ghost map with disc-shaped structuring element with radius $r_{\mathcal{B}}$ set to 0.5% of the image diagonal. The applied erosion successfully removes remaining ghosting artefacts at the edge of the can; cf. zooms in (b) and (d).

Erosion

Due to noise in the input images, the ghost map estimation might be also noisy and thus, some ghost-prone regions might be undetected; cf. e.g. boundary of the can in the zoom of Figure 5.13(b). As a remedy, we suggest to take into account neighbouring pixels for the decision if a pixel is prone to produce a ghost or not. To this end, we apply the morphological *erosion* operation on the previously computed ghost maps:

$$\tilde{w}_i(\boldsymbol{x}) \;=\; \inf\{w_i(\boldsymbol{x}+\boldsymbol{y}) \mid \boldsymbol{y} \in \mathcal{B}\}\,, \tag{5.39}$$

where $\mathcal{B} \subset \mathbb{R}^2$ is a so-called *structuring element*. It describes the neighbourhood region that is considered for determining the infimum. In this work, we apply a disc-shaped structuring element whose radius depends on the image diagonal. Figure 5.13 illustrates a result of this technique. The ghost regions are enlarged such that less wrong information will be used in the fusion process. Obviously, there is a trade-off between discarding too much information and keeping ghost-prone regions. To this end, the threshold parameter θ and the radius of the structuring element need to be adapted adequately; cf. upcoming Section 5.4.4.

At this point, we want to note that one could also think of an *opening* operation that is less dissipative in the sense that less pixels are indicated as ghost-prone. However, we follow a conservative approach in this work, i.e. we prefer to indicate some ghost-free regions as ghost-prone as vice-versa. Thus, we apply erosion instead of opening in this setting. Due to similar reasons, we also refrain from a patch-based consistency

check in this work, where the decision of a single pixel depends on its neighbourhood.

Intensity Exchange and Adjacent Ghost Removal

Very dark and very bright images contain large under- and over-saturated regions, respectively. Due to this, the BTF estimation by means of histogram specification might provide unreasonable results. In this regard, the bidirectional approach can be seen as a remedy. Nevertheless, the BTF estimation might be difficult in both directions in certain cases. As remedy, we make use of the observation that two adjacent images (close in brightness) allow the best BTF estimation. Hence, we propose an *adjacent deghosting* approach: We start with the reference image f_r and compute ghost maps for the adjacent images f_{r-1} and f_{r+1}, as discussed above. Here, we assume that the exposure series is ordered with increasing overall brightness. Next, we exchange the ghost-prone regions, i.e. where $w_i(\boldsymbol{x}) = 0$, with the help of the transferred reference image. Subsequently, we continue with this procedure where the ghost-corrected image $r-1$ serves as reference for the image $r-2$. Similarly, the ghost-corrected image $r+1$ is the reference for $r+2$. We repeat this procedure until the ghost maps for every image are computed.

The discussed intensity exchange has a further important benefit: Our ghost-corrected images can directly serve as input for exposure fusion. There is no need to adapt the existing fusion technique. More specifically, it is not required to include the actual ghost maps in the fusion approach. In fact, the deghosted images can be seen as a standard input series, free of ghosts. This is an essential advantage of our approach compared to related ghost removal techniques, which simplifies its usage.

5.4.4 Experiments

In Figure 5.14, we juxtapose the exposure fusion results of the previously considered exposure sets without (Figure 5.14 (a)) and with (Figure 5.14(b)) the discussed ghost removal procedure. Especially, the zooms illustrate the usefulness of the proposed technique. The ghosting artefacts are reduced significantly, and the final fusion results offer a good quality.

Next, we additionally test our ghost removal approach for three different image stacks of the deghosting benchmark set of Tursun et al. [TAEE15]. Figure 5.15 shows the corresponding input images and our ghost-corrected versions. Here, the images are separated by ± 2 *exposure values* (EVs). Indeed, our deghosted images demonstrate that our approach provides results of high quality such that the ghost-corrected images are ready to be used in any exposure fusion method. Due to space restrictions, we depict here only three example exposure sets. However, we want to note that our ghost removal technique generalises well to other image sets. Moreover, we applied the same parameter setting for all experiments: The threshold parameter θ is set to 3^2, and we set the radius $r_{\mathcal{B}}$ of the structuring element to 0.5% of the image diagonal.

[2]Please note the we measure here the pixel difference in the CIE-Lab colour space; cf. Section 2.2.2.

(a) without ghost removal (b) with proposed ghost removal

Figure 5.14: Comparison of exposure fusion results (a) without and (b) with proposed ghost removal. Especially the zooms indicated by the red rectangles illustrate the good performance of our technique.

Figure 5.16 compares the corresponding exposure fusion results using the uncorrected input images (Figure 5.16(a)) and our ghost-corrected versions (Figure 5.16(b)). The fusion results illustrate the sound performance of our ghost removal technique. While the fusion of the original images produces severe ghosting artefacts, the fusion of the processed images is free of such undesirable ghosts; cf. zooms in Figure 5.16.

5.4.5 Limitations and Discussion

As discussed, we apply a rather selective ghost removal produce. In fact, we prefer to indicate ghost-free regions as likely to produce ghosts as vice-versa. On the one hand, this allows to produce fusion results that are free of any ghosting artefacts. On the other hand, this strategy might be too dissipative, i.e. too much information might be lost. One remedy to tackle this problem is a careful selection of the parameters θ and $r_{\mathcal{B}}$ such that this information loss is minimised. Moreover, we apply a binary ghost decision: If some pixel might produce a ghost, it is completely excluded in the fusion process. Certainly, this is the right way to go for creating ghost-free fusion

results. However, it might be beneficial to include the confidence about this decision into the ghost maps. In this regard, ghost maps with continuous values that reflect the probability of producing a ghost seem to be reasonable. This is part of future work.

Moreover, our ghost removal approach is tailored to exposure series where we can assume a global brightness change caused by changing the exposure time or the CRF. However, e.g. in the context of multispectral image fusion, it might be appropriate to also model *local* brightness changes. One possibility to implement this could be a similar clustering scheme as proposed in [DSV+14].

Furthermore, we have seen that the discussed bidirectional BTF-based consistency check enables a desirably robust ghost removal technique. Hence, a natural idea would be to transfer this concept to an alignment strategy. In this regard, we also implemented and evaluated a variational optic flow approach that incorporates a bidirectional BTF-based constancy assumption; cf. also [HG11,BUC15]. However, the computed optic flow fields offered in most cases an inferior quality compared to the CRT-based approach presented in Section 5.3. On the one hand, this again illustrates the generally high performance of our CRT-based alignment. On the other hand, this shows the need for a more sophisticated handling of saturated image regions when transforming pure intensity values; cf. also [BML15]. Nevertheless, this is a promising topic for future work. Apart from that, we propose a conceptually different approach in the upcoming section of this chapter, which can be seen as a successful alternative to such a hand-tailored technique.

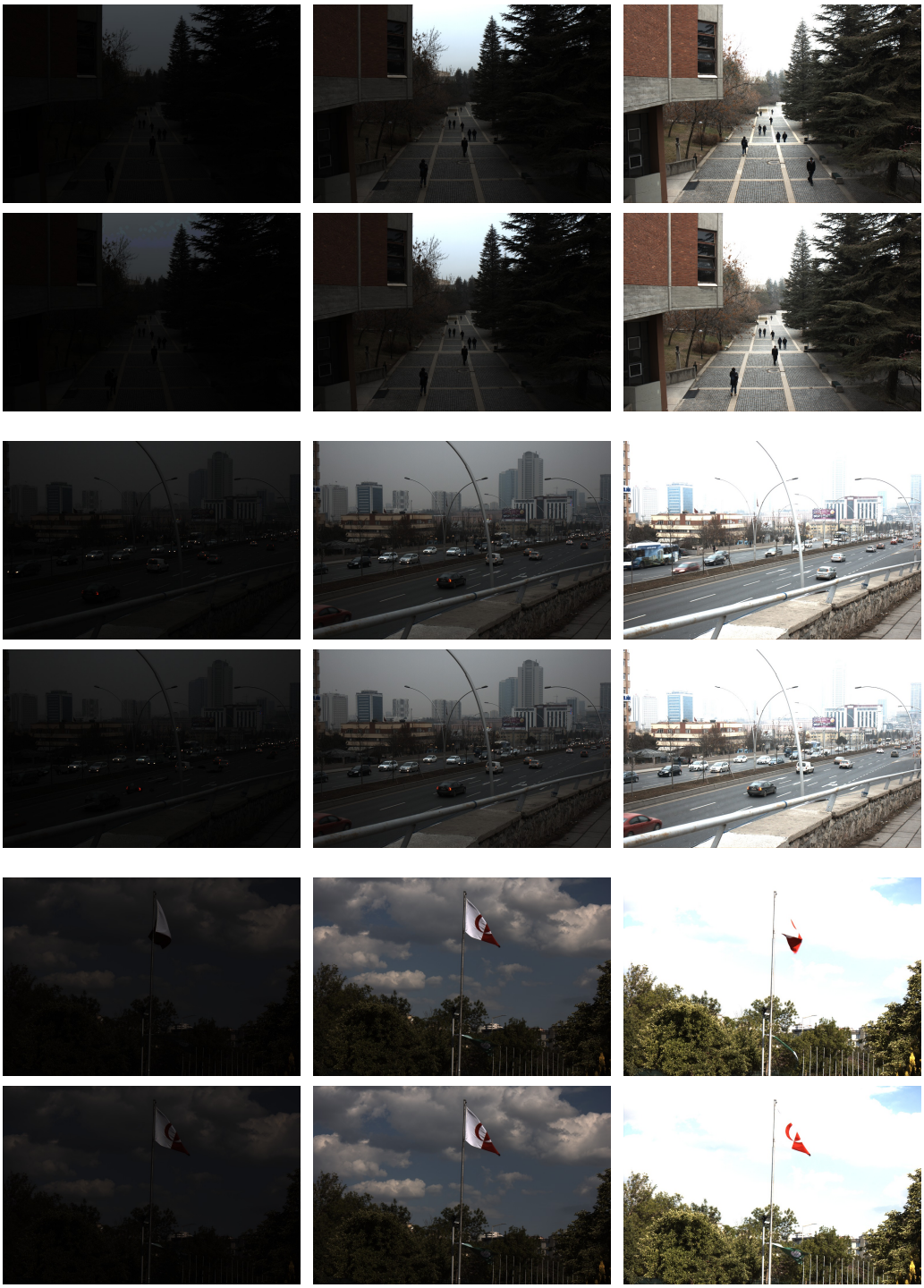
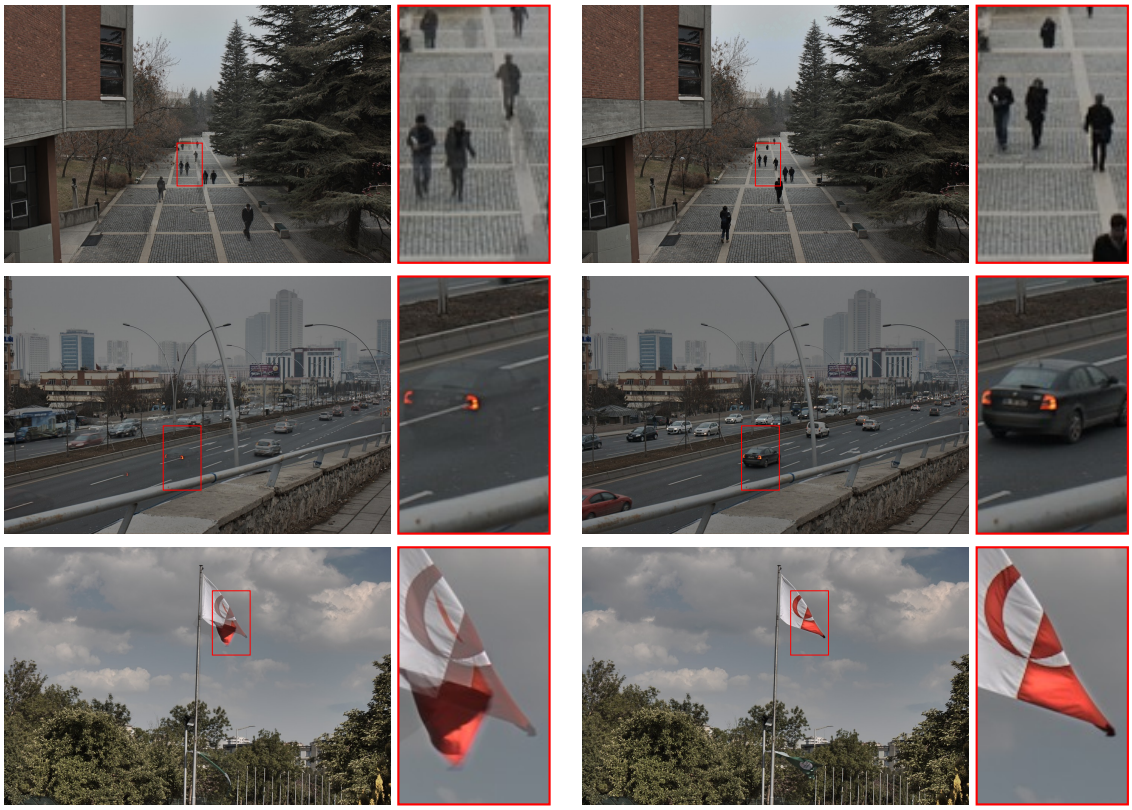


Figure 5.15: Ghost removal results for three exposure series of the benchmark dataset of Tursun et al. [TAEE15]. *Odd rows*: Input images, separated by ±2 EV. *Even rows*: Ghost-corrected images produced by our proposed ghost removal technique. Here, the central image is chosen as reference.



(a) without ghost removal (b) with proposed ghost removal

Figure 5.16: Comparison of exposure fusion results with and without proposed ghost removal. (a) Fusion of original input images (*odd rows* in Figure 5.15). Ghosts are clearly visible. (b) Fusion of ghost-corrected images (*even rows* in Figure 5.15). Ghost artefacts are eliminated; cf. zooms indicated by red rectangles.

5.5 Simultaneous HDR and Optic Flow Computation

Many modern cameras have an *auto exposure bracketing* (AEB) mode that creates a set of images with varying exposure times. This option of capturing differently exposed *low dynamic range* (LDR) images can be exploited to recover a larger dynamic range of the scene that goes beyond the contrast range of the sensor. The resulting *high dynamic range* (HDR) images are of great importance for many visual computing applications. First of all, it allows to visualise real HDR content on specific HDR monitors that offer a visual perception close to the real-world. Moreover, the HDR information generally allows to produce LDR images with an arbitrary exposure time and camera response function in a postprocessing step. Similarly, tone mapping techniques enable the photographer to produce overall well-exposed LDR images with a nice artistic appearance based on the HDR content. Also several image processing applications such as image deconvolution can profit from HDR information since the brightness values are ensured to behave linearly and saturation problems are reduced or even eliminated. A further application that benefits from HDR information is image-based modelling, where HDR values are important to recover specific surface properties. One example is the estimation of the *bidirectional reflectance distribution function* (BRDF) by means of several images; see e.g. [LKG+03]. For a more detailed discussion of HDR imaging, the displaying of HDR data and its usefulness for visual computing applications we refer the interested reader to the extensive textbook of Reinhard et al. [RHD+10].

To produce HDR composite images, standard multi-exposure methods in the literature such as [MP95, DM97, MN99] require perfectly aligned input images. As we have seen, in a practical scenario this requirement is often not fulfilled: Even small camera shakes or moving objects may lead to severe artefacts in the final result. In Section 5.3, we have already discussed optic flow-based alignment strategies that do not make use of the exposure times or the camera response function. Consequently, such approaches are especially useful in the exposure fusion setting where those quantities are assumed to be unknown. On the contrary, in this section we present a method that is specifically tailored to the multi-exposure HDR reconstruction setting. In particular, we propose a technique that makes efficient use of this additional knowledge about the imaging process.

Main Contributions. Most previous research on HDR imaging of dynamic scenes tries to compensate for the present motion by preprocessing the LDR images *before* combining them. However, since the alignment task and the HDR reconstruction highly influence each other, we propose to solve them in a *joint* ansatz that benefits from mutual interactions. Moreover, in contrast to existing methods, we are not only interested in a nice-looking image with high contrast, but also in an accurate representation of the scene motion. On the one hand, this opens the possibility for further computer vision tasks that are based on precise motion estimations such as inter-frame computation or scene analysis. On the other hand, in this way we aim

at only merging pixels in the composite that belong to the same object. This is particularly important if the physical correctness of the computed HDR irradiance values is required.

Our goal is to pursuit this joint approach by presenting the first fully coupled approach that simultaneously computes

(i) an aligned HDR irradiance map,

(ii) dense and accurate optic flow fields for each input image.

In particular, we propose an energy functional that is simultaneously minimised w.r.t. these two quantities. Although our functional is composed of relatively simple and intuitive assumptions, the resulting overall model turns out to be very powerful. Indeed, we show that the inherent feedback among the HDR irradiance map and the optic flow is highly beneficial for the quality of *both* estimates.

Section Outline. We start with a discussion of the relation to previous work in Section 5.5.1. Next, we present our variational approach for the simultaneous computation of the HDR image and the optic flow fields in Section 5.5.2. After explaining our algorithmic realisation in Section 5.5.3, we evaluate the performance of our approach in Section 5.5.4. Finally, we discuss possible limitations in Section 5.5.5.

5.5.1 Relation to Previous Work

The pure HDR reconstruction problem with perfectly aligned LDR images has been widely researched in the last two decades; see e.g. [MP95, DM97, MN99, RBS03, GN03, GN04, GAW+10] or the book of Reinhard et al. [RHD+10] for a broad overview. More recently, there has also been a lot of research tackling the problem of motion in the HDR acquisition pipeline; cf. discussion of related work in Section 5.2.

In particular, our approach can be attributed to the image alignment methods that try to register all parts of the images. This generally allows to merge all available information in the HDR composite, and in this way to make optimal use of the input data. Contrary to early approaches (e.g. [War03]) that assume a single global transformation for each image pair, we perform a local registration of the images under generally arbitrary motion patterns. Most related to our idea is the patch matching approach of Sen et al. [SKY+12]. In contrast to a decoupled pre-alignment, they are the first to *jointly* estimate the aligned input images and the HDR composite. However, in contrast to their method we do not only output aligned images. In fact, we also compute dense accurate representations of the scene motion (*optic flow*) which may serve as input for further tasks. Moreover, our model incorporates the camera-specific response function explicitly. First, this allows us to refrain from preliminary transformations of the LDR images to the linear domain as required in the method of Sen et al. [SKY+12]. Second, and even more importantly, it allows our optimisation to adapt to the shape of this function in an accurate way.

5.5.2 Variational Model

Observation Model

We assume that the unknown HDR irradiance image $u\colon \Omega \to \mathbb{R}$ is related to the observed LDR images f_i $(i = 1, \dots, n_f)$ via the *camera response function* (CRF) Φ in the following way:

$$f_i(\boldsymbol{x}) \;=\; \Phi(t_i \cdot u(\boldsymbol{x}))\,, \tag{5.40}$$

where $\boldsymbol{x} = (x_1, x_2)^\mathsf{T}$ denotes the position on the 2D rectangular image domain $\Omega \subset \mathbb{R}^2$, and t_i is the exposure time applied for capturing the LDR image f_i. As mentioned in the introduction, varying the exposure times and merging the differently exposed images allows to reconstruct the HDR content of the scene; cf. Figure 1.8. In the following we assume that the CRF Φ maps the incident light energy $t_i \cdot u(\boldsymbol{x})$ to the interval $[0, 1]$. After this mapping, the values are quantised and stored with a certain amount of bits per brightness value.

Obviously, Equation (5.40) only holds for a stack of perfectly registered input images. Even small camera shakes or moving objects break this assumption. Hence, to account for such motion, we modify our model in the following way:

$$f_i(\boldsymbol{x} + \boldsymbol{v}_i(\boldsymbol{x})) \;=\; \Phi(t_i \cdot u(\boldsymbol{x}))\,, \tag{5.41}$$

where the optic flow vector field $\boldsymbol{v}_i = (v_{i1}, v_{i2})^\mathsf{T}\colon \Omega \to \mathbb{R}^2$ allows for each position $\boldsymbol{x}$ in the irradiance map u to specify the corresponding position $\boldsymbol{x} + \boldsymbol{v}_i(\boldsymbol{x})$ in the unaligned input image f_i. Thus, once the optic flow has been computed, the registration of the input images is straightforward. For the sake of readability, we will from now on omit the argument $\boldsymbol{x}$ of the optic flow $\boldsymbol{v}_i$ and the HDR image u.

Energy Functional

Let us now embed the discussed observation model as data term into a variational framework. To this end, we develop an energy functional whose minimiser yields the irradiance map u as well as the optic flow fields $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{n_f}$. In particular, our energy consists of three main components:

$$E(u, \boldsymbol{v}_1, \dots, \boldsymbol{v}_{n_f}) \;=\; \sum_{i=1}^{n_f} D_i(u, \boldsymbol{v}_i) \;+\; \alpha \cdot \sum_{i=1}^{n_f} R_{\boldsymbol{v}}(\boldsymbol{v}_i) \;+\; \beta \cdot R_u(u)\,. \tag{5.42}$$

With the discussed observation model in (5.41), every LDR image yields one data term

$$D_i(u, \boldsymbol{v}_i) \;=\; \int_\Omega \Psi\Big(\big(f_i(\boldsymbol{x} + \boldsymbol{v}_i) \;-\; \Phi(t_i\, u)\big)^2\Big)\, \mathrm{d}\boldsymbol{x} \tag{5.43}$$

that rewards small cost if the measured brightness value f_i at position $\boldsymbol{x} + \boldsymbol{v}_i$ is close to the *predicted* brightness value that is computed by $\Phi(t_i\, u)$. Furthermore, we penalise the difference between those two quantities with the subquadratic function

$$\Psi(z^2) \;=\; \sqrt{z^2 + \varepsilon^2}\,, \qquad \varepsilon > 0\,. \tag{5.44}$$

This provides a robustness w.r.t. outliers, caused e.g. by occlusions or noise [Hub81, BA91, Coh93]. Please note that we apply a separate penalisation of each individual data term $D_i(u, v_i)$. In this way, we account for the fact that the positions of such outliers are often not correlated between the different input frames. In particular, the applied penalisation provides a beneficial behaviour w.r.t. ghosting artefacts. We deepen this discussion later by analysing the Euler-Lagrange equations.

Next, the regularisation terms

$$R_{\boldsymbol{v}}(\boldsymbol{v}_i) \;=\; \int_\Omega \left(|\boldsymbol{\nabla} v_{i1}|^2 + |\boldsymbol{\nabla} v_{i2}|^2\right) \mathrm{d}\boldsymbol{x} \qquad \text{and} \qquad R_u(u) \;=\; \int_\Omega |\boldsymbol{\nabla} u|^2 \,\mathrm{d}\boldsymbol{x} \tag{5.45}$$

reward smooth flow fields and a smooth irradiance map, respectively. Here, $\boldsymbol{\nabla} \coloneqq (\partial_x, \partial_y)^\mathsf{T}$ denotes the gradient operator. Moreover, the positive parameters α and β allow to steer the respective amount of smoothing.

We select the image with the fewest saturated pixels to be the reference, to which the final irradiance map should be aligned. Consequently, the flow field $\boldsymbol{v}_r$ corresponding to the reference frame f_r is obviously identical to $\mathbf{0}$ everywhere.

Euler-Lagrange Equations

According to the calculus of variations (cf. Section 2.3), the minimiser of the energy in (5.42) must necessarily fulfil the Euler-Lagrange equations. With the abbreviation

$$\Psi_i' \;\coloneqq\; \Psi'\big((f_i(\boldsymbol{x} + \boldsymbol{v}_i) - \Phi(t_i\, u))^2\big) \tag{5.46}$$

they are for the flow variables v_{i1} and v_{i2} $(i = 1, \ldots, n_f)$ given by

$$\Psi_i' \cdot \left(f_i(\boldsymbol{x} + \boldsymbol{v}_i) - \Phi(t_i u)\right) \cdot \partial_x f_i(\boldsymbol{x} + \boldsymbol{v}_i) - \alpha\, \Delta v_{i1} \;=\; 0\,, \tag{5.47}$$

$$\Psi_i' \cdot \left(f_i(\boldsymbol{x} + \boldsymbol{v}_i) - \Phi(t_i u)\right) \cdot \partial_y f_i(\boldsymbol{x} + \boldsymbol{v}_i) - \alpha\, \Delta v_{i2} \;=\; 0\,. \tag{5.48}$$

For the irradiance part, we obtain

$$\sum_{i=1}^{n_f} \Big(\Phi'(t_i u) \cdot t_i \cdot \Psi_i' \cdot \left(\Phi(t_i u) - f_i(\boldsymbol{x} + \boldsymbol{v}_i)\right)\Big) \;-\; \beta\, \Delta u \;=\; 0\,. \tag{5.49}$$

The corresponding homogeneous Neumann boundary conditions read

$$\boldsymbol{n}^\mathsf{T} \boldsymbol{\nabla} v_{i1} \;=\; 0\,, \quad \boldsymbol{n}^\mathsf{T} \boldsymbol{\nabla} v_{i2} \;=\; 0\,, \quad \text{and} \quad \boldsymbol{n}^\mathsf{T} \boldsymbol{\nabla} u \;=\; 0\,, \tag{5.50}$$

where $\boldsymbol{n}$ is the outer normal vector at the boundary of Ω.

Discussion

In contrast to previous HDR reconstruction methods, we define our data terms (5.43) in the *intensity domain* and not in the irradiance domain, i.e. we do *not* apply the inverse camera response function. This seems at first glance like a minor difference. However, this is in fact an important design choice as we will see in the following paragraphs.

Selective Weighting. To compute the irradiances u in (5.49), the term $\Phi'(t_i\, u) \cdot t_i$ provides an intuitive weighting of the different LDR images. First, the derivative of the camera response function Φ' is a natural confidence measure of the intensity values that accounts for quantisation noise and saturated pixels (cf. [MP95]): The smaller the derivative of the CRF, the larger the quantisation noise and the less the resulting weight of the considered brightness value. Second, the additional weighting by the exposure time t_i accounts for the fact that images taken with longer exposure times contain less sensor noise and thus, should get a higher weight (cf. [RBS03]). In other words, larger exposure times stretch the irradiance range in such a way that small values are quantised more finely, which consequently yields less under-saturated pixels. Contrary to the irradiance computation, the discussed weighting term $\Phi'(t_i\, u) \cdot t_i$ does *not* show up in the Euler-Lagrange equations of the optic flow in (5.47) and (5.48). In fact, this is an advantageous property that naturally arises from the carefully designed energy, too: If the estimated irradiance or rather the predicted brightness value $\Phi(t_i\, u)$ suggests that a pixel should be saturated in the i-th input image, we do not discard this information. Instead, we make optimal use of this knowledge by implicitly enforcing the matching of under- and over-saturated pixels.

Inherent Ghost Removal. Besides the described discriminative weighting of the intensity values, the term $\Phi'(t_i\, u) \cdot t_i$ plays an additional important role in the irradiance computation. In fact, it intrinsically performs a ghost removal by an intuitive *plausibility check*: The estimated irradiance u yields the predicted light energy $t_i\, u$. If this prediction is larger than 1, Φ' vanishes and we do not consider the corresponding brightness value f_i in the irradiance computation since it is physically not meaningful. In addition, this ghost removal behaviour is substantially supported by the robustified data terms (5.43). From an optic flow point of view, the penaliser Ψ increases the robustness w.r.t. outliers. For instance, such outliers might be caused by noise or occluded regions, i.e. parts of the reference image that are not visible in the other frames or vice versa. However, in our setting, the nonlinear terms Ψ_i' also appear in the computation of the irradiance map in (5.49). Here, they play a different but nevertheless important role: If the motion estimation is incorrect in some parts of the image, the brightness value $f_i(\boldsymbol{x} + \boldsymbol{v}_i)$ differs significantly from its prediction $\Phi(t_i\, u)$. In fact, the proposed penalisation function accounts for these cases and weights those terms down that would cause artefacts in the HDR image. This leads to ghost-free images which further help to improve the motion estimation, and the other way around.

Both features, i.e. the selective weighting and the inherent deghosting, share similarities to existing HDR imaging and deghosting approaches that are tailored to those tasks explicitly. In contrast, in our case, those features naturally follow from our carefully designed energy functional. In this way, both successful features are inherently incorporated in our model, and support the alignment as well as the HDR reconstruction process. In particular, instead of performing the alignment and the ghost removal in two subsequent steps (cf. Section 5.3 and 5.4), our joint approach benefits form the induced feedback between both tasks.

Anisotropic Modification

The optic flow regularisation terms $R_{\boldsymbol{v}}(\boldsymbol{v}_i)$ in (5.45) lead to the Laplacian terms Δv_{i1} and Δv_{i2} of the Euler-Lagrange equations in (5.47) and (5.48). They provide a linear isotropic diffusion of the flow fields that is space-invariant and equal in all directions. However, such a smoothing is not always desirable. In fact, we want to reduce smoothing across object structures if the motion differs between the objects. To this end, we propose to adapt the amount of smoothing to the local structure in a nonlinear anisotropic way. Inspired by [ZBW11b], we present a *joint irradiance- and flow-driven* smoothing that is steered by 2×2 diffusion tensors $\boldsymbol{D}_{\boldsymbol{v}}(u, \boldsymbol{v}_i)$. More specifically, we construct these tensors by the normalised vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ that point along and across edges of the evolving HDR image u. To this end, we compute those vectors as the orthonormal eigenvectors of the irradiance structure tensor (cf. [FG87])

$$G_\rho * \big(\boldsymbol{\nabla}(G_\sigma * u) \, \boldsymbol{\nabla}(G_\sigma * u)^\mathsf{T} \big) \,, \tag{5.51}$$

where $G_\rho *$ and $G_\sigma *$ denote convolutions with a Gaussian of standard deviation ρ and σ, respectively. The corresponding eigenvalues of $\boldsymbol{D}_{\boldsymbol{v}}(u, \boldsymbol{v}_i)$ determine the amount of smoothing. With the Charbonnier diffusivity [CBAB94]

$$g(z^2) \;=\; \frac{1}{\sqrt{1 + z^2/\lambda^2}} \,, \qquad \lambda > 0 \tag{5.52}$$

we compute those eigenvalues as follows:

$$g\big((\boldsymbol{r}_j^\mathsf{T} \boldsymbol{\nabla} v_{i1})^2 + (\boldsymbol{r}_j^\mathsf{T} \boldsymbol{\nabla} v_{i2})^2 \big) \quad \text{where} \quad j \in \{1, 2\} \,. \tag{5.53}$$

Finally we exchange the homogeneous diffusion terms Δv_{i1} as well as Δv_{i2} by their anisotropic counterparts

$$\operatorname{div}(\boldsymbol{D}_{\boldsymbol{v}}(u, \boldsymbol{v}_i) \, \boldsymbol{\nabla} v_{i1}) \qquad \text{and} \qquad \operatorname{div}(\boldsymbol{D}_{\boldsymbol{v}}(u, \boldsymbol{v}_i) \, \boldsymbol{\nabla} v_{i2}) \,, \tag{5.54}$$

where div denotes the divergence operator. The reliance on the evolving HDR image instead of the LDR input images (irradiance- vs. image-driven) is highly beneficial. In fact, it allows for adapting to object structures that are hardly visible in the single LDR images, but clearly distinctive in the HDR composite which combines all information from the whole exposure sequence.

In a similar way, we also apply such an anisotropic modification to the smoothness term of the HDR irradiance map:

$$\operatorname{div}(\boldsymbol{D}_u(u) \, \boldsymbol{\nabla} u) \,, \tag{5.55}$$

where the eigenvectors of the irradiance diffusion tensor $\boldsymbol{D}_u(u)$ are, as above, given by eigenvectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. Here, we compute the corresponding eigenvalues by

$$g\big((\boldsymbol{r}_j^\mathsf{T} \boldsymbol{\nabla} u)^2 \big) \qquad \text{where} \quad j \in \{1, 2\} \,. \tag{5.56}$$

Thanks to its edge-preserving property, this anisotropic diffusion of the irradiances is particularly beneficial in the presence of noise or under a high light sensitivity setting of the camera. In fact, our regularisation approach can be interpreted as an extension of the work by Rameshan et al. [RCV11] who showed that an incorporation of (isotropic) regularisation in the HDR fusion process is preferable to a decoupled denoising of the input images before composing the HDR image.

Colour Images

For didactic reasons, we have restricted ourselves to greyscale images so far. However, a modification of our method to colour images is straightforward: First, we perform a joint robustification of all colour channels, i.e. we sum up the channels in the argument of Ψ in (5.43). Second, we use the combined structure tensor to determine the eigenvectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ of the diffusion tensors (cf. [Di 86, WS99]):

$$\sum_{\mathrm{C}\in\{\mathrm{R,G,B}\}} G_\rho * \left(\boldsymbol{\nabla}(G_\sigma * u_\mathrm{C})\, \boldsymbol{\nabla}(G_\sigma * u_\mathrm{C})^\mathsf{T}\right) . \tag{5.57}$$

Moreover, we add up the squared directional derivatives of all colour channels in the argument of the diffusivity function g in (5.56) to compute the eigenvalues of $\boldsymbol{D}_u$.

5.5.3 Minimisation

We discretise the Euler-Lagrange equations in (5.47), (5.48) and (5.49) with finite differences on a rectangular grid with uniform grid sizes. Here, we apply the approach of Weickert et al. [WWW13] with $\alpha_{\text{stencil}} = 0.4$ and $\gamma_{\text{stencil}} = 1.0$ in order to discretise the diffusion expressions that originate form our anisotropic modification; cf. also Section 5.A. This results in a nonconvex and nonlinear system of equations. Similar to [BBPW04], we transform this nonconvex problem to a series of convex subproblems. To avoid being trapped in local minima and to handle large motions, we embed this series into a coarse-to-fine pyramid-based approach; cf. also Section 5.3.4. Consequently, in each step we solely compute small incremental values of v_1, v_2, and u. These increments are then successively used to update the solutions from the previous iterations. Hereby, all optic flow vectors are initialised with $\mathbf{0}$ while we determine the initialisation of the irradiance map by means of the reference frame f_r. With this procedure, we have to solve sparse nonlinear equations systems on each pyramid level. To this end, we apply our FSI scheme from Chapter 3. Moreover, we constrain the irradiance values to the physically plausible positive range, which additionally counteracts possible linearisation errors. The resulting solver is well-suited for fast parallel computing on the GPU.

5.5.4 Experiments

Our experimental evaluation consists of two main parts. First, we consider a synthetic HDR scene. The main reason for the use of such synthetic data is the availability of

ground truth data that is particularly required for a quantitative comparison of the optic flow fields. In the second part, we demonstrate the performance of our method on a real-world exposure series. In order to appropriately visualise the computed HDR irradiance maps, we apply the popular tone mapping operator of Fattal et al. [FLW02][3]. Furthermore, we applied a common nonlinear CRF and appropriate exposure times for the synthetic scene to ensure that the produced LDR images are good representatives of real LDR images. For the real-world scene, we calibrated the CRF in advance with a set of perfectly aligned LDR images following the method of Grossberg and Nayar [GN04].

Synthetic Data Set

The three input images of our synthetic *Balloon* data set are depicted in Figure 5.8(a–c). During the acquisition, the camera is rotating clock-wise while the balloon is moving towards the upper right corner. Moreover, the exposure times of the images vary by ± 3 *exposure values* (EVs), and every LDR image contains large over- and under-saturated regions, respectively. As discussed in Section 5.3.5, we created this scene in *Blender* and computed the ground truth flow fields with our *Blender* plug-in.

Let us now compare our approach to state-of-the-art HDR alignment methods from the literature. In our evaluation, we want to focus on methods that do not only aim at rejecting moving objects but try to align all parts of the input images. Otherwise, it would not be possible to reconstruct the HDR content of moving objects. Additionally, we want to stress that we are also especially interested in the underlying scene motion: We consider the motion estimation not only as a side-product, but rather as a valuable information about the scene. In particular, we compare our method to the optic flow-based technique of Zimmer et al. [ZBW11a] and the patch-based methods of Sen et al. [SKY+12] and Hu et al. [HGPS13]. Since Zimmer et al. and Hu et al. do not output an HDR image directly, we used their aligned images as input for the HDR reconstruction algorithm of Robertson et al. [RBS03][3].

Let us first compare our approach with the state-of-the-art method for optic flow-based HDR registration by Zimmer et al. [ZBW11a]. Figure 5.17 shows that while the method of Zimmer et al. is not able to detect the motion of the balloon, our approach yields reliable flow fields. This observation is underlined by Table 5.4 that rates the quality of the computed flow fields in terms of the *average angular error* (AAE) [BFB94] (cf. Equation (5.30)). Here, we also see that our joint method clearly outperforms the CRT-based optic flow method discussed in Section 5.3 which is specifically tailored to such illumination changes. It is worth mentioning that also further pre-alignment tests with top-ranked optic flow methods produced unsatisfactory results. This demonstrates that our proposed *coupled computation of the HDR irradiances and the optic flow is clearly preferable to a decoupled optic flow-based pre-alignment*. In fact, our joint approach marks the state-of-the-art of optic flow computation in this HDR setting.

[3]implementation provided by *pfstools* [MKMS07]

Table 5.4: Quantitative comparison of the computed optic flow fields for the synthetic data set in Figure 5.17. Best result is written in bold face.

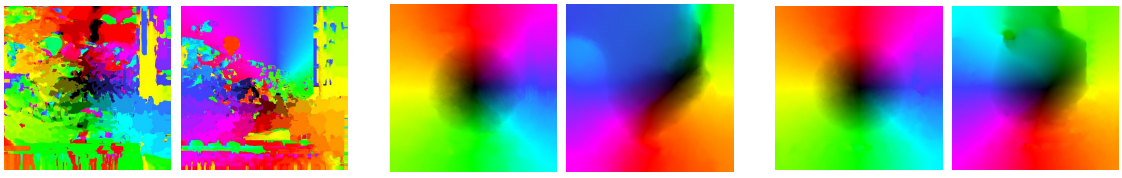
method	AAE (in degree)
Sen et al. [SKY+12]	not available
Hu et al. [HGPS13]	44.20
Zimmer et al. [ZBW11a]	12.60
CRT-based flow (Section 5.3)	5.17
proposed	**3.47**

Concerning the irradiance maps in Figure 5.18, we observe that an additional ghost removal as postprocessing step is necessary for the approach of Zimmer et al. In contrast, our approach contains such deghosting features inherently. This is very advantageous: Obviously, it removes ghosts in the evolving HDR images. On top of that, ghost-free HDR irradiance maps yield better motion estimations; consider for instance the irradiance-driven smoothing or the explicit enforcing of the mapping of saturated pixels. In this way, our joint optimisation and the induced mutual interactions successively improve both estimates and produce accurate final results.

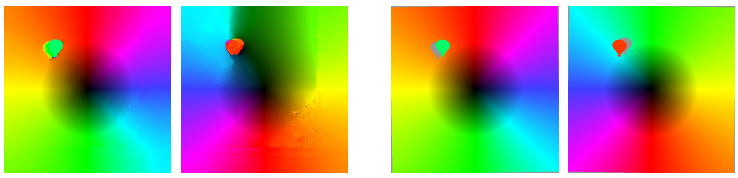
As mentioned in Section 5.5.1, Sen et al. [SKY+12] also follow the idea of a joint computation of the HDR image and aligned LDR images. However, due to the patch-based nature of their approach they do not compute a meaningful displacement map with subpixel precision. Thus, it is not possible to accurately specify the scene motion which may serve as input for further computer vision tasks. Regarding the HDR image, we observe artefacts especially in regions that are saturated in one of the LDR images like the sky region; cf. last column in Figure 5.18. Here, their algorithm produces unpleasant results with low contrast.

Hu et al. [HGPS13] do not explicitly exploit the benefits of a joint alignment and irradiance computation since their main goal is the pure alignment of the image stack — not the HDR composite. The final image looks visually similar to our result. Nevertheless, differences can again be found especially in the sky region. Furthermore, the noisy displacement maps in Figure 5.17(a) illustrate that the method does not always combine the correct pixels. This may result in visually attractive HDR images but not in accurate, physically meaningful irradiances.

Last but not least, it is important to note that Sen et al. [SKY+12] and Hu et al. [HGPS13] highly weight down the influence of all other input images whenever the reference image is not saturated. On the one hand, this increases the robustness of their algorithms. On the other hand, this discards valuable information which results in less accurate HDR maps. Hence, we refrain from such a high preference of the reference image and aim at merging all available data to recover a maximal dynamic range.



(a) Hu et al. [HGPS13] (b) Zimmer et al. [ZBW11a] (c) CRT flow (Section 5.3)



(d) proposed (e) ground truth

Figure 5.17: Computed flow fields that describe the motion from Figure 5.9(a,c) to the reference frame in Figure 5.9(b). Compared to all other methods, the proposed approach produces the most accurate displacement maps. Thanks to our joint optimisation, we are able to estimate the motion of the balloon (besides the camera motion) with high precision even though it is hardly visible in some of the LDR images.

Real-World Data Set

Let us now experimentally verify our observations on real-world data. To this end, we consider a sequence of five images each separated by one exposure value. In Figure 5.9(a–c), we show three of those five images where again camera and object motions are present. While all competing methods have problems especially with the specularities on the window sill, our final HDR image shows less artefacts; cf. Figure 5.19. In particular, we observe a severe drawback of patch-based approaches: The algorithm of Sen et al. [SKY+12] clones content from the background region to the specular region on the sill, which leads to unrealistic results. This clearly illustrates that optic flow is important to produce a reliable representation of scene motion and thereby to correctly combine corresponding pixels. In this regard, our flow fields in Figure 5.20 again show highly accurate motion estimations.

5.5.5 Limitations and Discussion

Similar to Section 5.3, there is one general limitation of all warping-based optic flow methods that unavoidably carries over to our approach: If the sequence is undersampled in time such that small objects undergo a too large displacement, the motion estimation tends to fail. Clearly, it is the task of future research to overcome this drawback. To address this issue, an interesting way might be the design of appropriate hardware with sufficiently small time intervals between the differently exposed images. Other approaches incorporate sparse feature matching to handle such large displacements; see e.g. [BM11]. For this purpose, an application of our illumination-robust CRT features might be promising.

Moreover, we have seen how to benefit from the exposure times and the camera

response function. On the one hand, this allows for a joint model with inherent features that benefit from the induced mutual feedback. On the other hand, this limits the applicability of the proposed technique since the input images are assumed to follow the HDR imaging process, and the CRF and exposure times have to be known. Hence, depending on the input data available prior knowledge either the more generally applicable two step approach of CRT-based alignment and the discussed ghost removal step or the presented joint approach is to prefer.

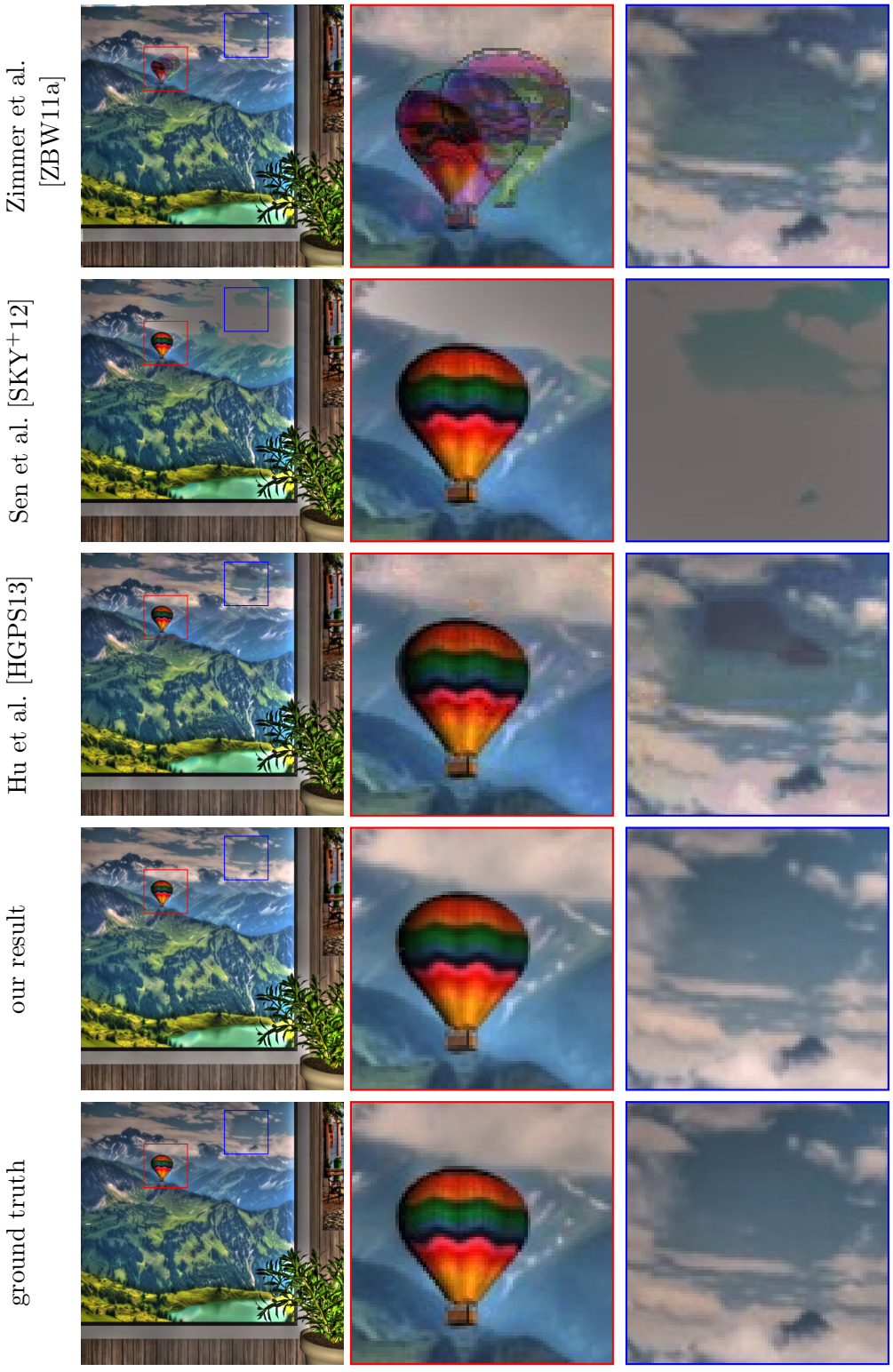


Figure 5.18: Visual comparison of the HDR images. The full HDR irradiance maps (*first column*) with their zoom-ins (*second and third column*) illustrate that our method produces the most accurate irradiances.

(a) Zimmer et al. [ZBW11a]

(b) Hu et al. [HGPS13]

(c) Sen et al. [SKY+12]

(d) our result

Figure 5.19: Visual comparison of the HDR results for the real-world example from Figure 5.9. While the competing methods have especially problems with the highlights on the window sill, our method yields a convincing irradiance map.

(a) flow from reference to short exposure

(b) flow from reference to long exposure

Figure 5.20: Computed optic flow fields for the real-world example from Figure 5.9. Compared to the CRT-based flow from Section 5.3 (cf. Figure 5.9(d,e)), our joint approach allows for more precise displacement maps. In particular, with our joint approach we are able to recover the movement of the small-scale rabbit's ears accurately. This posed a big challenge for the CRT-based flow.

5.6 Summary and Conclusions

5.6.1 Alignment with Complete Rank Transform

In the first part of this chapter, we have presented a variational approach for the general alignment of differently exposed frames. To this end, we have introduced two novel morphologically invariant descriptors, namely the *complete rank transform* (CRT) and the *complete census transform* (CCT). Compared to their predecessors, both transforms carry as much local image information as possible without leaving this class of invariance. Our descriptors are well suited for pattern matching applications where highest accuracy is desired, such as optic flow estimation. We have demonstrated this within a variational framework, where we achieved state-of-the-art results for the KITTI benchmark. In this regard, our experiments illustrated a comparable performance of CCT and CRT. Hence, due to its higher computationally efficiency, we recommend our CRT-based flow as method of choice. It not only provides high quality results in case of drastic illumination changes, but it is also competitive if only mild brightness changes are present. In particular, we recommend it as a general alternative to census-based optic flow approaches. Last but not least, we have illustrated the performance of our technique for the alignment of exposure series.

As we have seen, our CRT-based approach provides not only good results under strong illumination changes, but it also yields competitive flow fields when no brightness changes are present. Nevertheless, an adaptive spatially varying decision about which constancy assumption to choose seems to be a promising direction for future work. This would allow to adapt the amount of invariance to the underlying illumination conditions.

From a more general point of view, we have designed an optic flow approach that is robust to illumination changes by incorporating morphologically invariant features. In a similar way, the systematic design and embedding of features that are robust to different degradations such rotations or scale changes would be highly interesting to investigate in future work.

5.6.2 Ghost Removal with Brightness Transfer Function

While our CRT-based optic flow approach provides accurate displacement maps for most image regions, due to violations of the model assumptions such as occlusions or under- and over-saturated brightness values we cannot expect perfect alignment results everywhere. Unfortunately, even small misalignments may cause severe ghosting artefacts when fusing the images to an overall well-exposed composite image. As remedy, we have proposed a ghost removal technique that identifies image regions that are likely to produce ghosts in the output image. More specifically, we based our approach on the concept of a brightness transfer function that relates the intensity values from frame to frame. We determined those BTFs by means of image histograms which provide a robustness w.r.t. to small scene motion (cf. [GN03]). With the computed BTFs, we performed a bidirectional consistency check to identify ghosts. We

further increased the robustness of our approach by applying erosion and an adjacent ghost removal strategy. Last but not least, the proposed intensity exchange leads to output images that can directly severe as input for a fusion approach without modification. In this regard, we intentionally refrain from the knowledge of the exposure times and the camera response function such that our method is well-suited for the general exposure fusion setting.

Our ghost removal scheme provides robustly good results for many image sequences. However, to systematically compare our results to related ghost removal approaches, a more extensive evaluation would be desirable. Certainly, this has to be best performed on the basis of perceptual user studies; see e.g. [TAEE15]. Additionally, an evaluation of different deghosting results by means of objective quality metrics is an important step for future work. Recently, Tursun et al. [TAEE16] introduced such a quality metric for the HDR setting. It would be interesting to analyse our approach with this metric, and for instance to automatically adapt the parameters of our ghost removal technique accordingly.

5.6.3 Simultaneous HDR and Optic Flow Computation

Provided further knowledge about the image set, we proposed an alternative approach to the discussed two-step procedure of optic flow-based pre-alignment and subsequent ghost removal. More specifically, we have presented the first fully coupled approach that simultaneously computes an HDR irradiance map as well as accurate dense displacement fields. Especially the introduced feedback in this joint computation of the irradiances and the optic flow turned out to be highly beneficial. We have demonstrated that the proposed strategy is clearly preferable to a decoupled optic flow-based pre-alignment presented in previous work. In contrast to existing patch-based approaches, our dense displacement maps represent the underlying motion reliably. This is particularly of great importance for tasks that require accurate motion estimations with subpixel precision, such as object removal, artificial motion blur generation or bullet time effects in an HDR context. To conclude, we do not only produce a nice-looking HDR image, but we additionally compute accurate motion estimations of dynamic scenes. In summary, whenever optic flow is required in an HDR setting, we recommend to consider a *simultaneous* computation of the HDR image and the flow fields as demonstrated by this work.

Thanks to the reliance on optic flow we believe that our method is especially suited for continuous image sequences. Hence, an application to the production of HDR videos seems to be promising (cf. for instance [KUWS03, KSB+13]). In this regard, a modelling of a temporal smoothness in the sense of [Nag90, WS99, VBVZ11, BML15] could be profitable since we can assume a continuous scene and camera motion. Moreover, the explicit integration of the camera response function in our model generally allows to additionally compute such a CRF and/or estimate the exposure times on-the-fly if unknown. First experiments in this direction show very promising results. Last but not least, we want to point out the need for image deblurring in the HDR

setting as a possible future research topic. Since the usage of large exposure times inherently causes motion blur in the input images, it would be interesting to investigate possible adaptations of our joint HDR alignment approach to such a difficult scenario.

5.A Finite Differences for Anisotropic Diffusion

Following Weickert et al. [WWW13], we explain in this section a finite difference discretisation of the anisotropic diffusion term

$$\operatorname{div}(\boldsymbol{D}\,\boldsymbol{\nabla} u) \tag{5.58}$$

with a space-variant symmetric positive definite diffusion tensor

$$\boldsymbol{D} \;:=\; \begin{pmatrix} a(\boldsymbol{x}) & b(\boldsymbol{x}) \\ b(\boldsymbol{x}) & c(\boldsymbol{x}) \end{pmatrix}. \tag{5.59}$$

To obtain a suitable discretisation of (5.58), Weickert et al. [WWW13] propose to consider the following energy functional

$$E(\boldsymbol{u}) \;=\; \int_\Omega \boldsymbol{\nabla} u^\top \boldsymbol{D}\,\boldsymbol{\nabla} u \,\mathrm{d}\boldsymbol{x} \;=\; \int_\Omega \left(a\,u_{x_1}^2 + 2\,b\,u_{x_1}u_{x_2} + c\,u_{x_2}^2\right) \mathrm{d}\boldsymbol{x} \tag{5.60}$$

since its Euler-Lagrange equation is given by

$$-\operatorname{div}(\boldsymbol{D}\,\boldsymbol{\nabla} u) \;=\; 0\,. \tag{5.61}$$

In particular, with the abbreviations

$$\boxed{u_{\bullet\bullet}} \;:=\; \frac{u_{i+1,j} - u_{i,j}}{h_1} \qquad \text{and} \qquad \boxed{\overline{u}_{\bullet\bullet}} \;:=\; \frac{u_{i+1,j+1} - u_{i,j+1}}{h_1} \tag{5.62}$$

as well as

$$\boxed{u_{\bullet\bullet}} \;:=\; \frac{u_{i,j+1} - u_{i,j}}{h_2} \qquad \text{and} \qquad \boxed{u_{\bullet\bullet}} \;:=\; \frac{u_{i+1,j+1} - u_{i+1,j}}{h_2} \tag{5.63}$$

Weickert et al. discretise (5.60) at intermediate grid points $(i+\frac{1}{2}, j+\frac{1}{2})$ as follows:

$$(u_{x_1})^2_{i+\frac{1}{2},j+\frac{1}{2}} \;\approx\; \frac{1-\alpha}{2} \cdot \left(\boxed{u_{\bullet\bullet}}\,\boxed{u_{\bullet\bullet}} + \boxed{\overline{u}_{\bullet\bullet}}\,\boxed{\overline{u}_{\bullet\bullet}}\right) + \alpha \cdot \boxed{u_{\bullet\bullet}}\,\boxed{\overline{u}_{\bullet\bullet}}, \tag{5.64}$$

$$(u_{x_2})^2_{i+\frac{1}{2},j+\frac{1}{2}} \;\approx\; \frac{1-\alpha}{2} \cdot \left(\boxed{u_{\bullet\bullet}}\,\boxed{u_{\bullet\bullet}} + \boxed{u_{\bullet\bullet}}\,\boxed{u_{\bullet\bullet}}\right) + \alpha \cdot \boxed{u_{\bullet\bullet}}\,\boxed{u_{\bullet\bullet}}, \tag{5.65}$$

$$(u_{x_1}u_{x_2})_{i+\frac{1}{2},j+\frac{1}{2}} \;\approx\; \frac{1-\beta}{4} \cdot \left(\boxed{u_{\bullet\bullet}}\,\boxed{u_{\bullet\bullet}} + \boxed{\overline{u}_{\bullet\bullet}}\,\boxed{u_{\bullet\bullet}}\right) + \frac{1+\beta}{4} \cdot \left(\boxed{u_{\bullet\bullet}}\,\boxed{u_{\bullet\bullet}} + \boxed{\overline{u}_{\bullet\bullet}}\,\boxed{u_{\bullet\bullet}}\right). \tag{5.66}$$

Next, a discretisation of (5.58) is given by differentiation of the discrete energy functional. Consequently, this discretisation results in the following stencil for a pixel (i,j) and its 3×3 neighbourhood:

$$\frac{1}{2h^2}\cdot \begin{array}{|c|c|c|} \hline [(\beta-1)b+\alpha(a+c)]_{i-\frac{1}{2},j+\frac{1}{2}} & \begin{array}{c}[(1-\alpha)c-\alpha a-\beta b]_{i+\frac{1}{2},j+\frac{1}{2}} \\ +\,[(1-\alpha)c-\alpha a-\beta b]_{i-\frac{1}{2},j+\frac{1}{2}}\end{array} & [(\beta+1)b+\alpha(a+c)]_{i+\frac{1}{2},j+\frac{1}{2}} \\ \hline \begin{array}{c}[(1-\alpha)a-\alpha c-\beta b]_{i-\frac{1}{2},j+\frac{1}{2}} \\ +\,[(1-\alpha)a-\alpha c-\beta b]_{i-\frac{1}{2},j-\frac{1}{2}}\end{array} & \begin{array}{c}-\,[(1-\alpha)(a+c)-(\beta-1)b]_{i+\frac{1}{2},j+\frac{1}{2}} \\ -\,[(1-\alpha)(a+c)-(\beta+1)b]_{i+\frac{1}{2},j-\frac{1}{2}} \\ -\,[(1-\alpha)(a+c)-(\beta+1)b]_{i-\frac{1}{2},j+\frac{1}{2}} \\ -\,[(1-\alpha)(a+c)-(\beta-1)b]_{i-\frac{1}{2},j-\frac{1}{2}}\end{array} & \begin{array}{c}[(1-\alpha)a-\alpha c-\beta b]_{i+\frac{1}{2},j+\frac{1}{2}} \\ +\,[(1-\alpha)a-\alpha c-\beta b]_{i+\frac{1}{2},j-\frac{1}{2}}\end{array} \\ \hline [(\beta+1)b+\alpha(a+c)]_{i-\frac{1}{2},j-\frac{1}{2}} & \begin{array}{c}[(1-\alpha)c-\alpha a-\beta b]_{i+\frac{1}{2},j-\frac{1}{2}} \\ +\,[(1-\alpha)c-\alpha a-\beta b]_{i-\frac{1}{2},j-\frac{1}{2}}\end{array} & [(\beta-1)b+\alpha(a+c)]_{i+\frac{1}{2},j-\frac{1}{2}} \\ \hline \end{array} \tag{5.67}$$

where we assume an equal grid size $h := h_1 = h_2$ in both dimensions. Please note that we refer to the parameters α and γ as α_{stencil} and γ_{stencil} throughout this work (except for this appendix to shorten the notation).

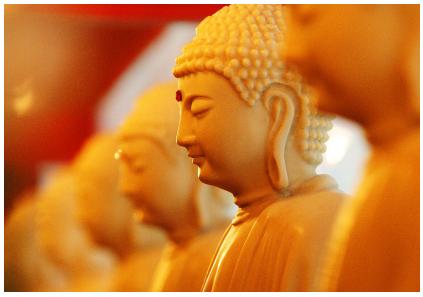
Weickert et al. show that stable discretisations are obtained with $\alpha \in [0, \frac{1}{2}]$ and $|\beta| \leq 1 - 2\alpha$. In the general case, the parameters α and β can also be space-variant. In particular, it turns out be to beneficial to adapt β to the sign of b, i.e. $\beta = \gamma(1 - 2\alpha)\,\text{sign}(b)$, where $\gamma \in [-1, 1]$. In fact, this discretisation of the anisotropic diffusion can be seen as a generalisation of various discretisation approaches. Moreover, it opens the door to superior discretisations. For more details, we refer the interested reader to the paper of Weickert et al. [WWW13].

Chapter 6

Depth from Focus with Anisotropic Second Order Regularisation

"Where your focus goes,
your energy flows."

Les Brown



pixabay.com

Contents

6.1 Introduction 131
6.2 Variational Depth from Focus 132
6.3 Anisotropic Second Order Coupling Model 157
6.4 Summary and Conclusions 171

Main parts of this chapter base on our work published in [BHW13, BHW15, HSW15].

6.1 Introduction

Especially in optical microscopy and macro photography, a typical problem is the limited depth of field of common imaging devices. Due to this, it is often not possible to capture a single image that is entirely sharp. A common remedy to address this problem is to take several photographs while varying the focal settings. In this way, the acquired image stack contains all required information to produce a single image that is desirably sharp. The task of combining the images of such a focal stack to an all-in-focus composite is referred to as *focus fusion*; cf. Section 4.4.4.

Similar to the previous chapter, further knowledge about the image stack allows to infer additional information about the captured scene. In particular, knowledge about the (relative) distances of the focal planes is highly beneficial. A focal plane describes the distance to the camera where scene objects are captured sharply. Hence, such

knowledge allows to infer depth information in addition to an all-in-focus composite image. In this regard, approaches that compute a depth map by means of a focal stack are referred to as *depth from focus* or *depth from defocus* techniques; cf. for instance the textbook of Chaudhuri and Rajagopalan [CR99][1]. More specifically, depth from focus methods make use of the in-focus regions, i.e. regions that are captured sharply, to determine the depth values. In contrast, depth from defocus techniques additionally infer depth information from the amount of blur. In fact, they model how the amount of blur relates to the distance to the focal plane. Incorporating such a relation requires to rely on sophisticated camera models that mimic the out-of-focus effect in a realistic way. Unfortunately, this usually renders depth from defocus methods rather complex and computationally expensive. In this regard, depth from focus approaches offer a better efficiency and applicability. Hence, we prefer such a variational depth from focus approach in this work. As it turns out, it is robust and computationally efficient while providing high quality results.

Moreover, in the context of depth map computation, several applications have demonstrated the benefits of second order regularisation. This is due to the fact that many real-world scenes can be well described by (piecewise) planar objects. Hence, we additionally introduce a novel second order regularisation strategy that adapts to the local image and depth structure in a beneficial anisotropic way. In this way, we are able to improve the results of our variational depth from focus approach even further.

Chapter Outline. This chapter consists of two main parts: In Section 6.2, we present our variational framework for depth from focus. Next, we introduce our novel anisotropic second order regularisation technique in Section 6.3. We conclude this chapter with a summary and outlook in Section 6.4.

6.2 Variational Depth from Focus

In this section, we present our variational approach for depth from focus. Given a set of differently focused images, we aim at an all-in-focus composite as well as an accurate depth map of the captured scene.

Main Contributions. The crucial point of all depth from focus methods is the decision about the in-focus areas. To this end, we investigate different in-focus measures that lead to initial depth maps. Such initial depth maps are prone to outliers and noise. Hence, we present a general variational framework that introduces a modern regularisation strategy on those depth maps. In fact, we assume that neighbouring pixels in the fused image belong to similar depth layers. Following this assumption, we smooth the depth maps with a sophisticated anisotropic diffusion process combined

[1]The terms *shape from focus* or *shape from defocus* are also frequently used in the literature for such approaches. In this work, we stick to the terms *depth from focus* or *depth from defocus*.

with robust data fidelity terms that are capable of incorporating multiple in-focus measures. Finally, the computed depth maps allow to fuse the input images in a straightforward way. Here, the explicit modelling of smooth transitions in depth provides not only more accurate depth maps, but also counteracts unpleasant seams in the final composite images. In fact, in this way, we are able to adequately handle pixels that are never captured totally in-focus since they lie between two focal planes.

Section Outline. We start with a survey of related work in Section 6.2.1. After outlining the general structure of our framework in Section 6.2.2, we discuss its individual components in Section 6.2.3, 6.2.4, and 6.2.5. Next, we evaluate our approach by means of experiments, and compare it to related techniques in Section 6.2.6. Finally, we discuss possible limitations in Section 6.2.7.

6.2.1 Related Work

Focus Fusion

An early example of focus fusion can already be found in the article of Adelson et al. [AAB+84] from 1984. Here, Laplacian pyramids [BA83] are applied to select the in-focus regions and to fuse the input images accordingly. Later, Burt and Kolczynski [BK93] generalise this technique to alternative pyramid representations. Related to such a pyramid-based approach are focus fusion methods that base on wavelet transforms; see e.g. [LMM95, LKTW04, FVB+04, DC06, LON+07]. Such techniques apply a multiscale transformation of the complete image stack. Next, all transformed images are combined by selecting the transform-based coefficients that have the highest probability of belonging to in-focus areas. Then, the composed multiresolution representation is transformed back to the spatial domain, which finally yields the desired all-in-focus image. In [ZB99], Zhang and Blum present a generic framework for multiscale image fusion and compare different approaches.

In contrast to such multiscale approaches, several methods directly work in the spatial domain. Here, the basic idea is to first select regions from all frames that are in focus, and then to combine them to one composite. To this end, Ghoshtasby [Gos06] performs a local blending of blocks of the input images. Similarly, Eltoukhy and Kavusi [EK03] apply a smoothness of the in-focus map w.r.t. two neighbouring frames. Moreover, several focus fusion methods employ machine learning techniques: For instance, Wang et al. [WMG10] consider a neural network, Wu et al. [WYP+13] propose a hidden Markov model, and Wan et al. [WZQ13] perform a principal component analysis for the focus fusion task. Furthermore, Wang et al. [WSF08] and Yuan et al. [YMG+14] explicitly model a smoothness of the output image within a variational framework. Comparably, Pop et al. [PLTB07] propose a PDE-based approach for focus fusion. Another successful idea is to segment the in-focus map such that each segment corresponds to one input frame. This is for instance accomplished by graph cut methods [ADA+04, XTAA04, LY08, LJW+14] or level set approaches [vCF06]. Af-

ter segmentation, the input images are combined appropriately e.g. by a gradient domain fusion technique such as in [ADA+04].

Depth from Focus

Indeed, depth from focus methods are closely related to focus fusion approaches. However, besides an all-in-focus composite image, depth from focus methods additionally aim at a reliable depth map of the captured scene. Already in 1968, Horn [Hor68] presented a method to measure the distance to the captured object by finding the focus setting where the image region is sharpest. In particular, he determined the sharpness of an image by means of the Fourier transform. Similar to Adelson et al. [AAB+84], Darrell and Wohn [DW88] consider Laplacian pyramids to estimate the depth of the captured scene. In the meantime, various sharpness measures have been proposed. A recent evaluation of such sharpness measures can be found in [PPG13a]. In [PPG13b], Pertuz et al. additionally discuss the reliability of such measures which reflects the probability of computing an accurate depth. Several of those measures are inspired by autofocus techniques that search for the best focus setting of the camera such that the produced image is desirably sharp. In this context, Tenenbaum [Ten70] considers the gradient magnitude as measure for sharpness. This measure is in the literature often referred to as *Tenegrad*. Jarvis [Jar67, Jar83] suggests the entropy, the variance, and the sum of absolute differences to measure sharpness. Analyses and evaluations of such sharpness criteria in the context of automatic focusing can be found in [SSNW83, GYL85, Kro88, SCN93, ST98, MXv14].

Finding the image where the sharpness of a specific image region is maximal yields a depth estimate for this region. Obviously, due to several distortions such as noise or edge bleeding those depth estimates might be inaccurate in some image parts; see e.g. the discussion in [NS92]. As a remedy, several approaches model a smoothness of the underlying depth map.[2] For instance, Nayar and Nakagawa [NN94] apply a Gaussian interpolation technique to smooth initial depth maps that are created with the sum of modified Laplacian as sharpness measure. In [SC95, YC99], the authors explicitly model piecewise planar and piecewise curved surfaces, respectively. Goldsmith [Gol00] smooth the depth map by iterative applications of median filters. Comparably, Shim and Choi [SC10] iteratively apply averaging filters to remove outliers. Moreover, several approaches employ machine learning techniques to produce smooth 3D shapes. Examples are neural networks [AC01], kernel regression [MC10], or genetic programming [MMC11]. In [Mah13, MBSC15], regularisation of the depth map is modelled by means of *total variation* (TV).

[2] Please note that such a smoothness assumption of the depth map along the image stack implicitly requires a knowledge of the relative distances of the focal planes. Otherwise, it would be unclear how to relate the images of the focal stack in terms of depth measurements. In other words, the knowledge of the relative focal plane distances allows to define the grid size in the depth dimension that is required for depth regularisation. This is important to keep in mind since it constitutes a major difference to focus fusion methods that refrain from such an assumption.

Related to a regularisation of the depth map, Aydin and Akgul [AMMS08] apply a bilateral filter-based technique to smooth the initial sharpness measures. Similarly, Mahmood and Choi [MC12a] consider 3D diffusion process for this task. Moreover, several approaches investigate the acquisition of the focal stack in order to perform an efficient depth sampling; see e.g. [XS93, VGT+11, MC12b, PGP15].

Depth from Defocus

Already Pentland [Pen87] and Grossmann [Gro87] discussed the possibility of not only considering sharp image information but to additionally infer depth information from the amount of blur in unsharp regions. To this end, so-called *depth from defocus* approaches explicitly model the relation between the focal setting of the camera and the amount of blur; see e.g. [EL93]. Variational depth from defocus methods can for instance be found in [JFS02, AVU08, PSSW14]. On the one hand, such depth from defocus approaches are able to extract more information from the input images than depth from focus techniques. This might lead to better depth maps while less input images are required. On the other hand, depth from defocus methods are more computationally demanding and often require additional information about the image acquisition system such as the point spread function. In this regard, depth from focus approaches as presented in this work offer a greater efficiency and applicability.

6.2.2 Depth from Focus Framework

Let $f_1, \dots, f_{n_f}$ denote the n_f differently focused input images. Here, we assume that the focal plane distance increases equidistantly from frame to frame. Our goal is to find a depth map $d\colon \Omega \to \mathbb{R}$ that selects for each location $\boldsymbol{x} \in \Omega$ the frame that is most in focus. To this end, our depth from focus approach consists of three main parts: In the first step (Section 6.2.3), we identify sharp image regions by means of specific in-focus measures. This allows to construct initial depth maps. However, due to possibly misleading in-focus cues and noise in the input data those initial depth maps might be unreliable in some image regions. Hence, in the core part of our method (Section 6.2.4), we propose a variational technique to smooth these initial depth maps. To this end, we model an anisotropic image- and depth-driven regularisation technique, and combine it with robust data fidelity terms. Finally, the computed depth map d allows to fuse the input images to an all-in-focus composite $u\colon \Omega \to \mathbb{R}$ in a straightforward way. In fact, we select for each position $\boldsymbol{x}$ the input image that corresponds to the computed depth $d(\boldsymbol{x})$; cf. Section 6.2.5.

6.2.3 Initial Depth Map Computation

To determine which image regions are in focus, various sharpness measures have been proposed; see e.g. [Ten70, SSNW83, SI85, Kro88, NN94, SCN93, XTAA04, HJ07, AK09, TCMY11, PPG13a] and references therein. All of those approaches share the idea that an extremum of a sharpness measure corresponds to the image that is most in

Table 6.1: Overview of applied in-focus measures. See text for further details.

in-focus measure	formula
m_1: gradient magnitude	$\lvert\boldsymbol{\nabla} f_\sigma\rvert$
m_2: trace of structure tensor	$\operatorname{tr}(\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma))$
m_3: determinant of structure tensor	$\det(\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma))$
m_4: Laplacian magnitude	$\lvert\Delta f_\sigma\rvert$
m_5: Frobenius norm of Hessian	$\lvert\boldsymbol{\mathcal{H}} f_\sigma\rvert_{\mathrm{F}}$
m_6: variance	$\frac{1}{\lvert\mathcal{N}(\boldsymbol{x})\rvert}\int_{\mathcal{N}(\boldsymbol{x})}\left(f_\sigma(\boldsymbol{y})-\bar{f}_\sigma(\boldsymbol{x})\right)^2\,\mathrm{d}\boldsymbol{y}$

focus. We store this information in an initial depth map $\tilde{d}\colon \Omega \to \mathbb{R}$. In particular, it specifies for each pixel a depth value that corresponds to the focal plane distance of the considered image.

In-Focus Measures

Our method is not limited to a specific sharpness measure. To illustrate this, we consider six different in-focus measures that are summarised in Table 6.1. To tackle noise, we presmooth the input images with a Gaussian of standard deviation σ.

Gradient Magnitude. Our first sharpness criterion is the gradient magnitude (see e.g. [Ten70]):

$$m_1(\boldsymbol{x}) \;\coloneqq\; \lvert\boldsymbol{\nabla} f_\sigma(\boldsymbol{x})\rvert\,, \tag{6.1}$$

where f_σ denotes a presmoothed input image, $\boldsymbol{\nabla} \coloneqq (\partial_{x_1}, \partial_{x_2})^\mathsf{T}$ the gradient operator, and $\lvert\cdot\rvert$ the Euclidean norm.

Trace of Structure Tensor. Following [LMW10, LLM11] we also consider the structure tensor [FG87]

$$\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma) \;\coloneqq\; G_\rho * \left(\boldsymbol{\nabla} f_\sigma\; \boldsymbol{\nabla} f_\sigma^\mathsf{T}\right)\,, \tag{6.2}$$

where $G_\rho *$ denotes convolution with a Gaussian of standard deviation ρ. To identify sharp image regions, we determine the trace of the structure tensor, i.e.

$$m_2(\boldsymbol{x}) \;\coloneqq\; \operatorname{tr}(\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma))\,. \tag{6.3}$$

We can interpret $\operatorname{tr}(\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma))$ as the sum of the eigenvalues of $\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma)$. Thus, this measure provides useful information about image features such as edges and corners.

Determinant of Structure Tensor. Similarly to the previous measure, we additionally consider the determinant of the structure tensor as sharpness criterion:

$$m_3(\boldsymbol{x}) \;\coloneqq\; \det(\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma))\,. \tag{6.4}$$

This can be seen as the product of the eigenvalues of the structure tensor. Hence, both eigenvalues should be nonzero for a high sharpness value. Furthermore, since two large eigenvalues of the structure tensor represent corners, this in-focus measure can be interpreted as a corner detector.

Laplacian Magnitude. The sharpness criteria discussed above base on first order image derivatives. In contrast, a measure of the second order variation is the absolute value of the Laplacian:

$$m_4(\boldsymbol{x}) \coloneqq |\Delta f_\sigma(\boldsymbol{x})| = |\partial_{x_1x_1} f_\sigma + \partial_{x_2x_2} f_\sigma| \,. \tag{6.5}$$

For completeness, we also want to mention the frequently applied *modified Laplacian* (see e.g. [NN94]) that is given by

$$|\partial_{x_1x_1} f_\sigma| + |\partial_{x_2x_2} f_\sigma| \,. \tag{6.6}$$

Obviously, this measure it not rotationally invariant. Hence, we refrain from the modified Laplacian as in-focus measure in this work.

Frobenius Norm of Hessian. As an additional second order sharpness criterion we investigate the Frobenius norm $|\cdot|_{\mathrm{F}}$ of the 2×2 Hessian $\boldsymbol{\mathcal{H}} f_\sigma$:

$$m_5(\boldsymbol{x}) \coloneqq |\boldsymbol{\mathcal{H}} f_\sigma|_{\mathrm{F}} = \sqrt{(\partial_{x_1x_1} f_\sigma)^2 + 2\,(\partial_{x_1x_2} f_\sigma)^2 + (\partial_{x_2x_2} f_\sigma)^2} \,. \tag{6.7}$$

It is a measure of local second order variation and is related to the image curvature.

Variance. Another frequently applied sharpness criterion is the local variance in some neighbourhood $\mathcal{N}(\boldsymbol{x})$ (see e.g. [SI85]):

$$m_6(\boldsymbol{x}) \coloneqq \frac{1}{|\mathcal{N}(\boldsymbol{x})|} \int_{\mathcal{N}(\boldsymbol{x})} \big(f_\sigma(\boldsymbol{y}) - \bar{f}_\sigma(\boldsymbol{x})\big)^2 \,\mathrm{d}\boldsymbol{y} \,, \tag{6.8}$$

where $\bar{f}_\sigma(\boldsymbol{x})$ is the mean value in $\mathcal{N}(\boldsymbol{x})$, i.e. $\bar{f}_\sigma(\boldsymbol{x}) \coloneqq \frac{1}{|\mathcal{N}(\boldsymbol{x})|} \int_{\mathcal{N}(\boldsymbol{x})} f_\sigma(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}$.

We implement those six sharpness measures in a straightforward way with central finite difference approximations. Subsequently, we determine for each pixel the frame where the in-focus measure takes its maximal value. For each criterion, this yields an initial depth map $\tilde{d}_i \colon \Omega \to \mathbb{R}$.

Confidence Function

Two types of locations cause severe problems while construction those initial depth maps: homogeneous regions that hardly have any texture and regions that are never captured in focus. Here, the sharpness criteria attain small values in all frames. Thus, the decision of the depth value is highly influenced by noise. To tackle this problem,

we propose a confidence function $w\colon \Omega \to \mathbb{R}$ that specifies the quality of the computed depth map. In particular, we consider the binary confidence function

$$w(\boldsymbol{x}) \;:=\; \begin{cases} 1 & \text{if } m(\boldsymbol{x}) \geq \theta\,, \\ 0 & \text{else.} \end{cases} \tag{6.9}$$

Accordingly, we treat values of a measure m that are larger or equal to the threshold parameter $\theta \geq 0$ as reliable. In contrast, smaller values are discarded since we assume they are too much influenced by noise, and thus not reliable.

Discussion. Obviously, also real-valued confidence functions that seem to offer a better discrimination of the computed sharpness measures appear to be a reasonable choice. However, the proposed *binary* function turned out to provide superior results in most experiments. Moreover, we also investigated further confidence indicators such as the variation (variance) of a specific sharpness criterion over the focal stack. However, also here the proposed confidence function that is based on the maximal value of the sharpness measure led to better results. Due to those reasons, we stick in the following to the simple confidence function presented in Equation (6.9).

6.2.4 Depth Map Regularisation

Indeed, the discussed confidence function helps to specify the quality of the initial depth maps, and in this way to detect unreliable estimations. However, we still need to determine which depth values should be used at pixels that are identified as unreliable. Moreover, the confidence function is only an indicator of the quality. The depth map may still be erroneous or noisy although indicated as reliable.

To address these issues, we propose a diffusion-based approach that allows to compute a dense and smooth depth map d, given the initial maps $\tilde{d}_i$ and their confidence functions w_i. In particular, we assume a spatial continuity between parts selected from different frames, i.e. neighbouring pixels most probably should be chosen from a similar depth level. As a first step to explain our model, we consider the following energy functional:

$$E(d) \;=\; \int_\Omega \Big(\sum_{i=1}^{n_{\tilde{d}}} w_i \cdot \Psi\Big((d-\tilde{d}_i)^2\Big) \;+\; \alpha \cdot |\boldsymbol{\nabla} d|^2 \Big)\, \mathrm{d}\boldsymbol{x}\,, \tag{6.10}$$

where the regularisation parameter α allows to steer the impact of the smoothness term $|\boldsymbol{\nabla} d|^2$, and $n_{\tilde{d}}$ describes the number of considered initial depth maps. The data term $\sum_{i=1}^{n_{\tilde{d}}} w_i \cdot \Psi\Big((d-\tilde{d}_i)^2\Big)$ models our similarity assumptions of the solution d and the initial maps $\tilde{d}_i$. Here, the use of multiple in-focus measures or rather initial depth maps is particularly beneficial when the selected measures supplement each other, i.e. when each performs best in different image regions. The summands of the data term are weighted with the corresponding confidence functions w_i. In addition,

in order to reduce the influence of outliers in the initial depth maps, we apply the subquadratic penalisation function

$$\Psi(z^2) \;=\; \sqrt{z^2+\varepsilon^2}\,, \tag{6.11}$$

with the small positive constant ε. As discussed, each initial depth map might offer a different quality in a specific image region; i.e. we cannot necessarily assume a strong correlation between the individual initial depth maps $\tilde{d}_i$. This motivates our *separate* penalisation of the individual data term parts in (6.10).

Euler-Lagrange Equation

Following the calculus of variations (cf. Section 2.3), the minimiser of the energy functional in (6.10) satisfies the Euler-Lagrange equation

$$\sum_{i=1}^{n_{\tilde{d}}} w_i \cdot \Psi'\Big((d-\tilde{d}_i)^2\Big) \cdot (d-\tilde{d}_i) \;-\; \alpha \cdot \Delta d \;=\; 0 \tag{6.12}$$

with the homogeneous Neumann boundary condition

$$\boldsymbol{n}^{\mathsf{T}} \boldsymbol{\nabla} d \;=\; 0\,, \tag{6.13}$$

where $\boldsymbol{n}$ is the outer normal vector on the image boundary $\partial\Omega$.

Anisotropic Modification

The smoothness term $|\boldsymbol{\nabla} d|^2$ leads to the linear isotropic diffusion term Δd in (6.12). It provides a smoothing that is space-invariant and equal in all directions. However, similar to Section 5.5.2, it may be beneficial to adapt the diffusion process to the local structure of the image. Thus, inspired by [NE86, Wei96, ZBW+09], we replace the discussed linear isotropic diffusion term by a nonlinear anisotropic variant. In fact, we propose an anisotropic *joint image- and depth-driven* regularisation technique: First, we determine the smoothing directions by the structures of the evolving all-in-focus image u, because this image is supposed to provide a richer directional information than the individual unsharp input images. Second, we adjust the amount of smoothing by the gradients of the computed depth map d.

More specifically, we determine the smoothing directions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ as the normalised eigenvectors of the structure tensor (cf. [FG87])

$$\boldsymbol{J}_\rho(\boldsymbol{\nabla} u_\sigma) \;=\; G_\rho * \big(\boldsymbol{\nabla} u_\sigma \; \boldsymbol{\nabla} u_\sigma^{\mathsf{T}}\big)\,. \tag{6.14}$$

In our upcoming experiments we set the standard deviations of the applied Gaussians to $\sigma = 1.0$ and $\rho = 1.3$. The vector $\boldsymbol{r}_1$ corresponds to the largest eigenvalue, and consequently points across edges of u. Accordingly, $\boldsymbol{r}_2$ points along edges. These vectors build the basis of our diffusion tensor

$$\boldsymbol{D} \;=\; (\boldsymbol{r}_1 \;\; \boldsymbol{r}_2) \begin{pmatrix} g\left((\boldsymbol{r}_1^{\mathsf{T}} \boldsymbol{\nabla} d)^2\right) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{r}_1^{\mathsf{T}} \\ \boldsymbol{r}_2^{\mathsf{T}} \end{pmatrix}, \tag{6.15}$$

where $g\colon \mathbb{R} \to [0,1]$ denotes the Charbonnier diffusivity [CBAB94]

$$g(z^2) = \frac{1}{\sqrt{1 + z^2/\lambda^2}} \tag{6.16}$$

with some contrast parameter $\lambda > 0$. The diffusion tensor in (6.15) effects a reduced smoothing across image structures (along $\boldsymbol{r}_1$) and a strong smoothing along them (along $\boldsymbol{r}_2$). Replacing Δd in (6.12) by its anisotropic counterpart $\operatorname{div}(\boldsymbol{D}\,\boldsymbol{\nabla} d)$ yields our final model

$$\sum_{i=1}^{n_{\tilde{d}}} w_i \cdot \Psi'\Big((d - \tilde{d}_i)^2\Big) \cdot (d - \tilde{d}_i) \;-\; \alpha \cdot \operatorname{div}(\boldsymbol{D}\,\boldsymbol{\nabla} d) \;=\; 0 \tag{6.17}$$

with the natural boundary condition

$$\boldsymbol{n}^{\mathsf{T}} \boldsymbol{D}\,\boldsymbol{\nabla} d = 0\,. \tag{6.18}$$

Implementation Details

We discretise Equation (6.17) on a rectangular pixel grid with uniform grid sizes. Accordingly, we apply a finite difference scheme for the space discretisation of the divergence expression. In particular, we compute for each pixel a 3×3 stencil following the approach of Weickert et al. [WWW13] with parameters $\alpha_{\text{stencil}} = 0.4$ and $\gamma_{\text{stencil}} = 1.0$; cf. also Section 5.A. This results in a large nonlinear system of equations which we solve with our FSI scheme from Chapter 3. As discussed, this scheme is well-parallelisable and hence well-suited for a fast GPU implementation. Furthermore, the runtime of the core part of our algorithm, the anisotropic depth map smoothing, is almost independent of the number of images in the focus stack. Favourably, this provides beneficial scaling properties of our algorithm to large image sets; cf. Section 6.2.6.

Extension to Colour Images

For didactic reasons, we have restricted ourselves to grey-valued images so far. However, the extension of our model to colour images is straightforward. Instead of using the structure tensor in (6.14) for a single colour channel, we apply the combined structure tensor (cf. [Di 86, WS99])

$$G_\rho * \sum_{\mathrm{C} \in \{\mathrm{R,G,B}\}} \boldsymbol{\nabla} u_{\sigma,\mathrm{C}}\; \boldsymbol{\nabla} u_{\sigma,\mathrm{C}}{}^{\mathsf{T}} \tag{6.19}$$

to determine the eigenvectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. Here, $u_{\sigma,\mathrm{C}}$ represents the colour channel C of the smoothed fused RGB image $\boldsymbol{u}_\sigma$. Please note that we still determine *one* joint depth map d for all channels.

6.2.5 Image Fusion

With the computed depth map d we are able to directly fuse the input colour images $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{n_f}$ to an all-in-focus composite $\boldsymbol{u} \colon \Omega \to \mathbb{R}^3$. In fact, we select for each position $\boldsymbol{x}$ the input image that corresponds to the computed depth value $d(\boldsymbol{x})$. Since our variational approach yields real-valued depth maps, we apply linear interpolation between the nearest input frames:

$$\boldsymbol{u}(\boldsymbol{x}) \;=\; (1 - \epsilon(\boldsymbol{x})) \cdot \boldsymbol{f}_{\lfloor d(\boldsymbol{x}) \rfloor} \;+\; \epsilon(\boldsymbol{x}) \cdot \boldsymbol{f}_{\lfloor d(\boldsymbol{x}) \rfloor + 1} \,, \tag{6.20}$$

where $\lfloor d(\boldsymbol{x}) \rfloor$ denotes the largest integer smaller or equal to $d(\boldsymbol{x})$, and $\epsilon(\boldsymbol{x}) \coloneqq d(\boldsymbol{x}) - \lfloor d(\boldsymbol{x}) \rfloor$.

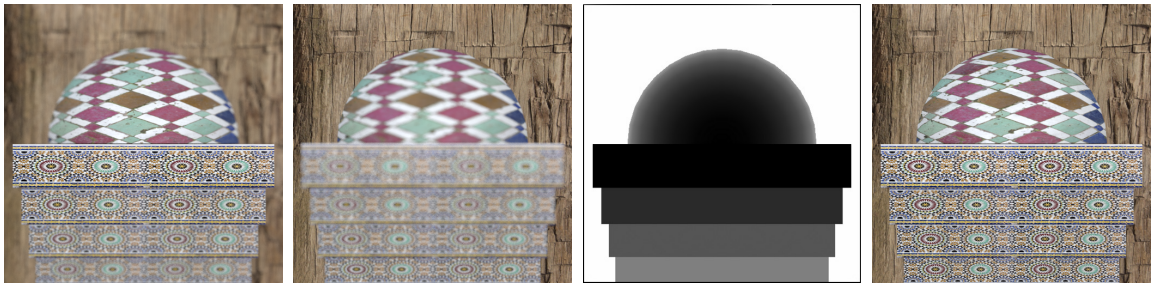
6.2.6 Experiments

Our experimental section consists of the following five main parts: First, we evaluate the discussed in-focus measures and demonstrate improvements provided by the proposed depth map smoothing technique. Second, we show that our anisotropic modification further enhances the quality of the fused images. Next, we illustrate that our method beneficially combines the information from multiple in-focus measures, and in this way yields even more accurate results. Moreover, we compare the performance of our approach to competing methods from the literature. Finally, we evaluate the runtime of our algorithm.

Evaluation of In-Focus Measures

To systematically evaluate the discussed sharpness criteria in terms of quality measures, we generated a synthetic focal stack with ground truth data; cf. Figure 6.1. To this end, we developed a *Blender* [Ble16] plug-in that makes efficient use of the internal rendering engine. This engine implements a realistic depth of field effect by a suitable camera model and ray casting techniques. The resulting focal stack consists of thirteen images with size 512×512, where the focal plane distance increases from frame to frame. In addition, we provide the ground truth depth map (Figure 6.1(c)) and the ground truth all-in-focus image (Figure 6.1(d)). In this way, we are able to measure the difference between the resulting fused images and the ground truth all-in-focus image that we produced with a pinhole camera model.

Results with Initial Depth Maps. First, we compare the quality of the initial depth maps created with the discussed in-focus measures; cf. Table 6.1. We depict the resulting depth maps and the fused images in Figure 6.2. In addition, the second column of Table 6.2 lists the corresponding quantitative results. In particular, we analyse the *mean squared error* (MSE) between the fused image and its ground truth. Here, we apply optimised parameters. We observe that the gradient (m_1), the Laplacian (m_4) and the Hessian (m_5) perform worst, while the trace (m_2) and the determinant (m_3) of the structure tensor as well as the variance (m_6) produce the most accurate



(a) Frame 1 (b) Frame 13 (c) depth map (d) all-in-focus image

Figure 6.1: Synthetic data set modelled and generated with our plug-in for *Blender* [Ble16]. *From left to right*: (a) Frame 1 with shortest focal plane distance. (b) Frame 13 with largest focal plane distance. (c) Ground truth depth map. The grey value indicates the distance to the camera; i.e. the brighter, the larger the distance (black boundary added for visibility reasons). (d) Ground truth all-in-focus image.

initial depth maps. This can be explained as follows: The sharpness criteria m_1, m_4, and m_5 are local, pointwise measures, while m_2, m_3, and m_6 inherently contain rich neighbourhood information; consider the outer Gaussian convolution in m_2 and m_3 and the neighbourhood size in m_6. Hence, the latter criteria already include valuable spatial information that leads to more reliable initial depth maps.

Results with Smoothed Depth Maps. In this paragraph, we investigate the behaviour of the sharpness measures when combined with the proposed regularisation techniques. To this end, Figure 6.3 depicts the final depth maps and the fused images after smoothing the initial depth maps with our linear isotropic model; cf. Equation (6.12). We observe that all depth maps improve substantially. The final depth maps are much more reliable. This observation is underlined by the error measures in Table 6.2. The MSEs of the fused images are significantly reduced for all in-focus measures, which clearly illustrates the usefulness of our depth map regularisation. Moreover, we see that the variation between the different in-focus measures is decreased compared to the initial depth maps. This is caused by our regularisation technique that incorporates rich spatial information in *every* depth computation.

Evaluation of Anisotropic Regularisation

Let us now compare the linear isotropic to the proposed nonlinear anisotropic regularisation strategy. To this end, we depict in Figure 6.4 the results with our anisotropic joint image- and depth-driven technique. The corresponding MSEs in Table 6.2 clearly demonstrate the benefits compared to the linear isotropic approach. In fact, it provides improvements of up to 24%. Moreover, we see that the trace of the structure tensor (m_2) and the Hessian (m_5) provide the best overall results in this setting.

Table 6.2: Results with initial and smoothed depth maps. The parameters are optimised w.r.t. the *mean squared error* (MSE) between the fused image and its ground truth. The best results for each measure are written in bold face.

in-focus measure	initial	isotropic	anisotropic
m_1: gradient magnitude	95.35	8.66	**8.39**
	($\sigma = 0.7$)	($\theta = 30$, $\alpha = 2$)	($\theta = 37$, $\alpha = 6$)
m_2: trace of structure tensor	3.59	3.57	**3.48**
	($\sigma = 0.2$, $\rho = 1.7$)	($\theta = 0$, $\alpha = 1$)	($\theta = 0$, $\alpha = 4$)
m_3: determinant of structure tensor	4.74	4.45	**4.17**
	($\sigma = 0.6$, $\rho = 1.8$)	($\theta = 0$, $\alpha = 0.5$)	($\theta = 0$, $\alpha = 8$)
m_4: Laplacian magnitude	23.08	4.61	**3.53**
	($\sigma = 0.5$)	($\theta = 48$, $\alpha = 1$)	($\theta = 60$, $\alpha = 2$)
m_5: Frobenius norm of Hessian	10.55	4.19	**3.47**
	($\sigma = 0.5$)	($\theta = 37$, $\alpha = 1$)	($\theta = 41$, $\alpha = 3$)
m_6: variance	4.74	4.00	**3.98**
	($\sigma = 0$, 9×9 patch)	($\theta = 0$, $\alpha = 8$)	($\theta = 0$, $\alpha = 9$)

Multiple In-Focus Measures

As discussed, our general approach is not restricted to the use of a single in-focus measure. In fact, a combination of multiple in-focus measures allows to improve the quality of our depth from focus approach even further. In this regard, Figure 6.5(d) shows the resulting depth map and the corresponding fused image when combining the initial depth maps created with the trace of the structure tensor (m_2), the Laplacian (m_4), and the Hessian (m_5) in the data term of our energy functional. Here, we also apply our anisotropic regularisation technique. The corresponding MSE is 3.25 with the smoothness parameter $\alpha = 2$. This demonstrates the benefits of the combination of multiple in-focus measures. We want to note that this result might be even further improved by introducing weights for the individual data term parts and optimising for them.

Comparison to Competing Methods

Let us now compare our method to competing methods from the literature. As multi-scale transformation-based method, we select the approach of Forster et al. [FVB$^+$04] that performs a complex wavelet decomposition. As a representative of segmentation-based techniques we choose the method of Agarwala et al. [ADA$^+$04]. In addition, in particular to evaluate our depth map estimation, we compare our results to the depth from defocus method of Aguet et al. [AVU08]. They explicitly model the physical image acquisition process and jointly optimise for the sharp image and the depth map. We computed the results of Forster et al. and Aguet et al. with *ImageJ* plug-ins provided by the authors; see [Bio15]. To ensure a fair evaluation we optimised the involved model parameters. For the method of Agarwala et al. we considered the

Table 6.3: Comparison to competing methods in terms of the *mean squared error* (MSE), the correlation, and the *structural similarity* (SSIM) index between the fused image and its ground truth. Best results are written in bold face.

method	MSE	correlation	SSIM
Forster et al. [FVB+04]	152.12	0.95	0.98
Agarwala et al. [ADA+04]	135.97	0.98	0.98
Aguet et al. [AVU08]	113.73	0.98	0.98
proposed	**3.25**	**1.00**	**1.00**

provided executable [ADA+04], where we computed all results with the *maximum contrast image objective* setting and gradient-domain fusion.

We depict the results for our synthetic data set in Figure 6.5. The visual comparison shows that our technique provides the most accurate depth map. To objectively evaluate the results, we apply the following three quality measures: the MSE, the correlation, and the *structural similarity* (SSIM) [WBSS04][3]. While smaller MSE values indicate a better performance, larger values of the correlation and the SSIM are superior. Table 6.3 lists the resulting quality measures between the fused images and the ground truth all-in-focus image. Also here, our method yields the best results w.r.t. all three quality measures.

Besides tests with synthetic data, we additionally evaluate our approach on two commonly available real-world data sets in Figure 6.6 and 6.7.

The *Insect* data set in Figure 6.6 consists of thirteen frames with size 1344×1021 while the focal plane distance increases from frame to frame. We observe that our computed depth map is less noisy and more reliable than the other depth maps. The zooms into the fused images help to provide more details: The quality of our approach outperforms the methods of Forster et al. and Aguet et al. The method of Agarwala et al. is comparable to our method. Both methods preserve fine-scale details in a desirable way.

The *Clock* data set in Figure 6.7 consists of two frames with size 480×480. Again, we see that our technique creates a more reliable depth map compared to the competing approaches. Moreover, the zooms into the fused images illustrate that our result provides sharp small-scale details without visible artefacts.

In addition, we evaluate our depth from focus technique for several synthetic and real-world focal stacks provided by Mahmood and Choi [MC12a]. To this end, we depict our results for the synthetic sets in Figure 6.8. The computed depth maps in Figure 6.8(e) convincingly resemble the ground truth depth maps in Figure 6.8(c). Also the fused images in Figure 6.8(d) are desirably sharp. In Table 6.4, we provide a quantitative comparison to the method of Mahmood and Choi [MC12a] in terms of

[3]We compute the SSIM with the code provided by the authors using the default parameter setting [WBSS04].

Table 6.4: Comparison to the method of Mahmood and Choi [MC12a] in terms of the *root mean squared error* (RMSE). Best results are written in bold face.

data set	Mahmood and Choi [MC12a]	proposed
slope	0.67	**0.12**
sinusoidal	3.61	**0.46**
wave	5.32	**0.58**

Table 6.5: Runtimes of our depth from focus approach on an *NVIDIA GeForce GTX 460*.

number of images	image resolution	runtime in seconds
2	256×256	0.5
2	512×512	1.2
2	1024×1024	4.1
13	256×256	0.5
13	512×512	1.3
13	1024×1024	4.2

the *root mean squared error* (RMSE). We outperform their method for all three data sets.

Furthermore, we present our resulting depth maps and all-in-focus images for the real-world data sets of Mahmood and Choi [MC12a] in Figure 6.9. Once again, we observe precise depth estimates that lead to the desired image fusion results. Indeed, the produced depth maps are piecewise smooth while preserving important depth structures, and the fused images are desirably sharp.

To further demonstrate the general good performance of our technique, we depict our results for additional real-world data sets in Figure 6.10. Thanks to our anisotropic edge-preserving regularisation strategy the resulting depth maps are of high quality and the fused images provide the desired sharpness.

Runtime Evaluation

Finally, let us investigate the efficiency of our method. Our reference implementation is written in *CUDA* and runs on an *NVIDIA GeForce GTX 460* graphics card. Table 6.5 lists the runtimes for different focal stacks that differ in the number of images and the image resolution. First of all, we provide desirably fast runtimes of less than five seconds for realistic data sets with an image resolution of up to 1024×1024. Moreover, the main part of our technique is the regularisation of the depth map. Except for the computation of the diffusion tensor, this step is independent of the number of images in the focal stack. Hence, increasing the number of images does not significantly increase the runtime of our algorithm. This is a clear advantage of our depth-based approach compared to related focus fusion methods.

6.2.7 Limitations and Discussion

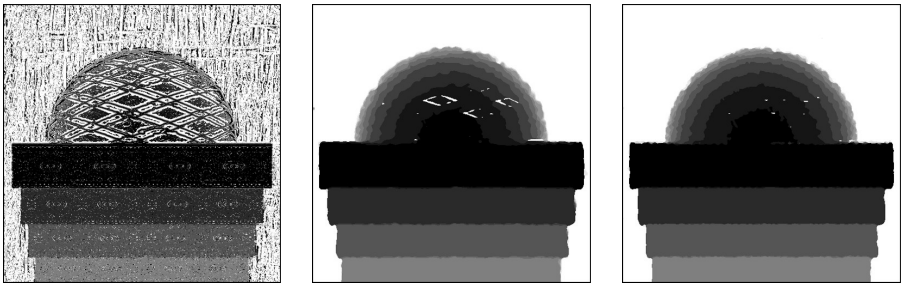
The conducted experiments on various focal stacks demonstrate that our variational technique provides robust and efficient solutions for the depth from focus task. Both, the computed depth maps and the fused all-in-focus images are of high quality and compare favourably to competing approaches from the literature.

As an intrinsic drawback of depth from focus methods, we want to mention that the quality of the output depth maps inherently depends on the number of input images or, in other words, on the depth sampling by means of the different focal plane positions. Since we only consider image regions that are in-focus to compute our initial depth maps, few number of images do not allow precise depth estimates. In this regard, depth from defocus approaches offer a greater potential since also the information of blurred image features is taken into account. In fact, an accurate modelling of the relation between the amount of blur and the distance to the focal plane theoretically allows to reconstruct the depth of surface points that are never captured sharply. To a certain extent, we also provide this feature by filling-in depth values by means of neighbouring points. However, this regularisation approach would certainly benefit from additional data term assumptions in such image regions.

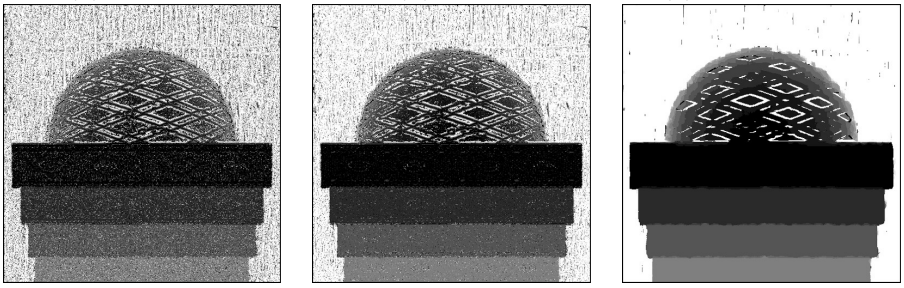
An interesting idea to incorporate such assumptions is to directly model a data term that opts for a maximal sharpness criterion instead of forcing the similarity to an initial depth map. In this way, also information about the blurriness is included such that such a strategy can be interpreted as an intermediate approach between depth from focus and depth from defocus. Indeed, first experiments in this direction yield promising results. However, in all conducted experiments the computed depth maps were less accurate that the one obtained with the presented simpler depth from focus approach. This can be explained by the fact that a direct optimisation for a maximal sharpness measure is highly nonconvex, and thus we are not guaranteed to find the optimal solution. Nevertheless, this approach might be a fruitful starting point for future work. Main tasks in this regard are the selection of measures with reasonable sharpness curves over the image stack and suitable optimisation techniques that may approximate those sharpness curves by sufficiently smooth functions and/or apply appropriate linearisation strategies. A method that goes in a similar direction can be found in [MBSC15]. Such a technique could offer a desirable tradeoff between the incorporation of blurred image features and the high complexity of standard depth from defocus approaches.

Besides extensions of the data term, also more advanced regularisation techniques might be an option for further quality improvements. Indeed, the discussed depth from focus approach models a first order smoothness of the depth maps. However, several applications have uncovered the need for second order regularisation in the context of depth map computation since real-world scenes can often be well described by piecewise planar objects. In fact, we introduce such a second order regularisation technique in Section 6.3 where we also attach great importance to the incorporation of directional image information to steer the smoothing process. In this sense, this

approach can be seen as a natural extension of the presented anisotropic first order regularisation technique to higher order.



(a) gradient (b) trace (c) determinant

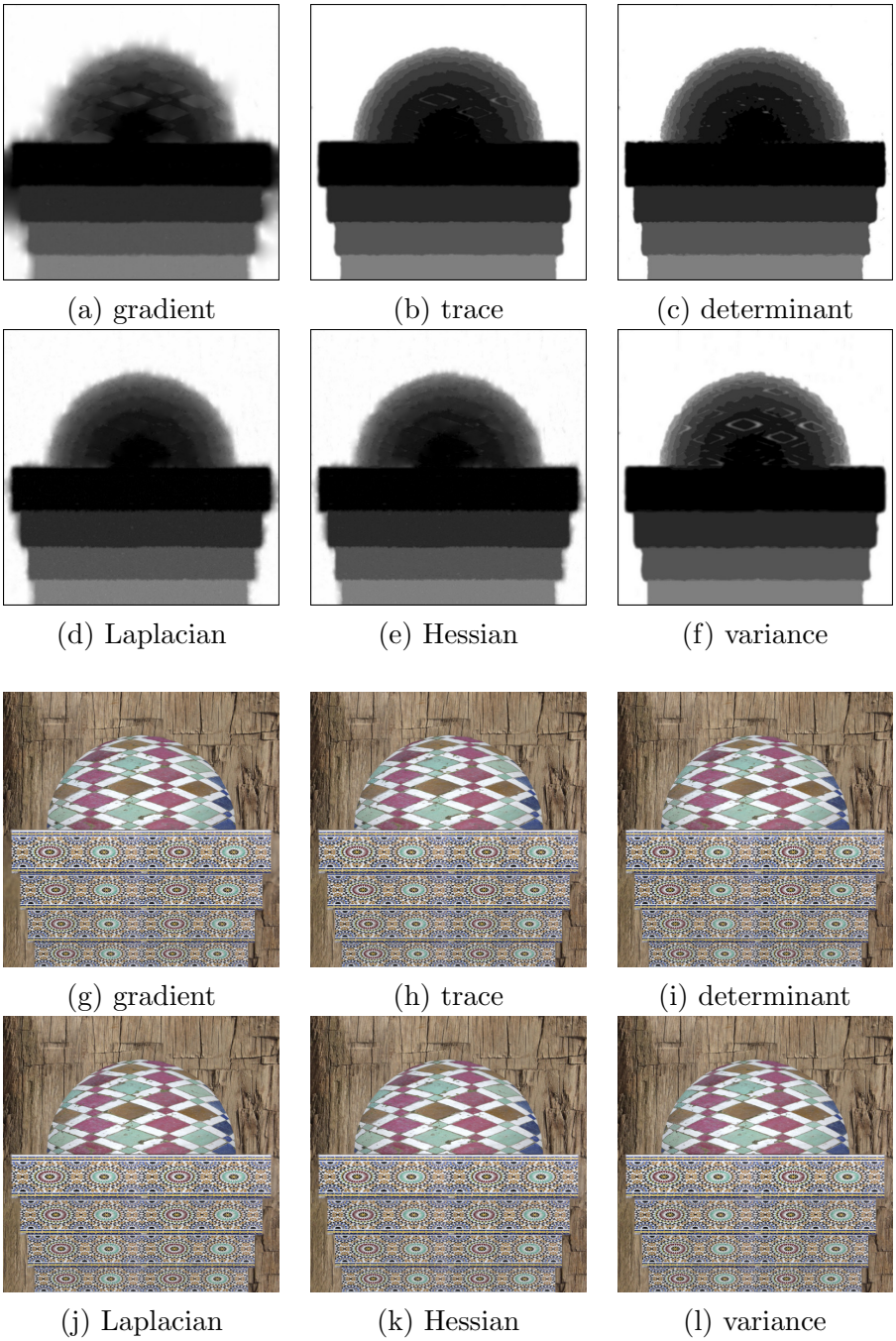


(d) Laplacian (e) Hessian (f) variance

(g) gradient (h) trace (i) determinant

(j) Laplacian (k) Hessian (l) variance

Figure 6.2: Comparison of different in-focus measures by means of the *initial* depth maps in (a–f) and the fused images in (g–l). The corresponding parameters and error measures can be found in Table 6.2.



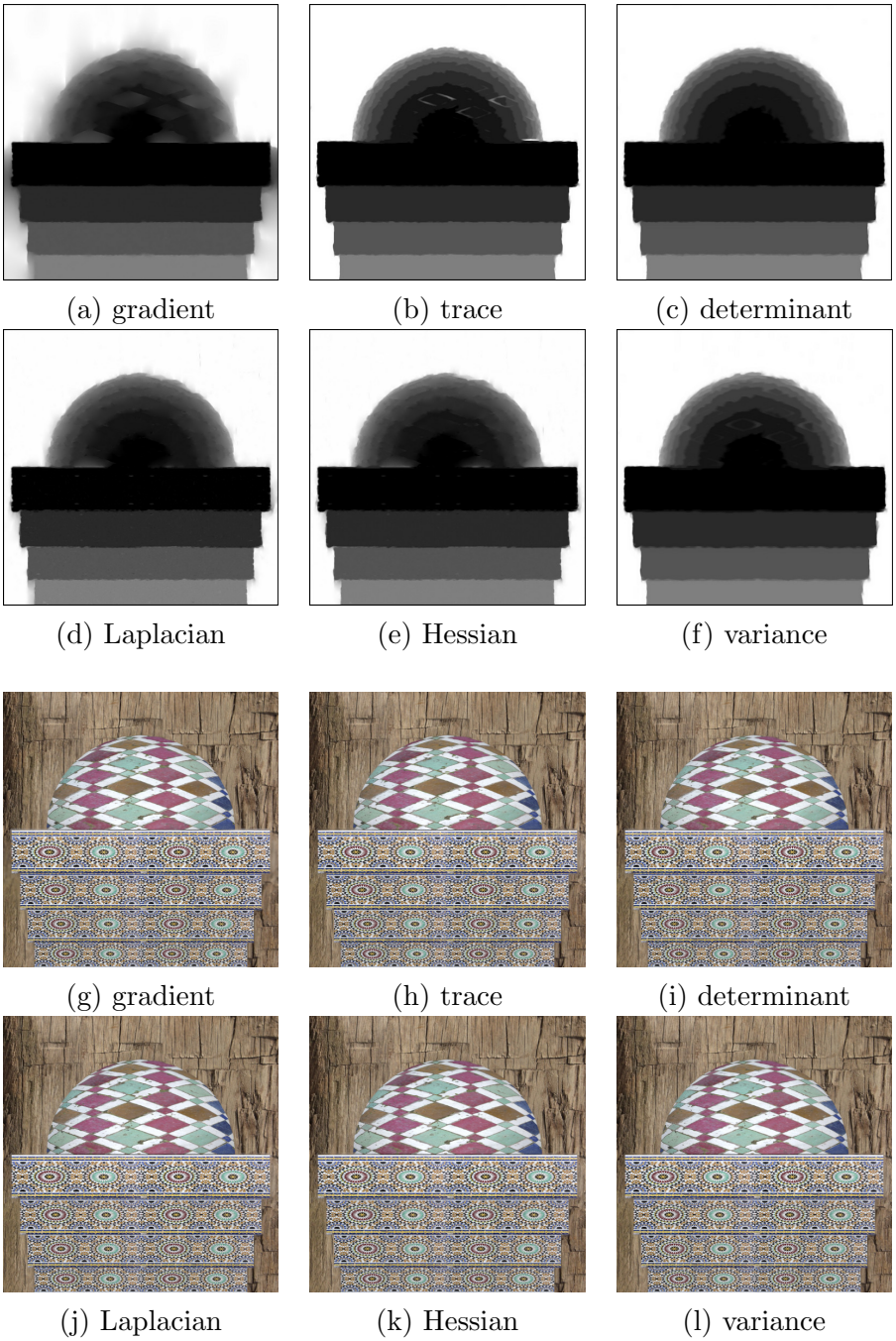
(a) gradient (b) trace (c) determinant

(d) Laplacian (e) Hessian (f) variance

(g) gradient (h) trace (i) determinant

(j) Laplacian (k) Hessian (l) variance

Figure 6.3: Comparison of different in-focus measures by means of the smoothed depth maps in (a–f) and the fused images in (g–l) with the *isotropic* approach. The corresponding parameters and error measures can be found in Table 6.2.



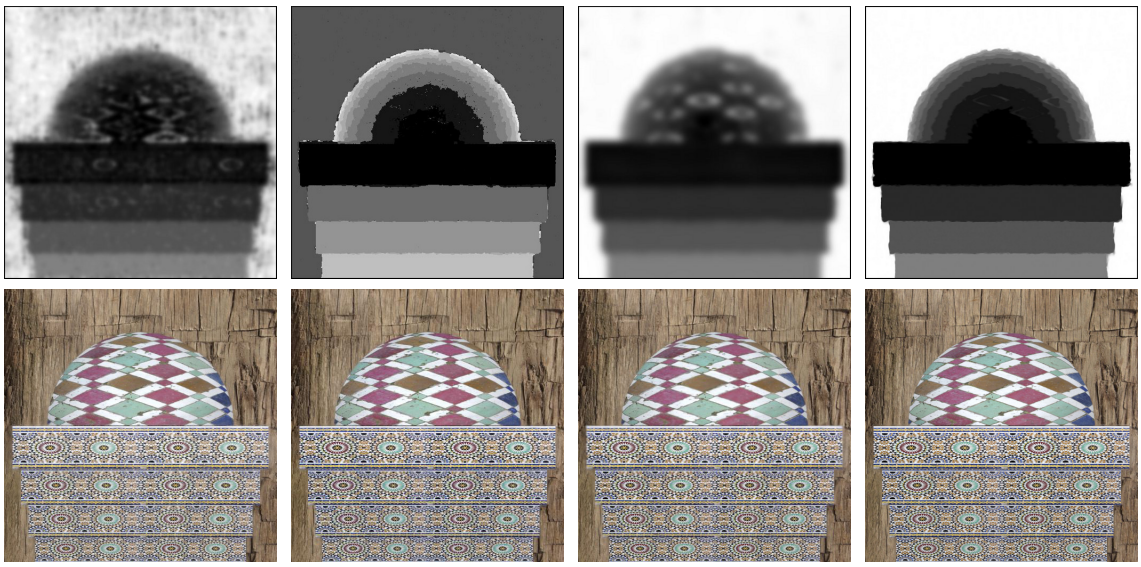
(a) gradient (b) trace (c) determinant

(d) Laplacian (e) Hessian (f) variance

(g) gradient (h) trace (i) determinant

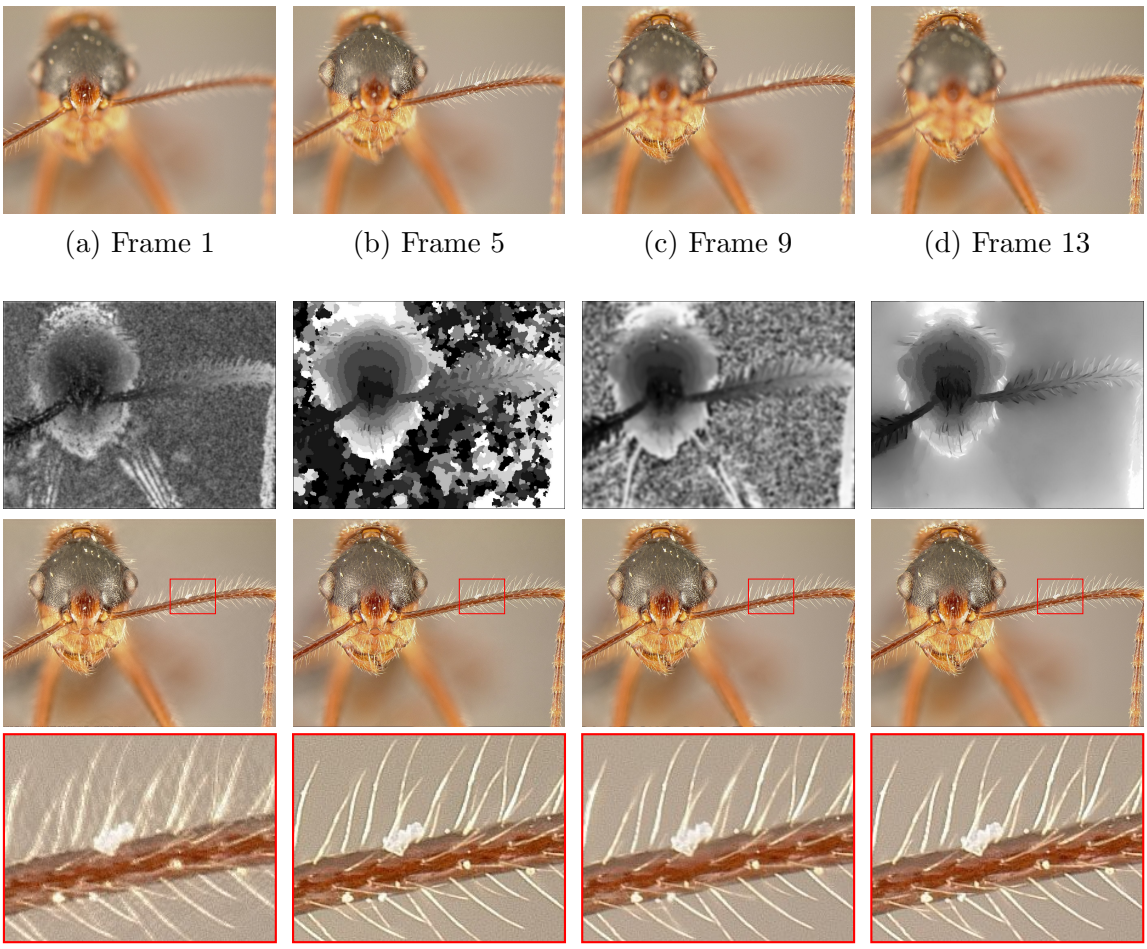
(j) Laplacian (k) Hessian (l) variance

Figure 6.4: Comparison of different in-focus measures by means of the smooth depth maps in (a–f) and the fused images in (g–l) with the *anisotropic* approach. The corresponding parameters and error measures can be found in Table 6.2.



(a) Forster et al. (b) Agarwala et al. (c) Aguet et al. (d) proposed

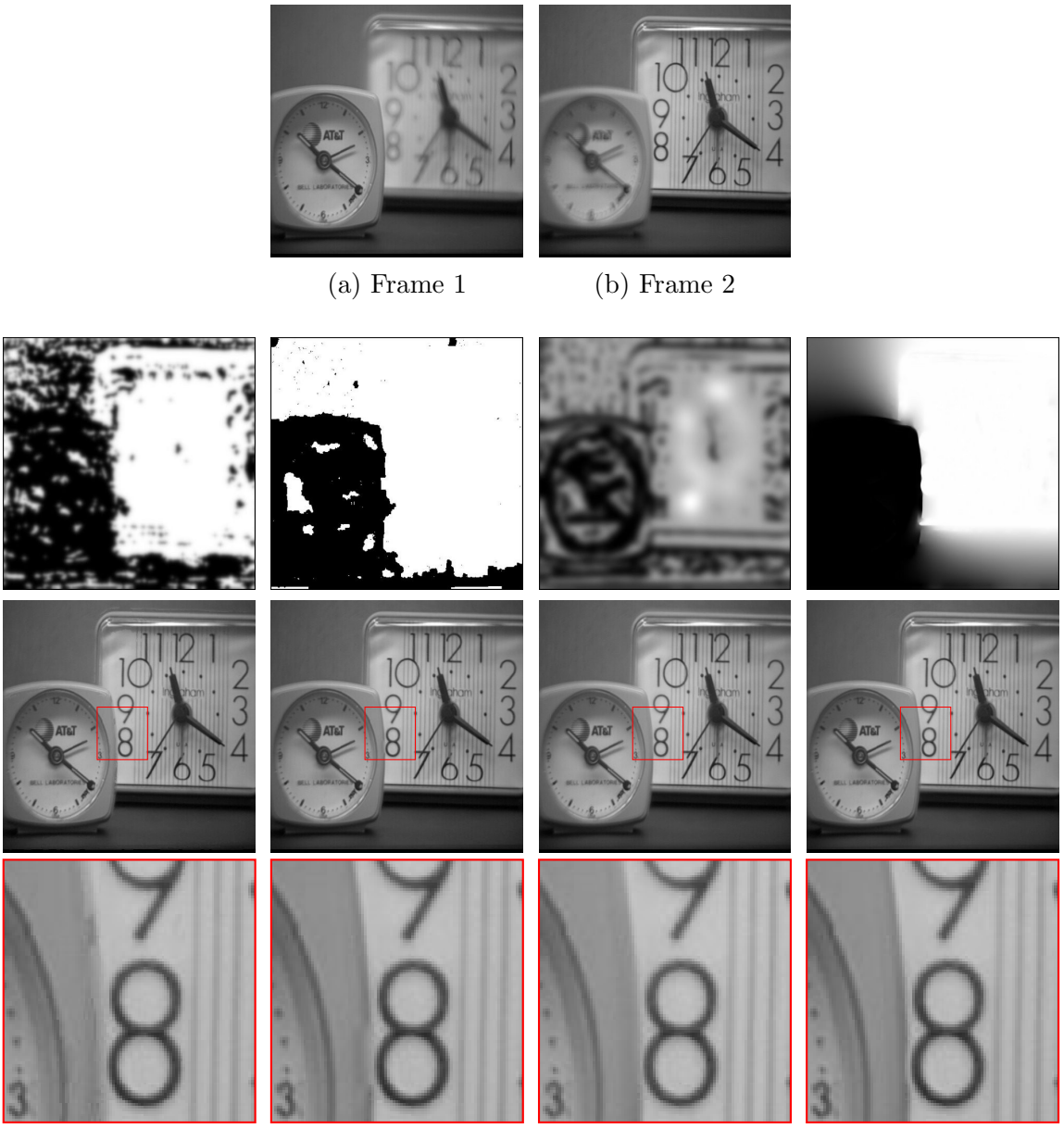
Figure 6.5: Visual comparison for the synthetic data set from Figure 6.1. *Top*: Computed depth maps. *Bottom*: Fused all-in-focus images. Here, we use multiple in-focus measures, namely the trace of the structure tensor (m_2), the Laplacian (m_4), and the Hessian (m_5) to compute our result in (d).



(a) Frame 1 (b) Frame 5 (c) Frame 9 (d) Frame 13

(e) Forster et al. (f) Agarwala et al. (g) Aguet et al. (h) proposed

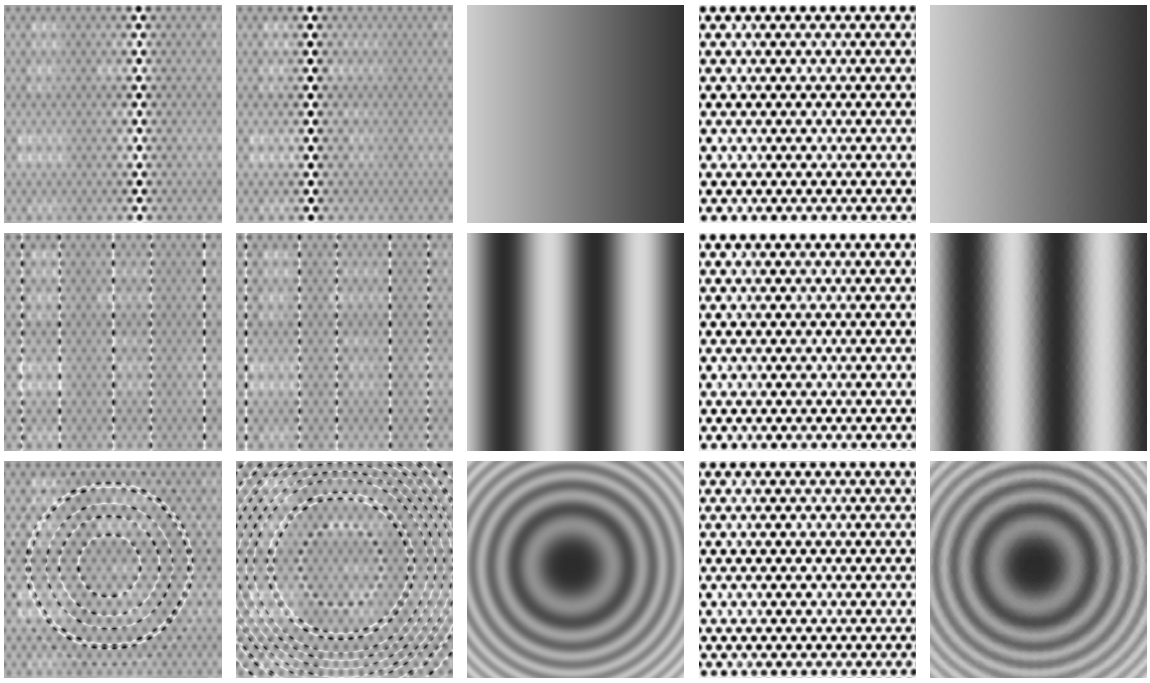
Figure 6.6: Results for the *Insect* data set [ADA+04]. *From top to bottom*: Input frames with increasing focal plane distance, computed depth maps, and fused images with zooms. In this experiment, we use the Frobenius norm of the Hessian as in-focus measure with the parameters $\sigma = 0.5$, $\theta = 60$, and $\alpha = 5$ to compute our result in (h).



(a) Frame 1 (b) Frame 2

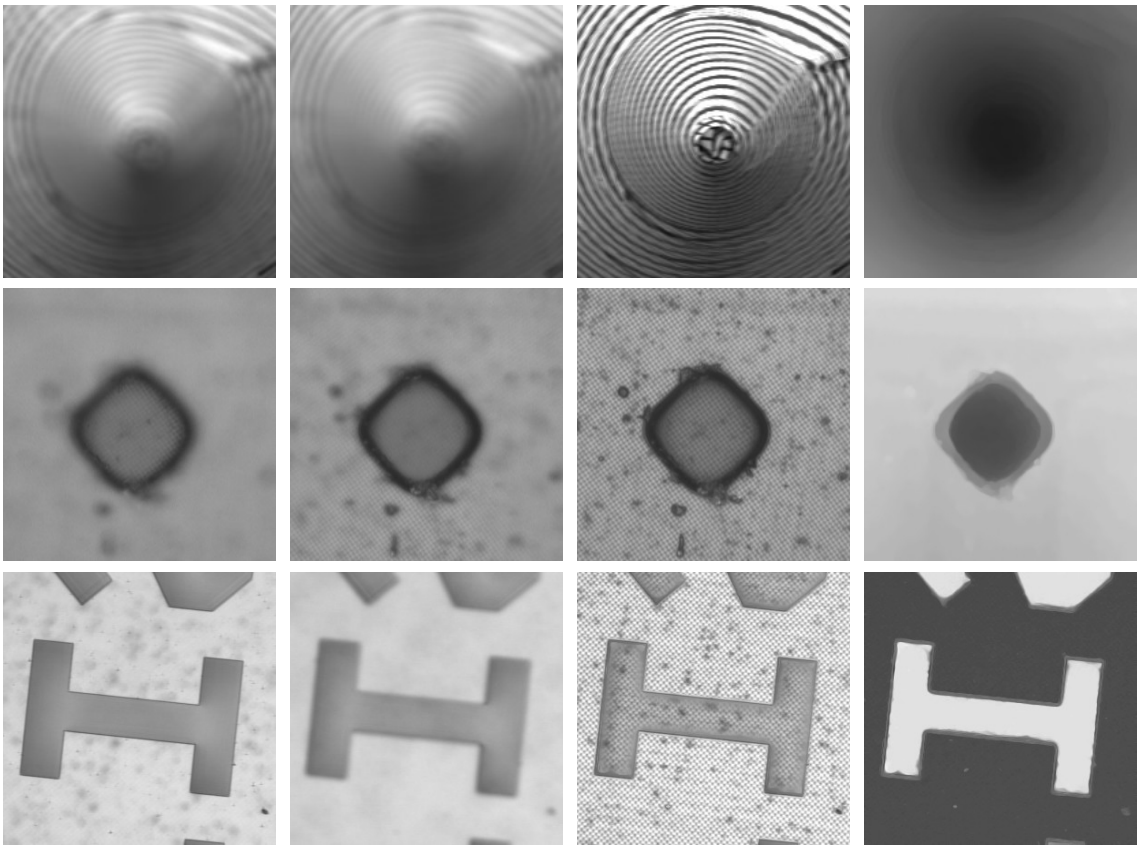
(c) Forster et al. (d) Agarwala et al. (e) Aguet et al. (f) proposed

Figure 6.7: Results for the *Clock* data set available at [BO10]. *From top to bottom*: Input frames with different focal planes, computed depth maps, and fused images with zooms. In this experiment, we use the Frobenius norm of the Hessian as in-focus measure with parameters $\sigma = 1.0$, $\theta = 6$, and $\alpha = 100$ to compute our result in (f).



(a) near focus (b) far focus (c) ground truth (d) our image (e) our depth

Figure 6.8: Our results for the synthetic data sets of Mahmood and Choi [MC12a]. *From top to bottom*: *Slope* data set ($\sigma = 2.5$, $\theta = 0$, $\alpha = 42$), *Sinusoidal* data set ($\sigma = 0.5$, $\theta = 46$, $\alpha = 2$), and *Wave* data set ($\sigma = 0.4$, $\theta = 31$, $\alpha = 1$). Each set consist of 60 frames with size 300×300. *From left to right*: (a,b) Two input frames of the focal stack with a short and long focal plane distance. (c) Provided ground truth depth map. (d) Our computed all-in-focus image. (e) Our computed depth map. For all sets, we use the Frobenius norm of the Hessian as in-focus measure.



(a) near focus (b) far focus (c) fused image (d) depth map

Figure 6.9: Our results for the real-world data sets of Mahmood and Choi [MC12a]. *From top to bottom*: *Cone* data set with 97 frames of size 200 × 200 ($\sigma = 1.0$, $\theta = 0$, $\alpha = 10$), *TFT-LCD color filter* data set with 60 frames of size 300 × 300 ($\sigma = 0.5$, $\theta = 0$, $\alpha = 7$), and *Letter I* data set with 60 frames of size 300 × 300 ($\sigma = 0.2$, $\theta = 15$, $\alpha = 1$). *From left to right*: (a,b) Two input frames of the focal stack with a short and long focal plane distance. (c) Our computed all-in-focus image. (d) Our computed depth map. For all sets, we apply the Frobenius norm of the Hessian as in-focus measure.

(a) fused images (b) final depth maps

Figure 6.10: Results for real-world data sets from [BO10]. *From top to bottom*: *Watch* data set with 395 frames of size 640×427 ($\theta = 10$, $\alpha = 5$), *Fly* data set with 20 frames of size 1024×682 ($\theta = 10$, $\alpha = 5$), *Coffee* data set with 91 frames of size 1024×683 ($\theta = 10$, $\alpha = 25$), and *Flower* data set with 20 frames of size 1024×683 ($\theta = 5$, $\alpha = 30$). Here, we apply the Frobenius norm of the Hessian as in-focus measure with presmoothing parameter $\sigma = 1.0$.

6.3 Anisotropic Second Order Coupling Model

Second order regularisation has become a powerful tool in various visual computing applications. For example, it is well-suited for the computation of depth maps since many real-world scenes can be well described by (piecewise) planar objects. In a variational context, there exist three popular approaches to model such a second order smoothness assumption: (*i*) The most intuitive one is to directly penalise second order derivatives; see e.g. [Gri81, Ter83]. However, such a direct penalisation only allows to model second order variations of the solution explicitly. It does not give direct access to its first order derivative, which is for instance required to model jumps of the solution. (*ii*) In this regard, coupling models offer an interesting alternative for higher order regularisation; see e.g. [Har87, BKP10] or related infimal convolution approaches [CL97]. In the second order case, such a coupling model consists of two terms: The first term (*coupling term*) couples the gradient of the unknown with some auxiliary vector field, while the second one (*smoothness term*) enforces smoothness of this vector field. The combination of both terms allows to treat both, first and second order variations in the solution explicitly. (*iii*) A related idea is to locally parametrise the unknown by affine functions, and to optimise for the introduced parameters with suitable smoothness constraints; see e.g. [NBK08]. However, this does not allow such an explicit access to both derivative orders; cf. also [TPCB08].

Concerning first order regularisation, several approaches have demonstrated the benefits of incorporating anisotropy in the smoothing process; see e.g. [NE86,ZBW11b, SZV+12, BHW15, PWM+15]. Consequently, it seems to be a fruitful idea to also apply anisotropic concepts in second order regularisation. For instance, Lenzen et al. [LBL13] incorporate directional information into a direct second order approach. Unfortunately, as discussed above, such a direct approach constrains the degree of freedom in the modelling. Also the nonlocal coupling model of Ranftl et al. [RBP14] can be seen as related to our work. However, we aim at a fully local model that allows a natural modelling of the anisotropy in terms of image and depth derivatives. In this way, we can provide a natural, systematic transition from anisotropic first to anisotropic second order approaches. In a local framework, Ranftl et al. [RGPB12] and Ferstl et al. [FRR+13] propose a coupling model that incorporates directional image information, but the anisotropy is restricted to the coupling term. To summarise, first steps to include anisotropy into second order models have been done. However, existing approaches do not exploit successful anisotropic ideas to the full extent.

Main Contributions. The main goal of our work is to close this gap and to systematically incorporate well-established anisotropic ideas from first order approaches into second order coupling models. Here, we make maximal use of the directional information by introducing anisotropy both, into the coupling as well as into the smoothness term. In addition, we propose a joint image- and depth-driven technique that allows a different amount of coupling and smoothing along and across image structures. Contrary to previous work, we apply a direction-dependent penalisation that is important

for good inpainting results. Last but not least, we demonstrate the performance of our anisotropic second order technique in the context of depth from focus (cf. also Section 6.2).

Section Outline. We start with a survey of related second order regularisation approaches in Section 6.3.1. In Section 6.3.2, we present an isotropic coupling model that marks the starting point of our work. Next, we explain how to extract directional information from images in Section 6.3.3. We incorporate this information into the coupling and smoothness term in Section 6.3.4 and Section 6.3.5, respectively. This leads to our fully anisotropic coupling model in Section 6.3.6. After explaining our numerical realisation in Section 6.3.7, we evaluate our technique and compare it to related approaches in Section 6.3.8. Section 6.3.9 illustrates the performance of our method for depth from focus. Finally, we discuss possible limitations of our work in Section 6.3.10.

6.3.1 Related Work

Direct Models

Let us first discuss methods that apply a direct penalisation of second order derivatives. As one of the first approaches in image processing and computer vision, Grimson [Gri81, Gri82, Gri83] investigated several second order regularisers in the context of surface interpolation. In particular, he proposed a penalisation of the Laplacian and the Frobenius norm of the Hessian, which he referred to as *quadratic variation*. The surfaces that minimise this penalisation term are so-called *thin plate splines*; cf. e.g. [Duc76]. Related early regularisation techniques have been described in [Ter83, KLB85, Bou86].

In [NFD97], Nielsen et al. consider general higher order Tikhonov regularisation terms and discuss recursive implementations under specific boundary conditions. In order to preserve discontinuities, Geman and Reynolds [GR92] propose nonlinear higher order smoothness terms. Comparably, Scherzer [Sch98] presents higher order variants of bounded variation regularisers for the task of image denoising. Related to that are approaches that consider the space of bounded Hessians; see e.g. [HS06, BP10, PS14]. Mathematical analyses of such direct methods are conducted in [SDN05, PS08, DFLM09]. Moreover, Steidl [Ste06] discusses the dual treatment of such higher order regularisation functionals. In this regard, Chan et al. [CEP10] consider a dual method for higher order regularisation with application to texture extraction and image smoothing.

Lysaker et al. [LLT03] suggest a nonlinear penalisation of the Frobenius norm of the Hessian and a separate nonlinear penalisation of the diagonal entries of the Hessian (*modified Laplacian*), respectively. Note that the latter one lacks rotationally invariance. Motivated by second order directional derivatives, the authors in [DLBU11, LBU12] motivate the penalisation of the spectral and Frobenius norm of the Hessian matrix in the context of image deblurring in biomedical applications. Lai et

al. [LTC13] penalise the Frobenius norm of the Hessian in a nonlinear way for 3D surface reconstruction. Trobin et al. [TPCB08] present a second order regulariser for motion estimation that is intended to eliminate a bias of conventional second order approaches w.r.t. certain affine flow fields.

In [CMM00, LT06], the authors combine a first order *total variation* (TV) term and second order terms to counteract staircasing artefacts that are characteristic for pure TV regularisation. In a similar way, Lenzen et al. [LSG11] combine anisotropic first order smoothness with an isotropic second order penalisation of the Hessian to denoise time-of-flight data. Here, the weight of both smoothness terms adapts locally.

In the context of shape from shading, Vogel et al. [VBWD07] propose a Hessian-based regularisation that is guided the image and depth structure. In [LBL13], Lenzen et al. present an adaptive second order TV regulariser that adjusts to the local image structure in an anisotropic way. Second order TV-like methods for cyclic data are discussed by Bergmann et al. [BLSW14]. Furthermore, Lellmann et al. [LPSS15] analyse and apply a nonlocal variant of the Hessian for image smoothing.

Alternative ideas are second order div-curl regularisation terms. They are for example successfully applied in the context of optic flow computation; see e.g. [Sut94, GP96, YSS07, YSM07]. Moreover, several approaches consider curvature-based regularisation. For instance, Terzopoulos [Ter83, Ter84] proposes a strain energy term that incorporates the mean and Gaussian curvature of the surface. Masnou and Morel [MM98] present an elastica-based image inpainting process. Such an approach is analysed by Shen et al. in [SKC03], and Tai et al. [THC11] present a fast method for elastica-based regularisation using an augmented Lagrangian approach. An approximation of Euler's elastica energy was recently proposed by Bredies et al. [BPW15]. Fischer and Modersitzki [FM03] present a curvature-based image registration method. Brito-Loeza and Chen [BLC10] discuss extensions of curvature-based models to vector-valued images, and Zhu and Chan [ZC12b] perform image denoising by minimising the mean curvature of the image surface.

Variational models with direct second order penalisation lead to fourth order PDEs. In this regard, there exist several approaches that directly model such higher order evolution processes. For example, Wei [Wei99] presents a generalisation of the Perona-Malik evolution [PM90]. Also the so-called *low curvature image simplifier* (LCIS) by Tumblin and Turk [TT99] constitutes such an evolution process, where an analysis of this process can be found in [BG04]. Shen and Chan [SC02] present a biharmonic image inpainting by means of Green's functions. You and Kaveh [YK00] propose fourth order PDEs for noise removal. Related to those approaches, Greer and Betrozzi [GB04a] consider travelling wave solutions of fourth order PDEs for image processing. The authors also conduct a mathematical analysis of such nonlinear fourth order PDEs in [GB04b]. A further modelling and analysis of higher order nonlinear PDEs can be found in [DWB05, DWB09]. In the master's thesis of Roth [Rot09], several anisotropic variants of fourth order PDEs are investigated. Another idea, in the context of image inpainting, is to consider the higher order Cahn-Hilliard equation [BEG07a, BEG07b, BHS09] or fourth order total variation flow [SBBH09].

Coupling Models

As discussed above, those direct approaches allow to model second order variations of the solution explicitly. This might be useful to guide discontinuities of the slope of the solution which correspond to kinks. However, it does not give explicit access to the first derivative which is for instance required to model jumps. Hence, researchers came up with coupled higher order regularisation techniques. To the best of our knowledge, a first appearance of such a model is given by the coupled depth/slope approach for surface reconstruction by Harris [Har86, Har87]. Motivated by the human visual system, he presented a general coupling framework that allows to treat each derivative order explicitly. Further early approaches that are inspired by this work are the method of Horn [Hor90] in the context of shape from shading and the stereo reconstruction method of Belhumeur [Bel93]. In [Sut91], Suter discusses a generalisation of Harris' approach by means of augmented Lagrangian formulations. A more recent coupling model for arbitrary order in the sense of [Har86, Har87] was proposed by Hewer et al. [HWS$^+$13]. It is applied to compute the Lagrangian strain tensor from image sequences in a direct way.

To penalise the curvature of an image in the context of image inpainting, Ballester et al. [BBC$^+$01] couple the normal directions of the level lines with an auxiliary variable and penalise its divergence. In the area of motion estimation, Corpetti et al. [CMP02] couple the divergence and the curl of the flow field with auxiliary variables to apply a second order div-curl penalisation. Two methods that can also be seen as related to coupling models were presented by Tasdizen et al. [TWBO03] and Lysaker et al. [LOT04]. In the context of surface reconstruction, the authors propose a two-step approach where they first smooth the surface normals, and in a second step fit a surface to the smoothed normal field.

Also the *infimal convolution* (IC) approach by Chambolle and Louis [CL97] and the *total generalised variation* (TGV) method of Bredies et al. [BKP10] can be interpreted in the sense of a coupling model. Comparisons and evaluations of IC- and TGV-based approaches are presented in [BBBM13, Mül13, PB15, Ste15, DSV17].

The infimal convolution technique [CL97] formulates an unknown function u as a sum of two auxiliary functions u_1 and u_2. Then, different regularisation terms are applied to both functions; e.g. modelling a first order smoothness assumption on u_1 and second order smoothness assumptions on u_2. In fact, a reformulation of this approach uncovers a close connection to coupling models in the form of [Har87]. As a major difference, here the auxiliary coupling field is required to be integrable. Infimal convolution has found numerous applications and extensions. For instance, Chan et al. [CEP07] apply a variant of it for decomposing an image into its structural and textural components. In [SS08, SST11], the authors study discrete IC functionals. Holler and Kunisch [HK14] apply an IC-type approach for image and video reconstruction. Moreover, Balle et al. [BEF$^+$15] consider it for analysing the deformation of metallic materials. Recent mathematical analyses of such IC methods are conducted by Burger et al. in [BPPS16a, BPPS16b]. A generalisation of IC to manifold-valued images is

proposed by Bergmann et a. [BFPS17].

In [BKP10], Bredies et al. introduced total generalised variation that can be seen as a higher order generalisation of first order TV smoothness to arbitrary order. A second order variant of it is analysed by Bredies and Valkonen [BV11], where also applications to image denoising and deconvolution are presented. Further applications include the fusion of range images [PZB11], magnetic resonance imaging [KBPS11], stereo and optic flow computation [RGPB12, RBP14], depth map upsampling [FRR+13], and diffusion tensor imaging [VBK13]. An extension to vector-valued images is addressed in [LWP+15]. Moreover, there are several mathematical analyses that investigate the properties of TGV-based approaches. Theoretical properties in the 1D case are considered in [BKV13, PS15, PB15]. Bredies and Holler [BH14] study TGV for regularisation in inverse problems. In [BH15a, BH15b], the authors present and analyse a TGV-based framework for image decompression, zooming, and reconstruction. While the first part [BH15a] focusses on a mathematical analysis, the second part [BH15b] mainly deals with numerical aspects. An asymptotic behaviour of TGV is investigated by Papafitsoros and Valkonen [PV15]. Moreover, an analysis of related coupling models was recently conducted by Bildhauer et al. [BFW16].

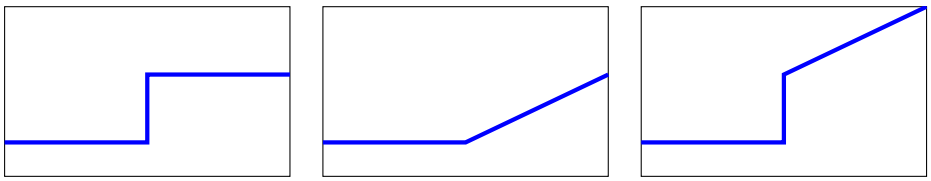
This large variety of second order approaches and in particular of second order coupling models demonstrates the need and success of second order regularisation in recent years. Moreover, as we have seen, there are already a few techniques that also include some kind of anisotropy into higher order smoothing processes; see e.g. [Wei99, TT99, Rot09, RGPB12, LBL13, FRR+13, RBP14, LPSS15]. However, as mentioned above, existing approaches do not exploit successful anisotropic ideas to the full extent. To close this gap, we systematically introduce a novel anisotropic second order coupling model in this work.

6.3.2 Isotropic Coupling Model

Compared to direct implementations of higher order regularisation, coupled formulations as in [Har87, BKP10, HWS+13] offer several advantages: First, they do not require the explicit estimation and implementation of higher order derivatives. Second and even more importantly, they allow to individually penalise each derivative order explicitly, which enables an explicit modelling of jumps and kinks of the solution. Hence, we base our anisotropic second order regulariser on the following isotropic coupling model that replaces a direct second order smoothness term of a depth map d by

$$R_{\mathrm{I}}(d) \;=\; \inf_{\boldsymbol{p}} \left\{ \int_{\Omega} \Big(\Psi\big(|\boldsymbol{\nabla} d - \boldsymbol{p}|^2\big) \;+\; \beta \cdot \Psi\big(|\boldsymbol{\mathcal{J}} \boldsymbol{p}|_{\mathrm{F}}^2\big) \Big) \,\mathrm{d}\boldsymbol{x} \right\}, \tag{6.21}$$

where $\Psi(z^2) = \sqrt{z^2 + \varepsilon^2}$ is a subquadratic function with a small positive constant ε, $|\cdot|$ denotes the Euclidean norm, and $|\cdot|_{\mathrm{F}}$ the Frobenius norm. Here, the auxiliary vector field $\boldsymbol{p} = (p_1, p_2)^{\mathsf{T}}$ can be seen as an approximation of the gradient $\boldsymbol{\nabla} d$ of the



(a) jump (b) kink (c) jump and kink

Figure 6.11: 1D illustration of jumps and kinks.

unknown d. Furthermore, $\boldsymbol{\mathcal{J}p}$ denotes the Jacobian of this vector field $\boldsymbol{p}$, i.e.[4]

$$\boldsymbol{\mathcal{J}p} = \begin{pmatrix} \partial_{x_1} p_1 & \partial_{x_2} p_1 \\ \partial_{x_1} p_2 & \partial_{x_2} p_2 \end{pmatrix} . \tag{6.22}$$

Since the first term in (6.21) inherently couples $\boldsymbol{\nabla} d$ to $\boldsymbol{p}$, we refer to it as *coupling term*. The second term provides smoothness of the auxiliary vector field $\boldsymbol{p}$. Hence, we refer to it as *smoothness term*. Here, the parameter $\beta > 0$ allows to steer the importance of both terms.

Next, we explain the meaning and interplay of both terms in more detail: As discussed, the coupling term forces a similarity of the gradient of d and the auxiliary vector field $\boldsymbol{p}$. For didactic reasons, let us first assume a hard coupling, i.e. $\boldsymbol{p}$ is identical to $\boldsymbol{\nabla} d$. Since the smoothness term penalises first order variations of $\boldsymbol{p}$, it implicitly penalises second order variations of d. In other words, it allows to model the smoothness of the slope of the solution. In particular, the nonlinear function Ψ provides a discontinuity preserving behaviour such that discontinuities of the slope are preserved. Such slope discontinuities correspond to *kinks* of the solution; cf. Figure 6.11(b).

With that in mind, let us now investigate the behaviour of the coupling term. With the nonlinear function Ψ, it provides a discontinuity preserving coupling, i.e. it allows sparse deviations of the gradient of d from the smooth vector field $\boldsymbol{p}$. This shows that the coupling term allows to model discontinuities of the solution directly. Such discontinuities correspond to *jumps*; cf. Figure 6.11(a).

Summing up, the discussed coupling model provides direct access to both derivative orders and in this way, allows to model jumps and kinks of the unknown function d by the coupling and smoothness term, respectively. For small ε, the coupling model in (6.21) resembles *total generalised variation* (TGV) of second order [BKP10]. In many image processing and computer vision applications, such isotropic coupling models have led to high quality results. However, they do not make use of any directional information which is important for a variety of applications such as the one that we consider in this work.

[4] One could also think of considering the so-called symmetrised Jacobian, as for instance discussed in [BKP10]. However, this renders the actual implementation more difficult, and did not lead to superior results in basic experiments. Hence, we consider the standard Jacobian (6.22) in this work.

6.3.3 Extracting Directional Information

As for instance demonstrated by Nagel and Enkelmann [NE86] in the context of optic flow estimation, it is highly beneficial to guide the regularisation of the unknown by the local structure of the given input images in an anisotropic way. In fact, this allows to apply a different kind of smoothing along and across image structures. In this work, we transfer this successful concept from first to second order regularisation, and in particular to the discussed coupling model. To this end, let us first determine a way to identify the structures of an image or more specifically the directions across and along them.

Let f denote a given guidance image. In the case of depth from focus we take the evolving all-in-focus image as guidance. Then, we calculate the smoothing directions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ as the normalised eigenvectors of the structure tensor [FG87]

$$G_\rho * \big(\boldsymbol{\nabla}(G_\sigma * f)\boldsymbol{\nabla}(G_\sigma * f)^\mathsf{T}\big)\,, \tag{6.23}$$

where $*$ describes a convolution, and G_σ and G_ρ are Gaussians with standard deviation σ and ρ, respectively. The computed eigenvectors form an orthonormal system where the vector $\boldsymbol{r}_1$, which belongs to the dominant eigenvalue, points across image structures and $\boldsymbol{r}_2$ along them.

In the following sections, we describe how to successfully incorporate this directional information into the isotropic coupling model (6.21). To make optimal use of the directional information, we introduce anisotropy both in the coupling (Section 6.3.4) and in the smoothness term (Section 6.3.5).

6.3.4 Anisotropic Modification of Coupling Term

The isotropic coupling term from (6.21) can be rewritten as

$$C_\mathrm{I}(d,\boldsymbol{p}) \;=\; \Psi\big(|\boldsymbol{\nabla} d - \boldsymbol{p}|^2\big) \;=\; \Psi\Big(\sum_{i=1}^{2}\big(\boldsymbol{e}_i^\mathsf{T}(\boldsymbol{\nabla} d - \boldsymbol{p})\big)^2\Big)\,, \tag{6.24}$$

where $\boldsymbol{e}_1 = (1,0)^\mathsf{T}$ and $\boldsymbol{e}_2 = (0,1)^\mathsf{T}$. This reformulation of the coupling term in terms of the unit vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ allows to incorporate the directional information as follows: First, we exchange $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ in Equation (6.24) by the smoothing directions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. Second, we penalise both directional components differently to introduce an anisotropic behaviour, i.e. we exchange the position of the penalisation function Ψ and the summation $\sum_{i=1}^{2}$. In the context of first order regularisation, this idea has been proposed by Weickert and Schnörr in 2001 [WS01]. Finally, this results in our anisotropic coupling term

$$C_\mathrm{A}(d,\boldsymbol{p}) \;=\; \sum_{i=1}^{2}\Psi_i\Big(\big(\boldsymbol{r}_i^\mathsf{T}(\boldsymbol{\nabla} d - \boldsymbol{p})\big)^2\Big)\,. \tag{6.25}$$

Here, we apply different penalisation functions Ψ_i along and across image structures. This allows for instance to enforce a full coupling along edges by setting the corresponding $\Psi_2(z^2) = z^2$, and to relax the coupling constraint in the orthogonal direction with $\Psi_1(z^2) = 2\varepsilon\sqrt{z^2+\varepsilon^2}$ such that $\Psi_1'(z^2)$ is the Charbonnier diffusivity [CBAB94]. To analyse the introduced anisotropy in a better way, let us take a look at the Euler-Lagrange equations of (6.25) w.r.t. d and $\boldsymbol{p}$:

$$\operatorname{div}(\boldsymbol{D}\,\boldsymbol{\nabla} d) \;=\; \operatorname{div}(\boldsymbol{D}\,\boldsymbol{p})\,, \tag{6.26}$$

$$\boldsymbol{D}\,\boldsymbol{\nabla} d \;=\; \boldsymbol{D}\,\boldsymbol{p}\,. \tag{6.27}$$

Equation (6.26) is similar to the Poisson equation $\Delta d = \operatorname{div}(\boldsymbol{p})$ that builds the basis of gradient domain methods; see e.g. [FLW02, PGB03, BZCC10]. However, here the tensor

$$\boldsymbol{D}(d,\boldsymbol{p}) \;=\; \sum_{i=1}^{2} \Psi_i'\Big(\big(\boldsymbol{r}_i^{\mathsf{T}}(\boldsymbol{\nabla} d - \boldsymbol{p})\big)^2\Big)\cdot \boldsymbol{r}_i\boldsymbol{r}_i^{\mathsf{T}} \tag{6.28}$$

steers this process in an anisotropic way (cf. also [ARC06]). Moreover, this equation shows a nice feature of our model: When fixing the coupling variable $\boldsymbol{p}$ to $\boldsymbol{0}$, our second order coupling model comes down to a first order anisotropic diffusion process on the unknown d; see e.g. [Wei98] and references therein. Please note that for $\boldsymbol{p} = \boldsymbol{0}$ the smoothness term vanishes since in this trivial case $|\boldsymbol{\mathcal{J}}\boldsymbol{p}|_{\mathrm{F}}^2$ is equal to 0 everywhere.

Furthermore, Equation (6.27) implements a similarity of $\boldsymbol{p}$ and $\boldsymbol{\nabla} d$ that is enforced along edges ($\boldsymbol{r}_2$) and relaxed across ($\boldsymbol{r}_1$). This becomes obvious when considering the tensor $\boldsymbol{D}$ in (6.28) that adapts the amount of similarity in a directional dependent way. This is achieved by a solution-driven scaling of the eigenvalues of $\boldsymbol{D}$, where its eigenvectors are given by $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$.

6.3.5 Anisotropic Modification of Smoothness Term

Let us now introduce anisotropy into the smoothness term in a similar way. To this end, we first rewrite the isotropic variant from (6.21) by means of the unit vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$:

$$S_{\mathrm{I}}(\boldsymbol{p}) \;=\; \Psi\big(|\boldsymbol{\mathcal{J}}\boldsymbol{p}|_{\mathrm{F}}^2\big) \;=\; \Psi\Big(\sum_{i=1}^{2}\sum_{j=1}^{2}\big(\boldsymbol{e}_j^{\mathsf{T}}\boldsymbol{\mathcal{J}}\boldsymbol{p}\,\boldsymbol{e}_i\big)^2\Big)\,. \tag{6.29}$$

The term $\boldsymbol{e}_j^{\mathsf{T}}\boldsymbol{\mathcal{J}}\boldsymbol{p}\,\boldsymbol{e}_i$ can be interpreted as a pendant of the second order directional derivative $\partial_{\boldsymbol{e}_j\boldsymbol{e}_i} d = \boldsymbol{e}_j^{\mathsf{T}}\boldsymbol{\mathcal{H}} d\,\boldsymbol{e}_i$ where $\boldsymbol{\mathcal{H}} d$ represents the Hessian of d. Our goal is to penalise this term differently along and across image structures. Hence, similarly to the anisotropic modification of the coupling term, we modify Equation (6.29) by exchanging $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ with $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, and swapping the positions of the penalisation function Ψ and the summation $\sum_{i=1}^{2}$. This leads to our anisotropic smoothness term

$$S_{\mathrm{A}}(\boldsymbol{p}) \;=\; \sum_{i=1}^{2}\Psi_i\Big(\sum_{j=1}^{2}\big(\boldsymbol{r}_j^{\mathsf{T}}\boldsymbol{\mathcal{J}}\boldsymbol{p}\,\boldsymbol{r}_i\big)^2\Big)\,, \tag{6.30}$$

where we again apply different penalisations Ψ_i in both directions. Also here, let us shed light on the introduced anisotropy by analysing the associated Euler-Lagrange equation of (6.30) that is given by

$$\mathbf{div}(\mathcal{J}\boldsymbol{p}\,\boldsymbol{T}) \;=\; \begin{pmatrix}\operatorname{div}(\boldsymbol{T}\,\boldsymbol{\nabla} p_1)\\ \operatorname{div}(\boldsymbol{T}\,\boldsymbol{\nabla} p_2)\end{pmatrix} \;=\; \mathbf{0}\,, \tag{6.31}$$

where **div** is a common extension of the standard divergence operator div. In particular, **div** applies the divergence operator div to the rows of a matrix-valued function and thus yields a column vector with two components. Equation (6.31) can be seen as an anisotropic diffusion of the coupling variable $\boldsymbol{p}$. Here, the diffusion tensor

$$\boldsymbol{T}(\boldsymbol{p}) \;=\; \sum_{i=1}^{2} \Psi_i'\Big(\sum_{j=1}^{2} \big(\boldsymbol{r}_j^{\mathsf{T}}\,\mathcal{J}\boldsymbol{p}\,\boldsymbol{r}_i\big)^2\Big)\cdot \boldsymbol{r}_i\boldsymbol{r}_i^{\mathsf{T}} \tag{6.32}$$

effects this anisotropic behaviour: We smooth the coupling variable $\boldsymbol{p}$ differently across and along image structures, where the amount of smoothness is determined by the eigenvalues of $\boldsymbol{T}$.

6.3.6 Anisotropic Coupling Model

With the proposed anisotropic coupling (6.25) and smoothness term (6.30), our fully anisotropic coupling model is given by

$$R_{\mathrm{A}}(d) \;=\; \inf_{\boldsymbol{p}} \left\{ \int_\Omega \Big(C_{\mathrm{A}}(d,\boldsymbol{p}) + \beta\cdot S_{\mathrm{A}}(\boldsymbol{p})\Big)\,\mathrm{d}\boldsymbol{x}\right\}. \tag{6.33}$$

As in the isotropic case (6.21), the coupling term $C_{\mathrm{A}}(d,\boldsymbol{p})$ is responsible for handling jumps whereas the smoothness term $S_{\mathrm{A}}(\boldsymbol{p})$ is responsible for handling kinks. However, contrary to the isotropic model, our novel anisotropic model now effectively incorporates directional information to steer this coupling and smoothing process.

Furthermore, for scenarios where jumps and kinks of the unknown function highly correlate with edges of the guidance image, it is beneficial to include also the strength of an image edge in addition to its direction. To this end, we scale both summands of the coupling term (6.25) and of the smoothness term (6.30) with $g_i\big((\boldsymbol{r}_i^{\mathsf{T}}\boldsymbol{\nabla} f_\sigma)^2\big)$, where $g_i(z^2)$ is a decreasing function with $g_i(0) = 1$, and $f_\sigma = G_\sigma * f$ a smoothed version of the guidance image f. This further reduces coupling and smoothing across image edges while enforcing it along them. Referring to Section 6.3.4 and 6.3.5, this solely causes an additional scaling of the eigenvalues of the tensors $\boldsymbol{D}$ and $\boldsymbol{T}$ in Equation (6.28) and (6.32), respectively.

In Table 6.6 we summarise different regularisation terms that result from our model with specific parameter choices. We will evaluate those regularisers in Section 6.3.8.

Table 6.6: Overview of regularisers covered by our model. Note that $\boldsymbol{D}$ and $\boldsymbol{T}$ degenerate to the identity matrix $\boldsymbol{I}$ if $\Psi_i(z^2) = z^2$ and $g_i(z^2) = 1$ for $i \in \{1, 2\}$.

	regularisation model		$\boldsymbol{p}$	$\boldsymbol{D}$	$\boldsymbol{T}$
(FI)	1st order	isotropic	fixed to $\mathbf{0}$	$\boldsymbol{I}$	$\boldsymbol{I}$
(FA)	1st order	anisotropic	fixed to $\mathbf{0}$	Eq. (6.28)	Eq. (6.32)
(CI)	coupled	isotropic	optimised	$\boldsymbol{I}$	$\boldsymbol{I}$
(CA)	coupled	anisotropic	optimised	Eq. (6.28)	Eq. (6.32)

6.3.7 Minimisation

Combining the proposed regularisation term (6.33) with the data term (6.10) of our variational depth from focus approach leads to the energy functional

$$E(d) \;=\; \int_\Omega \sum_{i=1}^{n_{\tilde{d}}} w_i \cdot \Psi\Big((d - \tilde{d}_i)^2\Big)\,\mathrm{d}\boldsymbol{x} + \alpha \inf_{\boldsymbol{p}} \left\{ \int_\Omega \big(C_\mathrm{A}(d, \boldsymbol{p}) + \beta \cdot S_\mathrm{A}(\boldsymbol{p})\big)\,\mathrm{d}\boldsymbol{x} \right\} \tag{6.34}$$

with the smoothness weight $\alpha > 0$. A minimiser of this energy functional has to fulfil the Euler-Lagrange equations

$$\sum_{i=1}^{n_{\tilde{d}}} w_i(\boldsymbol{x})\, \Psi'\Big(\big(d(\boldsymbol{x}) - \tilde{d}_i(\boldsymbol{x})\big)^2\Big)\,\big(d(\boldsymbol{x}) - \tilde{d}_i(\boldsymbol{x})\big) - \alpha\, \mathrm{div}(\boldsymbol{D}(\boldsymbol{\nabla} d - \boldsymbol{p})) \;=\; 0\,, \tag{6.35}$$

$$\boldsymbol{D}(\boldsymbol{p} - \boldsymbol{\nabla} d) - \beta\, \mathbf{div}(\boldsymbol{\mathcal{J}}\boldsymbol{p}\boldsymbol{T}) \;=\; \mathbf{0}\,. \tag{6.36}$$

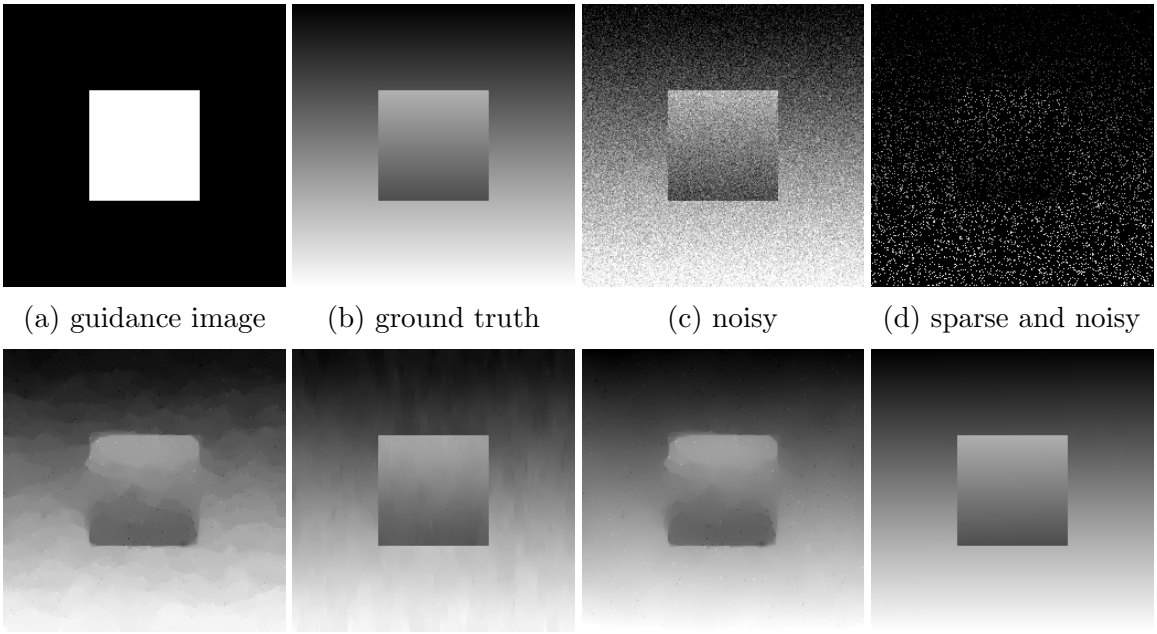
With $\boldsymbol{n}$ as outer normal vector on the image boundary $\partial\Omega$, the corresponding boundary conditions read

$$(\boldsymbol{\nabla} d - \boldsymbol{p})^\mathsf{T} \boldsymbol{D}\,\boldsymbol{n} \;=\; 0 \qquad \text{and} \qquad \boldsymbol{\mathcal{J}}\boldsymbol{p}\,\boldsymbol{T}\,\boldsymbol{n} \;=\; \mathbf{0}\,. \tag{6.37}$$

We discretise Equation (6.35) and (6.36) on a uniform rectangular grid and approximate the derivatives at intermediate grid points. Accordingly, we appropriately discretise the divergence expressions with the approach of Weickert et al. [WWW13] using the stencil parameters $\alpha_\mathrm{stencil} = 0.4$ and $\gamma_\mathrm{stencil} = 1$; cf. also Section 5.A. We solve the resulting system of equations with the discussed FSI scheme from Chapter 3.

6.3.8 Evaluation

In this section we evaluate the proposed regularisation model and compare it to the baseline methods from Table 6.6. To this end, we consider a synthetic data set where ground truth is available. In Figure 6.12(a–d), we depict the input guidance image, the ground truth depth map that consists of two segments with a linear slope in vertical direction, a noisy depth map, and a sparse version of it. The last one serves as input for our evaluation. More specifically, we generated the input depth map



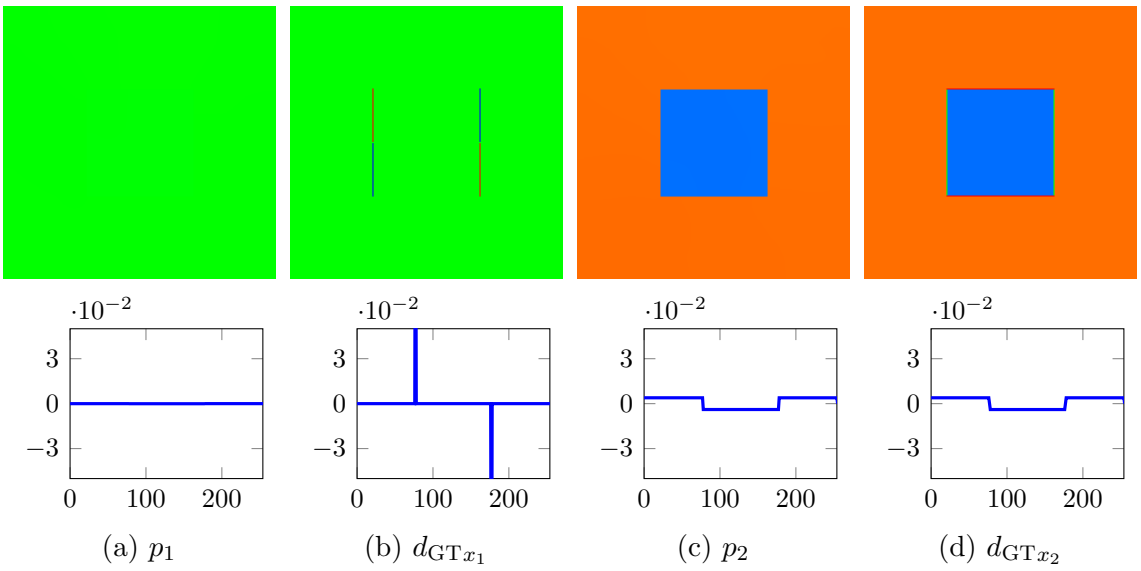
(a) guidance image (b) ground truth (c) noisy (d) sparse and noisy

(e) FI (2.85) (f) FA (1.96) (g) CI (2.60) (h) CA (**0.44**)

Figure 6.12: Synthetic experiment. (a) Input guidance image. (b) Ground truth depth map. (c) Noisy version of (b). (d) Sparse and noisy version of (b) that serves as input depth map. (e–h) Computed depth maps with different regularisation techniques; cf. Table 6.6. We state the *root mean squared error* (RMSE) between the computed and the ground truth depth map in brackets under the corresponding results ($\times 10^{-2}$).

in the following way: First we add Gaussian noise of standard deviation 0.1 to the ground truth depth map, where the initial depth values range from 0 to 1. Next, we randomly select 10% of this noisy version to obtain the final sparse and noisy input depth map that we consider as input for our approach.

In Figure 6.12(e–h), we show the resulting depth maps that are computed with *first order isotropic* (FI), *first order anisotropic* (FA), *coupled isotropic* (CI), and *coupled anisotropic* (CA) regularisation; cf. also Table 6.6. For each approach the regularisation parameters α and β are optimised w.r.t. the *root mean squared error* (RMSE). The resulting RMSEs between the ground truth and the computed depth maps are listed right below the corresponding results in Figure 6.12. First, this experiment demonstrates that incorporating directional information from the guidance image is highly beneficial. Both first and coupled anisotropic regularisers outperform their isotropic counterparts. With anisotropic regularisation the edges of the computed depth maps are desirably sharp, while the isotropic variants cannot provide this quality. Second, the assumption of piecewise affine functions is much more suited than assuming piecewise constant depth maps in this case. Accordingly, both second order coupling models yield superior results than their corresponding first order variants. It is clearly visible that the latter ones lead to piecewise constant patches, which is not desirable in the considered scenario. Here, the difference between the assumption



(a) p_1 (b) $d_{\mathrm{GT}x_1}$ (c) p_2 (d) $d_{\mathrm{GT}x_2}$

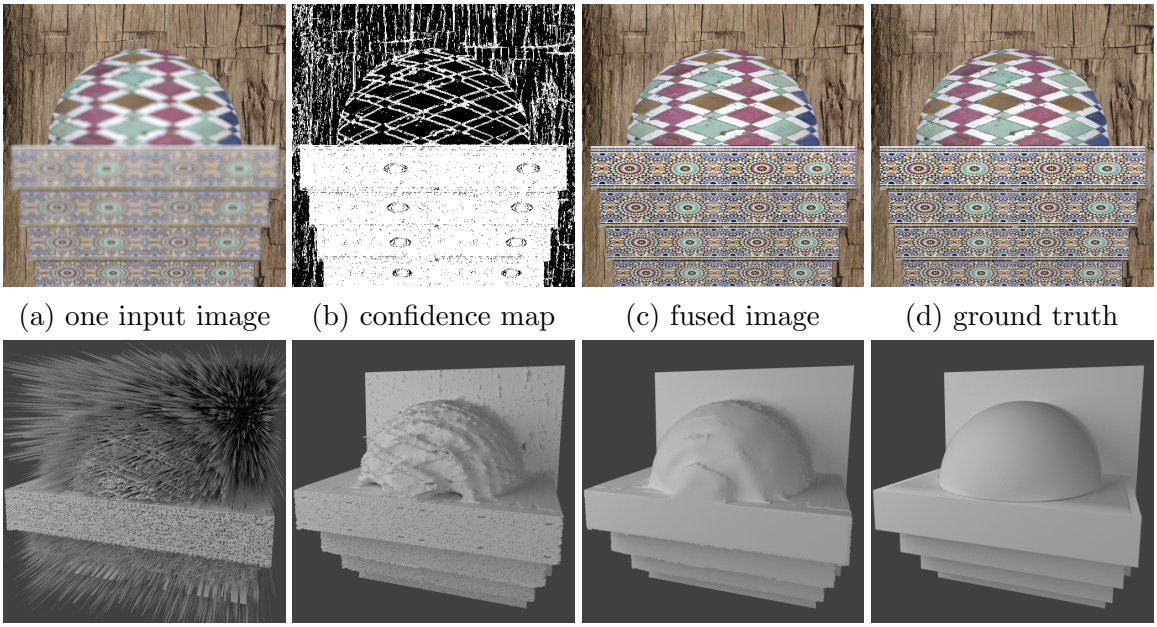
Figure 6.13: Comparison of the computed coupling vector field $\boldsymbol{p} = (p_1, p_2)^\mathsf{T}$ and the gradient of the ground truth depth map $\boldsymbol{\nabla} d_{\mathrm{GT}} = (d_{\mathrm{GT}x_1}, d_{\mathrm{GT}x_2})^\mathsf{T}$ from Figure 6.12(b). *Top*: Colour coded values from $-0.5 \cdot 10^{-2}$ (blue) over 0 (green) to $0.5 \cdot 10^{-2}$ (red). *Bottom*: Horizontal scanlines of row 100 where the image size is 256×256.

of piecewise constant and piecewise affine depth maps is especially noticeable in the background region. Last but not least, the proposed coupled anisotropic regulariser provides the best result, both visually and in terms of the RMSE.

Moreover, let us illustrate an important feature of the proposed anisotropic coupling model in more detail. To this end, Figure 6.13 juxtaposes the computed coupling vector field $\boldsymbol{p}$ to the gradient $\boldsymbol{\nabla} d_{\mathrm{GT}}$ of the ground truth depth map d_{GT} from Figure 6.12(b). Across an edge of the guidance image f, i.e. in $\boldsymbol{r}_1$-direction, the coupling term allows a deviation of $\boldsymbol{p}$ from $\boldsymbol{\nabla} d_{\mathrm{GT}}$. In fact, there are horizontal jumps in the computed depth map d, but these discontinuities are not contained in p_1; cf. Figure 6.12(h) and Figure 6.13(a,b). In other words, p_1 can be constant while still allowing a jump in d. This demonstrates that the computed $\boldsymbol{p}$ does not have to resemble the gradient of the ground truth depth map d_{GT} everywhere to get a good approximation. In fact, deviations in a specific direction are allowed and even beneficial to get a smaller energy and consequently better depth estimates.

6.3.9 Application to Depth from Focus

In this section, we want to demonstrate the performance of our novel anisotropic second order coupling model for the task of depth from focus; cf. also Section 6.2. To this end, let us first consider our synthetic focal stack from Figure 6.1. It contains ground truth data that allows for a comparison in terms of the *mean squared error* (MSE). Figure 6.14(a) shows one of the thirteen focal stack images. In Figure 6.14(b) we depict the confidence map w from Section 6.2 that is computed with the Frobenius



(a) one input image (b) confidence map (c) fused image (d) ground truth

(e) initial depth (f) Section 6.2 (g) proposed (h) ground truth

Figure 6.14: Results for synthetic data set from Figure 6.1. *Top*: (a) One of the thirteen unsharp input images. (b) Confidence map. (c) Our fused image. (d) Ground truth image. *Bottom*: Rendered depth maps. The initial depth map in (e) is created with the Frobenius norm of the Hessian as sharpness measure with $\sigma = 0.5$ (cf. Table 6.2). In (f), we show the corresponding final depth map computed with the first order anisotropic approach from Section 6.2. The depth map in (g) is computed by means of our novel anisotropic second order coupling model, and (h) depicts the ground truth depth map for comparison.

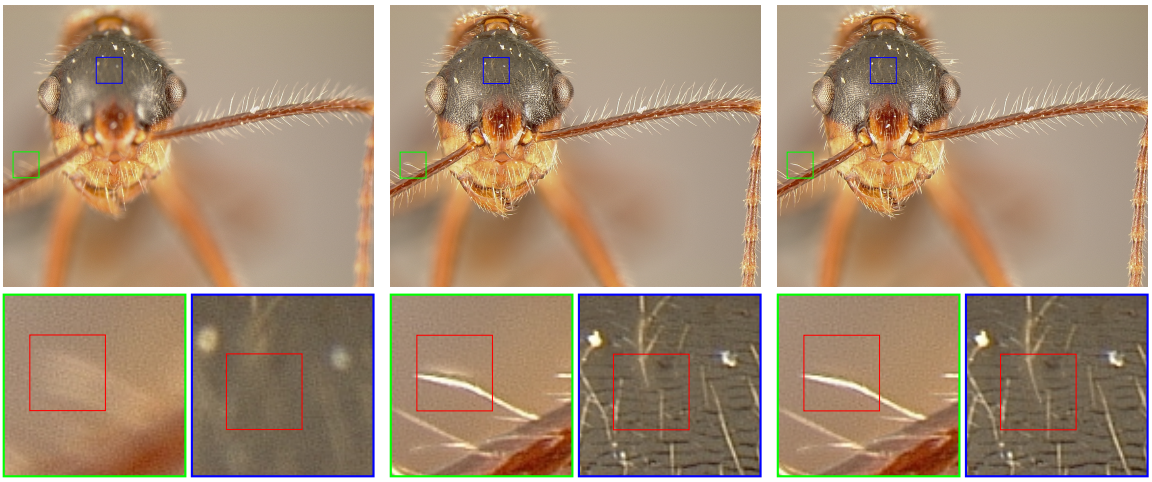
norm of the Hessian as sharpness measure using the parameters $\sigma = 0.5$ and $\theta = 41$; cf. Table 6.2. In Figure 6.14(e–h), we compare the corresponding initial depth, the result from Section 6.2, and our novel depth estimation to the ground truth. We see that both smoothing techniques are able to improve the initial depth map effectively. Moreover, our novel anisotropic second order approach leads to significantly less staircasing artefacts than the first order smoothness process from Section 6.2. In fact, it resembles the ground truth depth map convincingly. Also our corresponding fused image in Figure 6.14(c) offers a high quality and is close to the ground truth all-in-focus image in Figure 6.14(d). These findings are underlined by Table 6.7, where we compare our results in terms of the MSE between the fused image and its ground truth. Using the initial depth map to fuse the images yields an error of 10.55. This is improved by the first order approach from Section 6.2 which leads to an MSE of 3.47. Exchanging the first order regularisation technique by our novel anisotropic second order approach provides a further improvement with an MSE of 3.08.

The comparison to further state-of-the-art approaches in Table 6.7 illustrates the usefulness of our technique for the task of depth from focus.

In Figure 6.15, we additionally demonstrate the performance of our approach by

Table 6.7: *Mean squared error* (MSE) between computed and ground truth all-in-focus image (cf. also Table 6.3). Best result is written in bold face.

Forster et al. [FVB+04]	Agarwala et al. [ADA+04]	Aguet et al. [AVU08]	1st order anisotropic	coupled anisotropic
152.12	135.97	113.73	3.47	**3.08**



(a) one input image (b) Section 6.2 (c) coupled anisotropic

Figure 6.15: Results for real-world focal stack of an insect from Figure 6.6. *From left to right*: (a) One of the thirteen unsharp input images. (b) Fused image computed with the anisotropic first order approach from Section 6.2; cf. also Figure 6.6(h). (c) Our fused image computed with our novel anisotropic second order coupling model. The zooms into the images provide more details. Here, the red rectangles indicate important differences.

means of a real-world experiment with a focal stack of an insect; cf. also Figure 6.6. Since no ground truth data is available, we have to restrict ourselves to a visual comparison. To this end, we depict one unsharp input image of the focal stack in Figure 6.15(a). The resulting fused image obtained with the first order approach from Section 6.2 is shown in Figure 6.15(b), and the one computed with our novel anisotropic second order coupling model in Figure 6.15(c). Especially the zooms illustrate that our new all-in-focus image contains less errors and more small scale details than the one produced with the first order approach from Section 6.2.

6.3.10 Limitations and Discussion

We have seen that the proposed anisotropic second order coupling model is well-suited for piecewise affine functions and that it beneficially incorporates directional information from a guidance image. In this way, we were able to improve the quality of our depth from focus approach from Section 6.2 even further.

As a main drawback, we want to mention that the gain in quality comes at the price of a larger runtime of the algorithm. While the first order approach offers runtimes

of less than five seconds for image sizes such as 1024×1024, the proposed second order approach requires more than twice as much. In this regard, we also need to take into account the additional smoothness parameter β. On the one hand, this parameter allows to steer the smoothness of the solution in a desired way. However, on the other hand, it is an additional parameter that has to be chosen carefully to provide good results. In this context, the bilevel parameter learning approach for higher order models of De Los Reyes et al. [DSV17] might be an interesting starting point for future work.

6.4 Summary and Conclusions

6.4.1 Variational Depth from Focus

In the first part of this chapter (Section 6.2), we have presented our variational framework for depth form focus. Our method incorporates modern techniques such as robustified data fidelity terms and joint image- and depth-driven anisotropic regularisation. Moreover, our flexible approach is able to combine the information from multiple in-focus measures. In this way, we provide state-of-the-art results. We demonstrate this with several in-focus measures and evaluate our final results w.r.t. competing methods from the literature. Not only the all-in-focus composites benefit from the proposed approach, but also the computed depth maps provide valuable information that may serve as input for further computer vision tasks. Last but not least, the proposed algorithmic realisation on the GPU leads to desirably fast algorithms.

As discussed in Section 6.2.7, a direct optimisation for a maximal sharpness measure in the data term of our energy functional might be a promising direction for future work. Moreover, there exist several approaches that estimate a blur map based on a single input image; see e.g. [BD07,SXJ15]. This for instance allows to a certain extent to refocus the captured image in a postprocessing step. It would be interesting to investigate how such approaches benefit from the anisotropic regularisation techniques presented in this work. Vice versa, the applied blurriness measures might be also useful for the proposed depth from focus technique; cf. our discussion in Section 6.2.7.

6.4.2 Anisotropic Second Order Coupling Model

We have seen that anisotropic techniques allow to obtain results of highest quality when using first order regularisation. On the other hand, recent developments have rendered higher order regularisation very attractive. Hence, in the second part of this chapter (Section 6.3), we built a bridge between both approaches and systematically combined anisotropic ideas and higher order regularisation. As a result, our novel anisotropic second order coupling model allows to steer the preferred direction of jumps and kinks by means of local image structures. To achieve this, we have introduced a direction-dependent behaviour both in the coupling and the smoothness term. We have experimentally shown that this yields superior results compared to

baseline approaches, i.e. first order anisotropic and second order isotropic regularisation. Moreover, we have demonstrated the usefulness of the proposed regularisation technique for the task of depth from focus.

In this regard, we believe that our novel anisotropic second order smoothness term is advantageous for further computer vision applications such as stereo or optic flow computation as well. In fact, first applications of our model can for instance be found in [MJBB16, MSV+17, MSB17]. Moreover, an extension of our model to even higher order in the sense of [Har86, BKP10, HWS+13] might be an interesting topic for further research. Besides, the presented method is by construction well-suited for applications that intrinsically come along with a so-called guidance image. If such a guidance image is not available, a determination of the smoothing directions by means of the evolving solution in the sense of edge-enhancing anisotropic diffusion processes (cf. [Wei98]) might be beneficial, and thus interesting to investigate in future work.

Chapter 7

Conclusions and Outlook

"Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning."

Albert Einstein

pixabay.com

Contents

7.1 Summary and Conclusions 173

7.2 Outlook 176

7.1 Summary and Conclusions

As discussed in Chapter 1, the main goal of our work was the fusion of multiple images to one composite that offers more information than any of the inputs. To achieve this goal, we approached the individual fusion tasks within a variational framework, since such a variational framework provides a transparent modelling with a solid mathematical background.

Fast Semi-Iterative Schemes. In Chapter 3, we introduced a novel solver class for the solution of such variational problems, our so-called *Fast Semi-Iterative* (FSI) schemes. In fact, they are well-suited for various diffusion-like processes, elliptic problems, and constrained convex optimisation. In particular, we considered a recursion relation between box filtering and an explicit scheme for 1D homogeneous diffusion. Generalisations of this concept led to our FSI schemes that extrapolate a basic solver iteration by means of the older iterate. The resulting algorithms are simple to implement and well-suited for parallel computing. We have demonstrated their benefits for image processing tasks and Nesterov's worst case problems for convex and strongly convex optimisation. In fact, we applied our FSI schemes as efficient solvers for our image fusion applications in Chapter 4, 5, and 6.

Variational Image Fusion. In Chapter 4, we have presented our general variational image fusion technique. To this end, we have exploited the observation that many fusion tasks aim for a composite image with a high local contrast. Hence, we based on important findings in variational contrast enhancement, and designed an image fusion technique that directly opts for an optimal composite image. This output-driven idea was the key for our versatile fusion method that compares favourably to specifically tailored approaches in various application areas. Thorough evaluations demonstrate the performance of our technique for its three main application areas, namely multispectral imaging, decolourisation, and exposure fusion. Indeed, our general method produces results of high quality, and even outperforms competing approaches for various example images. On top of that, all components of our variational model have an intuitive meaning and in this way, enable the user to individually adjust the output image to their personal preferences. We believe that our technique is generally suitable for applications that require the fusion of multiple images. It provides composite images that condense the most important information from the input stacks in a desirable way, and offer important features such a local contrast adaptation.

Deghosting and HDR Imaging. Already small movements of the camera or moving scene objects during the acquisition of the exposure set might result in severe ghosting artefacts in the fused composites. To compensate for such motion, we have proposed two conceptually different approaches in Chapter 5.

First, we have considered a two-step approach that consists of an image alignment step followed by a ghost removal technique as postprocessing. Obviously, severe brightness changes between the differently exposed images of the exposure stack render the alignment task highly difficult. In a general setting, we have to cope with any kind of monotonically increasing intensity rescalings. In this context, a feasible constancy assumption is that the intensity relations within a certain neighbourhood remain unchanged. In fact, based on this observation, we introduced the *complete rank transform* (CRT) that makes maximal use of this assumption. More specifically, it stores for every pixel a signature vector that encodes the intensity relations (ranks) in a neighbourhood patch. We embedded those CRT signatures as constancy assumption into a variational approach for optic flow computation. Our experiments demonstrated benefits w.r.t. related methods from the literature. Moreover, we illustrated the performance of our optic flow approach for the alignment of exposure series. In this context, violations of the model assumptions such as highly over- or under-saturated image regions might lead to inaccurate flow vectors in certain parts of the image. Unfortunately, already small misalignments may cause severe ghosting artefacts when fusing the images to an overall well-exposed composite image. Hence, as a remedy we proposed an additional ghost removal step as postprocessing. To this end, we based on the concept of a *brightness transfer function* (BTF) that relates the brightness values between different frames. We computed those BTFs by means of image histograms that provide a robustness w.r.t. small scene motion. With the help of the computed BTFs we performed bidirectional consistency checks to iden-

tify and to eliminate remaining ghosting artefacts. The proposed approach provides consistently good results for various data sets as demonstrated by our experiments.

Provided further knowledge about the exposure stack, we presented an alternative approach to the above two-step method of optic flow-based pre-alignment and subsequent ghost removal. In fact, we have proposed the first fully coupled approach that simultaneously computes an aligned HDR irradiance map as well as dense displacement fields. The introduced feedback in this joint computation of the irradiances and the optic flow turned out to be highly beneficial. In this way, our approach offers important features such as an inherent ghost removal and an intuitive weighting of the intensity values in the HDR reconstruction. Our results demonstrated advantages to an optic flow-based pre-alignment. In fact, our joint approach marks the state-of-the-art for optic flow computation in the HDR reconstruction setting.

Depth from Focus with Anisotropic Second Order Regularisation. Also in the context of focal stacking, we made profit from prior information about the image set. In particular, knowledge about the relative distances of the camera to the focal planes additionally allowed us to infer depth information about the captured scene. To this end, we have presented a variational framework for the task of depth from focus in Chapter 6. In the first step, we determined initial depth maps based on specific sharpness measures. Next, we smoothed those depth maps with our variational technique. In particular, we combined robust data fidelity terms with an anisotropic regularisation strategy to achieve state-of-the-art results. In this way, not only the fused output images but also the computed depth maps are highly accurate and provide valuable information for further applications.

Moreover, several approaches have uncovered the need for second order regularisation in the context of depth map computation. This is due to the fact that many real-world scenes can be well described by (piecewise) planar objects. In this regard, we have presented the first fully anisotropic coupling model that incorporates directional information from the evolving all-in-focus image. More specifically, we have systematically combined anisotropic ideas and higher order regularisation. As a result, our proposed anisotropic second order coupling model allows to guide jumps and kinks of the solution in a direction-dependent way. On top of that, with this novel regularisation technique we improved the quality of our depth from focus approach even further.

General Conclusions. We have seen how rather intuitive concepts such as box filtering, contrast enhancement, and direction-dependent smoothing can lead to highly flexible algorithms that provide state-of-the-art results. In this regard, we have designed our models in a transparent variational framework with an established mathematical background. In this way, we beneficially coupled individual tasks in joint models that highly profit from the induced mutual feedback. Overall, we preferred simple and transparent models with solid mathematical foundations to sophisticated approaches that are tuned to specific benchmarks. In this way, we uncovered the main

ingredients and ideas that are most responsible for qualitatively good results.

Moreover, we have seen how to exploit additional prior knowledge about the input stacks, e.g. the exposure times or the relative distances of the focal planes. In this way, we could extract further information about the captured scenes such as HDR irradiance or depth values. This shows a rather general principle: Any information about the imaging process should be considered when designing image analysis methods. This may require drastic model changes, but ensures an optimal use of the provided input data. A similar principle is valid for the design of invariant features; cf. Chapter 5. Here, one should aim at a feature that carries as much information as possible within the desired class of invariance. Obviously, here the meaning of information differs for each application.

7.2 Outlook

In fact, we have already discussed several modifications and extensions of our techniques in the individual chapters of this thesis. At this point, we want to discuss possible directions for future work on a more general level.

Joint Fusion Approaches. In this work, we have considered HDR imaging (Chapter 5) and depth from focus (Chapter 6) as two independent research tasks. However, provided suitable input stacks, it might be also possible to approach both problems in a joint manner. In this regard, changing the aperture of a camera from frame to frame creates images that are at the same time differently exposed *and* differently focused. Hence, they generally allow both, HDR and depth computation at the same time. Obviously, this problem setting is harder than the individual problems since there are several ambiguities. Moreover, changing the aperture effects the intensity values differently from changing the exposure time. In a similar way, an aperture change causes focus changes that are more difficult to predict than those caused by focal plane distance variations. However, first approaches for such a setting show promising results; see e.g. [HK07, Has08]. Hence, this problem poses an interesting topic for future work.

In a similar way, it might be possible to consider the task of stereo computation and depth from focus in a joint ansatz. First approaches that go into such a direction can be found in [FG06, LSWY10, THS13, THMR13]. Due to their intrinsic advantages and shortcomings, this might be a fruitful way to increase the quality of depth estimation in difficult scenarios. One application area for such an approach might be for instance smartphones that provide two slightly shifted cameras with different focal settings or light field cameras; see e.g. [Lev06, MM16].

Fusion of Images with Variable Resolution. The fusion tasks considered in this work aim at fusing images of the same size. Moreover, also the computed flow fields, HDR irradiance maps, and depth maps are of the same size as the input images.

Obviously, in some scenarios it might be appropriate to fuse only certain image parts or to combine images (or related information) that have different resolutions.

In this regard, as a first example application we want to mention image cloning. Here, the task is to clone parts of one image into another. A popular technique to approach such a seamless image cloning task are gradient domain methods. Here, the idea is to modify the image content in the gradient domain, and to later integrate the resulting gradient field to an image that is free of visible seams; see e.g. [PGB03]. Interestingly, our variational approach from Chapter 4 as well as our second order coupling model from Chapter 6 share strong similarities to nonlocal gradients and anisotropic gradient domain methods, respectively. Hence, it would be highly interesting to modify and evaluate our techniques for such an image cloning task and related image processing applications. A similar idea to gradient domain approaches is to first manipulate a so-called drift vector field, and to later integrate this field to a smooth final result; see e.g. [WHBV13]. Also here, such modifications might be beneficial.

Another popular example for the fusion of images with variable resolution is super-resolution. Here, the task is to combine several low resolution images to a single composite with high resolution. Related problems are the upsampling of low resolution multispectral data by means of a high resolution *panchromatic* (PAN) image or the upsampling of a low resolution depth map by means of a guidance image with higher resolution. For all problems, there exist successful variational approaches, see e.g. [UPWB10, BCI+06, FRR+13], that might further benefit from ideas presented in this work.

Closing the loop to joint fusion approaches, also in such settings it could be advantageous to combine several tasks. As an example, let us consider Microsoft's Kinect sensor [Mic16]. It provides a visible spectrum image and a near-infrared image with high resolution. Additionally, the user has access to a low resolution depth map. Obviously, it seems to be a fruitful idea to upsample this depth map by means of the high resolution images. In addition, it might be advantageous to combine this upsampling task with a cross-modal stereo computation between the visible spectrum and the near-infrared image; see e.g. [CBF11]. This also provides valuable depth estimates. In such a joint model, both tasks could highly benefit from the induced mutual feedback, and in this way lead to accurate high resolution depth maps.

Under the guise of image fusion we have contributed with this work to several research fields in image processing and computer vision: efficient numerical schemes, multispectral imaging, decolourisation, exposure fusion, high dynamic range imaging, stereo and optic flow computation, ghost removal, depth from focus, and higher order regularisation. Accordingly, we hope that the presented concepts as well trigger further research into several directions.

Appendix A

Notation

Before we give a detailed list of the symbols used in this thesis, let us first comment on our general notation conventions. We denote scalars and vectors in lower-case, where vectors are additionally written in bold face. Furthermore, we represent matrices by upper-case and bold face letters. As an example, the i-th entry of a vector $\boldsymbol{a}$ is given by a_i. The entry in the i-th row and j-th column of a matrix $\boldsymbol{A}$ is denoted by a_{ij}. Moreover, we write the channels of colour spaces in a non-italic font style.

a	entry of diffusion tensor $\boldsymbol{D}$
a_i	coefficient of polynomial function
$\boldsymbol{a}$	arbitrary vector
$\boldsymbol{A}$	symmetric negative semi-definite matrix
b	distance between transferred and observed intensity value (Section 5.A: entry of diffusion tensor $\boldsymbol{D}$)
$\boldsymbol{b}$	arbitrary vector
B_{2n+1}	box filter of length $(2n+1)h$
$\mathcal{B}$	structuring element
$\boldsymbol{B}$	symmetric positive definite matrix
c	entry of diffusion tensor $\boldsymbol{D}$
C	place holder for colour channel
$\bar{\mathrm{C}}$	transformed colour channel C
C	coupling term
$\mathcal{C}$	closed convex set
$\mathcal{C}_{\mathrm{ball}}$	closed convex unit ball set
$\mathcal{C}_{\mathrm{box}}$	closed convex box set
$\mathcal{C}_{\mathrm{simp}}$	closed convex simplex set
$\boldsymbol{C}$	lower triangular matrix with real and positive diagonal entries
d	depth map
$\tilde{d}$	initial depth map
$\boldsymbol{d}$	known right hand side
D	data term
D^ℓ	data term on level ℓ
$\bar{D}^\ell$	linearised data term on level ℓ
$\mathcal{D}$	range of signature digit
$\boldsymbol{D}$	diffusion tensor

det	determinant
div	divergence
div	divergence of matrix-valued function that applies div to rows
$\boldsymbol{dv} = (dv_1, dv_2)^\mathsf{T}$	optic flow increment
E	energy functional
E^ℓ	energy functional on level ℓ
$\boldsymbol{e}^k$	error vector at time level k
$\boldsymbol{e}_1$, $\boldsymbol{e}_2$	unit vector in horizontal and vertical direction
f	grey-valued input image
$f_{i,j}$	grey-value at pixel (i, j)
f_σ	image smoothed with Gaussian of standard deviation σ
$\bar{f}$	mean value in neighbourhood
$\tilde{f}$	attachment image
$\boldsymbol{f}$	colour-valued input image
$\mathfrak{f}$	vector representing image patch
F	convex function
$\mathcal{F}$	integrand of energy functional
g	diffusivity function
G_σ	Gaussian with standard deviation σ
$\boldsymbol{G}$	smooth strictly maximal monotone map
h	grid size in 1D
h_1, h_2	grid size in horizontal and vertical direction
H	cumulative histogram
$\boldsymbol{\mathcal{H}}$	Hessian matrix
i	index variable
I	identity operator
$\boldsymbol{I}$	identity matrix
j	index variable
$\boldsymbol{J}$	structure tensor
$\boldsymbol{\mathcal{J}}$	Jacobian matrix
k	time level or iteration number
K	patch size
ℓ	warping level (Chapter 3: strong convexity parameter)
L, a, b	lightness channel and colour opponent dimensions of CIE-Lab colour space
L	discrete Laplacian in 1D
$\mathscr{L}$	Lipschitz constant
$\boldsymbol{L}$	symmetric negative semi-definite matrix implementing the 1D Laplacian
m	in-focus measure (Chapter 3: cycle number)
M	number of unknowns with $M \ll N$

n	number of iterations
	(Chapter 2: dimensionality of unknown vector)
n_1, n_2	number of pixels in horizontal and vertical direction
n_a	number of coefficients
$n_{\tilde{d}}$	number of initial depth maps
n_f	number of input images
n_u	number of unknown functions
$\boldsymbol{n}$	outer normal vector on $\partial\Omega$
N	number of unknowns
$\mathcal{N}$	neighbourhood
$\mathcal{N}_1$, $\mathcal{N}_2$	neighbourhood in horizontal and vertical direction
$\mathbb{N}$	set of natural numbers
$\mathcal{O}$	Landau symbol
$\boldsymbol{O}$	zero matrix
$\boldsymbol{p} = (p_1, p_2)^\mathsf{T}$	coupling vector field
	(Chapter 3: dual variable)
$\mathcal{P}_\mathcal{C}$	projection onto closed convex set $\mathcal{C}$
$\boldsymbol{P}$	positive definite diagonal preconditioning matrix
q	index variable
r	index of reference image
$r_\mathcal{B}$	radius of disc-shaped structure element $\mathcal{B}$
$\boldsymbol{r}_1$, $\boldsymbol{r}_2$	eigenvectors of diffusion tensor
R, G, B	red, green, and blue channel of RGB colour space
R	regularisation term
R^ℓ	regularisation term on level ℓ
$\mathbb{R}$	set of real numbers
$\boldsymbol{s}$	feature vector
S	smoothness term
t	diffusion time
t_i	exposure time for image f_i
T	stopping time of diffusion process
$T_{B_{2n+1}}$	stopping time corresponding to application of box filter B_{2n+1}
T_2	Chebyshev polynomial of the first kind with order 2
$\boldsymbol{T}$	diffusion tensor
tr	trace
u	grey-valued output image
u^k	u evaluated at time level k
u_{GT}	ground truth solution
u_{ref}	reference solution
u_{x_1}	abbreviation of $\partial_{x_1} u$
$\boldsymbol{u}$	colour-valued output image
	(Chapter 2: unknown vector or vector-valued function)
$\check{\boldsymbol{u}}$	projected vector $\boldsymbol{u}$
$\boldsymbol{u}^*$	(unknown) exact solution

U_n	Chebyshev polynomial of the second kind with order n
$\boldsymbol{v} = (v_1, v_2)^\mathsf{T}$	optic flow field (Chapter 2: arbitrary vector)
w	weight map
$\tilde{w}$	eroded weight map
x	1D spatial coordinate
$\boldsymbol{x} = (x_1, x_2)^\mathsf{T}$	2D spatial coordinate
X, Y, Z	X, Y, and Z channel of CIE-XYZ colour space
$\mathrm{X_n}$, $\mathrm{Y_n}$, $\mathrm{Z_n}$	X, Y, and Z value of reference white point
Y, Cb, Cr	luminance, blue-yellow chroma, and red-cyan chroma channel of YCbCr colour space
z	arbitrary scalar
$\boldsymbol{z}$	arbitrary vector
α	weight of regularisation term (Section 5.A: discretisation parameter; cf. α_{stencil})
α_{HB}	α parameter of heavy ball method
α_k	FSI extrapolation parameter
α_{stencil}	α parameter of anisotropic discretisation
β	regularisation weight (Section 5.A: discretisation parameter)
β_{HB}	β parameter of heavy ball method
γ	weight of contrast term (Section 5.A: discretisation parameter; cf. γ_{stencil})
γ_{stencil}	γ parameter of anisotropic discretisation
$\tilde{\gamma}$	parameter of gamma correction
δ	weight of dispersion term part
Δ	Laplacian
ϵ	value in $[0, 1]$
ε	parameter of nonlinear function Ψ
ζ	signature length
η	downsampling factor
θ	thresholding parameter
ϑ	weight of saturation term
Θ	colour transfer function for CIE-XYZ to CIE-Lab conversion
κ	parameter of Nesterov's convex optimisation problem
$\varkappa$	parameter of Nesterov's strongly convex optimisation problem
λ	parameter of nonlinear function
μ	parameter of dispersion term
μ_1, μ_2	eigenvalues of diffusion tensor
ρ	standard deviation of a Gaussian
ϱ	spectral radius
σ	standard deviation of a Gaussian
σ_n^2	variance of a box filter B_{2n+1}
τ	time step size

τ_k	varying time step sizes of FED scheme
$\tau_{\text{E-score}}$	τ value of E-score evaluation
$\varphi_{i\to j}$	brightness transfer function from image f_i to image f_j
Φ	camera response function
Φ'	derivative of camera response function Φ
$\chi_{(z_1<z_2)}$	indicator function
Ψ	nonlinear function
Ψ_λ	nonlinear function with parameter λ
Ψ'	derivative of nonlinear function Ψ
ω	relaxation parameter
Ω	rectangular image domain
$*$	convolution operator
$\boldsymbol{\nabla}$	gradient operator
$\partial\Omega$	boundary of rectangular image domain Ω
∂_{x_1}	first order partial derivative w.r.t. x_1
$\partial_{x_1x_2}$	second order partial derivative w.r.t. x_1 and x_2
$\perp$	perpendicular
$\parallel$	parallel
$\lfloor\cdot\rfloor$	floor operator
$\lvert\Omega\rvert$	cardinality of set Ω
$\lvert z\rvert$	absolute value of z
$\lvert\boldsymbol{z}\rvert$	Euclidean norm of $\boldsymbol{z}$
$\lvert\cdot\rvert_{\text{F}}$	Frobenius norm
$\lVert\cdot\rVert_\infty$	supremum norm

Appendix B

Abbreviations

A	anisotropic
AAE	average angular error
AEB	auto exposure bracketing
AEE	average endpoint error
BPE	bad pixel error
BRDF	bidirectional reflectance distribution function
BRIEF	binary robust independent elementary features
BTF	brightness transfer function
CA	coupled anisotropic
CCT	complete census transform
cf.	confer
CI	coupled isotropic
CIFT	contrast invariant feature transform
CPG	cyclic projected gradient
CRF	camera response function
CRT	complete rank transform
CT	census transform
D	dimensional
DFG	Deutsche Forschungsgemeinschaft
DRIM	dynamic range independent metric
ED	explicit diffusion
e.g.	exempli gratia (for example)
Eq.	equation
et al.	et alii or et aliae (and others)
et seqq.	et sequentes (and the following)
EV	exposure value
F	Frobenius
FA	first order anisotropic
FED	fast explicit diffusion
FI	first order isotropic
FSI	fast semi-iterative
GPU	graphics processor unit
GT	ground truth

HB	heavy ball
HDR	high dynamic range
HOG	histogram of oriented gradients
I	isotropic
IC	infimal convolution
i.e.	id est (that is)
LCIS	low curvature image simplifier
LDP	local directional patterns
LDR	low dynamic range
LIOP	local intensity order pattern
MIA	mathematical image analysis
MSE	mean squared error
MTB	median threshold bitmap
NIR	near-infrared
OBN	ordered Bell number
PAN	panchromatic
PDE	partial differential equation
PG	projected gradient
Ph.D.	Doctor of Philosophy
RMSE	root mean squared error
RT	rank transform
SIFT	scale-invariant feature transform
SSIM	structural similarity
s.t.	subject to
SURF	speeded-up robust features
TGV	total generalised variation
TV	total variation
w.r.t.	with respect to

Appendix C

Own Publications

1. M. Boshtayeva, D. Hafner, and J. Weickert. Focus fusion with anisotropic depth map smoothing. In A. Bors, E. Hancock, W. Smith, and R. Wilson, editors, *Computer Analysis of Images and Patterns*, volume 8048 of *Lecture Notes in Computer Science*, pages 67–74. Springer, Berlin, 2013.

2. M. Boshtayeva, D. Hafner, and J. Weickert. A focus fusion framework with anisotropic depth map smoothing. *Pattern Recognition*, 48(11):3310–3323, November 2015. Invited paper.

3. O. Demetz, D. Hafner, and J. Weickert. The complete rank transform: A tool for accurate and morphologically invariant matching of structures. In T. Burghardt, D. Damen, W. Mayol-Cuevas, and M. Mirmehdi, editors, *Proc. British Machine Vision Conference*, pages 50.1–50.12, Bristol, UK, September 2013. BMVA Press. Awarded the Maria Petrou Prize for Invariance in Computer Vision.

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

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

6. D. Hafner, O. Demetz, and J. Weickert. Simultaneous HDR and optic flow computation. In *Proc. 22nd International Conference on Pattern Recognition*, pages 2065–2070, Stockholm, Sweden, August 2014.

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

8. D. Hafner, P. Ochs, J. Weickert, M. Reißel, and S. Grewenig. FSI schemes: Fast semi-iterative solvers for image processing and computer vision. In B. Andres

and B. Rosenhahn, editors, *Pattern Recognition*, volume 9796 of *Lecture Notes in Computer Science*, pages 91–102. Springer, Berlin, 2016. Awarded the GCPR 2016 Best Paper Award.

9. D. Hafner, C. Schroers, and J. Weickert. Introducing maximal anisotropy into second order coupling models. In J. Gall, P. Gehler, and B. Leibe, editors, *Pattern Recognition*, volume 9358 of *Lecture Notes in Computer Science*, pages 79–90. Springer, Berlin, 2015.

10. D. Hafner and J. Weickert. Variational exposure fusion with optimal local contrast. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale-Space and Variational Methods in Computer Vision*, volume 9087 of *Lecture Notes in Computer Science*, pages 425–436. Springer, Berlin, 2015.

11. D. Hafner and J. Weickert. Variational image fusion with optimal local contrast. *Computer Graphics Forum*, 35(1):100–112, February 2016.

12. C. Schroers, D. Hafner, and J. Weickert. Multiview depth parameterisation with second order regularisation. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale-Space and Variational Methods in Computer Vision*, volume 9087 of *Lecture Notes in Computer Science*, pages 551–562. Springer, Berlin, 2015.

Bibliography

[AAB+84] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden. Pyramid methods in image processing. *RCA Engineer*, 29(6):33–41, 1984.

[AAG96] V. Alexiades, G. Amiez, and P.-A. Gremaud. Super-time-stepping acceleration of explicit schemes for parabolic problems. *Communications in Numerical Methods in Engineering*, 12(1):31–42, January 1996.

[AAHB11] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert. Image and video decolorization by fusion. In R. Kimmel, R. Klette, and A. Sugimoto, editors, *Computer Vision – ACCV 2010*, volume 6492 of *Lecture Notes in Computer Science*, pages 79–82. Springer, Berlin, 2011.

[AC01] M. Asif and T.-S. Choi. Shape from focus using multilayer feedforward neural networks. *IEEE Transactions on Image Processing*, 10(11):1670–1675, November 2001.

[ADA+04] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Transactions on Graphics*, 23(3):294–302, August 2004. `http://grail.cs.washington.edu/projects/photomontage/` (last accessed August 2016).

[AGLM93] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations in image processing. *Archive for Rational Mechanics and Analysis*, 123:199–257, September 1993.

[AK09] V. Aslantas and R. Kurban. A comparison of criterion functions for fusion of multi-focus noisy images. *Optics Communications*, 282(16):3231–3242, August 2009.

[Aky11] A. O. Akyüz. Photographically guided alignment for HDR images. In *Proc. EUROGRAPHICS 2011 (Areas Papers)*, pages 73–74, Llandudno, UK, April 2011.

[ALKC11] J. An, S. H. Le, J. G. Kuk, and N. I. Cho. A multi-exposure image fusion algorithm without ghost effect. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 1565–1568, Prague, Czech Republic, May 2011.

[AMMS08] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Dynamic range independent image quality assessment. *ACM Transactions on Graphics*, 27(3):69:1–69:10, August 2008.

[ARC06] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field? In A. Leonardis, H. Bischof, and A. Pinz, editors, *Computer Vision – ECCV 2006, Part I*, volume 3951 of *Lecture Notes in Computer Science*, pages 578–591. Springer, Berlin, 2006.

[ARS12] A. Ayvaci, M. Raptis, and S. Soatto. Sparse occlusion detection with optical flow. *International Journal of Computer Vision*, 97(3):322–338, May 2012.

[AS72] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 9th edition, 1972.

[Aub02] S. Aubenas. *Gustave Le Gray: 1820–1884*. Getty Publications, Los Angeles, September 2002.

[AV94] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10:1217–1229, December 1994.

[AVU08] F. Aguet, D. Van De Ville, and M. Unser. Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy. *IEEE Transactions on Image Processing*, 17(7):1144–1153, July 2008.

[AW95] V. Aurich and J. Weule. Non-linear Gaussian filters performing edge preserving diffusion. In G. Sagerer, S. Posch, and F. Kummert, editors, *Mustererkennung*, pages 538–545. Springer, Berlin, 1995.

[BA83] P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image codec. *IEEE Transactions on Communications*, 31(4):531–540, April 1983.

[BA91] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 292–302, Maui, HI, June 1991.

[BB15] T. Batard and M. Bertalmío. Duality principle for image regularization and perceptual color correction models. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale-Space and Variational Methods in Computer Vision*, volume 9087 of *Lecture Notes in Computer Science*, pages 449–460. Springer, Berlin, 2015.

[BBBM13] M. Benning, C. Brune, M. Burger, and J. Müller. Higher-order TV methods – enhancement via Bregman iteration. *Journal of Scientific Computing*, 54(2):269–310, February 2013.

[BBC+01] C. Ballester, M. Bertalmío, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Transactions on Image Processing*, 10(8):1200–1211, August 2001.

[BBPW04] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004, Part IV*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, 2004.

[BCI+06] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé. A variational model for P+XS image fusion. *International Journal of Computer Vision*, 69(1):43–58, August 2006.

[BCP09] M. Bertalmío, V. Caselles, and E. Provenzi. Issues about retinex theory and contrast enhancement. *International Journal of Computer Vision*, 83(1):101–119, June 2009.

[BCPR07] M. Bertalmío, V. Caselles, E. Provenzi, and A. Rizzi. Perceptual color correction through variational techniques. *IEEE Transactions on Image Processing*, 16(4):1058–1072, April 2007.

[BD07] S. Bae and F. Durand. Defocus magnification. *Computer Graphics Forum*, 26(3):571–579, September 2007.

[BE04] R. Bala and R. Eschbach. Spatial color-to-grayscale transform preserving chrominance edge information. In *Proc. Twelfth Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, volume 5, pages 82–86, Scottsdale, AZ, November 2004.

[BEF+15] F. Balle, D. Eifler, J. H. Fitschen, S. Schuff, and G. Steidl. Computation and visualization of local deformation for multiphase metallic materials by infimal convolution of TV-type functionals. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale-Space and Variational Methods in Computer Vision*, volume 9087 of *Lecture Notes in Computer Science*, pages 385–396. Springer, Berlin, 2015.

[BEG07a] A. Betrozzi, S. Esedoḡlu, and A. Gillette. Analysis of a two-scale Cahn-Hilliard model for binary image inpainting. *Multiscale Modeling & Simulation*, 6(3):913–936, 2007.

[BEG07b] A. Betrozzi, S. Esedoḡlu, and A. Gillette. Inpainting of binary images using the Cahn-Hilliard equation. *IEEE Transactions on Image Processing*, 16(1):285–291, January 2007.

[Bel93] P. N. Belhumeur. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. In *Proc. Fourth International Conference on Computer Vision*, pages 431–438, Berlin, Germany, May 1993.

[Ber14] M. Bertalmío. From image processing to computational neuroscience: A neural model based on histogram equalization. *Frontiers in Neuroscience*, 8:71:1–71:9, June 2014.

[BETV08] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, June 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.

[BFPS17] R. Bergmann, J. H. Fitschen, J. Persch, and G. Steidl. Infimal convolution coupling of first and second order differences on manifold-valued images. In F. Lauze, Y. Dong, and A. B. Dahl, editors, *Scale Space and Variational Methods in Computer Vision*, volume 10302 of *Lecture Notes in Computer Science*, pages 447–459. Springer, Berlin, 2017.

[BFW16] M. Bildhauer, M. Fuchs, and J. Weickert. An alternative approach towards the higher order denoising of images. analytical aspects. *Zapiski Nauchnykh Seminarov POMI*, 444(45):47–88, 2016.

[BFY00] M. J. Black, D. J. Fleet, and Y. Yacoob. Robustly estimating changes in image appearance. *Computer Vision and Image Understanding*, 78(1):8–31, April 2000.

[BG04] A. L. Betrozzi and J. B. Greer. Low-curvature image simplifiers: Global regularity of smooth solutions and Laplacian limiting schemes. *Communications on Pure and Applied Mathematics*, 57(6):764–790, June 2004.

[BH14] K. Bredies and M. Holler. Regularization of linear inverse problems with total generalized variation. *Journal of Inverse and Ill-posed Problems*, 22(6):871–913, December 2014.

[BH15a] K. Bredies and M. Holler. A TGV-based framework for variational image decompression, zooming, and reconstruction. Part I: Analytics. *SIAM Journal on Imaging Sciences*, 8(4):2814–2850, 2015.

[BH15b] K. Bredies and M. Holler. A TGV-based framework for variational image decompression, zooming, and reconstruction. Part II: Numerics. *SIAM Journal on Imaging Sciences*, 8(4):2851–2886, 2015.

[BHS09] M. Burger, L. He, and C.-B. Schönlieb. Cahn-Hilliard inpainting and a generalization for grayvalue images. *SIAM Journal on Imaging Sciences*, 2(4):1129–1167, 2009.

[BHW13] M. Boshtayeva, D. Hafner, and J. Weickert. Focus fusion with anisotropic depth map smoothing. In A. Bors, E. Hancock, W. Smith, and R. Wilson, editors, *Computer Analysis of Images and Patterns*, volume 8048 of *Lecture Notes in Computer Science*, pages 67–74. Springer, Berlin, 2013.

[BHW15] M. Boshtayeva, D. Hafner, and J. Weickert. A focus fusion framework with anisotropic depth map smoothing. *Pattern Recognition*, 48(11):3310–3323, November 2015.

[Bio15] Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne. Imaging web demonstrations: Extended depth of field. `http://bigwww.epfl.ch/demo/edf/` (last accessed August 2016), 2015.

[BK93] P. J. Burt and R. J. Kolczynski. Enhanced image capture through fusion. In *Proc. Fourth International Conference on Computer Vision*, pages 173–182, Berlin, Germany, May 1993.

[BKP10] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, September 2010.

[BKV13] K. Bredies, K. Kunisch, and T. Valkonen. Properties of L^1-TGV^2: The one-dimensional case. *Journal of Mathematical Analysis and Applications*, 398(1):438–454, February 2013.

[BL13] M. Bertalmío and S. Levine. Variational approach for the fusion of exposure bracketed pairs. *IEEE Transactions on Image Processing*, 22(2):712–723, February 2013.

[BLC10] C. Brito-Loeaza and K. Chen. On high-order denoising models and fast algorithms for vector-valued images. *IEEE Transactions on Image Processing*, 19(6):1518–1527, June 2010.

[Ble16] Blender Foundation. Blender. `https://www.blender.org/` (last accessed August 2016), 2016.

[BLSW14] R. Bergmann, F. Laus, G. Steidl, and A. Weinmann. Second order differences of cyclic data and applications in variational denoising. *SIAM Journal on Imaging Sciences*, 7(4):2916–2953, December 2014.

[BM11] 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):500–5513, March 2011.

[BML15] T. Bengtsson, T. McKelvey, and K. Lindström. Optical flow estimation on image sequences with differently exposed frames. *Optical Engineering*, 54(9):093103:1–093103:17, September 2015.

[BN98] D. N. Bhat and S. K. Nayar. Ordinal measures for image correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):415–423, April 1998.

[BO10] O. Barth and M. Osthof. Project: Focal stack photography. `http://resources.mpi-inf.mpg.de/departments/d4/teaching/ss2010/ComputationalPhotography/projects/focalstack/html/results.html` (last accessed August 2016), 2010.

[Bog00] L. Bogoni. Extending dynamic range of monochrome and color images through fusion. In *Proc. 15th International Conference on Pattern Recognition*, volume 3, pages 7–12, Barcelona, Spain, September 2000.

[Bou86] T. E. Boult. Smoothness assumptions in human and machine vision, and their implications for optimal surface interpolation. Technical Report CUCS-244-86, Department of Computer Science, Columbia University, New York, NY, 1986.

[BP10] M. Bergounioux and L. Piffet. A second-order model for image denoising. *Set-Valued and Variational Analysis*, 18(3):277–306, December 2010.

[BPPS16a] M. Burger, K. Papafitsoros, E. Papoutsellis, and C.-B. Schönlieb. Infimal convolution regularisation functionals of BV and L^p spaces. Part I: The finite p case. *Journal of Mathematical Imaging and Vision*, 55(3):343–369, July 2016.

[BPPS16b] M. Burger, K. Papafitsoros, E. Papoutsellis, and C.-B. Schönlieb. Infimal convolution regularisation functionals of BV and L^p spaces. The case $p = \infty$. In L. Bociu, J.-A. Désidéri, and A. Habbal, editors, *System Modeling and Optimization*, IFIP Advances in Information and Communication Technology, pages 169–179. Springer, Berlin, 2016.

[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.

[BPW15] K. Bredies, T. Pock, and B. Wirth. A convex, lower semicontinuous approximation of Euler's elastica energy. *SIAM Journal on Mathematical Analysis*, 47(1):566–613, 2015.

[BS11] M. Brown and S. Süsstrunk. Multi-spectral SIFT for scene category recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1063–6919, Colorado Springs, CO, June 2011. `http://ivrl.epfl.ch/supplementary_material/cvpr11/` (last accessed August 2016).

[BSFG09] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, 28(3):24:1–24:10, August 2009.

[BSL+11] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, March 2011. `http://vision.middlebury.edu/flow/` (last accessed August 2016).

[Buc80] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1–26, July 1980.

[BUC15] F. Bouzaraa, O. Urfalioglu, and G. Cordara. Dual-exposure image registration for HDR processing. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 1553–1557, South Brisbane, Australia, April 2015.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, 2004.

[BV11] K. Bredies and T. Valkonen. Inverse problems with second-order total generalized variation constraints. In *Proc. 9th International Conference on Sampling Theory and Applications*, pages 1–4, Singapore, May 2011.

[BW05] A. Bruhn and J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proc. Tenth IEEE International Conference on Computer Vision*, volume 1, pages 749–755, Beijing, China, October 2005.

[BZCC10] P. Bhat, C. L. Zitnick, M. Cohen, and B. Curless. Gradientshop: A gradient-domain optimization framework for image and video filtering. *ACM Transactions on Graphics*, 29(2):10:1–10:14, March 2010.

[BZDB13] J. Braux-Zin, R. Dupont, and A. Bartoli. A general dense image matching framework combining direct and feature-based costs. In *Proc. IEEE International Conference on Computer Vision*, pages 185–192, Sydney, Australia, December 2013.

[Čad08] M. Čadík. Perceptual evaluation of color-to-grayscale image conversions. *Computer Graphics Forum*, 27(7):1745–1754, October 2008. `http://cadik.posvete.cz/color_to_gray_evaluation/` (last accessed August 2016).

[Can03] F. M. Candocia. Simultaneous homographic and comparametric alignment of multiple exposure-adjusted pictures of the same scene. *IEEE Transactions on Image Processing*, 12(12):1485–1494, December 2003.

[CBAB94] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proc. IEEE International Conference on Image Processing*, volume 2, pages 168–172, Austin, TX, November 1994.

[CBF11] W.-C. Chiu, U. Blanke, and M. Fritz. Improving the kinect by cross-modal stereo. In *Proc. British Machine Vision Conference*, pages 116.1–116.10, Dundee, Scotland, August 2011. BMVA Press.

[CEP07] T. F. Chan, S. Esedoglu, and F. E. Park. Image decomposition combining staircase reduction and texture extraction. *Journal of Visual Communication and Image Representation*, 18(6):464–486, December 2007.

[CEP10] T. F. Chan, S. Esedoglu, and F. Park. A fourth order dual method for staircase reduction in texture extraction and image restoration problems. In *Proc. IEEE International Conference on Image Processing*, pages 4137–4140, Hong Kong, China, September 2010.

[CGKC12] C.-H. Chan, B. Goswami, J. Kittler, and W. Christmas. Local ordinal contrast pattern histograms for spatiotemporal, lip-based speaker authentication. *IEEE Transactions on Information Forensics and Security*, 7(2):602–612, April 2012.

[CH04] W.-H. Cho and K.-S. Hong. Extending dynamic range of two color images under different exposures. In *Proc. 17th International Conference on Pattern Recognition*, volume 4, pages 853–856, Cambridge, UK, August 2004.

[CH06] L. Cerman and V. Hlaváč. Exposure time estimation for high dynamic range imaging with hand held camera. In *Proc. Computer Vision Winter Workshop 2006*, pages 76–81, Telč, Czech Republic, February 2006.

[Cha04] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, January 2004.

[CK83] N. Cornelius and T. Kanade. Adapting optical-flow to measure object motion in reflectance and X-ray image sequences. Technical Report 2502, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, 1983.

[CKZP13] J. Chen, V.-P. Kellokumpu, G. Zhao, and M. Pietikäinen. RLBP: Robust local binary pattern. In T. Burghardt, D. Damen, W. Mayol-Cuevas, and M. Mirmehdi, editors, *Proc. British Machine Vision Conference*, pages 122.1–122.10, Bristol, UK, September 2013. BMVA Press.

[CL97] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, April 1997.

[CLO+12] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua. BRIEF: Computing a local binary descriptor very fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1281–1298, July 2012.

[CMM00] T. F. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM Journal on Scientific Computing*, 22(2):503–516, 2000.

[CMP02] T. Corpetti, E. Memin, and P. Perez. Dense estimation of fluid flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):365–380, March 2002.

[Coh93] I. Cohen. Nonlinear variational method for optical flow computation. In *Proc. Eighth Scandinavian Conference on Image Analysis*, volume 1, pages 523–530, Tromsø, Norway, May 1993.

[Col99] Columbia Vision Laboratory (CAVE). RASCAL software. `http://www.cs.columbia.edu/CAVE/software/rascal/rrslrr.php` (last accessed August 2016), 1999.

[Cor70] T. N. Cornsweet. *Visual Perception*. Harcourt College Publishers, Fort Worth, June 1970.

[Cos12] A. Cosentino. Multispectral imaging documentation of "Madonna and Chield", Bergen Museum of Art. `http://www.gigapan.com/galleries/10008/` (last accessed August 2016), 2012.

[CP11] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, May 2011.

[CPOL12] W. Choe, S. Park, H. Oh, and S. Lee. Active motion high dynamic range imaging for digital still camera. In *Proc. 1st IEEE Global Conference on Consumer Electronics 2012*, pages 14–16, Tokyo, Japan, October 2012.

[CR96] D. Calvetti and L. Reichel. Adaptive Richardson iteration based on Leja points. *Journal of Computational and Applied Mathematics*, 71(2):267–286, July 1996.

[CR99] S. Chaudhuri and A. N. Rajagopalan. *Depth From Defocus: A Real Aperture Imaging Approach*. Springer, Berlin, April 1999.

[ČWNA08] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi. Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics*, 32(3):330–349, June 2008. `http://cadik.posvete.cz/tmo/` (last accessed August 2016).

[CWS15] G. M. Cárdenas, J. Weickert, and S. Schäffer. A linear scale-space theory for continuous nonlocal evolutions. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale-Space and Variational Methods in Computer Vision*, volume 9087 of *Lecture Notes in Computer Science*, pages 103–114. Springer, Berlin, 2015.

[Dav99] A. Davenport. *The History of Photography: An Overview*. University of New Mexico Press, Albuquerque, July 1999.

[DC06] I. De and B. Chanda. A simple and efficient algorithm for multifocus image fusion using morphological wavelets. *Signal Processing*, 86(5):924–936, May 2006.

[DD02] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 21(3):257–266, July 2002.

[Deu13] Deutsche Arbeitsgemeinschaft für Mustererkennung e. V. GCPR 2013 – Special Session on Robust Optical Flow. `http://www.dagm.de/symposien/special-sessions/` (last accessed August 2016), 2013.

[DFLM09] G. Dal Maso, I. Fonseca, G. Leoni, and M. Morini. A higher order model for image restoration: The one-dimensional case. *SIAM Journal on Mathematical Analysis*, 40(6):2351–2391, February 2009.

[DHW13] O. Demetz, D. Hafner, and J. Weickert. The complete rank transform: A tool for accurate and morphologically invariant matching of structures. In T. Burghardt, D. Damen, W. Mayol-Cuevas, and M. Mirmehdi, editors, *Proc. British Machine Vision Conference*, pages 50.1–50.12, Bristol, UK, September 2013. BMVA Press.

[DHW15] O. Demetz, D. Hafner, and J. Weickert. Morphologically invariant matching of structures with the complete rank transform. *International Journal of Computer Vision*, 113(3):220–232, July 2015.

[Di 86] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics and Image Processing*, 33:116–125, January 1986.

[DLBU11] Z. Dogan, S. Lefkimmiatis, A. Bourquard, and Michael Unser. A second-order extension of TV regularization for image deblurring. In *Proc. 18th IEEE International Conference on Image Processing*, pages 705–708, Brussels, Belgium, September 2011.

[DM97] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. SIGGRAPH '97*, pages 369–378, Los Angeles, CA, August 1997.

[DMM12] D. Dederscheck, T. Müller, and R. Mester. Illumination invariance for driving scene optical flow using comparagram preselection. In *Proc. IEEE Intelligent Vehicles Symposium*, pages 742–747, Alcala de Henares, Spain, June 2012.

[DN13] M. Drulea and S. Nedevschi. Motion estimation using the correlation transform. *IEEE Transactions on Image Processing*, 22(8):3260–3270, August 2013.

[DSV+14] O. Demetz, M. Stoll, S. Volz, J. Weickert, and A. Bruhn. Learning brightness transfer functions for the joint recovery of illumination changes and optical flow. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8689 of *Lecture Notes in Computer Science*, pages 455–471. Springer, Berlin, 2014.

[DSV17] J. C. De Los Reyes, C.-B. Schönlieb, and T. Valkonen. Bilevel parameter learning for higher-order total variation regularisation models. *Journal of Mathematical Imaging and Vision*, 57(1):1–25, January 2017.

[DT05] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, San Diego, CA, June 2005.

[Duc76] J. Duchon. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In W. Schempp and K. Zeller, editors, *Constructive Theory of Functions of Several Variables*, volume 571 of *Lecture Notes in Mathematics*, pages 85–100. Springer, Berlin, 1976.

[DW88] T. Darrell and K. Wohn. Pyramid based depth from focus. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 504–509, Ann Arbor, MI, June 1988.

[DWB05] S. Didas, J. Weickert, and B. Burgeth. Stability and local feature enhancement of higher order nonlinear diffusion filtering. In W. G. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 451–458. Springer, Berlin, 2005.

[DWB09] S. Didas, J. Weickert, and B. Burgeth. Properties of higher order nonlinear diffusion filtering. *Journal of Mathematical Imaging and Vision*, 35(3):208–226, March 2009.

[EK03] H. A. Eltoukhy and S. Kavusi. Computationally efficient algorithm for multifocus image reconstruction. In M. M. Blouke, N. Sampat, and R. J. Motta, editors, *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications IV*, volume 5017 of *Proceedings of SPIE*, pages 332–342. SPIE Press, Bellingham, 2003.

[EKB14] D. Eynard, A. Kovnatsky, and M. M. Bronstein. Laplacian colormaps: a framework for structure-preserving color transformations. *Computer Graphics Forum*, 33(2):215–224, May 2014.

[EL93] J. Ens and P. Lawrence. An investigation of methods for determining depth from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2):97–108, February 1993.

[EUS06] A. Eden, M. Uyttendaele, and R. Szeliski. Seamless image stitching of scenes with large motions and exposure differences. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2498–2505, New York, NY, June 2006.

[Fai07] M. D. Fairchild. The HDR photographic survey. In *Proc. Fifteenth Color Imaging Conference*, volume 6, pages 233–238, Albuquerque, NM, November 2007.

[Fai13] M. D. Fairchild. *Color Appearance Models*. Wiley-IS&T, Chichester, 3rd edition, 2013.

[FAR07] R. Fattal, M. Agrawala, and S. Rusinkiewicz. Multiscale shape and detail enhancement from multi-light image collections. *ACM Transactions on Graphics*, 26(3):51:1–51:9, July 2007.

[FBPC11] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles. An analysis of visual adaptation and contrast perception for tone mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):2002–2012, October 2011.

[FBPC12] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles. Generation of HDR images in non-static conditions based on gradient fusion. In *Proc. International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 31–37, Rome, Italy, February 2012.

[FDW12] M. M. Fouad, R. M. Dansereau, and A. D. Whitehead. Geometric image registration under arbitrarily-shaped locally variant illuminations. *Signal, Image and Video Processing*, 6(4):521–532, November 2012.

[FE04] B. Fröba and A. Ernst. Face detection with the modified census transform. In *Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 91–96, Seoul, South Korea, May 2004.

[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.

[FG06] C. Frese and I. Gheţa. Robust depth estimation by fusion of stereo and focus series acquired with a camera array. In *Proc. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 243–248, Heidelberg, Germany, September 2006.

[FLW02] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256, July 2002.

[FM03] B. Fischer and J. Modersitzki. Curvature based image registration. *Journal of Mathematical Imaging and Vision*, 18(1):81–85, January 2003.

[FRR+13] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof. Image guided depth upsampling using anisotropic total generalized variation. In *Proc. IEEE International Conference on Computer Vision*, pages 993–1000, Sydney, Australia, December 2013.

[FS08] C. Fredembach and S. Süsstrunk. Colouring the near-infrared. In *Proc. Sixteenth Color Imaging Conference*, pages 176–182, Portland, OR, November 2008.

[FVB+04] B. Forster, D. Van De Ville, J. Berent, D. Sage, and M. Unser. Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images. *Microscopy Research and Technique*, 65(1–2):33–42, September 2004.

[GAW+10] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. A. Lensch. Optimal HDR reconstruction with linear digital cameras. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 215–222, San Francisco, CA, June 2010.

[GB97] P. Golland and A. M. Bruckstein. Motion from color. *Computer Vision and Image Understanding*, 68(3):346–362, December 1997.

[GB04a] J. B. Greer and A. L. Bertozzi. Traveling wave solutions of fourth order PDEs for image processing. *SIAM Journal on Mathematical Analysis*, 36(1):38–68, 2004.

[GB04b] J. B. Greer and A. L. Betrozzi. H^1 solutions of a class of fourth order nonlinear equations for image processing. *Discrete and Continuous Dynamical Systems*, 10(1&2):349–366, January 2004.

[GD07] M. Grundland and N. A. Dodgson. Decolorize: Fast, contrast enhancing, color to grayscale conversion. *Pattern Recognition*, 40(11):2891–2896, June 2007.

[GF00] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Dover, New York, 2000.

[GFJ15] E. Ghadimi, H. R. Feyzmahdavian, and M. Johansson. Global convergence of the heavy-ball method for convex optimization. In *Proc. European Control Conference*, volume 1, pages 310–315, Linz, Austria, July 2015.

[GG07] M. Gevrekci and B. K. Gunturk. On geometric and photometric registration of images. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 1261–1264, Honolulu, HI, April 2007.

[GGC+09] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. In *Proc. IEEE International Conference on Computational Photography*, pages 11499067:1–11499067:7, San Francisco, CA, April 2009.

[GIN13] M. Gupta, D. Iso, and S. K. Nayar. Fibonacci exposure bracketing for HDR imaging. In *Proc. IEEE International Conference on Computer Vision*, pages 1473–1480, Sydney, Australia, December 2013.

[GK86] G. H. Golub and R. Kannan. Convergence of a two-stage Richardson process for nonlinear equations. *BIT Numerical Mathematics*, 26(2):209–216, June 1986.

[GKTT13] M. Granados, K. I. Kim, J. Tompkin, and C. Theobalt. Automatic noise modeling for ghost-free HDR reconstruction. *ACM Transactions on Graphics*, 32(6):201:1–201:10, November 2013.

[GLU12] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, Providence, RI, June 2012. `http://www.cvlibs.net/datasets/kitti/` (last accessed August 2016).

[GN87] M. A. Gennert and S. Negahdaripour. Relaxing the brightness constancy assumption in computing optical flow. Technical Report 975, Artificial Intelligence Laboratory, Massachusetts Instituite of Technology, Cambridge, MA, June 1987.

[GN03] M. D. Grossberg and S. K. Nayar. Determining the camera response from images: What is knowable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1455–1467, November 2003.

[GN04] M. D. Grossberg and S. K. Nayar. Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282, October 2004.

[GNV86] M. H. Gutknecht, W. Niethammer, and R. S. Varga. k-step iterative methods for solving nonlinear systems of equations. *Numerische Mathematik*, 48(6):699–712, November 1986.

[GO08] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling and Simulation*, 7(3):1005–1028, November 2008.

[Gol00] N. T. Goldsmith. Deep focus; a digital image processing technique to produce improved focal depth in light microscopy. *Image Analysis & Stereology*, 19(3):163–167, 2000.

[Gos05] A. A. Goshtasby. Fusion of multi-exposure images. *Image and Vision Computing*, 23(6):611–618, June 2005.

[Gos06] A. A. Goshtasby. Fusion of multifocus images to maximize image information. In K. L. Priddy and E. Ertin, editors, *Intelligent Computing: Theory and Applications IV*, volume 6229 of *Proceedings of SPIE*, pages 62290L:1–62290L:10. SPIE Press, Bellingham, 2006.

[GOTG05] A. A. Gooch, S. C. Olsen, J. Tumblin, and B. Gooch. Color2Gray: Salience-preserving color removal. *ACM Transactions on Graphics*, 24(3):634–639, July 2005.

[GP96] S. N. Gupta and J. L. Prince. Stochastic models for DIV-CURL optical flow methods. *IEEE Signal Processing Letters*, 3(2):32–34, February 1996.

[GR92] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):367–383, March 1992.

[Gre13] S. Grewenig. *Fast Explicit Methods for PDE-Based Image Analysis*. PhD thesis, Department of Mathematics, Saarland University, Saarbrücken, Germany, 2013.

[Gri81] W. E. L. Grimson. *From Images to Surfaces: A Computational Study of the Human Early Visual System*. MIT Press, Cambridge, MA, 1981.

[Gri82] W. E. L. Grimson. A computational theory of visual surface interpolation. *Philosophical Transactions of the Royal Society of London, Series B*, 298(1092):395–427, September 1982.

[Gri83] W. E. L. Grimson. An implementation of a computational theory of visual surface interpolation. *Computer Vision, Graphics, and Image Processing*, 22(1):39–69, April 1983.

[Gro87] P. Grossmann. Depth from focus. *Pattern Recognition Letters*, 5(1):63–69, January 1987.

[Gro06] T. Grosch. Fast and robust high dynamic range image generation with camera and object movement. In L. Kobbelt, T. Kuhlen, T. Aach, and R. Westermann, editors, *Vision, Modeling, and Visualization*, pages 277–284. AKA, Berlin, 2006.

[GS78] W. Gentzsch and A. Schlüter. Über ein Einschrittverfahren mit zyklischer Schrittweitenänderung zur Lösung parabolischer Differentialgleichungen. *Zeitschrift für Angewandte Mathematik und Mechanik*, 58:T415–T416, 1978. In German.

[GSL08] M. Granados, H.-P. Seidel, and H. P. A. Lensch. Background estimation from non-time sequence images. In *Proc. Graphics Interface*, pages 33–40, Windsor, Canada, May 2008.

[Gut89] M. H. Gutknecht. Stationary and almost stationary iterative (k, l)-step methods for linear and nonlinear systems of equations. *Numerische Mathematik*, 56(2):179–213, February 1989.

[Gut15] M. H. Gutknecht. Revisiting (k, ℓ)-step methods. *Numerical Algorithms*, 69(2):455–469, June 2015.

[GV61] G. H. Golub and R. S. Varga. Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods, Part I. *Numerische Mathematik*, 3(1):147–156, December 1961.

[GV96] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 3rd edition, October 1996.

[GVPB15] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío. Enhanced variational image dehazing. *SIAM Journal on Imaging Sciences*, 8(3):1519–1546, July 2015.

[GW07] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, Upper Saddle River, 3rd edition, June 2007.

[GWW+08] I. Galić, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H.-P. Seidel. Image compression with anisotropic diffusion. *Journal of Mathematical Imaging and Vision*, 31(2):255–269, July 2008.

[GYL85] F. C. A. Groen, I. T. Young, and G. Ligthart. A comparison of different focus functions for use in autofocus algorithms. *Cytometry*, 6(2):81–91, March 1985.

[Han05] J. Hannavy. *Encyclopedia of Nineteenth-Century Photography*. Taylor & Francis, New York, 2005.

[Har86] J. G. Harris. The coupled depth/slope approach to surface reconstruction. Technical Report AITR-908, MIT Artificial Intelligence Laboratory, Cambridge, MA, June 1986.

[Har87] J. G. Harris. A new approach to surface reconstruction: The coupled depth/slope model. In *Proc. 1st International Conference on Computer Vision*, pages 277–283, London, UK, June 1987.

[Has08] S. W. Hasinoff. *Multiple-Aperture Photography*. PhD thesis, University of Toronto, Department of Computer Science, Toronto, Canada, 2008.

[HB96] G. D. Hager and P. N. Belhumeur. Real-time tracking of image regions with changes in geometry and illumination. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 403–410, San Francisco, CA, June 1996.

[HCF02] G. Hermosillo, C. Chefd'Hotel, and O. Faugeras. Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50(3):329–343, December 2002.

[HDW13] D. Hafner, O. Demetz, and J. Weickert. Why is the census transform good for robust optic flow computation? In A. Kuijper, T. Pock, K. Bredies, and H. Bischof, editors, *Scale-Space and Variational Methods in Computer Vision*, volume 7893 of *Lecture Notes in Computer Science*, pages 210–221. Springer, Berlin, 2013.

[HDW14] D. Hafner, O. Demetz, and J. Weickert. Simultaneous HDR and optic flow computation. In *Proc. 22nd International Conference on Pattern Recognition*, pages 2065–2070, Stockholm, Sweden, August 2014.

[HDWR15] D. Hafner, O. Demetz, J. Weickert, and M. Reißel. Mathematical foundations and generalisations of the census transform for robust optic flow computation. *Journal of Mathematical Imaging and Vision*, 52(1):71–86, May 2015.

[Hel77] G. Hellwig. *Partial Differential Equations*. Teubner, Stuttgart, 1977.

[HF01] H. W. Haussecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):661–673, June 2001.

[HG11] I. Hossain and B. K. Gunturk. High dynamic range imaging of non-static scenes. In F. H. Imai and F. Xiao, editors, *Digital Photography VII*, volume 7876 of *Proceedings of SPIE*, pages 78760P:1–78760P:8. SPIE Press, Bellingham, 2011.

[HGP12] J. Hu, O. Gallo, and K. Pulli. Exposure stacks of live scene with hand-held cameras. In K. M. Lee, Y. Matsushita, J. M. Rehg, and Z. Hu, editors, *Computer Vision – ECCV 2012*, volume 3024 of *Lecture Notes in Computer Science*, pages 499–512. Springer, Berlin, 2012.

[HGPS13] J. Hu, O. Gallo, K. Pulli, and X. Sun. HDR deghosting: How to deal with saturation? In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1063–1170, Portland, OR, June 2013.

[HJ07] W. Huang and Z. Jing. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, 28(4):493–500, March 2007.

[HK07] S. W. Hasinoff and K. N. Kutulakos. A layer-based restoration framework for variable-aperture photography. In *Proc. IEEE 11th International Conference on Computer Vision*, pages 9848948:1–9848948:8, Rio de Janeiro, Brazil, October 2007.

[HK14] M. Holler and K. Kunisch. On infimal convolution of TV-type functionals and applications to video and image reconstruction. *SIAM Journal on Imaging Sciences*, 7(4):2258–2300, 2014.

[HLL+10] Y. S. Heo, K. M. Lee, S. U. Lee, Y. Moon, and J. Cha. Ghost-free high dynamic range imaging. In R. Kimmel, R. Klette, and A. Sugimoto, editors, *Computer Vision – ACCV 2010*, volume 6493 of *Lecture Notes in Computer Science*, pages 486–500. Springer, Berlin, 2010.

[HM07] E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. *Methods of Information in Medicine*, 46(3):292–299, 2007.

[Hor68] B. K. P. Horn. Focusing. Technical Report Memo No. 160, MIT Artificial Intelligence Laboratory, Cambridge, MA, May 1968.

[Hor90] B. K. P. Horn. Height and gradient from shading. *International Journal of Computer Vision*, 5(1):37–75, August 1990.

[HOW+16] D. Hafner, P. Ochs, J. Weickert, M. Reißel, and S. Grewenig. FSI schemes: Fast semi-iterative solvers for image processing and computer vision. In B. Andres and B. Rosenhahn, editors, *Pattern Recognition*, volume 9796 of *Lecture Notes in Computer Science*, pages 91–102. Springer, Berlin, 2016.

[HP10] J. Herwig and J. Pauli. An information-theoretic approach to multi-exposure fusion via statistical filtering using local entropy. In *Proc. Signal Processing, Pattern Recognition and Applications*, pages 50–57, Innsbruck, Austria, February 2010.

[HS52] M. R. Hestenes and E. Stiefel. Methods of conjugate gradient for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):406–436, December 1952.

[HS81] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1–3):185–203, August 1981.

[HS06] W. Hinterberger and O. Scherzer. Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing*, 76(1):109–133, January 2006.

[HS07] H. Hirschmüller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Proc. IEEE International Conference on Computer Vision Workshops*, pages 9738137:1–9738137:8, Minneapolis, MN, June 2007.

[HSGL11] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski. Non-rigid dense correspondence with applications for image enhancement. *ACM Transactions on Graphics*, 30(4):70:1–70:10, July 2011.

[HSW15] D. Hafner, C. Schroers, and J. Weickert. Introducing maximal anisotropy into second order coupling models. In J. Gall, P. Gehler, and B. Leibe, editors, *Pattern Recognition*, volume 9358 of *Lecture Notes in Computer Science*, pages 79–90. Springer, Berlin, 2015.

[Hub81] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[HW15] D. Hafner and J. Weickert. Variational exposure fusion with optimal local contrast. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale-Space and Variational Methods in Computer Vision*, volume 9087 of *Lecture Notes in Computer Science*, pages 425–436. Springer, Berlin, 2015.

[HW16] D. Hafner and J. Weickert. Variational image fusion with optimal local contrast. *Computer Graphics Forum*, 35(1):100–112, February 2016.

[HWS+13] A. Hewer, J. Weickert, T. Scheffer, H. Seibert, and S. Diebels. Lagrangian strain tensor computation with higher order variational models. In T. Burghardt, D. Damen, W. Mayol-Cuevas, and M. Mirmehdi, editors, *Proc. British Machine Vision Conference*, pages 129.1–129.10, Bristol, UK, September 2013. BMVA Press.

[HY04] L. A. Hagemann and D. M. Young. *Applied Iterative Methods*. Dover, New York, July 2004.

[Iij62] T. Iijima. Basic theory on normalization of pattern (in case of typical one-dimensional pattern). *Bulletin of the Electrotechnical Laboratory*, 26:368–388, January 1962. in Japanese.

[ILLP11] J. Im, S. Lang, S. Lee, and J. Paik. Geometrical transformation-based ghost artifacts removing for high dynamic range image. In *Proc. 18th IEEE International Conference on Image Processing*, pages 357–360, Brussels, Belgium, September 2011.

[Jar67] R. A. Jarvis. Focus optimisation criteria for computer image processing. *Microscope*, 24(2):163–180, 1967.

[Jar83] R. A. Jarvis. A perspective on range finding techniques for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):122–139, March 1983.

[JBL12] D. E. Jacobs, J. Baek, and M. Levoy. Focal stack compositing for depth of field control. Technical Report 2012-1, Stanford Computer Graphics Laboratory, Stanford, CA, October 2012.

[JFS02] H. Jin, P. Favaro, and S. Soatto. A variational approach to shape from defocus. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002, Part II*, volume 2351 of *Lecture Notes in Computer Science*, pages 18–30. Springer, Berlin, 2002.

[JKC10] T. Jabid, M. H. Kabir, and O. Chae. Local directional pattern (LDP) for face recognition. In *Proc. Digest of Technical Papers International Conference on Consumer Electronics*, pages 329–330, Las Vegas, NV, January 2010.

[JLW08] K. Jacobs, C. Loscos, and G. Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications*, 28(2):84–93, March 2008.

[JO12] T. Jinno and M. Okuda. Multiple exposure fusion for high dynamic range image acquisition. *IEEE Transactions on Image Processing*, 21(1):358–365, January 2012.

[Jof07] J. Joffre. Exposure series. `http://www.hdrsoft.com/examples2.html` (last accessed August 2016), 2003–2007.

[KAR06] E. A. Khan, A. O. Akyüz, and E. Reinhard. Ghost removal in high dynamic range images. In *Proc. IEEE International Conference on Image Processing*, pages 2005–2008, Atlanta, GA, October 2006.

[KBPS11] F. Knoll, K. Bredies, T. Pock, and R. Stollberger. Second order total generalized variation (TGV) for MRI. *Magnetic Resonance im Medicine*, 65(2):480–491, February 2011.

[KC11] K. Kotwal and S. Chaudhuri. An optimization-based approach to fusion of multi-exposure, low dynamic range images. In *Proc. 14th International Conference on Information Fusion*, pages 1942–1948, Chicago, IL, July 2011.

[KHC+06] W.-C. Kao, C.-C. Hsu, L.-Y. Chen, C.-C. Kao, and S.-H. Chen. Integrating image fusion and motion stabilization for capturing still images in high dynamic range scenes. In *Proc. IEEE Tenth International Symposium on Consumer Electronics*, pages 86–91, St. Petersburg, Russia, June 2006.

[KK06] Y.-H. Kim and A. C. Kak. Error analysis of robust optical flow estimation by least median of squares methods for the varying illumination model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1418–1435, September 2006.

[KLB85] J. R. Kender, D. Lee, and T. E. Boult. Information-based complexity applied to optimal recovery of 2&1/2-D sketch. In *Proc. IEEE Third Workshop on Computer Vision: Representation and Control*, pages 157–167, Bellaire, MI, October 1985.

[KLL13] T. H. Kim, H. S. Lee, and K. M. Lee. Optical flow via locally adaptive fusion of complementary data costs. In *Proc. IEEE International Conference on Computer Vision*, pages 3344–3351, Sydney, Australia, December 2013.

[KP04] S. J. Kim and M. Pollefeys. Radiometric alignment of image sequences. In *Proc. IEEE Conference Society on Computer Vision and Pattern Recognition*, volume 1, pages 645–651, Washington, DC, June 2004.

[Kro88] E. Krotkov. Focusing. *International Journal of Computer Vision*, 1(3):223–237, October 1988.

[KSB+13] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen. Patch-based high dynamic range video. *ACM Transactions on Graphics*, 32(6):202:1–202:8, November 2013.

[KUWS03] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Transactions on Graphics*, 22(3):319–325, July 2003.

[Lan52] C. Lanczos. Solution of systems of linear equations by minimized iterations. *Journal of Research of the National Bureau of Standards*, 49(1):33–53, July 1952.

[LBL13] F. Lenzen, F. Becker, and J. Lellmann. Adaptive second-order total variation: An approach aware of slope discontinuities. In A. Kuijper, K. Bredies, T. Pock, and H. Bischof, editors, *Scale Space and Variational*

Methods in Computer Vision, volume 7893 of *Lecture Notes in Computer Science*, pages 61–73. Springer, Berlin, 2013.

[LBU12] S. Lefkimmiatis, A. Bourquard, and M. Unser. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Transactions on Image Processing*, 21(3):983–995, March 2012.

[LC09] H.-Y. Lin and W.-Z. Chang. High dynamic range imaging for stereoscopic scene representation. In *Proc. 16th IEEE International Conference on Image Processing*, pages 1522–4880, Cairo, Egypt, November 2009.

[Le 57] G. Le Gray. The Great Wave, Sète. `http://www.metmuseum.org/toah/images/hb/hb_1976.646.jpg` (last accessed August 2016), 1857.

[Lev06] M. Levoy. Light fields and computational imaging. *Computer*, 39(8):46–55, August 2006.

[LHM11] C. Lau, W. Heidrich, and R. Mantiuk. Cluster-based color space optimizations. In *Proc. International Conference on Computer Vision*, pages 1172–1179, Barcelona, Spain, November 2011.

[LJW+14] Y. Liu, J. Jin, Q. Wang, Y. Shen, and X. Dong. Region level based multi-focus image fusion using quaternion wavelet and normalized cut. *Signal Processing*, 97:9–30, April 2014.

[LK81] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. 7th International Joint Conference on Artificial Intelligence*, volume 2, pages 674–679, Vancouver, Canada, August 1981.

[LKG+03] P. A. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics*, 22(2):234–257, April 2003.

[LKTW04] S. Li, J. T. Kwok, I. W. Tsang, and Y. Wang. Fusing images with different focuses using support vector machines. *IEEE Transactions on Neural Networks*, 15(6):1555–1561, November 2004.

[LLM11] F. Luo, B. Lu, and C. Miao. Multifocus image fusion with trace-based structure tensor. In T. Yoshizawa, P. Wei, and J. Zheng, editors, *Optoelectronic Imaging and Processing Technology*, volume 8200 of *Proceedings of SPIE*, pages 82001G:1–82001G:8. SPIE Press, Bellingham, 2011.

[LLT03] M. Lysaker, A. Lundervold, and X.-C. Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590, December 2003.

[LM71] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, January 1971.

[LMM95] H. Li, B. S. Manjunath, and S. K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3):235–245, May 1995.

[LMW10] B. Lu, C. Miao, and H. Wang. Pixel level image fusion based on linear structure tensor. In *Proc. IEEE Youth Conference on Information Computing and Telecommunications*, pages 303–306, Beijing, China, November 2010.

[LON+07] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah. Pixel- and region-based image fusion with complex wavelets. *Information Fusion*, 8(2):119–130, April 2007.

[Lot02] B. Lotto. Craik-O'Brien-Cornsweet effect. `http://purveslab.net/see-for-yourself/` (last accessed August 2016), 2002.

[LOT04] M. Lysaker, S. Osher, and X.-C. Tai. Noise removal using smoothed normals and surface fitting. *IEEE Transactions on Image Processing*, 13(10):1345–1357, October 2004.

[Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[LPC11] D.-K. Lee, R.-H. Park, and S. Chang. Improved histogram based ghost removal in exposure fusion for high dynamic range images. In *Proc. IEEE 15th International Symposium on Consumer Electronics*, pages 586–591, Singapore, June 2011.

[LPSS15] J. Lellmann, K. Papafitsoros, C. Schönlieb, and D. Spector. Analysis and application of a non-local Hessian. *SIAM Journal on Imaging Sciences*, 8(4):2161–2202, October 2015.

[LRZ+10] Z. Li, S. Rahardja, Z. Zhu, S. Xie, and S. Wu. Movement detection for the synthesis of high dynamic range images. In *Proc. IEEE International Conference on Image Processing*, pages 3133–3136, Hong Kong, China, September 2010.

[LSG11] F. Lenzen, H. Schäfer, and C. Garbe. Denoising time-of-flight data with adaptive total variation. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, S. Wang, K. Kyungnam, B. Benes, K. Moreland, C. Borst, S. DiVerdi, C. Yi-Jen, and J. Ming, editors, *Advances in Visual Computing*, volume 6938 of *Lecture Notes in Computer Science*, pages 337–346. Springer, Berlin, 2011.

[LSWY10] F. Li, J. Sun, J. Wang, and J. Yu. Dual-focus stereo imaging. *Journal of Electronic Imaging*, 19(4):043009:1–043009:12, December 2010.

[LT06] M. Lysaker and X.-C. Tai. Iterative image restoration combining total variation minimization and a second-order functional. *International Journal of Computer Vision*, 66(1):5–18, January 2006.

[LTC13] R. Lai, X.-C. Tai, and T. F. Chan. A ridge and corner preserving model for surface restoration. *SIAM Journal on Scientific Computing*, 35(2):A675–A695, 2013.

[Lus13] M. Luschkova. Exposure fusion for dynamic scenes. Master's thesis, Department of Computer Science, Saarland University, Saarbrücken, Germany, August 2013.

[LWP+15] W. Lu, G. Wang, Z. Pan, J. Duan, and W. Wei. Vectorised multichannel total generalized variation for color image regularisation and its split Bregman algorithm. *Journal of Computational Information Systems*, 11(20):7397–7404, October 2015.

[LXJ12] C. Lu, L. Xu, and J. Jia. Contrast preserving decolorization. In *Proc. IEEE International Conference on Computational Photography*, pages 1–7, Seattle, WA, April 2012.

[LXJ14] C. Lu, L. Xu, and J. Jia. Contrast preserving decolorization with perception-based quality metrics. *International Journal of Computer Vision*, 110(2):222–239, November 2014.

[LY08] S. Li and B. Yang. Multifocus image fusion using region segmentation and spatial frequency. *Image and Vision Computing*, 26(7):971–979, July 2008.

[LYT11] C. Liu, J. Yuen, and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):978–994, May 2011.

[Mah13] M. T. Mahmood. Shape from focus by total variation. In *Proc. IEEE 11th IVMSP Workshop: 3D Image/Video Technologies and Applications*, pages S–31:1–S–31:4, Seoul, South Korea, June 2013.

[Man00] S. Mann. Comparametric equations with practical applications in quantigraphic image processing. *IEEE Transactions on Image Processing*, 9(8):1389–1406, August 2000.

[Max05a] Max Planck Institute for Informatics. HDR source image gallery. `http://resources.mpi-inf.mpg.de/hdr/gallery.html` (last accessed August 2016), 2005.

[Max05b] Max Planck Institute for Informatics. Recovering an HDR image from LDR exposures. `http://resources.mpi-inf.mpg.de/hdr/calibration/pfs.html#hdrrecovery` (last accessed August 2016), 2005.

[MB10] A. Malviya and S. G. Bhirud. Visual infrared video fusion for night vision using background estimation. *Journal of Computing*, 2(4):66–69, April 2010.

[MBSC15] M. Moeller, M. Benning, C. Schönlieb, and D. Cremers. Variational depth from focus reconstruction. *IEEE Transactions on Image Processing*, 24(12):5369–5378, December 2015.

[MBW07] Y. Mileva, A. Bruhn, and J. Weickert. Illumination-robust variational optical flow with photometric invariants. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 152–162. Springer, Berlin, 2007.

[MC10] M. T. Mahmood and T.-S. Choi. 3D shape recovery from image focus using kernel regression in eigenspace. *Image and Vision Computing*, 28(4):634–643, April 2010.

[MC12a] M. T. Mahmood and T.-S. Choi. Nonlinear approach for enhancement of image focus volume in shape from focus. *IEEE Transactions on Image Processing*, 21(5):2866–2873, May 2012.

[MC12b] M. Muhammad and T.-S. Choi. Sampling for shape from focus in optical microscopy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):564–573, March 2012.

[McC89] J. J. McCann. The role of simple nonlinear operations in modeling human lightness and color sensations. In B. E. Rogowitz, editor, *Human Vision, Visual Processing, and Digital Display*, volume 1077 of *Proceedings of SPIE*, pages 355–363. SPIE Press, Bellingham, 1989.

[MCF10] J. Molnár, D. Chetverikov, and S. Fazekas. Illumination-robust variational optical flow using cross-correlation. *Computer Vision and Image Understanding*, 114(10):1104–1114, October 2010.

[MG07] N. Menzel and M. Guthe. Freehand HDR photography with motion compensation. In H.-P. Seidel, H. P. A. Lensch, B. Rosenhahn, P. Slusallek, and J. Weickert, editors, *Vision, Modelling, and Visualization*, pages 127–134. AKA, Berlin, 2007.

[MG10] S. Mangiat and J. Gibson. High dynamic range video with ghost removal. In A. G. Tescher, editor, *Applications of Digital Image Processing XXXIII*, volume 7798 of *Proceedings of SPIE*, pages 779812:1–779812:8. SPIE Press, Bellingham, 2010.

[MG15] M. Menze and A. Geiger. Object scene flow for autonomous vehicles. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 3061–3070, Boston, MA, June 2015. `http://www.cvlibs.net/datasets/kitti/` (last accessed August 2016).

[MH03] J. C. Mason and D. C. Handscomb. *Chebyshev Polynomials*. CRC Press, Boca Raton, 2003.

[Mic16] Microsoft. Microsoft Kinect. `https://developer.microsoft.com/en-us/windows/kinect` (last accessed October 2016), 2016.

[Mit10] H. B. Mitchell. *Image Fusion: Theories, Techniques and Applications*. Springer, Berlin, 2010.

[MJBB16] D. Maurer, Y. C. Ju, M. Breuß, and A. Bruhn. Combining shape from shading and stereo: A variational approach for the joint estimation of depth, illumination and albedo. In R. Wilson, E. Hancock, and W. Smith, editors, *Proc. British Machine Vision Conference*, pages 76.1–76.14, York, UK, September 2016. BMVA Press.

[MKMS07] R. Mantiuk, G. Krawczyk, R. Mantiuk, and H.-P. Seidel. High-dynamic range imaging pipeline: Perception-motivated representation of visual content. In B. E. Rogowitz, T. N. Pappas, and S. J. Daly, editors, *Human Vision and Electronic Imaging XII*, volume 6492 of *Proceedings of SPIE*, pages 649212:1–649212:12. SPIE Press, Bellingham, 2007. `http://pfstools.sourceforge.net/` (last accessed August 2016).

[MKV09] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum*, 28(1):161–171, March 2009. `research.edm.uhasselt.be/~tmertens/` (last accessed February 2016).

[MM98] S. Masnou and J.-M. Morel. Level lines based disocclusion. In *Proc. International Conference on Image Processing*, pages 259–263, Chicago, IL, October 1998.

[MM12] M. A. Mohamed and B. Mertsching. TV-L1 optical flow estimation with image details recovering based on modified census transform. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, C. Fowlkes, S. Wang, M.-H. Choi, S. Mantler, J. Schulze, D. Acevedo, K. Mueller, and M. Papka, editors, *Advances in Visual Computing*, volume 7431 of *Lecture Notes in Computer Science*, pages 482–491. Springer, Berlin, 2012.

[MM16] C. Molgaard and I. A. McAllister. Mobile camera system. US Patent 20160007008, January 2016.

[MMC11] M. T. Mahmood, A. Majid, and T.-S. Choi. Optimal depth estimation by combining focus measures using genetic programming. *Information Science*, 181(7):1249–1263, April 2011.

[MMF03] S. Mann, C. Manders, and J. Fung. The lightspace change constraint equation (LCCE) with practical application to estimation of the projectivity+gain transformation between multiple pictures of the same subject matter. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 481–484, Hong Kong, China, April 2003.

[MMS06] R. Mantiuk, K. Myszkowski, and H.-P. Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception*, 3(3):286–308, July 2006.

[MN99] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 374–380, Fort Collins, CO, June 1999.

[MP95] S. Mann and R. W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. IS&T's 48th Annual Conference*, volume 48, pages 442–448, Washington, DC, May 1995.

[MPC09] T.-H. Min, R.-H. Park, and S. Chang. Histogram based ghost removal in high dynamic range images. In *Proc. IEEE International Conference on Multimedia and Expo*, pages 530–533, New York, NY, June 2009.

[MR06] A. Mittal and V. Ramesh. An intensity-augmented ordinal measure for visual correspondence. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 849–856, New York, NY, June 2006.

[MRK+13] A. Manakov, J. F. Restrepo, O. Klehm, R. Hegedüs, E. Eisemann, H.-P. Seidel, and I. Ihrke. A reconfigurable camera add-on for high dynamic range, multi-spectral, polarization, and light-field imaging. *ACM Transactions on Graphics*, 32(4):47:1–47:14, July 2013.

[MRM+14] M. A. Mohamed, H. A. Rashwan, B. Mertsching, M. A. García, and D. Puig. Illumination-robust optical flow using a local directional pattern. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(9):1499–1508, September 2014.

[MRR+11] T. Müller, C. Rabe, J. Rannacher, U. Franke, and R. Mester. Illumination robust dense optical flow using census signatures. In R. Mester and M. Felsberg, editors, *Pattern Recognition*, volume 6835 of *Lecture Notes in Computer Science*, pages 236–245. Springer, Berlin, 2011.

[MRRF11] T. Müller, J. Rannacher, C. Rabe, and U. Franke. Feature- and depth-supported modified total variation optical flow for 3D motion field estimation in real scenes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1193–1200, Providence, RI, June 2011.

[MSB17] D. Maurer, M. Stoll, and A. Bruhn. Order-adaptive regularisation for variational optical flow: Global, local and in between. In F. Lauze, Y. Dong, and A. B. Dahl, editors, *Scale Space and Variational Methods in Computer Vision*, volume 10302 of *Lecture Notes in Computer Science*, pages 550–562. Springer, Berlin, 2017.

[MSV+17] D. Maurer, M. Stoll, S. Volz, P. Gairing, and A. Bruhn. A comparison of isotropic and anisotropic second order regularisers for optical flow. In F. Lauze, Y. Dong, and A. B. Dahl, editors, *Scale Space and Variational Methods in Computer Vision*, volume 10302 of *Lecture Notes in Computer Science*, pages 537–549. Springer, Berlin, 2017.

[MSZ+11] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang. On building an accurate stereo matching system on graphics hardware. In *Proc. IEEE International Conference on Computer Vision Workshops*, pages 467–474, Barcelona, Spain, November 2011.

[MTCL12] Y.-S. Moon, Y.-M. Tai, J. H. Cha, and S.-H. Lee. A simple ghost-free exposure fusion for embedded HDR imaging. In *Proc. IEEE International Conference on Consumer Electronics*, pages 9–10, Las Vegas, NV, January 2012.

[Muk90] N. Mukawa. Estimation of shape, reflection coefficients and illuminant direction from image sequences. In *Proc. Third International Conference on Computer Vision*, pages 507–512, Osaka, Japan, December 1990.

[Mül13] J. Müller. *Advanced Image Reconstruction and Denoising: Bregmanized (Higher Order) Total Variation and Application in PET*. PhD thesis, Mathematisch-Naturwissenschaftliche Fakultät, Westfälische Wilhelms-Universität Münster, Münster, Germany, July 2013.

[Mus16] M. Mustafina. Patchmatch for large displacement optic flow estimation without warping. Master's thesis, Department of Computer Science, Saarland University, Saarbrücken, Germany, July 2016.

[MXv14] H. Mir, P. Xu, and P. van Beek. An extensive empirical evaluation of focus measures for digital photography. In N. Sampat, R. Tezaur, S. Battiato, and B. A. Fowler, editors, *Digital Photography X*, volume 9023 of *Proceedings of SPIE*, pages 90230I:1–90230I:11. SPIE Press, Bellingham, 2014.

[Nag83] H.-H. Nagel. Displacement vectors derived from second-order intensity variations in image sequences. *Computer Vision, Graphics and Image Processing*, 21(1):85–117, January 1983.

[Nag90] H.-H. Nagel. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow. In O. Faugeras, editor, *Computer Vision – ECCV '90*, volume 427 of *Lecture Notes in Computer Science*, pages 139–148. Springer, Berlin, 1990.

[Nat08] National University of Singapore, School of Computing. Enhancing photographs with near infrared images. `http://www.comp.nus.edu.sg/~photo/projects/nir.html` (last accessed April 2016), 2008.

[NBK08] T. Nir, A. M. Bruckstein, and R. Kimmel. Over-parameterized variational optical flow. *International Journal of Computer Vision*, 76(2):205–216, February 2008.

[NČN07] L. Neumann, M. Čadík, and A. Nemcsics. An efficient perception-based adaptive color to gray transformation. In *Proc. Third Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, pages 73–80, Banff, Canada, August 2007.

[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(5):565–593, September 1986.

[Neg98] S. Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9):961–979, September 1998.

[Nes04] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87 of *Applied Optimization*. Kluwer, Boston, December 2004.

[NFD97] M. Nielsen, L. Florack, and R. Deriche. Regularization, scale-space, and edge detection filters. *Journal of Mathematical Imaging and Vision*, 7(4):291–307, October 1997.

[NN94] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831, August 1994.

[NS92] H. N. Nair and C. V. Stewart. Robust focus ranging. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 309–314, Champaign, IL, June 1992.

[OBP15] P. Ochs, T. Brox, and T. Pock. iPiasco: Inertial proximal algorithm for strongly convex optimization. *Journal of Mathematical Imaging and Vision*, 53(2):171–181, October 2015.

[OLTK15] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon. Robust high dynamic range imaging by rank minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1219–1232, June 2015.

[ON94] M. Otte and H.-H. Nagel. Optical flow estimation: Advances and comparisons. In J.-O. Eklundh, editor, *Computer Vision – ECCV '94*, volume 800 of *Lecture Notes in Computer Science*, pages 49–60. Springer, Berlin, 1994.

[OPH96] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, January 1996.

[PA10] P. Puxbaum and K. Ambrosch. Gradient-based modified census transform for optical flow. In G. Bebis, R. D. Boyle, B. Parvin, D. Koracin, R. Chung, R. I. Hammoud, M. Hussain, K.-H. Tan, R. Crawfis, D. Thalmann, D. Kao, and L. Avila, editors, *Advances in Visual Computing, Part I*, volume 6453 of *Lecture Notes in Computer Science*, pages 437–448. Springer, Berlin, 2010.

[PAB+17] F. Pierre, J.-F. Aujol, A. Bugeau, G. Steidl, and V.-T. Ta. Variational contrast enhancement of gray-scale and RGB images. *Journal of Mathematical Imaging and Vision*, 57(1):99–116, January 2017.

[Pan12] G. Panin. Mutual information for multi-modal, discontinuity-preserving image registration. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, C. Fowlkes, S. Wang, M.-H. Choi, S. Mantler, J. Schulze, D. Acevedo, K. Mueller, and M. Papka, editors, *Advances in Visual Computing, Part II*, volume 7432 of *Lecture Notes in Computer Science*, pages 70–81. Springer, Berlin, 2012.

[PB14] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, January 2014.

[PB15] K. Papafitsoros and K. Bredies. A study of the one dimensional total generalised variation regularisation problem. *Inverse Problems and Imaging*, 9(2):511–550, May 2015.

[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.

[PC14] E. Provenzi and V. Caselles. A wavelet perspective on variational perceptually-inspired color enhancement. *International Journal of Computer Vision*, 106(2):153–171, January 2014.

[Pen87] A. P. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, July 1987.

[PGB03] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Transactions on Graphics*, 22(3):313–318, July 2003.

[PGP15] S. Pertuz, M. A. Garcia, and D. Puig. Efficient focus sampling through depth-of-field calibration. *International Journal of Computer Vision*, 112(3):342–353, May 2015.

[PH08] M. Pedone and J. Heikkilä. Constrain propagation for ghost removal in high dynamic range images. In *Proc. 3rd International Conference on Computer Vision Theory and Applications*, pages 36–41, Funchal, Portugal, January 2008.

[PHZA11] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen. *Computer Vision Using Local Binary Patterns*. Springer, Berlin, 2011.

[Pie09] G. Piella. Image fusion for enhanced visualization: A variational approach. *International Journal of Computer Vision*, 83(1):1–11, June 2009.

[PK10] F. Pece and J. Kautz. Bitmap movement detection: HDR for dynamic scenes. In *Proc. Conference on Visual Media Production*, pages 1–8, London, UK, November 2010.

[PLTB07] S. Pop, O. Lavialle, R. Terebes, and M. Borda. A PDE-based approach for image fusion. In J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*, volume 4678 of *Lecture Notes in Computer Science*, pages 121–131. Springer, Berlin, 2007.

[PM90] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[PM16] F. Pierre and P. Migerditichan. Débrumage variationnel. In *Proc. XXVème Colloque Gretsi*, pages hal–01168553:1–hal–01168553:5, Lyon, France, September 2016. In French.

[Pol64] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[PPBC09] R. Palma-Amestoy, E. Provenzi, M. Bertalmío, and V. Caselles. A perceptually inspired variational framework for color enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):458–474, March 2009.

[PPC11] N. Papadakis, E. Provenzi, and V. Caselles. A variational model for histogram transfer of color images. *IEEE Transactions on Image Processing*, 20(6):1682–1695, June 2011.

[PPG13a] S. Pertuz, D. Puig, and M. A. Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432, May 2013.

[PPG13b] S. Pertuz, D. Puig, and M. A. Garcia. Reliability measure for shape-from-focus. *Image and Vision Computing*, 31(10):725–734, October 2013.

[Pra01] W. K. Pratt. *Digital Image Processing: PIKS Inside*. Wiley, New York, 3rd edition, 2001.

[PS08] C. Pöschel and O. Scherzer. Characterization of minimizers of convex regularization functionals. In D. R. Larson, P. Massopust, Z. Nashed, M. C. Nguyen, M. Papadakis, and A. Zayed, editors, *Frames and Operator Theory in Analysis and Signal Processing*, volume 451 of *Contemporary Mathematics*, pages 219–248. American Mathematical Society, Providence, 2008.

[PS14] K. Papafitsoros and C. B. Schönlieb. A combined first and second order variational approach for image reconstruction. *Journal of Mathematical Imaging and Vision*, 48(2):308–338, February 2014.

[PS15] C. Pöschel and O. Scherzer. Exact solutions of one-dimensional total generalized variation. *Communications in Mathematical Sciences*, 13(1):171–202, 2015.

[PSL99] D. Purves, A. Shimpi, and R. Beau Lotto. An empirical explanation of the Cornsweet effect. *The Journal of Neuroscience*, 19(19):8542–8551, October 1999.

[PSSW14] N. Persch, C. Schroers, S. Setzer, and J. Weickert. Introducing more physics into variational depth-from-defocus. In X. Jiang, J. Hornegger, and R. Koch, editors, *Pattern Recognition*, volume 8753 of *Lecture Notes in Computer Science*, pages 15–27. Springer, Berlin, 2014.

[PTVF07] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 3rd edition, 2007.

[PV15] K. Papafitsoros and T. Valkonen. Asymptotic behaviour of total generalised variation. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale-Space and Variational Methods in Computer Vision*, volume 9087 of *Lecture Notes in Computer Science*, pages 702–714. Springer, Berlin, 2015.

[PWM+15] P. Peter, J. Weickert, A. Munk, T. Krivobokova, and H. Li. Justifying tensor-driven diffusion from structure-adaptive statistics of natural images. In X.-C. Tai, E. Bae, T. F. Chan, and M. Lysaker, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 8932 of *Lecture Notes in Computer Science*, pages 263–277. Springer, Berlin, 2015.

[PZB11] T. Pock, L. Zedebin, and H. Bischof. TGV-fusion. In C. S. Calude, G. Rozenberg, and A. Salomaa, editors, *Rainbow of Computer Science*, volume 6570 of *Lecture Notes in Computer Science*, pages 245–258. Springer, Berlin, 2011.

[Qia99] N. Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151, January 1999.

[RBP14] R. Ranftl, K. Bredies, and T. Pock. Non-local total generalized variation for optical flow estimation. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014, Part I*, volume 8689 of *Lecture Notes in Computer Science*, pages 439–454. Springer, Berlin, 2014.

[RBS03] M. Robertson, S. Borman, and R. Stevenson. Estimation-theoretic approach to dynamic range improvement using multiple exposures. *Journal of Electronic Imaging*, 12(2):219–228, April 2003.

[RC07] S. Raman and S. Chaudhuri. A matte-less, variational approach to automatic scene compositing. In *Proc. IEEE 11th International Conference on Computer Vision*, pages 574–579, Rio de Janeiro, Brazil, October 2007.

[RC09] S. Raman and S. Chaudhuri. Bilateral filter based compositing for variable exposure photography. In *Proc. EUROGRAPHICS 2009 (Short Papers)*, pages 369–378, Munich, Germany, March 2009.

[RC11] S. Raman and S. Chaudhuri. Reconstruction of high contrast images for dynamic scenes. *The Visual Computer*, 27(2):1099–1114, December 2011.

[RCV11] R. M. Rameshand, S. Chaudhuri, and R. Velmurugan. High dynamic range imaging under noisy observations. In *Proc. 18th IEEE International Conference on Image Processing*, pages 1333–1336, Brussels, Belgium, September 2011.

[RGPB12] R. Ranftl, S. Gehrig, T. Pock, and H. Bischof. Pushing the limits of stereo using variational stereo estimation. In *Proc. IEEE Intelligent Vehicles Symposium*, pages 401–407, Alcalá de Henares, Spain, June 2012.

[RGW05] K. Rasche, R. Geist, and James Westall. Re-coloring images for gamuts of lower dimension. *Computer Graphics Forum*, 24(3):423–432, September 2005.

[RHD+10] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Elsevier, Oxford, 2nd edition, 2010.

[RHW86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1, chapter 8, pages 318–362. MIT Press, Cambridge, MA, 1986.

[Ric11] L. F. Richardson. The approximate arithmetical solution by finite differences of physical problems involving differential equation, with an application to the stresses in a masonry dam. *Philosophical Transactions of the Royal Society A*, 210:307–357, January 1911.

[RMG+13] H. A. Rashwan, M. A. Mohamed, M. A. García, B. Mertsching, and D. Puig. Illumination robust optical flow model based on histogram of oriented gradients. In J. Weickert, M. Hein, and B. Schiele, editors, *Pattern Recognition*, volume 8142 of *Lecture Notes in Computer Science*, pages 354–363. Springer, Berlin, 2013.

[RMVS07] A. A. Rad, L. Meylan, P. Vandewalle, and S. Süsstrunk. Multidimensional image enhancement from a set of unregistered differently exposed images. In C. A. Bouman, E. L. Miller, and I. Polak, editors, *Computational Imaging V*, volume 6498 of *Proceedings of SPIE*, pages 649808:1–649808:12. SPIE Press, Bellingham, 2007.

[Rob60] H. P. Robinson. On printing photographic pictures from several negatives. *British Journal of Photography*, 7(115):94–95, April 1860.

[Rob69] H. P. Robinson. *Pictorial Effect in Photography: Being Hints on composition and Chiaroscuro for Photographers*. Piper & Carter, London, 1869.

[Rob58] H. P. Robinson. Fading Away. `http://images.metmuseum.org/CRDImages/ph/web-large/FI_26_14.jpg` (last accessed August 2016), 1958.

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1–4):259–268, November 1992.

[Rot09] J. M. Roth. Higher order anisotropic smoothing of images. Master's thesis, Department of Computer Science, Saarland University, Saarbrücken, Germany, January 2009.

[RSSF02] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3):267–276, July 2002.

[Sau64] V. K. Saul'yev. *Integration of Equations of Parabolic Type by the Method of Nets*. Pergamon, Oxford, 1964.

[SBBH09] C.-B. Schönlieb, A. Bertozzi, M. Burger, and L. He. Image inpainting using a fourth-order total variation flow. In *Proc. International Conference on Sampling Theory and Applications*, pages 201–202, Marseille, France, May 2009.

[SC95] M. Subbarao and T. Choi. Accurate recovery of three-dimensional shape from image focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):266–274, March 1995.

[SC97] G. Sapiro and V. Caselles. Histogram modification via differential equations. *Journal of Differential Equations*, 135(2):238–268, April 1997.

[SC02] J. Shen and T. F. Chan. Mathematical models for local nontexture inpaintings. *SIAM Journal on Applied Mathematics*, 62(3):1019–1043, 2002.

[SC10] S.-O. Shim and T.-S. Choi. A novel iterative shape from focus algorithm based on combinatorial optimization. *Pattern Recognition*, 43(10):3338–3347, October 2010.

[SCB13] R. Shen, I. Cheng, and A. Basu. QoE-based multi-exposure fusion in hierarchical multivariate Gaussian CRF. *IEEE Transactions on Image Processing*, 22(6):2469–2478, June 2013.

[Sch93] C. Schnörr. On functionals with greyvalue-controlled smoothness terms for determining optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1074–1079, October 1993.

[Sch98] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60(1):1–27, March 1998.

[Sch16] D. Scharstein. Optical flow evaluation results. `http://vision.middlebury.edu/flow/eval/results/results-e1.php` (last accessed August 2016), 2016.

[SCN93] M. Subbarao, T. Choi, and A. Nikzad. Focusing techniques. *Optical Engineering*, 32(11):2824–2836, November 1993.

[SCSB11] R. Shen, I. Cheng, J. Shi, and A. Basu. Generalized random walks for fusion of multi-exposure images. *IEEE Transactions on Image Processing*, 20(12):3634–3646, December 2011.

[SDN05] G. Steidl, S. Didas, and J. Neumann. Relations between higher order TV regularization and support vector regression. In R. Kimmel, N. A. Sochen, and J. Weickert, editors, *Scale Space and PDE Methods in Computer Vision*, volume 3459 of *Lecture Notes in Computer Science*, pages 515–527. Springer, Berlin, 2005.

[SE84] R. Shapley and C. Enroth-Cugell. Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3:1567–1599, March 1984.

[SEM11] A. Sellent, M. Eisemann, and M. Magnor. Two algorithms for motion estimation from alternate exposure images. In D. Cremers, M. Magnor, M. R. Oswald, and L. Zelnik-Manor, editors, *Video Processing and Computational Video*, volume 7082 of *Lecture Notes in Computer Science*, pages 25–51. Springer, Berlin, 2011.

[Sho53] G. Shortley. Use of Tschebyscheff-polynomial operators in the numerical solution of boundary-value problems. *Journal of Applied Physics*, 24(4):392–396, April 1953.

[Shu89] D. Shulman. Regularization of discontinuous flow fields. In *Proc. Workshop on Visual Motion*, pages 81–86, Irvine, CA, March 1989.

[SI85] S. A. Sugimoto and Y. Ichioka. Digital composition of images with increased depth of focus considering depth information. *Applied Optics*, 24(14):2076–2080, July 1985.

[SKB14] H. Singh, V. Kumar, and S. Bhooshan. Weighted least squares based detail enhanced exposure fusion. *ISNR Signal Processing*, 2014:498762:1–498762:18, February 2014.

[SKC03] J. Shen, S. H. Kang, and T. F. Chan. Euler's elastica and curvature-based inpainting. *SIAM Journal on Applied Mathematics*, 63(2):564–592, 2003.

[SKY+12] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtma. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Transactions on Graphics*, 31(6):203:1–203:11, November 2012.

[SL12] S. Silk and J. Lang. High dynamic range image deghosting by fast approximate background modelling. *Computers & Graphics*, 36(8):1060–1071, September 2012.

[SL15] P. Smith and C. Lefley. *Rethinking Photography: Histories, Theories and Education*. Routledge, London, September 2015.

[SLTM08] K. Smith, P.-E. Landes, J. Thollot, and K. Myszkowski. Apparent greyscale: A simple and fast conversion to perceptually accurate images and video. *Computer Graphics Forum*, 27(2):193–200, April 2008.

[SMDH13] I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In *Proc. 30th International Conference on Machine Learning*, pages 1139–1147, Atlanta, GA, June 2013.

[Soa09] S. Soatto. Actionable information in vision. In *Proc. IEEE 12th International Conference on Computer Vision*, pages 2138–2145, Kyoto, Japan, September 2009.

[SP95] N. J. A. Sloane and S. Plouffe. *The Encyclopedia of Integer Sequences*. Academic Press, San Diego, March 1995.

[SPC09] 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, *Vision, Modeling, and Visualization Workshop*, pages 155–164. DNB, Frankfurt, 2009.

[SPLC13] H.-S. Sung, R.-H. Park, D.-K. Lee, and S. Chang. Feature based ghost removal in high dynamic range imaging. *International Journal of Computer Graphics & Animation*, 3(4):23–39, October 2013.

[SPS09] D. Sidibe, W. Puech, and O. Strauss. Ghost detection and removal in high dynamic range images. In *Proc. 17th European Signal Processing Conference*, pages 2240–2244, Glasgow, Scotland, August 2009.

[SS02] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42, April 2002.

[SS06] S. Shalev-Shwartz and Y. Singer. Efficient learning of label ranking by soft projections onto polyhedra. *Journal of Machine Learning Research*, 7:1567–1599, July 2006.

[SS08] S. Setzer and G. Steidl. Variational methods with higher-order derivatives in image processing. In M. Neamtu and L. L. Schumaker, editors, *Approximation XII*, volume 9087, pages 360–386. Nashboro Press, Brentwood, 2008.

[SS12] A. Srikantha and D. Sidibe. Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication*, 27(6):650–662, July 2012.

[SSM12] A. Srikantha, D. Sidibe, and F. Meriaudeau. An SVD-based approach for ghost detection and removal in high dynamic range images. In *Proc. 21st International Conference on Pattern Recognition*, pages 380–383, Tsukuba, Japan, November 2012.

[SSM13] S. Setzer, G. Steidl, and J. Morgenthaler. A cyclic projected gradient method. *Computational Optimization and Applications*, 54(2):417–440, March 2013.

[SSNW83] J. F. Schlag, A. C. Sanderson, C. P. Neuman, and F. C. Wimberly. Implementation of automatic focusing algorithms for a computer vision system with camera control. Technical Report CMU-RI-TR-83-14, Department of Electrical Engineering, Carnegie-Mellon University, Pittsburgh, PA, August 1983.

[SST11] S. Setzer, G. Steidl, and T. Teuber. Infimal convolution regularizations with discrete $\ell 1$-type functionals. *Communications in Mathematical Sciences*, 9(3):797–827, 2011.

[ST98] M. Subbarao and J.-K. Tyan. Selecting the optimal focus measure for autofocusing and depth-from-focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):864–870, August 1998.

[ST04] P. Sand and S. Teller. Video matching. *ACM Transactions on Graphics*, 22(3):592–599, August 2004.

[Sta08] T. Stathaki. *Image Fusion: Algorithms and Applications*. Academic Press, Elsevier, London, May 2008.

[STC+12] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang. Probabilistic exposure fusion. *IEEE Transactions on Image Processing*, 21(1):341–357, January 2012.

[Ste04] F. Stein. Efficient computation of optical flow using the census transform. In C. E. Rasmussen, H. H. Bülthoff, B. Schölkopf, and M. A. Giese, editors, *Pattern Recognition*, volume 3175 of *Lecture Notes in Computer Science*, pages 79–86. Springer, Berlin, 2004.

[Ste06] G. Steidl. A note on the dual treatment of higher-order regularization functionals. *Computing*, 76(1):135–148, January 2006.

[Ste15] G. Steidl. Combined first and second order variational approaches for image processing. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 117(2):113–160, June 2015.

[Sut91] D. Suter. Constraint networks in vision. *IEEE Transactions on Computers*, 40(12):1359–1367, December 1991.

[Sut94] D. Suter. Motion estimation and vector splines. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 939–942, Seattle, WA, June 1994.

[SXJ15] J. Shi, L. Xu, and J. Jia. Just noticeable defocus blur detection and estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 657–665, Boston, MA, June 2015.

[SZV+12] C. Schroers, H. Zimmer, L. Valgaerts, A. Bruhn, O. Demetz, and J. Weickert. Anisotropic range image integration. In A. Pinz, T. Pock, H. Bischof, and F. Leberl, editors, *Pattern Recognition*, volume 7476 of *Lecture Notes in Computer Science*, pages 73–82. Springer, Berlin, 2012.

[TAEE15] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem. The state of the art in HDR deghosting: A survey and evaluation. *Computer Graphics Forum*, 34(2):683–707, May 2015.

[TAEE16] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem. An objective deghosting quality metric for HDR images. *Computer Graphics Forum*, 35(2):139–152, May 2016.

[TCMY11] J. Tian, L. Chen, L. Ma, and W. Yu. Multi-focus image fusion using a bilateral gradient-based sharpness criterion. *Optics Communications*, 284(1):80–87, January 2011.

[Ten70] J. M. Tenenbaum. *Accommodation in Computer Vision*. PhD thesis, Computer Science Department, Stanford University, California, USA, October 1970.

[Ter83] D. Terzopoulos. Multilevel computational processes for visual surface reconstruction. *Computer Vision, Graphics, and Image Processing*, 24(1):52–86, October 1983.

[Ter84] D. Terzopoulos. Multilevel reconstruction of visual surfaces: Variational principles and finite-element representations. In A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis*, volume 12 of *Springer Series in Information Sciences*, pages 237–310. Springer, Berlin, 1984.

[THC11] X.-C. Tai, J. Hahn, and G. J. Chung. A fast algorithm for Euler's elastica model using augmented Lagrangian method. *SIAM Journal on Imaging Sciences*, 4(1):313–344, 2011.

[THMR13] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In

Proc. IEEE International Conference on Computer Vision, pages 673–680, Sydney, Australia, December 2013.

[THS13] Y. Takeda, S. Hiura, and K. Sato. Fusing depth from defocus and stereo with coded apertures. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 209–216, Portland, OR, June 2013.

[TKTS11] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen. A versatile HDR video production system. *ACM Transactions on Graphics*, 30(4):41:1–41:10, July 2011.

[TLCT09] F. Tang, S. H. Lim, N. L. Chang, and H. Tao. A novel feature descriptor invariant to complex brightness changes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 2631–2638, Miami, FL, June 2009.

[TM98] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. Sixth International Conference on Computer Vision*, pages 839–846, Bombay, India, January 1998.

[TM07] A. Tomaszewska and R. Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *Proc. 15th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pages 49–56, Plzeň, Czech Republic, January 2007.

[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.

[TPCB08] W. Trobin, T. Pock, D. Cremers, and H. Bischof. An unbiased second-order prior for high-accuracy motion estimation. In G. Rigoll, editor, *Pattern Recognition*, volume 5096 of *Lecture Notes in Computer Science*, pages 396–405. Springer, Berlin, 2008.

[TT99] J. Tumblin and G. Turk. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In *Proc. SIGGRAPH '99*, pages 83–90, Los Angeles, CA, August 1999.

[TWBO03] T. Tasdizen, R. Whitaker, P. Burchard, and S. Osher. Geometric surface processing via normal maps. *ACM Transactions on Graphics*, 22(4):1012–1033, October 2003.

[UGVT88] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, 60(2):79–87, December 1988.

[UPWB10] M. Unger, T. Pock, M. Werlberger, and H. Bischof. A convex approach for variational super-resolution. In M. Goesele, S. Roth, A. Kuijper, B. Schiele, and K. Schindler, editors, *Pattern Recognition*, volume 6376 of *Lecture Notes in Computer Science*, pages 313–322. Springer, Berlin, 2010.

[Var57] R. S. Varga. A comparison of the successive overrelaxation method and semi-iterative methods using Chebyshev polynomials. *Journal of the Society for Industrial and Applied Mathematics*, 5(2):39–46, June 1957.

[Var09] R. S. Varga. *Matrix Iterative Analysis*. Springer, Berlin, 2nd edition, December 2009.

[VBK13] T. Valkonen, K. Bredies, and K. Knoll. Total generalized variation in diffusion tensor imaging. *SIAM Journal on Imaging Sciences*, 6(1):487–525, 2013.

[VBVZ11] S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer. Modeling temporal coherence for optical flow. In *Proc. International Conference on Computer Vision*, volume 30, pages 1116–1123, Barcelona, Spain, November 2011.

[VBWD07] O. Vogel, A. Bruhn, J. Weickert, and S. Didas. Direct shape-from-shading with adaptive higher order regularisation. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 871–882. Springer, Berlin, 2007.

[vCF06] F. Šroubek, G. Cristóbal, and J. Flusser. Image fusion based on level set segmentation. In *Proc. 14th European Signal Processing Conference*, volume 1, pages 15024146:1–15024146:5, Florence, Italy, September 2006.

[vdH96] P. J. van der Houwen. The development of Runge-Kutta methods for partial differential equations. *Applied Numerical Mathematics*, 20(3):261–272, March 1996.

[vG04] J. van de Weijer and T. Gevers. Robust optical flow from photometric invariants. In *Proc. International Conference on Image Processing*, volume 3, pages 1835–1838, Singapore, October 2004.

[VGT+11] D. Vaquero, N. Gelfand, M. Tico, K. Pulli, and M. Turk. Generalized autofocus. In *Proc. IEEE Workshop on Applications of Computer Vision*, pages 511–518, Kona, HI, January 2011.

[VHS90] J. G. Verwer, W. H. Hundsdorfer, and B. P. Sommerijer. Convergence properties of the Runge-Kutta-Chebyshev method. *Numerische Mathematik*, 57(1):157–178, December 1990.

[VRS13] C. Vogel, S. Roth, and K. Schindler. An evaluation of data costs for optical flow. In J. Weickert, M. Hein, and B. Schiele, editors, *Pattern Recognition*, volume 8142 of *Lecture Notes in Computer Science*, pages 343–353. Springer, Berlin, 2013.

[vS80] P. J. van der Houwen and B. P. Sommeijer. On the internal stability of explicit, m-stage Runge-Kutta methods for large m-values. *Zeitschrift für Angewandte Mathematik und Mechanik*, 60(10):479–485, 1980.

[VW97] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, September 1997.

[War03] G. Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics, GPU, and Game Tools*, 8(2):17–30, 2003.

[WBSS04] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004. `https://ece.uwaterloo.ca/~z70wang/research/ssim/` (last accessed August 2016).

[Wei96] J. Weickert. Theoretical foundations of anisotropic diffusion in image processing. *Computing Supplement*, 11:221–236, 1996.

[Wei98] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.

[Wei99] G. W. Wei. Generalized Perona-Malik equation for image restoration. *IEEE Signal Processing Letters*, 6(7):165–167, July 1999.

[Wel86] W. M. Wells III. Efficient synthesis of Gaussian filters by cascaded uniform filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):234–239, March 1986.

[WFW11] Z. Wang, B. Fan, and F. Wu. Local intensity order pattern for feature description. In *Proc. International Conference on Computer Vision*, pages 603–610, Barcelona, Spain, November 2011.

[WGF+97] A. M. Waxman, A. N. Gove, D. A. Fay, J. P. Racamato, J. E. Carrick, M. C. Seibert, and E. D. Savoye. Color night vision: Opponent processing in the fusion of visible and IR imagery. *Neural Networks*, 10(1):1–6, January 1997.

[WGSB16] J. Weickert, S. Grewenig, C. Schroers, and A. Bruhn. Cyclic schemes for PDE-based image analysis. *International Journal of Computer Vision*, 118(3):275–299, July 2016.

[WHBV13] J. Weickert, K. Hagenburg, M. Breuß, and O. Vogel. Linear osmosis models for visual computing. In A. Heyden, F. Kahl, C. Olsson, M. Oskarsson, and X.-C. Tai, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 8081 of *Lecture Notes in Computer Science*, pages 26–39. Springer, Berlin, 2013.

[Wil80] H. R. Wilson. A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics*, 38(3):171–178, October 1980.

[WMG10] Z. Wang, Y. Ma, and J. Gu. Multi-focus image fusion using PCNN. *Pattern Recognition Letters*, 43(6):2003–2016, June 2010.

[WPB10] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with nonlocal total variation regularization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 2464–2471, San Francisco, CA, June 2010.

[WPZ+09] A. Wedel, T. Pock, C. Zach, D. Cremers, and H. Bischof. An improved algorithm for TV-L1 optical flow. In D. Cremers, B. Rosenhahn, A. L. Yuille, and F. R. Schmidt, editors, *Statistical and Geometrical Approaches to Visual Motion Analysis*, volume 5604 of *Lecture Notes in Computer Science*, pages 23–45. Springer, Berlin, 2009.

[WS99] J. Weickert and C. Schnörr. Räumlich-zeitliche Berechnung des optischen Flusses mit nichtlinearen flußabhängigen Glattheitstermen. In W. Förstner, J. M. Buhmann, A. Faber, and P. Faber, editors, *Mustererkennung 1999*, pages 317–324. Springer, Berlin, 1999. In German.

[WS00] G. Wyszecki and W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, New York, 2000.

[WS01] 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.

[WSF08] W.-W. Wang, P.-L. Shui, and X.-C. Feng. Variational models for fusion and denoising of multifocus images. *IEEE Signal Processing Letters*, 15:65–68, January 2008.

[WT13] C. Wang and C. Tu. An exposure fusion approach without ghost for dynamic scenes. In *Proc. 6th International Congress on Image and Signal Processing*, pages 904–909, Hangzhou, China, December 2013.

[WWW13] J. Weickert, M. Welk, and M. Wickert. L^2-stable nonstandard finite differences for anisotropic diffusion. In A. Kuijper, K. Bredies, T. Pock, and H. Bischof, editors, *Scale-Space and Variational Methods in Computer Vision*, volume 7893 of *Lecture Notes in Computer Science*, pages 380–391. Springer, Berlin, 2013.

[WXRL10] S. Wu, S. Xie, S. Rahardja, and Z. Li. A robust and fast anti-ghosting algorithm for high dynamic range imaging. In *Proc. IEEE International Conference on Image Processing*, pages 397–400, Hong Kong, China, September 2010.

[WYP+13] W. Wu, X. Yang, Y. Pang, J. Peng, and G. Jeon. A multifocus image fusion method by using hidden Markov model. *Optics Communications*, 287:63–72, January 2013.

[WZQ13] T. Wan, C. Zhu, and Z. Qin. Multifocus image fusion based on robust principal component analysis. *Pattern Recognition Letters*, 34(9):1001–1008, July 2013.

[XJM10] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1293–1300, San Francisco, CA, June 2010.

[XS93] Y. Xiong and S. A. Shafer. Depth from focusing and defocusing. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 68–73, New York, NY, June 1993.

[XTAA04] N. Xu, K. Tan, H. Arora, and N. Ahuja. Generating omnifocus images using graph cuts and a new focus measure. In *Proc. 17th International Conference on Pattern Recognition*, volume 4, pages 697–700, Cambridge, UK, August 2004.

[Yao11] S. Yao. Robust image registration for multiple exposure high dynamic range image synthesis. In J. T. Astola and K. O. Egiazarian, editors, *Image Processing: Algorithms and Systems IX*, volume 7870 of *Proceedings of SPIE*, pages 78700Q:1–78700Q:9. SPIE Press, Bellingham, 2011.

[YC99] J. Yun and T. S. Choi. Accurate 3-D shape recovery using curved window focus measure. In *Proc. International Conference on Image Processing*, volume 3, pages 910–914, Kobe, Japan, October 1999.

[YK00] Y.-L. You and M. Kaveh. Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing*, 9(10):1723–1730, October 2000.

[YMG+14] J. Yuan, B. Miles, G. Garvin, X.-C. Tai, and A. Fenster. Efficient convex optimization approaches to variational image fusion. *Numerical Mathematics: Theory, Methods and Applications*, 7(2):234–250, May 2014.

[You54] D. M. Young. On Richardson's method for solving linear systems with positive definite matrices. *Journal of Mathematics and Physics*, 32(1):243–255, April 1954.

[You89] D. M. Young. A historical overview of iterative methods. *Computer Physics Communications*, 53(1–3):1–17, May 1989.

[YSM07] J. Yuan, C. Schnörr, and E. Mémin. Discrete orthogonal decomposition and variational fluid flow estimation. *Journal of Mathematical Imaging and Vision*, 28(1):67–80, May 2007.

[YSS07] J. Yuan, C. Schnörr, and G. Steidl. Simultaneous higher-order optical flow estimation and decomposition. *SIAM Journal on Scientific Computing*, 29(6):2283–2304, 2007.

[Yua58] Yuan'Chzhao-Din. *Some difference schemes for the solution of the first boundary value problem for linear differential equations with partial derivatives*. PhD thesis, Moscow State University, Russia, 1958. In Russian.

[Yv95] I. T. Young and L. J. van Vliet. Recursive implementation of the Gaussian filter. *Signal Processing*, 44(2):139–151, June 1995.

[ZB99] Z. Zhang and R. S. Blum. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of the IEEE*, 87(8):1315–1326, August 1999.

[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*, volume 5681 of *Lecture Notes in Computer Science*, pages 207–220. Springer, Berlin, 2009.

[ZBW11a] H. Zimmer, A. Bruhn, and J. Weickert. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. *Computer Graphics Forum*, 30(2):405–414, April 2011.

[ZBW11b] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, July 2011.

[ZC12a] W. Zhang and W.-K. Cham. Gradient-directed multiexposure composition. *IEEE Transactions on Image Processing*, 21(4):2318–2323, April 2012.

[ZC12b] W. Zhu and T. Chan. Image denoising using mean curvature of image surface. *SIAM Journal on Imaging Sciences*, 5(1):1–32, 2012.

[Zhe11] Y. Zheng. *Image Fusion and Its Applications*. InTech, Rijeka, June 2011.

[Zim11] H. Zimmer. *Correspondence Problems in Computer Vision – Novel Models, Numerics, and Applications*. PhD thesis, Department of Computer Science, Saarland University, Saarbrücken, Germany, 2011.

[ZLZ+13] J. Zheng, Z. Li, Z. Zhu, S. Wu, and S. Rahardja. Hybrid patching for a sequence of differently exposed images with moving objects. *IEEE Transactions on Image Processing*, 22(12):1057–7149, December 2013.

[ZSM08] X. Zhang, T. Sim, and X. Miao. Enhancing photographs with near infrared images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 10140128:1–10140128:8, Anchorage, AK, June 2008.

[ZVB14] S. W. Zamir, J. Vazquez-Corral, and M. Bertalmío. Gamut mapping in cinematography through perceptually-based contrast modification. *IEEE Journal of Selected Topics in Signal Processing*, 8(3):490–503, June 2014.

[ZW94] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In J.-O. Eklundh, editor, *Computer Vision – ECCV '94, Part II*, volume 801 of *Lecture Notes in Computer Science*, pages 151–158. Springer, Berlin, 1994.