Multi-Frame Super-Resolution Reconstruction with Applications to Medical Imaging

Bildfolgenbasierte Verfahren zur Auflösungserhöhung mit Anwendungen in der Medizinischen Bildgebung

Der Technischen Fakultät der
Friedrich-Alexander-Universität Erlangen-Nürnberg

zur Erlangung des Grades

Doktor-Ingenieur (Dr.-Ing.)

vorgelegt von

Thomas Köhler

aus

Bamberg, Deutschland

Als Dissertation genehmigt von der
Technischen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung:	11.09.2017
Vorsitzender des Promotionsorgans:	Prof. Dr.-Ing. Reinhard Lerch
Gutachter:	Prof. Dr.-Ing. Joachim Hornegger
	Prof. Sina Farsiu, Ph.D.

Abstract

The optical resolution of a digital camera is one of its most crucial parameters with broad relevance for consumer electronics, surveillance systems, remote sensing, or medical imaging. However, resolution is physically limited by the optics and sensor characteristics. In addition, practical and economic reasons often stipulate the use of out-dated or low-cost hardware. Super-resolution is a class of retrospective techniques that aims at high-resolution imagery by means of software. Multi-frame algorithms approach this task by fusing multiple low-resolution frames to reconstruct high-resolution images. This work covers novel super-resolution methods along with new applications in medical imaging.

The first contribution of this thesis concerns computational methods to super-resolve image data of a single modality. The emphasis lies on motion-based algorithms that are derived from a Bayesian statistics perspective, where subpixel motion of low-resolution frames is exploited to reconstruct a high-resolution image. More specifically, we introduce a confidence-aware Bayesian observation model to account for outliers in the image formation, e. g. invalid pixels. In addition, we propose an adaptive prior for sparse regularization to model natural images appropriately. We then develop a robust optimization algorithm for super-resolution using this model that features a fully automatic selection of latent hyperparameters. The proposed approach is capable of meeting the requirements regarding robustness of super-resolution in real-world systems including challenging conditions ranging from inaccurate motion estimation to space variant noise. For instance, in case of inaccurate motion estimation, the proposed method improves the peak-signal-to-noise ratio (PSNR) by 0.7 decibel (dB) over the state-of-the-art.

The second contribution concerns super-resolution of multiple modalities in the area of hybrid imaging. We introduce novel multi-sensor super-resolution techniques and investigate two complementary problem statements. For super-resolution in the presence of a guidance modality, we introduce a reconstruction algorithm that exploits guidance data for motion estimation, feature driven adaptive regularization, and outlier detection to reliably super-resolve a second modality. For super-resolution in the absence of guidance data, we generalize this approach to a reconstruction algorithm that jointly super-resolves multiple modalities. These multi-sensor methodologies boost accuracy and robustness compared to their single-sensor counterparts. The proposed techniques are widely applicable for resolution enhancement in a variety of multi-sensor vision applications including color-, multispectral- and range imaging. For instance in color imaging as a classical application, joint super-resolution of color channels improves the PSNR by 1.5 dB compared to conventional channel-wise processing.

The third contribution transfers super-resolution to workflows in healthcare. As one use case in ophthalmology, we address retinal video imaging to gain spatio-temporal measurements on the human eye background non-invasively. In order to enhance the diagnostic usability of current digital cameras, we introduce a framework to gain high-resolution retinal images from low-resolution video data by exploiting natural eye movements. This framework enhances the mean sensitivity of automatic blood vessel segmentation by 10 % when using super-resolution for image preprocessing. As a second application in image-guided surgery, we

investigate hybrid range imaging. To overcome resolution limitations of current range sensor technologies, we propose multi-sensor super-resolution based on domain-specific system calibrations and employ high-resolution color images to steer range super-resolution. In ex-vivo experiments for minimally invasive and open surgery procedures using Time-of-Flight (ToF) sensors, this technique improves the reliability of surface and depth discontinuity measurements compared to raw range data by more than 24 % and 68 %, respectively.

Kurzübersicht

Die optische Auflösung einer Kamera ist eine ihrer wichtigsten Kenngrößen mit hohem Stellenwert für Unterhaltungselektronik, Überwachungssysteme, Fernerkundung oder medizinische Bildgebung. Jedoch ist die Auflösung durch Optik und Sensoren physikalisch beschränkt. Daneben bedingen praktische oder ökonomische Gründe den Einsatz veralteter oder preiswerter Hardware. Verfahren zur Auflösungserhöhung sind eine Klasse retrospektiver Techniken mit dem Ziel hochauflösende Bildgebung softwarebasiert zu gewährleisten. Bildfolgenbasierte Algorithmen ermöglichen dies durch Fusion mehrerer niedrigauflösender Bilder zur Rekonstruktion hochauflösender Bilder. Diese Arbeit behandelt neuartige Methoden zur Auflösungserhöhung, sowie neue Anwendungen für die medizinische Bildgebung.

Der erste Beitrag dieser Arbeit betrifft Verfahren zur Auflösungserhöhung von Bildern einer einzelnen Modalität. Den Schwerpunkt bilden bewegungsbasierte und mit Bayesscher Statistik hergeleitete Algorithmen, bei denen Subpixel-Verschiebungen zwischen niedrig auflösenden Bildern zur Rekonstruktion eines hochauflösenden Bildes genutzt werden. Konkret führen wir ein konfidenzgewichtetes Beobachtungsmodell zur Behandlung von Ausreißern, z. B. defekte Pixel, in der Bildaufnahme ein. Zusätzlich stellen wir eine neue adaptive Verteilungsfunktion für die Regularisierung zur adäquaten Modellierung natürlicher Bilder vor. Wir entwickeln ferner einen robusten Optimierungsalgorithmus mit diesem Modell, der Hyperparameter vollautomatisch auswählt. Der vorgestellte Ansatz zur Auflösungserhöhung erfüllt in der Praxis Anforderungen hinsichtlich Robustheit, welche schwierige Rahmenbedingungen von ungenauer Bewegungsschätzung bis ortsvariantem Rauschen umfassen. Im beispielhaften Fall einer ungenauen Bewegungsschätzung verbessert die vorgeschlagene Methode das Spitzen-Signal-Rausch-Verhältnis (PSNR) um 0.7 decibel (dB) gegenüber dem Stand der Technik.

Der zweite Beitrag betrifft Ansätze zur Auflösungserhöhung für mehrere Modalitäten in der hybriden Bildgebung. Wir führen hierfür neue Mehrsensor-Verfahren ein und untersuchen zwei gegensätzliche Problemstellungen. Für die Auflösungserhöhung unter Verwendung einer Führungsmodalität stellen wir einen Algorithmus vor, der diese zur Bewegungsschätzung, merkmalsbasierten adaptiven Regularisierung und Ausreißerdetektion zur zuverlässigen Auflösungserhöhung einer zweiten Modalität einsetzt. Für den Fall, dass Führungsdaten fehlen, verallgemeinern wir diesen Ansatz zu einem Algorithmus, der mehrere Modalitäten simultan verarbeitet. Diese Mehrsensor-Methodik steigert Genauigkeit und Robustheit gegenüber Einzelsensor-Ansätzen. Die neu eingeführten Techniken sind vielfältig für eine Auflösungserhöhung in zahlreichen Anwendungen von Mehrsensor-Bildgebung einsetzbar, was Farb-, Multispektral- sowie Tiefenbildgebung umfasst. Im Bereich der Farbbildgebung als Beispiel für ein klassisches Anwendungsfeld, verbessert die simultane Auflösungserhöhung von Farbkanälen das PSNR um 1.5 dB gegenüber einer konventionellen kanalweisen Verarbeitung.

Der dritte Beitrag überträgt Verfahren zur Auflösungserhöhung in die Medizin. Als Anwendung in der Ophthalmologie behandeln wir Videobildgebung zur nicht-invasiven, örtlich-zeitlichen Untersuchung der menschlichen Retina. Um den diagnostischen Nutzen aktueller Digitalkameras zu verbessern, stellen wir ein

Verfahren zur Gewinnung hochauflösender Retinabilder aus niedrigauflösenden Videodaten unter Ausnutzung natürlicher Augenbewegungen vor. Das Verfahren verbessert die mittlere Sensitivität einer automatischen Blutgefäßsegmentierung um 10 %, wenn eine Auflösungserhöhung zur Bildvorverarbeitung genutzt wird. Als eine weitere Anwendung in der bildgeführten Chirurgie untersuchen wir hybride Tiefenbildgebung. Um Auflösungsbeschränkungen heutiger Tiefensensoren zu überwinden, führen wir anwendungsspezifische Kalibrierverfahren ein und verwenden hochauflösende Farbbilder für Mehrsensor-Auflösungserhöhung auf Tiefendaten. In ex-vivo Experimenten für minimal-invasive und offene Chirurgie mit Time-of-Flight (ToF) Sensoren verbessert diese Technik die Zuverlässigkeit von Oberflächen- und Tiefenkantenmessungen um mehr als 24 % bzw. 68 % gegenüber Rohdaten.

Acknowledgment

I would especially like to thank Prof. Dr. Joachim Hornegger for the opportunity of doing my Ph.D. in the highly fascinating area of super-resolution within the inspiring research environment at the Pattern Recognition Lab. I am certainly grateful for the freedom that he gave me in my research, the confidence that he put in me as well as his continuous encouragement within the last years. Let me also deeply thank Prof. Dr. Andreas Maier for his great support and outstanding scientific advice in the final phase of preparing this thesis. I am also particularly grateful to Prof. Dr. Claudius Schnörr, who awakened my enthusiasm for image processing during my Master's studies and pointed me towards a Ph.D study.

I greatly appreciate the wonderful time at the lab over the past years. In particular, let me thank the former Range Imaging Group – Dr. Sebastian Bauer, Dr. Sven Haase, and Jakob Wasza – for discussing papers, the great advice in the field of range imaging, and the nice conference trips we had together. Special thanks go to Sven for the fruitful and friendly cooperation within his DFG project. This resulted in several exciting publications that built the foundation of this thesis. A big thanks also goes to Xiaolin Huang Ph.D. for his tremendous support on the convergence proof for the algorithm presented in Chapter 4. Many thanks also to my room mates André Aichert, Lennart Husvogt, and Martin Kraus for the relaxed working atmosphere, inspiring scientific discussions, and nice leisure activities. Let me also acknowledge André Aichert, Dr. Alexander Brost, Simone Gaffling, Wilhelm Haas, Dr. Sven Haase, Matthias Hoffmann, and the uncountable number of student tutors for their valuable assistance in teaching. Their awesome help ensured that I enjoyed an excellent balance between teaching and doing research.

I would like to express the immense gratitude for proofreading this thesis to my „editor team“: André Aichert, Dr. Martin Berger, Martin Gabriel, Dr. Sven Haase, Matthias Hoffmann, and Dr. Ralf Peter Tornow. Their thorough review and invaluable feedback greatly improved this manuscript.

Many parts of my research would have been impossible without appropriate data and I deeply thank all project partners that supported my experimental studies: Peter Fürsattel and Dr. Sven Haase for enabling the Time-of-Flight data acquisitions, Christiane Köhler, Prof. Dr. Georg Michelson, and Dr. Ralf Peter Tornow for providing clinical fundus images, and Dr. Johannes Jordan for providing the Gerbil framework to visualize multispectral images. Another big thank you goes to Michel Bätz, Farzad Naderi, and Dr. Christian Riess for their great effort to capture super-resolution benchmark datasets within our ongoing collaboration.

I would like to acknowledge all students that I supervised during the last years, especially Cosmin Bercea, Florin C. Ghesu, Axel Heinrich, Katja Mogalle, and Anja Kürten. Their excellent works contributed to several scientific publications.

Last but not least, I would like to deeply thank my whole family and my wife Magdalena. Thank you very much for the patience and emotional support within the ups and downs of the last years.

Thomas Köhler

Contents

CHAPTER 1

Introduction

The resolution of an imaging system characterizes the level of spatial detail at which it captures images and – besides the contrast resolution – it is considered as a major quality indicator. This is obvious in digital photography, where the camera resolution is directly related to the acquisition of fine textures in a scene. In remote sensing as another prominent example, one is interested in measuring information on a planet surface over long distances, which requires high-resolution cameras. Resolution is also crucial in the context of medical imaging to support interventional or diagnostic workflows. For instance, morphological imaging modalities need to provide precise information regarding human anatomy.

In most of these areas, a large effort has been made by researchers and system manufacturers to develop sensors and optical components that enable high-resolution imagery. However, resolution is inherently limited. Besides technological constraints, the use of improved hardware might lead to unacceptable high costs or sizes of commercial systems. In contrast to mass products with a limited life cycle that may allow the use of improved hardware in future product releases, a simple replacement of hardware components is often not feasible in existing long-lived systems, e. g. in remote sensing or medical imaging. In these cases one has to optimally use existing hardware and needs to employ techniques to enhance the actual image resolution. This thesis investigates *super-resolution* methods to reconstruct high-resolution images from low-resolution ones retrospectively.

This chapter provides an introduction to the physics of digital imaging as well as different paradigms of super-resolution. Finally, the scientific contributions of this work are elaborated and an outline of the different chapters is presented.

1.1 Resolution of Digital Imaging Systems

Let us first discuss the meaning of *optical resolution* in terms of imaging systems as it is the focus of any super-resolution method to improve this property. Optical resolution is an abstract term and depends on a variety of physical parameters. In this work, optical resolution is defined in accordance to optics literature and

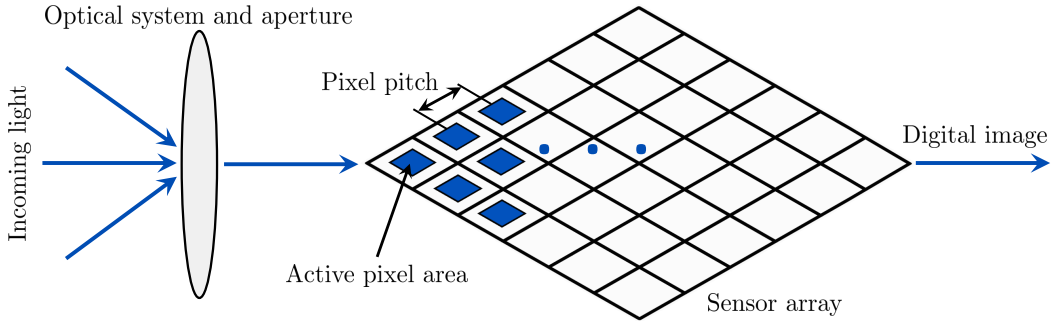


Figure 1.1: Illustration of the sensor array of a digital imaging system with square pixels. Incoming light passes an optical system and is integrated on the active pixel areas.

denotes the ability of an imaging system to capture spatial details. A classical approach to objectify this parameter are two-point criterions that measure the ability to resolve two point light sources without interference in the image [Dekk 97]. Examples for these objective criterions are the Rayleigh [Lord 79] or the Sparrow criterion [Spar 16]. As this thesis is focused on digital optical imaging, we consider an imaging system as the composition of optical components and a sensor array. There are two main aspects that influence the optical resolution[1].

Limitations of the Optics. In terms of optics, the optical resolution is inherently limited by *diffraction* that is related to the camera aperture size and the wavelength of light [Erso 06]. The diffraction barrier results in a spread of incoming light waves when passing a small aperture and leads to distortions of the light signal. Moreover, unavoidable manufacturing uncertainties of lenses cause additional distortions. For these reasons, point light sources cannot be captured as ideal points and appear blurred in an acquired image. These distortions limit the optical resolution and are modeled by the optical point spread function (PSF). This function denotes the impulse response of the optical system and causes a band-limitation in terms of spatial frequencies that can be actually resolved [Lind 12].

Limitations of the Sensor. In addition to optical effects, the utilized sensor technology influences the optical resolution. In digital imaging, resolution is affected by the discretization of incoming light according to the sensor geometry. Charge-coupled device (CCD) or complementary metaloxide semiconductor (CMOS) systems [El G 05] consist of picture elements (pixels) that represent the sampling positions for this discretization. Two major parameters of the sensor are the pixel pitch and the active pixel area, see Fig. 1.1. These parameters define the *pixel resolution* that denotes the number of pixels on the sensor array. The pixel resolution directly contributes to the optical resolution as long as the diffraction limit is not exceeded. According to the Nyquist-Shannon sampling theorem [Shan 48], the sensor needs to provide a sufficiently high resolution to avoid *aliasing* due to undersampling.

However, simply putting a higher number of pixels to the sensor array is often impracticable as pixels have a non-infinitesimal size and the maximum sensor area

[1] Notice that in this discussion we exclude conditions related to the scene, e. g. motion blur or atmospheric turbulence, that may also affect the overall optical resolution.

is bounded. Notice that only a percentage of the pixel area is light sensitive, which is quantified by the *fill factor* [El G 05]. The incoming light is integrated over this active area to gain a digital signal. This effectively applies a low-pass filter to the acquired images and leads to a loss of detail. Unfortunately, decreasing the fill factor would result in fewer photons being collected on each pixel. This attenuates the pixel sensitivity and causes an increase of shot noise [Chen 00]. For this reason, simply reducing the fill factor does not necessarily contribute to a higher optical resolution as noise can be seen as another resolution limiting property [Dekk 97].

1.2 Super-Resolution in this Work

Over the past decades, a variety of super-resolution techniques emerged in different scientific disciplines. Common to all of these approaches is the goal to enhance the optical resolution of an imaging system by engineering the resolution restricting aspects discussed in Section 1.1. To the best of our knowledge, there is no clear taxonomy regarding the meaning of super-resolution and the techniques may fundamentally differ [Drig 05]. This work distinguishes between *instrumental* and *computational* super-resolution according to Lindberg [Lind 12].

The instrumental approach, also known as optical super-resolution, is focused on an engineering of the optical PSF in order to increase the band-limitation of the system. Methods that are related to this class have been widely studied in physics and optical engineering and include, among others, the use of photoswitchable proteins in microscopy [Hofm 05] or superlenses [Zhan 08]. These techniques aim at breaking the diffraction limit as a resolution limiting property. However, it is in the very nature of the instrumental approach that it requires modifications on the underlying hardware, which is beyond the scope of this work.

Computational super-resolution [Bert 03] is a complementary approach and features resolution enhancement by means of software – without considerable effort regarding hardware modifications. This methodology is well-suited for low-cost imaging or workflows that do not allow changes on the system hardware. This area can be further divided into two domains: *diffractive* and *geometrical* approaches [Zale 10]. On the one hand, diffractive super-resolution aims at overcoming the diffraction barrier related to the optical system retrospectively [Garc 06, Zale 13, Ilov 14]. Geometrical super-resolution on the other hand has the goal to circumvent the limitations related to the sensor. One approach is to address the active pixel area as a resolution limiting factor [Bork 09, Bork 11]. This has the goal to alleviate the low-pass effect caused by spatial sampling with pixels of finite size. In contrast to these methods, the focus of this thesis lies in computational techniques that consider the pixel pitch as the limiting property. These methods have the objective to reconstruct images at finer pixel sampling from one or an entire sequence of undersampled images and have been widely investigated in image processing [Mila 10]. Undersampling refers to the fact that raw images are sampled below the Nyquist-Shannon frequency [Shan 48] and are therefore affected by aliasing. This improvement of the pixel sampling is due to redundancies or complementary information encoded in low-resolution images. As the primary goal in this area lies in an enhancement of the pixel sampling, we use the pixel resolution as a synonym

(a) Low-resolution frames **(b)** Super-resolved image

Figure 1.2: Example of multi-frame super-resolution by exploiting subpixel motion across a set of low-resolution frames. (a) Sequence of low-resolution frames. (b) Super-resolved image (4× magnification) gained from 17 frames using the method proposed in Chapter 4.

for the overall optical resolution. For the sake of brevity in the remainder of this thesis, *resolution* simply refers to the pixel resolution of the imaging system. Accordingly, super-resolution denotes the process of reconstructing high-resolution images from low-resolution ones by enhancing the pixel sampling.

This thesis further distinguishes super-resolution according to the type of information that is exploited for resolution enhancement. *Multi-frame* super-resolution obtains one or a set of high-resolution images from a sequence of low-resolution frames by using complementary information across the input frames. This can be achieved by utilizing relative motion [Park 03] as depicted in Fig. 1.2 or more seldom by defocusing across the frames [Raja 03]. The algorithms studied in this work mostly fall into the first category. *Single-image* super-resolution or upsampling recovers a high-resolution image from a single low-resolution one. This can be considered a special case of multi-frame super-resolution but the resulting reconstruction problem is highly underdetermined. State-of-the-art methods in this area are learning based [Yang 10, Free 00, Dong 14] or make use of redundancies within a single image [Glas 09]. These approaches need to be differentiated from the closely related *deconvolution* methods [Patr 16]. The goal of image deconvolution is to remove blur caused by diffraction, atmospheric turbulence, or motion but the pixel resolution of deblurred images remains the same.

1.3 Scientific Contributions

The major contributions of this thesis concern the theory and the development of computational methods for multi-frame super-resolution. In addition, new applications in different domains of digital optical imaging including workflows in healthcare are studied. Let us outline these contributions that cover three parts.

Numerical Methods for Multi-Frame Super-Resolution. The first contribution concerns the development of general-purpose techniques for multi-frame super-resolution. This part puts the emphasis on the design of robust numerical algo-

rithms that are well-suited under challenging conditions in real-world imaging systems, where super-resolution is prone to failure.

In the field of robust numerical algorithms, we propose a novel optimization method based on *space variant Bayesian modeling* of super-resolution. This formulation includes a *confidence-aware observation model* that considers *space variant noise and outliers* in the image formation process. Furthermore, we follow up on recent advances in the theory of compressed sensing [Cand 08] and introduce a *spatially adaptive prior distribution* to exploit *sparsity* of natural images in the gradient domain as prior knowledge for super-resolution. The numerical optimization under this model leads to an *iteratively re-weighted minimization* scheme, which facilitates the simultaneous reconstruction of high-resolution images along with the inference of latent model parameters. This approach can handle super-resolution for intensity images or 3-D range data under challenging conditions like inaccurate motion estimation, photometric variations, or space variant noise.

These methods have been originally published in a journal article [Kohl 16b].

Multi-Sensor Super-Resolution for Hybrid Imaging. The second contribution concerns the development of super-resolution methods for *hybrid imaging*. For this domain, we introduce novel *multi-sensor* super-resolution algorithms that are applicable to various imaging setups. These algorithms exploit the existence of a set of imaging modalities in contrast to conventional methods that consider only a single one. Overall, we study two problem statements:

- First, we investigate super-resolution of one imaging modality under the *guidance of a complementary modality*. To this end, we introduce a novel framework that exploits high-resolution *guidance* data to steer super-resolution on low-resolution *input* data. This comprises guidance data driven motion estimation, spatially adaptive regularization, as well as outlier detection. The merit of this formulation over conventional super-resolution is demonstrated in hybrid 3-D range imaging, where high-resolution color images are utilized to reliably super-resolve low-resolution range data.

  This methodology has been originally published in two conference proceedings [Kohl 13b, Kohl 14b] and one journal article [Kohl 15b].

- Second, we examine multi-sensor super-resolution for an arbitrary number of modalities and in the absence of reliable guidance data. We introduce a novel Bayesian model based on *linear regressions across the channels of multi-channel images* that represent the involved modalities. Furthermore, an energy minimization algorithm for the *joint reconstruction* of the different channels and latent hyperparameters of this model is presented. This method is applicable to many target applications, including color-, multispectral-, and hybrid range imaging. The proposed multi-channel reconstruction algorithm exploits mutual dependencies between different channels as a strong prior to boost the accuracy of super-resolution.

  This methodology has been originally published in two conference proceedings [Ghes 14, Kohl 15c].

Super-Resolution in Medical Imaging. The third contribution concerns the transfer of super-resolution algorithms to several fields in medical imaging with the goal to enhance medical workflows. The following applications are covered:

- In terms of diagnostic imaging, we study super-resolution for a non-invasive examination of the human eye background by means of retinal fundus images. In contrast to single-shot photography, we propose video imaging and use a tailored approach to reconstruct *high-resolution fundus images from low-resolution video frames*. This method utilizes natural eye motion during an examination as a cue for super-resolution. Furthermore, we introduce a *fully automatic noise and sharpness measure for fundus images* to steer the selection of latent model hyperparameters. The proposed framework enables high-resolution fundus imaging using mobile and cost-effective video hardware.

  These applications have been originally published in three conference proceedings [Kohl 13a, Kohl 14a, Kohl 16a].

- In the field of interventional workflows, we investigate super-resolution to facilitate image-guided surgery based on hybrid range imaging. We adopt multi-sensor super-resolution to *hybrid range imaging in 3-D endoscopy and image guided open surgery* that are studied as example applications. For this purpose, we present two *system calibration schemes* that are tailored for these applications in order to enable sensor data fusion of low-resolution range data and high-resolution color images. The proposed method enables high-resolution 3-D range measurements using current Time-of-Flight (ToF) sensors, which is studied in *ex-vivo* experiments for minimally invasive and open surgery procedures.

  These applications have been originally published in two conference proceedings [Kohl 13b, Kohl 14b] and one journal article [Kohl 15b].

To foster reproducible research and future work of other groups, source code of the developed algorithms have been made publicly available in a *multi-frame super-resolution toolbox* for MATLAB[2]. This work also led to the publication of the *Super-Resolution Erlangen* (SupER) benchmark [Kohl 17] – a comparative experimental validation of current super-resolution algorithms on a novel image database.

In addition to these research results, this work also contributed to related areas including multi-frame denoising [Kohl 12, Schi 17], blind deconvolution [Kohl 15d], joint image registration and super-resolution [Berc 16], or hardware acceleration of super-resolution [Wetz 13]. This thesis also contributed to research in medical image analysis including image-based tracking [Kurt 14], super-resolved segmentation [Haas 13c], and computer-assisted diagnostics [Kohl 15a].

1.4 Outline of this Thesis

This thesis is structured in an introductory part that covers the background of multi-frame super-resolution as well as three main parts as shown in Fig. 1.3.

[2] https://www5.cs.fau.de/research/software/multi-frame-super-resolution-toolbox/

Chapter 2 presents an in-depth analysis of super-resolution in the frequency domain. This includes a study of *single- and multi-channel sampling* along with a discussion of the *sampling theorem* according to Nyquist and Shannon. The reconstruction of well sampled signals from undersampled ones as the core of super-resolution is modeled as an inverse problem based on the multi-channel theory. A mathematical discussion regarding the *uniqueness* of this reconstruction and the *effective magnification factor* obtainable by super-resolution is presented.

The main body is divided into three parts in accordance with the contributions outlined in Section 1.3. Part I covers general-purpose methods for multi-frame super-resolution. In Chapter 3, we introduce the *computational framework for multi-frame super-resolution* that is widely employed in the algorithms presented in the remainder of this work. This chapter concerns the fundamentals of state-of-the-art algorithms ranging from the mathematical modeling of the image formation process to Bayesian methods that formulate super-resolution as a statistical parameter estimation problem. Subsequently, Chapter 4 introduces *robust super-resolution with sparse regularization* that extends this framework by space variant observation and prior distributions. In this chapter, we present *iteratively re-weighted minimization* for simultaneous super-resolution and model parameter estimation.

Part II covers super-resolution for hybrid imaging and comprises the two multi-sensor techniques developed in this work. In Chapter 5, we introduce *multi-sensor super-resolution using guidance images*. This framework augments conventional super-resolution (Chapter 3) by motion estimation, adaptive regularization, as well as outlier detection techniques that are steered by guidance data to leverage the reconstruction of a complementary modality. Subsequently, Chapter 6 introduces *multi-sensor super-resolution using locally linear regression* to jointly super-resolve a set of imaging modalities without explicitly using one of them as a guidance. This provides a generalization of *guidance image based super-resolution* (Chapter 5) and employs *iteratively re-weighted minimization* (Chapter 4) for numerical optimization. We present several potential applications of both methodologies ranging from color and multispectral imaging to hybrid 3-D range imaging.

Part III is focused on novel applications of super-resolution to facilitate diagnostic and interventional medical imaging. In Chapter 7, we address *super-resolution for retinal fundus video imaging*. This chapter is concerned with a new method to reconstruct high-resolution fundus images from multiple low-resolution video frames by exploiting natural human eye movements. This method is presented in the field of low-cost imaging to gain fundus images at high spatial resolution by means of low-priced and mobile video camera systems. Chapter 8 covers *super-resolution for image-guided surgery* using hybrid range imaging. This chapter adopts the previously introduced *guidance image based super-resolution* (Chapter 5) and presents the system calibration schemes to make it accessible for the desired application. The proposed framework is presented in the context of hybrid 3-D endoscopy and image-guided open surgery.

In Chapter 9, we draw a conclusion and summarize the main findings of this thesis. Finally, Chapter 10 provides an outlook regarding promising directions for future research.

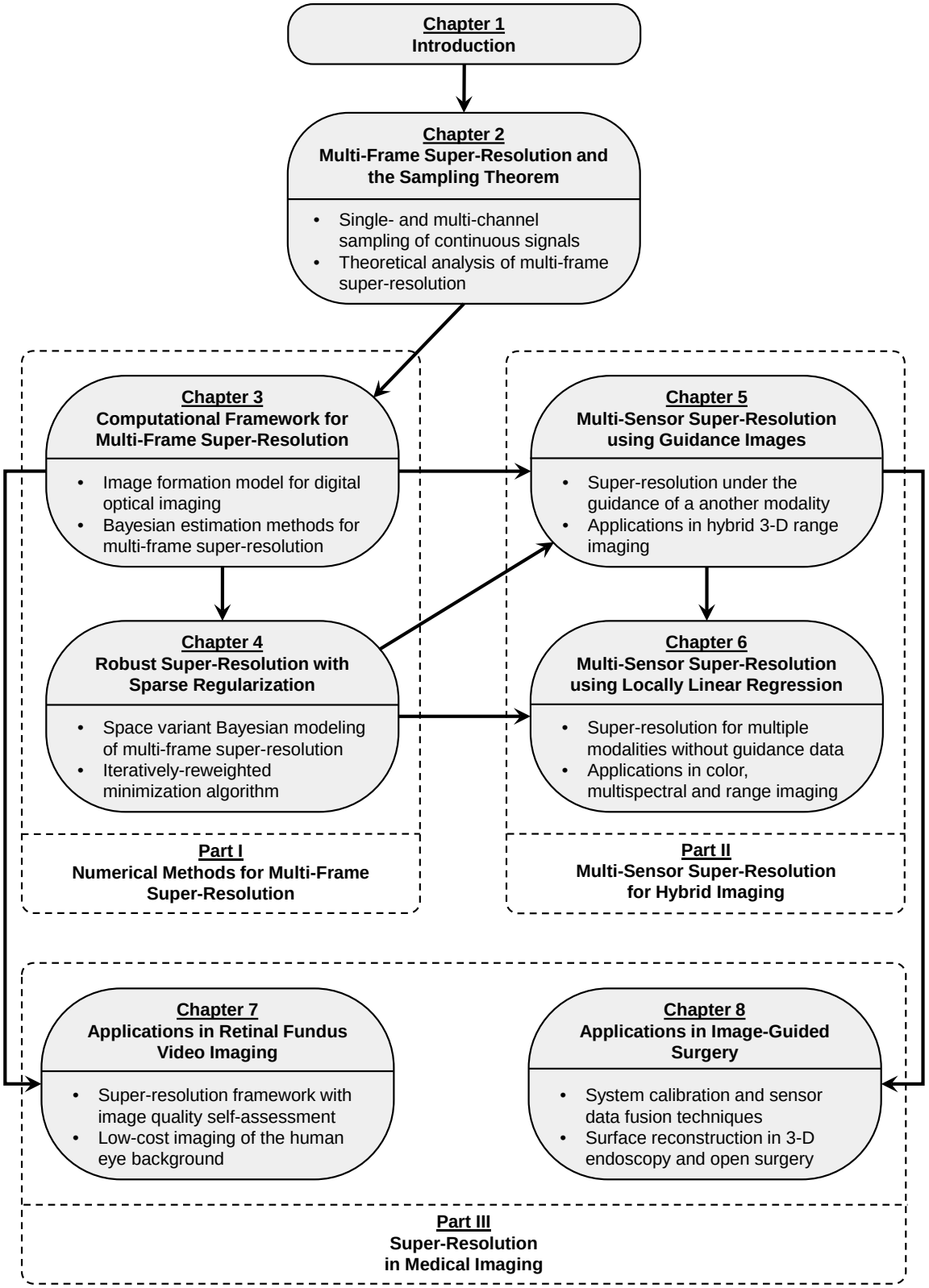


Figure 1.3: Structure of this thesis and relationship among the individual chapters. The background part provides a theoretical discussion of the relationship between super-resolution and the Nyquist-Shannon sampling theorem. The main body covers numerical methods for multi-frame super-resolution algorithms (Part I), multi-sensor super-resolution for hybrid imaging (Part II), as well as applications of super-resolutions in medical imaging (Part III).

CHAPTER 2

Multi-Frame Super-Resolution and the Sampling Theorem

This chapter is devoted to the relationship between the Nyquist-Shannon sampling theorem [Shan 48] as fundamental principle underlying the acquisition of digital signals and super-resolution reconstruction. In order to formulate this theoretical framework, single-channel sampling is generalized to the multi-channel case, where a continuous signal is sampled multiple times to capture a set of discrete signals. Based on a Fourier domain analysis, signal reconstruction to recover the original, continuous signal from multiple sampled channels is formulated as a linear inverse problem. It is shown that a solution of this linear problem yields a super-resolved signal to overcome the constraints stated by the sampling theorem. Finally, inherent limitations of super-resolution regarding the maximum magnification and the uniqueness of the reconstruction are derived.

The analysis of multi-channel sampling presented in this chapter is based on the pioneering work on image super-resolution algorithms [Tsai 84, Kim 90, Teka 92] formulated in the frequency domain. A similar analysis is also presented in the work of Vandewalle [Vand 06a, Vand 07], where the more general concept of finite dimensional Hilbert spaces is used as mathematical tool.

2.1 Introduction

In digital imaging, a continuous description of the real world, i. e. geometry or texture of objects, is discretized to provide a digital representation. In terms of signal processing, one major parameter of such a system is how this sampling is performed and whether the resulting image is sampled appropriately. In case of a digital camera, the number of pixels on a sensor array as well as the spacing between the pixels are relevant system parameters related to the sampling process,

see Section 1.1. The Nyquist-Shannon sampling theorem [Shan 48] states inherent requirements regarding an appropriate sampling in order to capture a digital representation without loss of information. These requirements concern the sampling frequency as well as the spectral properties of the continuous signal. Violating the sampling theorem such that the sampling frequency is chosen too small with regard to the signal's spectral properties leads to *aliasing*. In case of aliasing, low frequency components of the continuous signal are superimposed by its high frequency components resulting in signal distortions.

In presence of aliasing, a further analysis of the sampled signal is prone to errors and restoration techniques are required to overcome undersampling. In this context, super-resolution aims at reconstructing a digital signal that is free of aliasing artifacts from an undersampled signal. This methodology is based on the concept of *multi-channel* sampling, where a continuous signal is sampled multiple times as opposed to classical *single-channel* sampling. Super-resolution can be seen as a fusion of multiple channels in order to overcome the limitations stated by the sampling theorem for a single channel. This is feasible by exploiting complementary information across the channels. In case of digital imaging, these channels correspond to different frames of an image sequence taken from the same scene, whereas each frame contains a complementary view.

The remainder of this chapter is organized as follows. Section 2.2 covers the fundamentals of sampling under ideal and real conditions along with the Nyquist-Shannon theorem. In Section 2.3, we extend single-channel sampling to the multi-channel case. Accordingly, super-resolution is formulated as an inverse linear problem based on the multi-channel sampling theory. Section 2.4 presents a numerical algorithm to solve this inverse problem in the frequency domain. Section 2.5 covers fundamental properties of super-resolution regarding the effective magnification factor and the uniqueness of the signal reconstruction. Finally, Section 2.6 provides a summary and a conclusion of these concepts.

2.2 Single-Channel Sampling Theory

The sampling of a continuous signal is first modeled for single channels, where one set of discrete samples is obtained from the original signal [Mall 99]. For convenience, but without loss of generality, this process is modeled for one-dimensional signals only. As the different dimensions of multidimensional signals are separable, the underlying theory can be extended and applied to each individual dimension. Since common physical measurements such as digital images are real-valued, we limited the following analysis to real-valued signals.

Let $x : \mathbb{R} \to \mathbb{R}$ be a real-valued, continuous signal denoted by $x(t)$. For discretization in the domain $t \in \mathbb{R}$, $x(t)$ is sampled in equidistant steps with *sampling pitch* T. The sampled signal defined as continuous function is denoted by $y(t) = \mathcal{D}_T\{x(t)\}$, where $\mathcal{D}_T\{\cdot\}$ denotes the sampling operator. Then, the discretization $y[n]$ associated with $x(t)$ is obtained according to $y[n] := x(nT)$, where $n \in \mathbb{Z}$ denotes the sample index. This process is examined for two different situations including *ideal sampling* as well as *real sampling* as a reasonable model in the context of digital imaging.

2.2.1 Ideal Single-Channel Sampling

In case of ideal sampling, the sampling operator is modeled by Dirac delta impulses. Then, $y(t)$ is obtained from a product of the continuous signal $x(t)$ and a Dirac comb [Mall 99] as depicted in Fig. 2.1. More formally, single-channel sampling is modeled by:

$$y(t) = \mathcal{D}_T\{x(t)\} := \sum_{m=-\infty}^{\infty} x(t)\delta(t - mT), \tag{2.1}$$

where $m \in \mathbb{Z}$ and the discrete Dirac delta is defined as:

$$\delta(t) := \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{otherwise} \end{cases}. \tag{2.2}$$

The resulting signal $y(t)$ is continuous and represents the discrete values of $y[n]$ for $t = nT$. In order to present the sampling theorem, we model the sampling process in the frequency domain. Let $X(f) = \mathcal{F}\{x(t)\}$ be the continuous Fourier transform (CFT) [Rahm 11] of the signal $x(t)$ defined as:

$$X(f) = \mathcal{F}\{x(t)\} := \int_{-\infty}^{\infty} x(t) \cdot \exp(-j2\pi ft)dt, \tag{2.3}$$

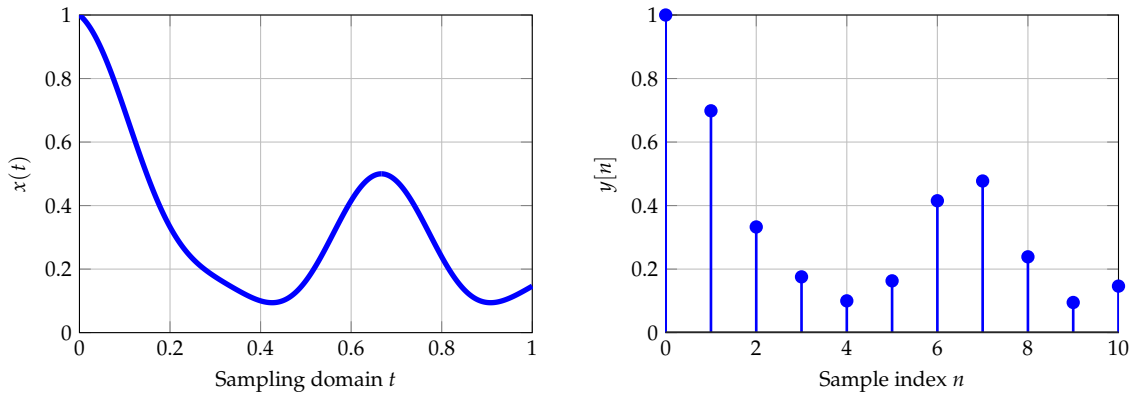
where j is the imaginary unit. Using the Fourier transform, linearity, and the convolution theorem, Eq. (2.1) can be written in the frequency domain according to:

$$\begin{aligned} Y(f) &= \mathcal{F}\left\{ \sum_{m=-\infty}^{\infty} x(t)\delta(t - mT) \right\} \\ &= \sum_{m=-\infty}^{\infty} \mathcal{F}\{x(t)\delta(t - mT)\} \\ &= \sum_{m=-\infty}^{\infty} \mathcal{F}\{x(t)\} \star \mathcal{F}\{\delta(t - mT)\} = \frac{1}{T} \sum_{m=-\infty}^{\infty} X(f) \star \Delta\left(f - \frac{m}{T}\right) \end{aligned} \tag{2.4}$$

where $\star$ denotes the convolution, $Y(f)$ denotes the CFT associated with the sampled signal $y(t)$, and $\Delta(f)$ is the CFT of the Dirac delta $\delta(t)$ [Mall 99]. According to the definition of the sampling pitch, we define the *sampling frequency* $f_s = \frac{1}{T}$. Thus, ideal sampling can be described as:

$$\begin{aligned} Y(f) &= \sum_{m=-\infty}^{\infty} X(f) \star f_s\Delta(f - mf_s) \\ &= f_s \sum_{m=-\infty}^{\infty} X(f - mf_s). \end{aligned} \tag{2.5}$$

Accordingly, the sampling of $x(t)$ by the frequency f_s corresponds to a periodic summation of $X(f)$, where the periodic length is given by the sampling frequency f_s. One key question is under which conditions the continuous signal $x(t)$ can be fully characterized by its discrete samples $y[n]$ without loss of information. The possibility of this reconstruction depends on the Fourier properties of $x(t)$ and is studied for *band-limited* signals that are defined as follows.



(a) Continuous signal **(b)** Discrete samples

Figure 2.1: Ideal sampling to obtain discrete samples $y[n]$ from a continuous signal $x(t)$.

Definition 2.1 (Band-limited signal)**.** *A continuous signal $x(t)$ is band-limited if there exists a cut-off frequency f_0 such that $X(f) = \mathcal{F}\{x(t)\}$ fulfills $X(f) = 0$ for $|f| \geq f_0$.*

Let us assume that $x(t)$ is band-limited. Depending on the sampling frequency f_s, three situations for the sampling process can be distinguished [Vand 10], see Fig. 2.2. In order to verify if $x(t)$ can be fully reconstructed from $y[n]$, the properties of the periodic summation in Eq. (2.5) are analyzed.

Nyquist Sampling. If the sampling frequency f_s is chosen according to $f_s \geq 2f_0$, adjacent parts of the spectrum $Y(f)$ calculated by the periodic summation of $X(f)$ in Eq. (2.5) are non-overlapping. For this situation that is depicted in Fig. 2.2b and termed as *Nyquist sampling*, it is feasible to recover $x(t)$ from its discrete samples $y[n]$. This fact is explained by the Nyquist-Shannon sampling theorem for band-limited signals [Shan 48].

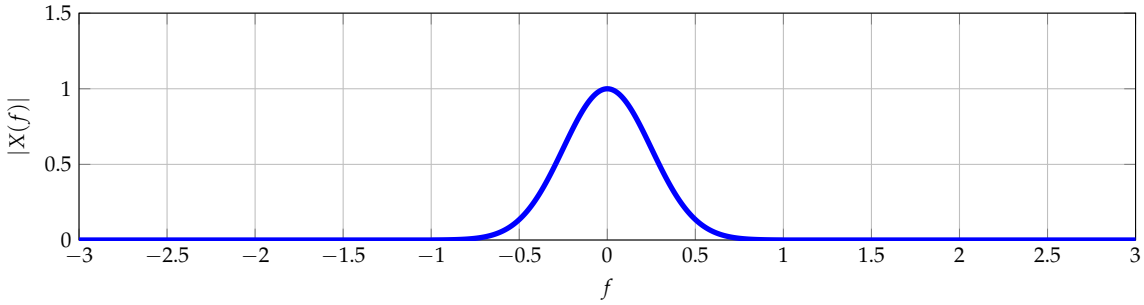
Theorem 2.1 (Nyquist-Shannon sampling theorem)**.** *Let $x(t)$ be a band-limited continuous signal with cut-off frequency f_0. Then, $x(t)$ is completely described by its discrete samples $y[n]$ obtained by the sampling frequency f_s if $f_s \geq 2f_0$.*

In case of Nyquist sampling, the signal $x(t)$ can be reconstructed from $y[n]$ using low-pass filtering to remove the periodic parts of $Y(f)$. We can obtain $x(t)$ based on the inverse CFT $\mathcal{F}^{-1}\{\cdot\}$ according to:

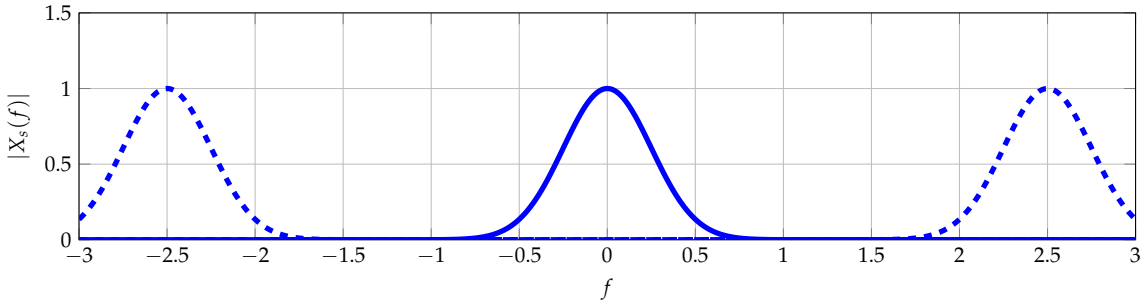
$$x(t) = \mathcal{F}^{-1}\Big\{Y(f) \cdot H_{\text{reco}}(f)\Big\}, \tag{2.6}$$

where $H_{\text{reco}}(f)$ denotes the CFT of a reconstruction low-pass filter to suppress the periodic parts. For example, one can employ the ideal low-pass filter:

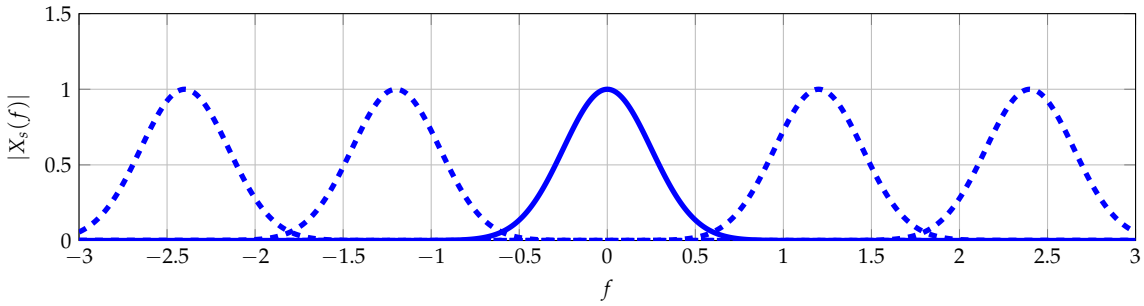
$$H_{\text{reco}}(f) = \begin{cases} 1 & \text{if } |f| \leq f_0 \\ 0 & \text{otherwise} \end{cases}. \tag{2.7}$$



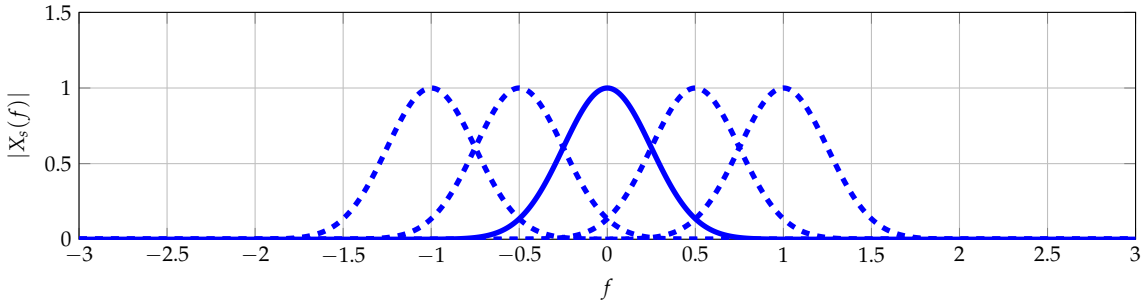
(a) CFT of the continuous signal $x(t)$



(b) CFT of the sampled signal $y(t)$ under Nyquist sampling ($f_s \geq 2f_0$)

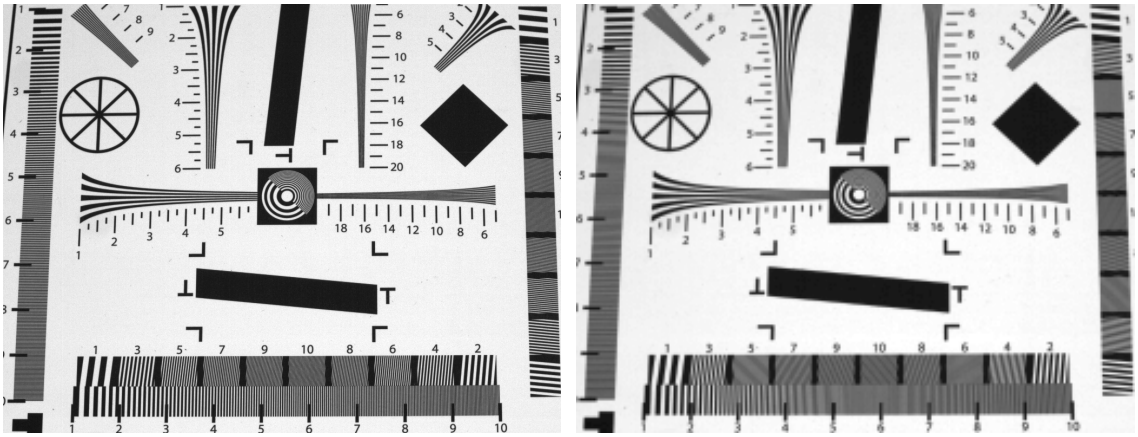


(c) CFT of the sampled signal $y(t)$ for undersampling with partial aliasing ($f_0 \leq f_s < 2f_0$)



(d) CFT of the sampled signal $y(t)$ for undersampling with total aliasing ($f_s < f_0$)

Figure 2.2: The sampling of the continuous, band-limited signal $x(t)$ with frequency f_s corresponds to a periodic summation of the CFT $X(f)$. For the sake of visualization, the frequencies f are normalized w. r. t. the band-limitation f_0 of $x(t)$. According of [Vand 10], three situations for the sampling process can be distinguished. Depending on f_s, the samples $y[n]$ are acquired at the Nyquist rate, partial aliased or total aliased.



(a) Full resolution **(b)** 4×4 hardware binning

Figure 2.3: Aliasing in digital imaging depicted on the ISO 12233:2000 resolution chart [ISO 00]. (a) Acquisition of the resolution chart at the full pixel resolution (2048×1088 px) of a Basler acA2000-50gm camera. (b) Acquisition of the same chart with 4×4 hardware binning relative to the full pixel resolution. Notice that structures with high spatial frequencies relative to the sampling frequency, e. g. line pairs with small spacings, are distorted by a Moiré pattern caused by undersampling.

Undersampling with Partial Aliasing. While $x(t)$ can be fully reconstructed if the sampling theorem holds true, this is no longer feasible in case of undersampling. Let us assume that $f_0 < f_s < 2f_0$. Then, in the periodic summation in Eq. (2.5), parts of $X(f)$ are superimposed as depicted in Fig. 2.2c. In this situation, the samples $y[n]$ are considered as *partial aliased*.

In digital imaging, a violation of the sampling theorem leads to aliasing artifacts that are visible as Moiré pattern. This effect is visualized on the ISO 12233:2000 resolution chart [ISO 00] in Fig. 2.3 that was acquired with a Basler acA2000-50gm complementary metaloxide semiconductor (CMOS) camera[1]. The resolution chart was captured at the full pixel resolution of the camera (Fig. 2.3a) as well as at a reduced resolution using 4×4 hardware binning [Kohl 17] of pixels on the sensor array (Fig. 2.3b). Here, line pairs that have small spacings compared to the sampling frequency are distorted by aliasing.

Notice that a reconstruction of $x(t)$ using low-pass filtering would be distorted due to undersampling if aliasing is not considered. One can remove aliasing in the design of the reconstruction low-pass filter $\tilde{H}_{\text{reco}}(f)$:

$$\tilde{H}_{\text{reco}}(f) = \begin{cases} 1 & \text{if } |f| \leq f_s - f_0 \\ 0 & \text{otherwise} \end{cases}. \tag{2.8}$$

Unfortunately, removing the signal distortions caused by undersampling, one also loses the high-frequency content present in the original signal $x(t)$. Another strategy to overcome aliasing is an artificial increase of the sampling frequency by means of super-resolution reconstruction as discussed below.

[1] http://www.baslerweb.com/en

Undersampling with Total Aliasing. A situation that is even more severe appears if the sampling frequency is chosen as $f_s < f_0$. In this case, the interference in the spectrum $Y(f)$ results in a *total aliased* discretization $y[n]$ as depicted in Fig. 2.2d. As opposed to the aforementioned partial aliasing, all parts of the spectrum $Y(f)$ and thus the entire signal $y[n]$ are distorted. Notice that a simple reconstruction of the continuous signal $x(t)$ by a removal of the aliasing artifacts using low-pass filtering is no longer possible in this situation. However, similar as for partial aliased signals, we will show how super-resolution can be utilized to perform a reconstruction of $x(t)$.

2.2.2 Real Single-Channel Sampling

Up to now, the sampling process was considered to be ideal, such that a Dirac delta can be used to model the sampling operator $\mathcal{D}_T$. Theoretically, this would result in an arbitrarily high resolution as long as the sampling frequency is chosen such that the sampling theorem is fulfilled. However, this simplistic assumption is never feasible in practice. In case of real sampling, the sampled signal is a blurred version of the original one due to the fact that the impulse response of the acquisition device deviates from the Dirac delta, see Section 1.1. Recently, non-ideal sampling models gained attention for signal reconstruction. This results in a deconvolution problem rather than interpolation as in the case of ideal sampling [Elda 06, Rama 08, Guev 10]. Modeling of real sampling becomes important for imaging systems, where diffraction, manufacturing uncertainties of lenses, and the summation of light photons on a finite pixel area of the sensor array restrict an ideal sampling and introduce blur. An illustration of this issue is depicted for the ISO 12233:2000 resolution chart in Fig. 2.4. In this example, steep edges on the resolution chart appear blurred in digital images.

Mathematically, blur in the sampling model is considered by replacing the Dirac delta in Eq. (2.1) by a blur kernel $h(t, t_0)$. Using a linear blur model, the sampled signal $y(t)$ is given by:

$$y(t) = \sum_{m=-\infty}^{\infty} x(t) h(t, mT). \tag{2.9}$$

This models the blur at position t, whereas the kernel is evaluated at the sample positions mT. If the blur kernel is assumed to be linear shift invariant (LSI), i. e. $h(t, mT) = h(t - mT)$, the amount of blurring depends only on the distance $t - mT$. Then, the sampling process can be described according to:

$$\begin{aligned} y(t) &= \sum_{m=-\infty}^{\infty} \underbrace{\left(x(t) \star h(t)\right)}_{z(t)} \delta(t - mT) \\ &= \mathcal{D}_T\{x(t) \star h(t)\}, \end{aligned} \tag{2.10}$$

Thus, the sampling process can be described in two steps. First, a filtered version $z(t)$ of the continuous signal $x(t)$ according to the underlying kernel $h(t)$ is determined. Since $h(t)$ is the impulse response of a low-pass filter, the convolution to

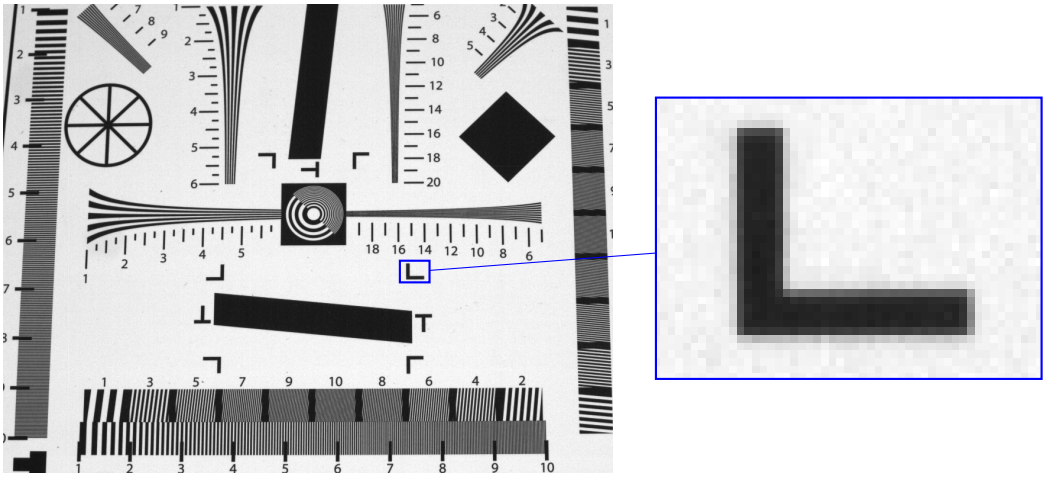


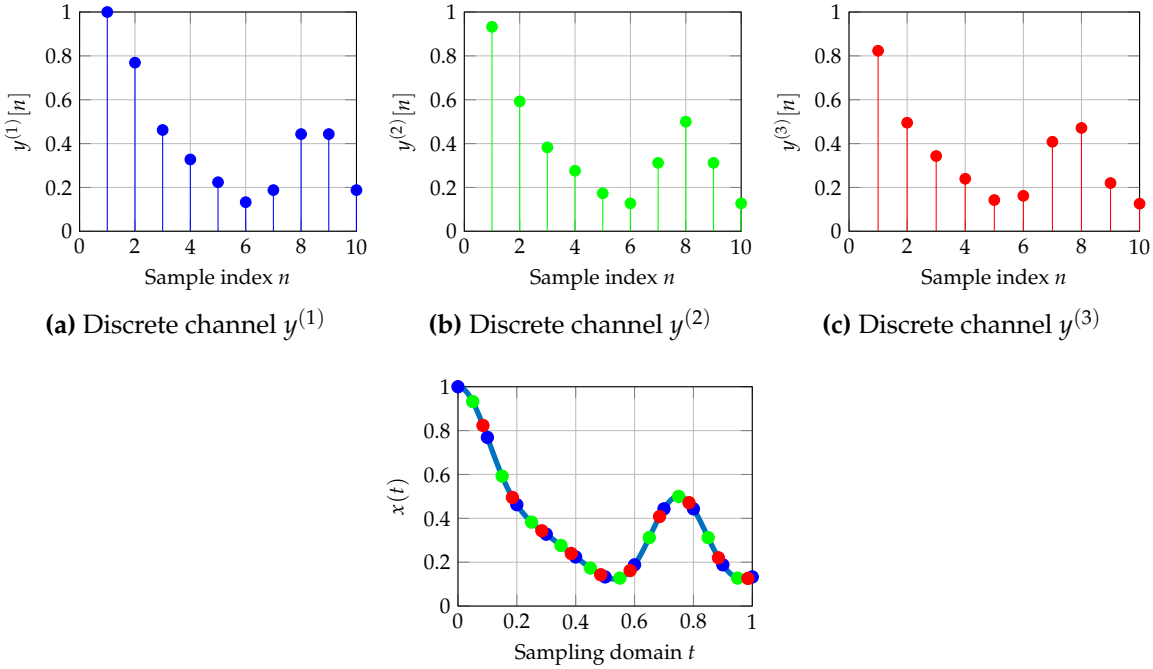
Figure 2.4: Illustration of sampling in digital imaging using the ISO 12233:2000 resolution chart [ISO 00]. Steep edges on the resolution chart are blurred in digital images due to diffraction, non-ideal lenses, and the finite pixel size of the camera.

determine $z(t)$ corresponds to an averaging of $x(t)$ over the support of the kernel, which explains the blurring of $x(t)$. Finally, to obtain a sampled signal $y(t)$ and thus the discretization $y[n]$, $z(t)$ is sampled with ideal Dirac deltas.

2.3 Multi-Channel Sampling Theory

In the previous section, the analysis was restricted to single-channel sampling, where the continuous signal $x(t)$ is sampled once to acquire a discrete signal $y[n]$. One situation interesting for the development of super-resolution algorithms is the case of multi-channel sampling [Papo 77, Unse 97], where $x(t)$ is sampled multiple times. This appears in digital imaging, where multiple images of a scene can be captured by moving the camera to different viewpoints while acquiring video data. This section presents multi-channel sampling based on the formulation of Tsai and Huang [Tsai 84]. In addition to the theoretical insights, this introduces the historically first multi-frame super-resolution algorithm proposed in [Tsai 84].

Multi-channel sampling of the continuous signal $x(t)$ can be modeled by a sequence of complementary channels determined from $x(t)$. If the sampling of these channels is not synchronized, the k-th channel $x^{(k)}(t)$, $k = 1, \ldots, K$ can be defined as a shifted version of $x(t)$ according to $x^{(k)}(t) = x(t - t_k)$, where $t_k \in \mathbb{R}$ denotes the *channel offset*. Without loss of generality, the first channel is defined as reference, i. e. $t_1 = 0$. The sampled signals obtained from the different channels are denoted as $y^{(k)}(t)$ and the corresponding discrete samples are given by $y^{(k)}[n] := y^{(k)}(nT_k)$, where T_k denotes the sampling pitch of the k-th channel. For the sake of notational brevity, let us assume that the sampling pitch is fixed, i. e. $T_k = T$ for all channels. The samples $y^{(k)}[n]$ contain complementary information about the underlying continuous signal $x(t)$. Consequently, $x(t)$ is sampled by a non-uniform scheme parametrized by the sampling frequency and the channel offsets as illustrated in Fig. 2.5. We analyze this process for two situations.



(a) Discrete channel $y^{(1)}$ **(b)** Discrete channel $y^{(2)}$ **(c)** Discrete channel $y^{(3)}$

(d) Continuous signal $x(t)$

Figure 2.5: Illustration of multi-channel sampling with constant sampling pitch for all channels. The continuous signal $x(t)$ is sampled K times ($K = 3$) according to the channel offsets t_k with $k = 1, \ldots, K$ as shown in (a) - (c). The fusion of the resulting discrete channels $y^{(k)}[n]$ leads to a non-uniform sampling in the domain t as shown in (d).

2.3.1 Ideal Multi-Channel Sampling

Let us first examine ideal multi-channel sampling, where the sampling operator $\mathcal{D}_T\{\cdot\}$ is modeled by the Dirac comb. Thus, the sampled version of the k-th channel is given by:

$$\begin{aligned} y^{(k)}(t) &= \sum_{m=-\infty}^{\infty} x^{(k)}(t)\delta(t - mT_k) \\ &= \sum_{m=-\infty}^{\infty} x(t - t_k)\delta(t - mT_k). \end{aligned} \tag{2.11}$$

Notice that the CFT $X^{(k)}(f) = \mathcal{F}\{x^{(k)}(t)\}$ for $k > 1$ is related to the CFT of the first channel $X(f) = \mathcal{F}\{x^{(1)}(t)\}$ using the shift property of the Fourier transform according to:

$$X^{(k)}(f) = \exp(-j2\pi f t_k)X(f). \tag{2.12}$$

In order to derive a relationship between $y^{(k)}[n]$ and $x(t)$, it is assumed that all discrete channels are defined by a finite number of N samples acquired over a finite interval. Then, the discrete samples $y^{(k)}[n]$ are represented by N complex-valued frequency coefficients using the discrete Fourier transform (DFT) [Oppe 99]:

$$\mathcal{Y}^{(k)}[n] = \sum_{m=0}^{N-1} \exp\left(-j2\pi n \frac{m}{N}\right) y^{(k)}[m], \tag{2.13}$$

where $n = 0, \dots, N-1$. The DFT coefficients $\mathcal{Y}^{(k)}[n]$ are related to the CFT $X^{(k)}(f)$ of the k-th channel. According to the aliasing property of the Fourier transform in Eq. (2.5) and the fact that the CFT and the associated DFT of a real-valued signal $x(t)$ are symmetric [Cool 69], this relationship is given by:

$$\mathcal{Y}^{(k)}[n] = f_s \sum_{m=-\infty}^{\infty} X^{(k)}\left(\frac{n}{N} f_s - m f_s\right). \tag{2.14}$$

Since the continuous signal $x(t)$ is assumed to be band-limited at a certain cut-off frequency, we have $X(f) = 0$ for $|f| \geq L f_s$ with a finite integer L. Hence, the aliasing property can be formulated for a finite interval in the Fourier domain instead of considering the infinite summation in Eq. (2.14). Moreover, using the Fourier shift theorem in Eq. (2.12), the aliasing property is formulated as:

$$\begin{aligned} \mathcal{Y}^{(k)}[n] &= f_s \sum_{m=-L}^{L-1} X^{(k)}\left(\frac{n}{N} f_s - m f_s\right) \\ &= f_s \sum_{m=-L}^{L-1} \exp\left(-j2\pi \left(\frac{n}{N} f_s - m f_s\right) t_k\right) X\left(\frac{n}{N} f_s - m f_s\right). \end{aligned} \tag{2.15}$$

Following the derivation of Kim et al. [Kim 90], this condition is written in terms of the linear system:

$$\underbrace{\begin{pmatrix} \mathcal{Y}^{(k)}[1] \\ \mathcal{Y}^{(k)}[2] \\ \vdots \\ \mathcal{Y}^{(k)}[N] \end{pmatrix}}_{\boldsymbol{\mathcal{Y}}^{(k)}} = \underbrace{\begin{pmatrix} \boldsymbol{w}_1^{(k)} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{w}_2^{(k)} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{w}_N^{(k)} \end{pmatrix}}_{\boldsymbol{W}^{(k)}} \underbrace{\begin{pmatrix} X_{1,-L} \\ X_{1,-L+1} \\ \vdots \\ X_{N,L-1} \end{pmatrix}}_{\boldsymbol{X}}, \tag{2.16}$$

where $\boldsymbol{\mathcal{Y}}^{(k)} \in \mathbb{C}^N$ comprises the DFT coefficients of the k-th channel, $\boldsymbol{X} \in \mathbb{C}^{2LN}$ comprises the $2L$ samples of $X(f)$ abbreviated as $X_{n,m} := f_s X(n/N f_s - m f_s)$ and $\boldsymbol{w}_n^{(k)} := (W_{n,-L}^{(k)}, W_{n,-L+1}^{(k)}, \dots, W_{n,L-1}^{(k)})$ are the row vectors containing the non-zero elements of the system matrix $\boldsymbol{W}^{(k)} \in \mathbb{C}^{N \times 2LN}$. These elements can be computed according to:

$$W_{n,m}^{(k)} = \exp\left(-j2\pi \left(\frac{n}{N} f_s - m f_s\right) t_k\right). \tag{2.17}$$

The sampling process for a single channel with N samples is modeled by $\boldsymbol{W}^{(k)}$, which is fully determined by the channel offset t_k and the sampling frequency f_s. After concatenating the system matrices to $\boldsymbol{W} = (\boldsymbol{W}^{(1)}, \dots, \boldsymbol{W}^{(K)})^\top \in \mathbb{C}^{KN \times 2LN}$ and the DFT coefficients to $\boldsymbol{\mathcal{Y}} = (\boldsymbol{\mathcal{Y}}^{(1)}, \dots, \boldsymbol{\mathcal{Y}}^{(K)})^\top$, the aliasing property is given by the linear system:

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{W}\boldsymbol{X}. \tag{2.18}$$

This linear system explains how the CFT of the continuous signal $x(t)$ is related to the DFT of the sampled channels $y^{(1)}[n], \dots, y^{(K)}[n]$. Thus, it represents a generative model for multi-channel sampling in the Fourier domain. In order to reconstruct $x(t)$ from its sampled channels, the linear problem in Eq. (2.18) must

be solved w. r. t. the CFT coefficients in $\boldsymbol{X}$. Note that, since $X(f)$ and $\mathcal{Y}^{(k)}[n]$ are complex-valued in general, Eq. (2.18) provides one condition for the real part and one for the imaginary part and thus $2 \cdot NK$ equations in total. However, for real-valued signals, $X(f)$ as well as $\mathcal{Y}^{(k)}[n]$ are symmetric [Cool 69] and the linear system provides NK independent constraints. In this situation, it is sufficient to formulate the aliasing property for one half of the spectrum only.

2.3.2 Real Multi-Channel Sampling

In order to model non-ideal sampling, the blur kernel for each channel is assumed to be linear and shift invariant in accordance to the analysis of single-channel sampling in Section 2.2.2. Let us consider the general case of different blur kernels $h^{(k)}(t)$ associated with the set of channels. Then, the aliasing property introduced in Eq. (2.14) can be formulated as:

$$\mathcal{Y}^{(k)}[n] = f_s \sum_{m=-L}^{L-1} \exp\left(-j2\pi \left(\frac{n}{N} f_s - m f_s\right) t_k\right) H^{(k)}\left(\frac{n}{N} f_s - m f_s\right) X\left(\frac{n}{N} f_s - m f_s\right), \tag{2.19}$$

where $H^{(k)}(f) = \mathcal{F}\{h^{(k)}(t)\}$ is the CFT of the blur kernel associated with the k-th channel. Reformulating the aliasing property as a linear system yields the generalized version of Eq. (2.18):

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{H}\boldsymbol{X}. \tag{2.20}$$

The system matrix is assembled as $\boldsymbol{H} = (\boldsymbol{H}^{(1)}, \ldots, \boldsymbol{H}^{(K)})^\top \in \mathbb{C}^{KN \times 2LN}$ according to the block structure in Eq. (2.16) with the non-zero elements:

$$H_{n,m}^{(k)} = \exp\left(-j2\pi \left(\frac{n}{N} f_s - m f_s\right) t_k\right) H^{(k)}\left(\frac{n}{N} f_s - m f_s\right), \tag{2.21}$$

This linear system states the relationship between the CFT of a continuous signal and the DFT of the corresponding discrete channels with consideration of a blur kernel.

2.4 From Multi-Channel Sampling to Super-Resolution

The analysis of multi-channel sampling presented above yields a super-resolution algorithm in the frequency domain. This is the historically first multi-frame approach as proposed by Tsai and Huang [Tsai 84]. In this framework, the target of super-resolution is the reconstruction of the continuous signal $x(t)$ from the sampled channels $y^{(k)}[n]$. However, for practical implementation purposes, the solution of this inverse problem is restricted to the reconstruction of discrete samples $x[n]$ using an artificial sampling frequency f_s' that should be higher than the original frequency f_s. Note that if $x[n]$ is determined such that the aliasing present in $y^{(k)}[n]$ is removed and thus sampled at the Nyquist rate, the continuous signal $x(t)$ can be recovered from $x[n]$. Therefore, f_s' must be chosen such that it fulfills the Nyquist-Shannon sampling theorem.

In order to define a super-resolution algorithm, the gain in resolution provided by the algorithm needs to be quantified. For this purpose, the *magnification factor* denotes the enhancement of the sampling frequency achieved by super-resolution relative to the original sampling frequency. This parameter is defined as follows.

Definition 2.2 (Magnification factor)**.** *Let f_s be the sampling frequency that is used to obtain the discrete channels $y^{(k)}[n]$ and $f_s' \geq f_s$ be the sampling frequency provided by a super-resolution algorithm. Then, the super-resolution magnification factor is given by:*

$$s = \frac{f_s'}{f_s}. \tag{2.22}$$

In a computational super-resolution approach, sN discrete samples $x[n]$ associated with the continuous signal $x(t)$ are reconstructed from the sampled channels $y^{(k)}[n]$, $k = 1, \ldots, K$ consisting of KN samples. In the Fourier domain formulation, $x[n]$ is represented by the sN discrete frequency coefficients of $X(f)$. The key idea behind super-resolution is that the different channels $y^{(k)}[n]$ contain complementary information about $x(t)$. In summary, the main computational steps of super-resolution in the Fourier domain are as follows:

1. **Fourier transform:** For all discrete channels $y^{(k)}[n]$, $k = 1, \ldots, K$ given by NK samples, the associated DFT coefficients are determined efficiently by means of a Fast Fourier Transform (FFT) [Cool 65]. These coefficients form a complex-valued observation vector $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{KN}$.

2. **Reconstruction:** Once the offsets t_k for all channels are known, the samples of the CFT $X(f)$ reorganized in $\boldsymbol{X} \in \mathbb{C}^{2L \cdot KN}$ for $s = 2L$ are determined for the magnification factor s. This is done by solving the linear systems in Eq. (2.18) (in case of ideal sampling) or in Eq. (2.20) (in case of real sampling).

3. **Inverse Fourier transform:** Finally, the discrete samples $x[n]$ associated with the continuous signal $x(t)$ are recovered from $\boldsymbol{X}$ using an inverse FFT.

Existing frequency domain based super-resolution methods mainly differ in the implementation of the second step. In [Tsai 84], the most basic situation of ideal sampling and the absence of observation noise is considered. This is a simplistic assumption, particularly in digital imaging, where an optical system acts as a blur kernel in the sampling process and observation noise is caused by non-ideal sensors. Kim et al. [Kim 90] have proposed a generalization of the Tsai and Huang algorithm, where observation noise is taken into account. In this case, an estimate for $\boldsymbol{X}$ can be obtained by solving the linear problem given in Eq. (2.18) in a least square manner. Later, Tekalp et al. [Teka 92] have introduced a frequency domain approach that considers observation noise as well as a blur kernel involved in the sampling process. Super-resolution is then based on the more general linear problem given in Eq. (2.20).

2.5 Limits of Super-Resolution

This section analyzes the linear problems in Eq. (2.18) and Eq. (2.20) in case of ideal and non-ideal sampling, respectively. The properties of these relationships between discrete channels and the underlying continuous signal indicate if super-resolution is applicable and how well the behavior of an algorithm can be.

The following analysis covers two aspects. First, it is examined under which conditions super-resolution is profitable in order to reconstruct an aliasing-free signal from undersampled ones. This states that under certain conditions super-resolution cannot provide resolution enhancement beyond a given limit. In particular, upper bounds regarding the effective magnification factor that depends on the sampling parameters are derived. Second, necessary and sufficient conditions for the existence of unique solutions for Eq. (2.18) and Eq. (2.20) are studied. This states inherent theoretical limitations for super-resolution in the presence of degenerate situations, where the underlying linear system is underdetermined.

2.5.1 Effective Magnification Factor

We are interested in an effective sampling frequency f^* that can be achieved by means of super-resolution. In this context, the term *effective* means that super-resolution cannot recover aliased parts of the spectrum beyond f^*, which states an upper bound of the artificial sampling frequency as well as an effective magnification factor s^*. Applying super-resolution beyond this limit does not yield additional gains compared to a simple interpolation of discrete samples. The key idea regarding the following analysis is that even if aliasing is usually considered as undesirable effect, it is a prerequisite for super-resolution. This is due to the fact that super-resolution exploits aliased components in undersampled signals according to Eq. (2.14). We examine this property for two different situations.

Magnification Factor for Ideal Sampling. Let us first consider ideal sampling of $x(t)$ with band-limitation f_0 below the Nyquist rate, i. e. $f_s \geq 2f_0$. Thus, there is no aliasing present in the sampled channels $y^{(k)}[n]$. In this case the infinite summation in Eq. (2.14) comprises only of one non-zero term, as there are no superimpositions of periodically shifted versions $X^{(k)}(f)$. Then, it follows for the aliasing property:

$$\mathcal{Y}^{(k)}[n] = X^{(k)}\left(\frac{n}{N} f_s\right), \tag{2.23}$$

for all $k = 1, \ldots, K$. That is, each DFT $\mathcal{Y}^{(k)}[n]$ is a sampled version of the corresponding CFT without loss of information as the sampling theorem is fulfilled. It is obvious that a solution of the linear system in Eq. (2.18) cannot recover additional information. Consequently, it follows for the effective sampling frequency $f^* = f_s$. Moreover, it is obvious that super-resolution cannot reconstruct frequencies beyond the band-limitation f_0. By combining these findings, the effective sampling frequency that can be achieved by super-resolution is given by:

$$f^* = \begin{cases} f_s & \text{if } f_s \geq 2f_0 \\ 2f_0 & \text{otherwise.} \end{cases} \tag{2.24}$$

This yields an upper bound regarding the effective magnification factor:

$$s^* = \begin{cases} 1 & \text{if } f_s \geq 2f_0 \\ 2\frac{f_0}{f_s} & \text{otherwise.} \end{cases} \tag{2.25}$$

Magnification Factor for Real Sampling. If real sampling is considered, the presence of the blur kernel $h^{(k)}(t)$ limits the effective magnification factor. As derived in Section 2.2.2, $h^{(k)}(t)$ acts as a low-pass filter for the channel $x^{(k)}(t)$. Unfortunately, this blur kernel also performs anti-aliasing depending on its cut-off frequency f_h. If the cut-off frequency is above the band-limitation, i. e. $f_h \geq f_0$, the blur kernel does not affect the sampling process. However, if $f_h < f_0$, spectral components of $X^{(k)}(f)$ affected by aliasing are suppressed by the blur kernel. In the worst case where $f_h < f_0 - f_s$, aliasing is fully removed. However, these are exactly the signal components exploited by super-resolution. Hence, the cut-off frequency f_h limits the effective sampling frequency to:

$$f^* = \begin{cases} \min(f_s, f_h) & \text{if } f_s \geq 2f_0 \\ \min(2f_0, f_h) & \text{otherwise.} \end{cases} \tag{2.26}$$

Note that f_h is now an upper bound for the sampling frequency f^* that can be smaller than the original sampling frequency f_s. The effective sampling frequency f^* yields an upper bound regarding the magnification factor:

$$s^* = \begin{cases} \frac{1}{f_s}\min(f_s, f_h) & \text{if } f_s \geq 2f_0 \\ \frac{1}{f_s}\min(2f_0, f_h) & \text{otherwise.} \end{cases} \tag{2.27}$$

This barrier needs to be considered in digital imaging, where the optical point spread function (PSF) and the finite size of pixels on the sensor array act as a low-pass filter. This has the consequence that super-resolution cannot provide effective magnifications beyond the system band-limitation related to these properties. In Chapter 3 and 4, we study different regularization techniques in conjunction with super-resolution reconstruction to alleviate this limitation in practical applications.

2.5.2 Uniqueness of the Reconstruction

The uniqueness of super-resolution based on the linear problems in Eq. (2.18) in case of ideal sampling and Eq. (2.20) in case of real sampling depends on several parameters of the sampling process. For this analysis, a reconstruction is called *unique* iff the associated system matrix involved in the linear problem is non-singular. In this case, the complementary information encoded by multiple channels is sufficient to provide a super-resolved signal. If the system matrix is singular, super-resolution becomes underdetermined and does not enable a unique reconstruction. In general, non-uniqueness might be caused by degenerate settings in terms of the sampling parameters resulting in a major limitation of super-resolution in practical applications.

This sections analyzes the uniqueness of the underlying inverse problem and derives conditions for a unique reconstruction. These derivations lead to two relevant classes of super-resolution algorithms applicable in digital imaging.

Uniqueness for Ideal Sampling. The uniqueness of Eq. (2.18) is first studied for ideal sampling. For this purpose, the channel offsets are used as a cue to provide complementary information and to guarantee a unique reconstruction. Let us study the case that the magnification factor is given by $s = K$ for K channels. If we consider the real and imaginary parts of the complex-valued Fourier transforms in Eq. (2.18), the system matrix is quadratic and an exact solution of this inverse problem can be obtained[2]. The conditions regarding uniqueness of super-resolution in this situation are summarized in the following theorem.

Theorem 2.2 (Uniqueness for ideal sampling). *Let $s = K$ be the super-resolution magnification factor and K be the number of channels in a multi-channel sampling process, where t_i with $i = 1, \ldots, K$ and $t_1 = 0$ are the corresponding channel offsets and T is the sampling pitch. Then, the solution of the linear inverse problem in Eq. (2.18) is unique if and only if: $t_j \neq c_1 t_i + c_2 T$ for all $1 \leq i < j \leq K$ and $c_1, c_2 \in \mathbb{Z}$.*

Proof. The proof of this theorem is given in Appendix A.1.1. □

These conditions provide an intuitive approach to perform super-resolution reconstruction. The distinct and non-integer offsets w. r. t. the sampling pitch enable a non-uniform sampling at a higher frequency compared to single-channel sampling. Thus, a solution of Eq. (2.18) can be seen as a fusion of the complementary information encoded by single channels, see Fig. 2.5. In fact, choosing distinct and non-integer channel offsets are necessary and sufficient conditions for a unique reconstruction. This is a popular strategy for resolution enhancement in digital imaging, where channel offsets can be related to subpixel displacements among multiple images [Park 03]. The offsets required for this *motion-based* super-resolution can be provided by capturing a set of images of the underlying scene while moving a camera to slightly different viewpoints.

Uniqueness for Real Sampling. Theorem 2.2 states necessary conditions for a unique reconstruction if only the channel offsets are exploited. However, in case of real sampling according to Eq. (2.20), the blur kernel can be used as cue to achieve uniqueness even in the absence of channel offsets. The following theorem summarizes the conditions to achieve a unique reconstruction in this situation.

Theorem 2.3 (Uniqueness for real sampling). *Let $s = K$ be the super-resolution magnification factor and K be the number of channels in multi-channel sampling with offsets $t_i = 0$ for all $i = 1, \ldots, K$ and sampling pitch T. Each channel $x^{(i)}(t)$ is affected by a blur kernel $H^{(i)}(f)$ denoted by $\boldsymbol{H}^{(i)}$ in matrix notation. Then, the solution of the linear inverse problem in Eq. (2.20) is unique if and only if:*

1. $\sum_{i=1}^{K} c_i \boldsymbol{H}^{(i)} \neq \boldsymbol{0}$ *for all* $c_i \neq 0$ *and* $i = 1, \ldots, K$ *(linear independent blur kernels)*
2. $\sum_{i=1}^{K} \left| H^{(i)}\left(\frac{n}{N} f_s + m f_s\right) \right| \neq 0$ *for all* $m = -L, \ldots, L-1$ *(kernel cut-off frequency)*

Proof. The proof of this theorem is given in Appendix A.1.2. □

[2] For $s < K$, an approximation can be obtained by means of least-squares estimation [Teka 92].

Besides the use of distinct, non-integer channel offsets to make super-resolution reconstruction unique, an alternative approach is to exploit the properties of the blur kernel. Complementary information required for super-resolution is gained by utilizing independent kernels for multiple channels. Moreover, the kernel cut-off frequency needs to be above the Nyquist rate and at least one kernel needs to span the entire frequency range that should be super-resolved. In this situation, a unique solution can be provided according to Theorem 2.3. This approach has been widely studied in digital imaging. In [Elad 97], Elad and Feuer investigated *motion-free* spatial domain super-resolution, which shows that this approach is feasible. Rajagopalan and Kiran [Raja 03] proposed a Fourier domain method to perform super-resolution reconstruction from multiple defocused images corresponding to varying levels of blur in a set of channels.

2.6 Conclusion

This chapter presented single- and multi-channel sampling as theoretical framework for super-resolution. This theory considered ideal sampling as well as real sampling in the presence of a blur kernel. Super-resolution was formulated as linear inverse problem that states the relation between a continuous signal and discrete channels that are captured by sampling the continuous signal multiple times. In this context, super-resolution aims at reconstructing an aliasing-free signal from multiple undersampled channels.

In order to derive fundamental limits of super-resolution, the properties of the underlying inverse problem were analyzed. First, the relationship between super-resolution and the Nyquist-Shannon sampling theorem was discussed. It was shown that the effective magnification, achievable by super-resolution, is bounded by the band-limitation of the continuous signal as well as the cut-off frequency of the blur kernel in case of real sampling. Second, the uniqueness of super-resolution reconstruction was examined. This analysis shows that super-resolution requires complementary information in multiple channels to provide a unique solution. The necessary and sufficient conditions to gain complementary information involve properties of the channel offsets or the blur kernel.

Part I

Numerical Methods for Multi-Frame Super-Resolution

CHAPTER 3

Computational Framework for Multi-Frame Super-Resolution

This chapter presents the common computational framework that is employed in the remainder of this thesis. This introduction comprises three parts. First, a literature survey on different paradigms in the field of super-resolution is presented including a review of state-of-the art frequency domain and spatial domain algorithms. Second, a spatial domain model for optical imaging is derived. Third, based on generative modeling, multi-frame super-resolution is approached from a Bayesian perspective and different parameter estimation schemes are presented.

3.1 Introduction and Literature Survey

In Chapter 2, super-resolution was introduced as a multi-channel signal reconstruction problem. A continuous signal was assumed to be sampled by multiple channels to obtain a set of discrete signals. In case of linearly independent offsets among the channels, each of these discrete signals contains complementary information about the underlying continuous one. In terms of *motion-based* super-resolution as the major scope of this work, channel offsets are explained by subpixel displacements across multiple images showing the same scene from slightly different perspectives. Mathematically, this motion is described by image-to-image transformations on the image plane and can be induced by camera motion, object motion, or a combination of both. Given a sequence of undersampled images along with their associated subpixel motion, the goal of image super-resolution is to obtain a high-resolution image from low-resolution ones. This reconstruction is unique if the channel offsets related to the subpixel motion are independent and distinct to multiples of the sampling rate, see Section 2.5.2.

Let us first present a survey on super-resolution paradigms including frequency and spatial domain methods, see Fig. 3.1. For a more comprehensive overview, we refer to the review articles by Park et al. [Park 03], Farsiu et al. [Fars 04a], and Nasrollahi and Moeslund [Nasr 14], as well as the book by Milanfar [Mila 10].

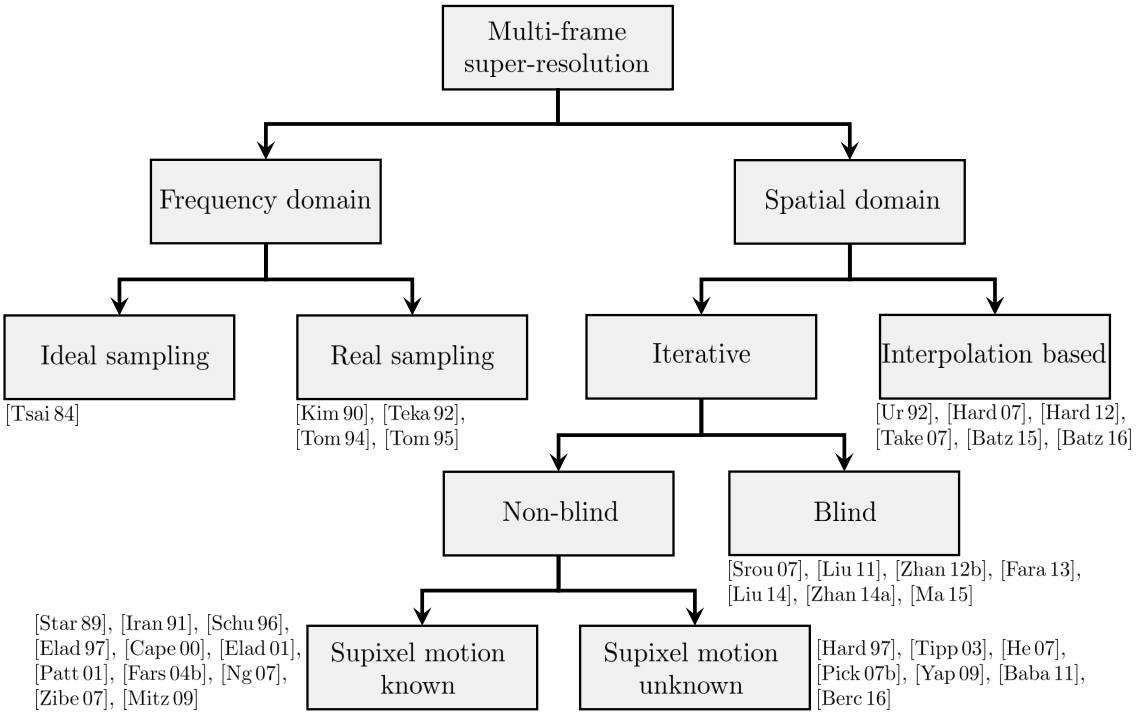


Figure 3.1: Classification of multi-frame super-resolution algorithms including references to seminal works and some of the most recent publications in the different domains.

3.1.1 Frequency Domain Reconstruction

Early approaches describe super-resolution reconstruction in the frequency domain based on the multi-channel sampling theory presented in Section 2.3. In this context, the method of Tsai and Huang [Tsai 84] employs the Fourier shift theorem to exploit translational subpixel motion in a sequence of low-resolution images. This yields a generative model for super-resolution in the Fourier domain. Translational motion is first determined by means of image registration, which can be performed in the Fourier domain using phase correlation. Then, super-resolution is implemented as inversion of the underlying model equations parametrized by the estimated motion, see Section 2.4. In the approaches of Kim et al. [Kim 90] and Tekalp et al. [Teka 92], this concept has been further extended to tackle sensor noise as well as blurring in the image formation. Tom et al. [Tom 94, Tom 95] have proposed simultaneous super-resolution and translation estimation in the Fourier domain. In [Rhee 99], Rhee and Kang have employed the discrete cosine transform (DCT) for super-resolution as alternative to the Fourier transform.

The frequency domain formulation provides valuable theoretical insights to super-resolution and the use of the FFT as a computational tool enables efficient implementations of these algorithms. However, only simple motion models can be used. For instance, the Fourier shift theorem in [Tsai 84] enables the description of translational motion but cannot model arbitrary displacements. In particular, it is not feasible to handle non-rigid motion. Additionally, blur needs to be described by LSI kernels to be tractable by means of the Fourier transform. This restricts the flexibility of these algorithms in terms of the underlying image formation model.

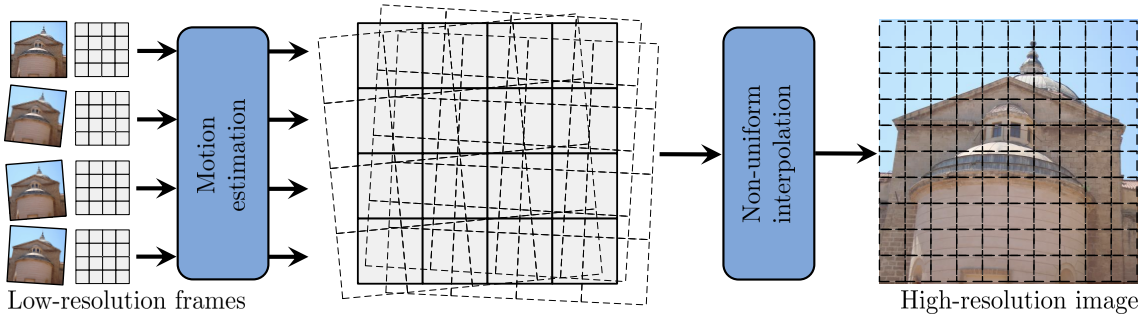


Figure 3.2: Spatial domain super-resolution using non-uniform interpolation. First, subpixel motion between multiple low-resolution frames is estimated and compensated. Then, the super-resolved image is interpolated from the motion-compensated frames.

3.1.2 Interpolation-Based Spatial Domain Reconstruction

Spatial domain reconstruction can be seen as a complementary trend in the design of super-resolution algorithms. These methods have the goal to enhance the flexibility regarding the choice of the image formation model. Ur and Gross [Ur 92] proposed interpolation-based reconstruction based on the multi-channel sampling theory of Papoulis [Papo 77] that has later been adopted in other spatial domain methods [Alam 00, Pham 06, Hard 07, Hard 12, Take 07, Batz 15]. Such algorithms differ in their implementations but share a similar conceptual structure. Their concept is described by a multi-stage procedure with the following steps, see Fig. 3.2:

1. Motion estimation by means of image registration determines the subpixel motion between multiple low-resolution frames.

2. Motion compensation transforms all low-resolution frames into a common high-resolution grid according to the motion estimate.

3. Non-uniform interpolation determines a super-resolved image based on the motion-compensated frames.

Some of the well known methods in this area include normalized convolution [Pham 06], kernel regression [Take 07], and adaptive Wiener filtering [Hard 07, Hard 12]. More recently, adaptive weighting schemes [Batz 16] as well as hybrid multi-frame and single-image reconstruction [Batz 15] have been proposed to alleviate the impact of motion estimation inaccuracies. These schemes can also be augmented by image deblurring [Patr 16] to remove blur after the interpolation.

A closely related class of algorithms employs *deep learning* that became a popular alternative to classical interpolation-based frameworks. These methods learn parts of the multi-stage procedure in Fig. 3.2 from pairs of low-resolution and high-resolution images. For instance, non-uniform interpolation [Kapp 16, Li 17] or motion compensation [Tao 17] can be learned via convolutional neural networks.

Unlike the aforementioned frequency domain methods that utilize a generative model, non-uniform interpolation aims at a direct reconstruction without considering such a model. These approaches can be implemented efficiently in terms of computational complexity. Moreover, the motion compensation is flexible and

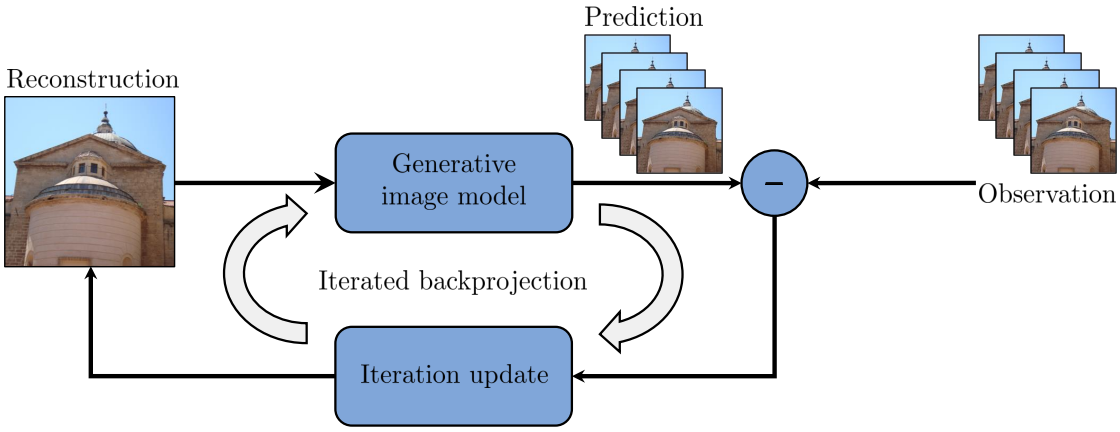


Figure 3.3: Spatial domain super-resolution using iterated backprojection [Iran 91]. The deviation between the acquired low-resolution observations and a prediction obtained from the current reconstruction under a generative image model is iteratively optimized.

can be applied under various types of subpixel motion. However, due to the sequential design, errors in one stage are propagated to following stages, e. g. from motion estimation to interpolation. This might lead to suboptimal reconstructions in terms of a global quality criterion [Park 03]. It is also difficult to model prior knowledge regarding the appearance of super-resolved images.

3.1.3 Iterative Spatial Domain Reconstruction

The vast majority of the state-of-art algorithms as well as the methods investigated in this work are formulated as iterative spatial domain reconstructions. In the same way as for the frequency domain methods, a generative model to describe the image formation is utilized. However, this model is formulated in the spatial domain to increase its flexibility. The basic idea is to iteratively refine an estimate for a super-resolved image such that it best explains the observed low-resolution data under the generative model. In literature, this concept has been first formalized in the *iterated backprojection* algorithm proposed by Irani and Peleg [Iran 91]. This approach is illustrated in Fig. 3.3 and we consider two realizations.

Non-Blind Reconstruction. In *non-blind* super-resolution, it is assumed that the parameters of the generative model are known a priori. In particular, it is assumed that the PSF is known either by system calibration, by automatic parameter selection [Nguy 01a], or by simply modeling it empirically with a realistic blur kernel.

One important class of these algorithms approaches super-resolution from a Bayesian statistics point of view that has also become a common tool in image denoising [Chen 07], restoration [Besa 91, Biou 06], and single-image expansion [Schu 94]. In [Elad 97], Elad and Feuer have proposed maximum likelihood (ML) estimation that has been used as a probabilistic framework in several follow-up works [Cape 00, Elad 01, Zibe 07, Mitz 09]. The subpixel motion that is exploited for this point estimation is determined via registration of low-resolution frames. Then, the most probable high-resolution image associated with the subpixel dis-

placed frames is reconstructed under a generative model. Maximum a-posteriori (MAP) estimation generalizes the ML approach by exploiting prior knowledge to regularize super-resolution. For this purpose, a Gaussian prior distribution to model the statistical appearance of images has been introduced in [Elad 97]. Later, other priors have been proposed [Schu 96, Fars 04b, Ng 07] that can better model the characteristics of natural images. These probabilistic models lead to energy minimization problems that can be solved iteratively. For details on these Bayesian methods, we refer to Section 3.3. A closely related technique is the projection onto convex sets (POCS) [Star 89, Patt 01]. POCS methods formulate prior knowledge by set theoretic constraints as opposed to probability distributions. Super-resolution is performed by iterative projections under these constraints.

Contrary to algorithms that estimate subpixel motion prior to super-resolution, there are also methods that treat the motion as hidden information. In a seminal work, Hardie et al. [Hard 97] proposed *joint* MAP estimation for both parameter sets based on alternating minimization. This avoids motion estimation on low-resolution data, which is error-prone due to undersampling [Vand 06b]. Notice that many algorithms that are formulated in the fashion of iterative spatial domain reconstruction, e. g. Gauss-Newton [He 07, Berc 16] or linear programming [Yap 09] schemes, fall into this category even if they are not explicitly derived in Bayesian frameworks. A related approach is to make use of Bayesian marginalization in the absence of a proper motion estimate. In [Tipp 03], Tipping and Bishop proposed marginalization over the domain of high-resolution images. A different approach has been developed by Pickup et al. [Pick 07b], where marginalization is performed over motion parameters to integrate them out from super-resolution reconstruction. As another technique in the field of Bayesian statistics, variational inference [Baba 11] has proven to be a valuable tool. As opposed to the ML and MAP schemes that provide point estimates, variational inference aims at determining full posterior probability distributions. This enables a joint estimation of the super-resolved image along with latent model parameters.

Blind Reconstruction. In terms of *blind* super-resolution, the blur kernel related to the camera PSF is assumed to be unknown, and so are the parameters of the generative image model. This class of algorithms is closely related to blind deconvolution [Patr 16] and treats the blur as a latent variable. Blind super-resolution is commonly implemented as an interlacing of non-blind reconstruction and blur estimation in joint optimization frameworks.

Super-resolution under an unknown PSF has been investigated by Sroubek et al. [Srou 07] and Faramarzi et al. [Fara 13]. Such methods combine multi-frame resolution enhancement and blind deconvolution to unified frameworks. These are formulated via iterative spatial domain reconstruction with known subpixel motion. Later, Zhang et al. [Zhan 12b] as well as Liu and Sun [Liu 11, Liu 14] have proposed to treat blur, subpixel motion and the latent high-resolution image as triple coupled variables to determine them simultaneously. This circumvents a direct motion estimation on low-resolution frames. More recently, the handling of motion blur has been studied for situations where optical and sensor blur are no appropriate models, e. g. fast camera-shake [Zhan 14a, Ma 15].

3.2 Modeling the Image Formation Process

One of the key components of multi-frame super-resolution is an appropriate mathematical modeling of image formation implemented by a digital imaging system. This section describes a spatial domain model that is widely applicable and employed to develop the super-resolution algorithms presented in the remainder of this work. This model can be seen as a spatial domain analog to the Fourier domain model derived in Chapter 2.

3.2.1 Continuous Image Formation Model

The image formation model used in this thesis is based on the work of Elad and Feuer [Elad 97] that has later been extended by Capel and Zisserman [Cape 03, Cape 04]. In recent years, this model has been utilized for the vast majority of super-resolution algorithms. It describes the physics of image acquisition in a forward process to explain how a digital image is obtained from a real-world scene.

For the derivation from a continuous point of view, let $x : \mathbb{R}^2 \to \mathbb{R}$ be the *irradiance light field* [Lin 04] obtained by an ideal projection of a 3-D scene onto the 2-D image plane of a digital camera. Mathematically, this projection can be described by a pinhole camera [Hart 04]. Since we limit ourselves to single-channel images, $x(\boldsymbol{u})$ denotes an intensity at the 2-D position $\boldsymbol{u} \in \mathbb{R}^2$. Using the assumption of an ideal projection to the image plane, $x(\boldsymbol{u})$ can be seen as a ground truth and is referred to as *ideal image* [Hard 07]. In particular, as no sampling is modeled and $x(\boldsymbol{u})$ is given as a continuous irradiance signal, it can be considered as a signal of infinite spatial resolution. In terms of an imaging system that acquires video data, one observes a set of K degraded frames given as continuous functions $y^{(1)}(\boldsymbol{u}), \ldots, y^{(K)}(\boldsymbol{u})$ associated with the ideal image $x(\boldsymbol{u})$. In order to describe the image formation process mathematically, the following operations are analyzed.

Motion Model. The ideal image $x(\boldsymbol{u})$ is assumed to be warped by a geometric transformation w. r. t. a certain coordinate reference for each frame to describe the acquisition of an image sequence. This geometric transformation encodes camera motion, object motion or a combination of both. For the sake of convenience, motion is described on the 2-D image plane instead of describing it by a 3-D transformation. The k-th warped version of $x(\boldsymbol{u})$ denoted by $x^{(k)}(\boldsymbol{u})$ is given by:

$$x^{(k)}(\boldsymbol{u}) = \mathcal{M}^{(k)}\{x(\boldsymbol{u})\}, \tag{3.1}$$

where $\mathcal{M}^{(k)}\{\cdot\}$ denotes the motion model for the k-th frame. Without any further assumption on the type of motion, this model is described by:

$$\mathcal{M}^{(k)}\{x(\boldsymbol{u})\} := x\big(\boldsymbol{u} + m^{(k)}(\boldsymbol{u})\big), \tag{3.2}$$

where $m^{(k)} : \mathbb{R}^2 \to \mathbb{R}^2$ denotes the displacement of $x(\boldsymbol{u})$ at position $\boldsymbol{u}$ relative to the reference frame. For a given reference frame $x^{(r)}(\boldsymbol{u})$ with $r \in \{1, \ldots, K\}$, $\mathcal{M}^{(r)}$ is the identity and $x^{(r)}(\boldsymbol{u})$ coincides with $x(\boldsymbol{u})$.

Sampling Model. Next, we need to describe the sampling process that explains how the imaging system discretizes the ideal image. In the most basic formulation of the image formation model, we limit ourselves to two common aspects.

First, each frame is affected by the PSF of the imaging system. This PSF is the impulse response of the entire system and describes how an ideal point object is captured on the image plane. We assume a LSI model that is characterized by a low-pass filter and define the PSF by a blur kernel $h(\boldsymbol{u})$. Then, a blurred version $\tilde{x}^{(k)}(\boldsymbol{u})$ of the warped image $x^{(k)}(\boldsymbol{u})$ is obtained by the convolution:

$$\begin{aligned}\tilde{x}^{(k)}(\boldsymbol{u}) &= x^{(k)}(\boldsymbol{u}) \star h^{(k)}(\boldsymbol{u}) \\ &= \int_{\mathbb{R}^2} x^{(k)}(\boldsymbol{v}) h^{(k)}(\boldsymbol{u}-\boldsymbol{v})\, d\boldsymbol{v}.\end{aligned} \tag{3.3}$$

Note that we assume the general case of a time variant blur kernel, i. e. $h^{(k)}(\boldsymbol{u})$ can be different for each frame. Following the analysis in [Bake 02], this blur kernel can be decomposed according to:

$$h^{(k)}(\boldsymbol{u}) = \left(h^{(k)}_{\text{optics}} \star h_{\text{sensor}}\right)(\boldsymbol{u}), \tag{3.4}$$

where $h^{(k)}_{\text{optics}}(\boldsymbol{u})$ models time variant blur caused by optical effects and $h_{\text{sensor}}(\boldsymbol{u})$ describes the time invariant blur caused the integration of light over a finite area on the sensor array corresponding to a pixel of the detector.

Finally, each frame is discretized on the sensor array that is modeled by two operations. First, each frame is sampled on the center positions of rectangular pixels. Since the integration of light on the sensor array is included in the PSF model, we describe the sampling by ideal Dirac impulses. Second, the sampled frame is disturbed by random measurement noise caused by an imperfect sensor. Mathematically, the sensor array is described by:

$$y^{(k)}(\boldsymbol{u}) = \mathcal{D}\left\{\tilde{x}^{(k)}(\boldsymbol{u})\right\} + \epsilon^{(k)}(\boldsymbol{u}), \tag{3.5}$$

where $\mathcal{D}\{\cdot\}$ denotes the sampling at the pixel positions and $\epsilon^{(k)}(\boldsymbol{u})$ is a stochastic signal to model measurement noise. Assuming a pixel pitch Δi in coordinate direction $i \in \{u, v\}$, the sampling operator is given by:

$$\mathcal{D}\left\{\tilde{x}^{(k)}(\boldsymbol{u})\right\} := \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \tilde{x}^{(k)}(u, v)\, \delta(u - m\Delta u, v - n\Delta v), \tag{3.6}$$

where $\delta(\boldsymbol{u})$ denotes the 2-D Dirac delta impulse.

Joint Motion and Sampling Model. The different physical effects that occur when acquiring a digital image from a real-world scene are combined in sequential order, see Fig. 3.4. In summary, the k-th frame $y^{(k)}(\boldsymbol{u})$ out of a set of K frames is related to the ideal image $x(\boldsymbol{u})$ according to:

$$y^{(k)}(\boldsymbol{u}) = \mathcal{D}\left\{\mathcal{M}^{(k)}\{x(\boldsymbol{u})\} \star h^{(k)}(\boldsymbol{u})\right\} + \epsilon^{(k)}(\boldsymbol{u}). \tag{3.7}$$

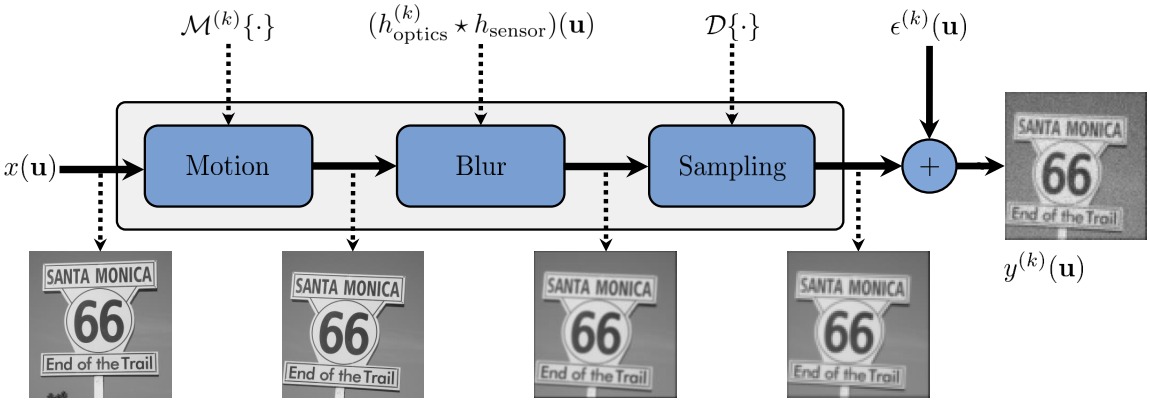


Figure 3.4: Illustration of the image formation model employed in this work.

This model is appropriate if the optical components and the sensor array have a dominant influence to the image formation while atmospheric effects need to be negligible. As shown in the following section, this scheme can be efficiently discretized to make it applicable for super-resolution.

Notice that there also exist related approaches with the order of the operations in Eq. (3.7) is reversed. One example is the blur-warping formulation with a reversed order of the motion and blur operators, which is superior to Eq. (3.7) in case of a high uncertainty regarding the motion as studied by Wang and Qi [Wang 04a].

3.2.2 Discretization of the Image Formation Model

In order to employ the image formation model defined in Eq. (3.7) in a computational super-resolution algorithm, a discretization of the model equation is required. For this discretization, two common assumptions must be met.

First, the only accessible information regarding a set of acquired images are the sampled intensity values at discrete pixel positions $\boldsymbol{u}$ in the domain of the input images denoted by $\Omega_y \subset \mathbb{R}^2$. For the sake of convenience, the intensity values at the pixel positions for each frame $y^{(k)}(\boldsymbol{u})$ of size $M_u \times M_v$ are reorganized into a vector using line-by-line scanning defined by:

$$\begin{aligned}\boldsymbol{y}^{(k)} := \big(y^{(k)}(\Delta u_y, \Delta v_y) \quad y^{(k)}(\Delta u_y, 2\Delta v_y) \quad \ldots \quad y^{(k)}(M_u \Delta u_y, M_v \Delta v_y)\big)^\top \\ \in \mathbb{R}^{M_u \cdot M_v},\end{aligned} \tag{3.8}$$

where Δu_y and Δv_y denote the pixel pitch in u- and v-direction, respectively. Since $\boldsymbol{y}^{(k)}$ is defined on the pixel grid of the acquired frames, $\boldsymbol{y}^{(k)}$ is referred to as a *low-resolution* frame. We denote by $\boldsymbol{y}$ the set of K low-resolution frames.

Second, as it is not feasible to reconstruct a continuous representation of the ideal image $x(\boldsymbol{u})$, we limit ourselves to the reconstruction of a digital image with finer spatial sampling in the domain $\Omega_x \subset \mathbb{R}^2$. The samples at pixel positions $\boldsymbol{u} \in \Omega_x$ in the image $x(\boldsymbol{u})$ of size $N_u \times N_v$ are reorganized to a vector according to:

$$\begin{aligned}\boldsymbol{x} := \big(x(\Delta u_x, \Delta v_x) \quad x(\Delta u_x, 2\Delta v_x) \quad \ldots \quad x(N_u \Delta u_x, N_v \Delta v_x)\big)^\top \\ \in \mathbb{R}^{N_u \cdot N_v},\end{aligned} \tag{3.9}$$

where Δu_x and Δv_x denote the pixel pitch in u- and v-direction, respectively. As these samples are defined on a finer pixel grid compared to the low-resolution frames, $\boldsymbol{x}$ is referred to as *high-resolution* image. Similar to the notation of low-resolution frames, the k-th warped version of the reference high-resolution image is denoted by $\boldsymbol{x}^{(k)}$. Assuming isotropic magnification, the magnification factor $s \in \mathbb{R}$ is given by $s = \sqrt{N/M}$, where $N = N_u \cdot N_v$ and $M = M_u \cdot M_v$ denote sizes of high-resolution and low-resolution images, respectively.

Discretization of the Motion Model

Let us next discretize the motion model. For this purpose, we consider discrete pixel positions $\boldsymbol{u} = (u, v)^\top \in \Omega_x$ in a warped high-resolution frame $\boldsymbol{x}^{(k)}$. Each point $\boldsymbol{u}$ in $\boldsymbol{x}^{(k)}$ is related to a transformed point $\boldsymbol{u}' = (u', v')^\top \in \Omega_x$ in $\boldsymbol{x}$, where $\boldsymbol{u}' = m^{(k)}(\boldsymbol{u})$. In order to describe subpixel motion, we introduce two motion models widely used within the algorithms developed in this work.

Parametric Motion. In case of a parametric model, $\boldsymbol{u}$ is assumed to be transformed by a global transformation that is characterized by a small number of parameters to describe image warping of $\boldsymbol{x}^{(k)}$ w. r. t. $\boldsymbol{x}$. We define this model via a *projective homography* in homogeneous coordinates [Hart 04] according to:

$$\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} \cong \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \boldsymbol{P} \begin{pmatrix} \boldsymbol{u} \\ 1 \end{pmatrix}, \tag{3.10}$$

where $\cong$ denotes equality up to a scale factor. In general, the homography $\boldsymbol{P}$ has full rank and is parametrized by eight degrees of freedom given by nine matrix elements minus a scale factor. In particular cases, the degrees of freedom can be reduced, which leads to a hierarchy of transformations [Hart 04]. The following cases are widely considered in literature. A homography $\boldsymbol{P}_{\text{affine}}$ is called *affine* if it can be parametrized by six degrees of freedom according to:

$$\boldsymbol{P}_{\text{affine}} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{t} \\ \boldsymbol{0}^\top & 1 \end{pmatrix}, \tag{3.11}$$

where $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$, $\boldsymbol{0} \in \mathbb{R}^2$ is an all-zero vector and $\boldsymbol{t} \in \mathbb{R}^2$ denotes a translation vector. An affine homography describes subpixel motion by rotation, anisotropic scaling and shearing as well as translation. This transformation does not preserve angles between lines and ratio of distances in an image but the parallelism of lines is invariant under an affine homography.

An affine homography $\boldsymbol{P}_{\text{rigid}}$ is called *rigid* if it can be parametrized by three degrees of freedom according to:

$$\boldsymbol{P}_{\text{rigid}} = \begin{pmatrix} \boldsymbol{R}(\varphi) & \boldsymbol{t} \\ \boldsymbol{0}^\top & 1 \end{pmatrix}, \tag{3.12}$$

where $\boldsymbol{R}(\varphi)$ is an orthogonal rotation matrix parametrized by the rotation angle φ. A rigid homography describes subpixel motion by rotation and translation without considering scaling or shearing. Notice that this transformation preserves angles between lines as well as the ratio of distances in an image.

These models are completely described by a few degrees of freedom but can only model 3-D motion under certain assumptions. One of their major limitations is that they are only applicable under rigid body motion. Moreover, even for rigid body motion, it can be shown that a homography is only valid for pure rotational camera motion or general camera motion in case of planar scenes due to occlusions of objects [Hart 04]. Nevertheless, this model is a reasonable approximation to describe rigid body motion for many applications of practical interest. One example is the acquisition of a static and non-planar scene from large distances such that it can be described as approximately planar, e. g. in remote sensing.

Non-Parametric Motion. A more flexible approach is the description by a dense displacement vector field according to:

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} u + m_u(\boldsymbol{u}) \\ v + m_v(\boldsymbol{u}) \end{pmatrix}. \tag{3.13}$$

The displacements $m(\boldsymbol{u}) = (m_u(\boldsymbol{u}), m_v(\boldsymbol{u}))^\top$ with $m_i : \Omega_y \to \mathbb{R}, i \in \{u, v\}$ describe the motion of $\boldsymbol{x}^{(k)}$ towards the reference image $\boldsymbol{x}$ at the pixel positions $\boldsymbol{u}$. Under the brightness constancy assumption, camera and object motion can be related to displacements on the image plane using the notion of *optical flow* [Horn 81].

This approach has the advantage that non-rigid deformations can be modeled along with rigid camera motion. In comparison to the parametric approach, it is also able to describe more general types of camera motion under projective distortions. However, as opposed to the parametric transformation in Eq. (3.10), the non-parametric model in Eq. (3.13) might be not bijective, e. g. due to occlusion. For this reason, one cannot obtain the motion of $\boldsymbol{x}$ relative to $\boldsymbol{x}^{(k)}$ by simply inverting the displacements $m(\boldsymbol{u})$ that describe the motion in the opposite direction.

Discretization of the Sampling Model

Next, we examine the sampling process and discretize Eq. (3.7). In this continuous equation, a single low-resolution frame $y^{(k)}(\boldsymbol{u})$ with blur kernel $h(\boldsymbol{u})$ and subpixel displacements $m(\boldsymbol{u})$ is related to $x(\boldsymbol{u})$ according to:

$$y^{(k)}(\boldsymbol{u}) = \mathcal{D}\{x(\boldsymbol{u} + m(\boldsymbol{u})) \star h(\boldsymbol{u})\} + \epsilon^{(k)}(\boldsymbol{u}). \tag{3.14}$$

The discretization of this relationship is derived in terms of the transformed image $x(\boldsymbol{u} + m(\boldsymbol{u}))$ and we set $\boldsymbol{u}' = \boldsymbol{u} + m(\boldsymbol{u})$. It is important to note that Eq. (3.14) defines a complicated non-linear relationship between $x(\boldsymbol{u})$ and $y^{(k)}(\boldsymbol{u})$ as we do not limit the motion model $m(\boldsymbol{u})$ to a simple linear transformation. To simplify the continuous formulation, we approximate Eq. (3.14) as:

$$\begin{aligned} y^{(k)}(\boldsymbol{u}) &\approx \mathcal{D}\{x(\boldsymbol{u}') \star h(\boldsymbol{u}')\} + \epsilon^{(k)}(\boldsymbol{u}) \\ &= \mathcal{D}\left\{\int_{\mathbb{R}^2} x(\boldsymbol{v}) h(\boldsymbol{u}' - \boldsymbol{v})\, d\boldsymbol{v}\right\} + \epsilon^{(k)}(\boldsymbol{u}), \end{aligned} \tag{3.15}$$

where we assumed that $h(\boldsymbol{u}')$ describing the blur kernel transformed according to $m(\boldsymbol{u})$ fulfills $h(\boldsymbol{u}') \approx h(\boldsymbol{u})$. Notice that under pure translational motion and an

arbitrary blur kernel or under rigid motion and a radially symmetric blur kernel, it follows that $h(\boldsymbol{u}') = h(\boldsymbol{u})$. Then, the motion and blur operations of the image formation model commutes [Fara 13]. In case of more general transformations, e. g. affine motion, the approximation $h(\boldsymbol{u}') \approx h(\boldsymbol{u})$ can be justified by the fact that $h(\boldsymbol{u})$ is typically a spatially smooth kernel. In particular, this approximation is sensible under small subpixel motion. In case of local or non-rigid motion, one needs to assure that these deformations are small compared to global motion.

In order to discretize the relationship between $x(\boldsymbol{u})$ and a single frame $y^{(k)}(\boldsymbol{u})$, we consider the vectorized versions of these images given by $\boldsymbol{x}$ and $\boldsymbol{y}^{(k)}$, respectively. Then, we implement Eq. (3.7) in matrix/vector notation as:

$$\boldsymbol{y}^{(k)} = \boldsymbol{W}^{(k)}\boldsymbol{x} + \epsilon^{(k)}, \tag{3.16}$$

where $\boldsymbol{W}^{(k)} \in \mathbb{R}^{M\times N}$ denotes the *system matrix* that comprises a discrete version of the motion model associated with the k-th frame, the blurring caused by the camera PSF as well as sampling on the sensor array. $\epsilon^{(k)} \in \mathbb{R}^M$ is a random vector to model additive noise. Given a sequence of K frames yields:

$$\underbrace{\begin{pmatrix}\boldsymbol{y}^{(1)}\\ \vdots \\ \boldsymbol{y}^{(K)}\end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix}\boldsymbol{W}^{(1)}\\ \vdots \\ \boldsymbol{W}^{(K)}\end{pmatrix}}_{\boldsymbol{W}} \boldsymbol{x} + \underbrace{\begin{pmatrix}\epsilon^{(1)}\\ \vdots \\ \epsilon^{(K)}\end{pmatrix}}_{\epsilon}, \tag{3.17}$$

where $\boldsymbol{W} \in \mathbb{R}^{KM\times N}$, $\boldsymbol{y} \in \mathbb{R}^{KM}$ and $\epsilon \in \mathbb{R}^{KM}$ denote the combined versions of the system matrices, the low-resolution frames and the noise vectors, respectively. In the sequel, we introduce two approaches to implement this relationship.

Implementation using Filter Operations. To avoid an explicit computation of the system matrix, the image formation can be implemented by discrete filter operations [Zome 00]. For this approach, the system matrix of the k-th frame is decomposed as:

$$\boldsymbol{W}^{(k)} = \boldsymbol{D}\boldsymbol{H}\boldsymbol{M}^{(k)}, \tag{3.18}$$

where $\boldsymbol{D} \in \mathbb{R}^{M\times N}$ denotes subsampling by Dirac impulses, $\boldsymbol{H} \in \mathbb{R}^{N\times N}$ models the blur kernel, and $\boldsymbol{M}^{(k)} \in \mathbb{R}^{N\times N}$ encodes the motion for the k-th frame in matrix notation. These operations are discrete versions of their continuous counterparts in Fig. 3.4 and can be implemented as follows. The motion operator $\boldsymbol{M}^{(k)}$ is modeled by geometric warping of $\boldsymbol{x}$ according to the underlying motion model. The blur operator $\boldsymbol{H}$ is implemented by means of a discrete convolution, whereas the filter kernel corresponds to a discrete version of $h(\boldsymbol{u})$. The subsampling operator $\boldsymbol{D}$ is implemented by a nearest-neighbor interpolation.

This approach is computationally efficient in terms of memory management as the system matrix does not need to be stored. However, due to the concatenation of the three operations interpolation artifacts in these stages are propagated. Here, image warping to implement $\boldsymbol{M}^{(k)}$ might result in aliasing artifacts due to resampling that is required to obtain the intermediate image $\boldsymbol{M}^{(k)}\boldsymbol{x}$. These artifacts are propagated to blurring and subsampling but are not physically meaningful.

Implementation using Matrix Operations. The approach that is employed in this thesis is based on the work of Tipping and Bishop [Tipp 03]. For this implementation, the system matrix is constructed without decomposition in discrete filters. This requires that the PSF is modeled by a narrow kernel $h(\boldsymbol{u}')$ and its continuous version is used to determine the matrix elements. This is reasonable since the integration of photons per pixel on the sensor array is performed over a finite area and one does not need to consider the entire detector surface. Hence, the convolution can be replaced by an integration over a circular neighborhood $\omega_{\text{PSF}}(\boldsymbol{u}')$ centered at $\boldsymbol{u}'$, where $\omega_{\text{PSF}}(\boldsymbol{u}') = \{\boldsymbol{v} : ||\boldsymbol{v} - \boldsymbol{u}'||_2 \leq N_{\text{PSF}}\}$ and N_{PSF} denotes the PSF radius. Then, the frame $y^{(k)}(\boldsymbol{u})$ is related to the high-resolution image $x(\boldsymbol{u})$ according to:

$$\begin{aligned} y^{(k)}(\boldsymbol{u}) &\approx \mathcal{D}\left\{ \int_{\omega_{\text{PSF}}(\boldsymbol{u}')} x(\boldsymbol{v}) h\left(\boldsymbol{u}' - \boldsymbol{v}\right) d\boldsymbol{v} \right\} + \epsilon^{(k)}(\boldsymbol{u}) \\ &= \mathcal{D}\left\{ \sum_{\boldsymbol{v} \in \omega_{\text{PSF}}(\boldsymbol{u}')} x(\boldsymbol{v}) h\left(\boldsymbol{u}' - \boldsymbol{v}\right) \right\} + \epsilon^{(k)}(\boldsymbol{u})\,. \end{aligned} \tag{3.19}$$

Thus, the intensity at the pixel position $\boldsymbol{u}$ is described by a weighted sum of the intensities in $\omega_{\text{PSF}}(\boldsymbol{u}')$ and the weights are expressed in terms of $h(\boldsymbol{u}')$. For the k-th frame, these weights are encoded in the system matrix $\boldsymbol{W}^{(k)}$. The matrix element at position (m, n) is determined with normalized row sums according to:

$$W_{mn} = \begin{cases} \frac{1}{\sum_{i=1}^{N} h(\boldsymbol{u}'_m - \boldsymbol{v}_i)} h\left(\boldsymbol{u}'_m - \boldsymbol{v}_n\right) & \boldsymbol{v}_n \in \omega_{\text{PSF}}(\boldsymbol{u}'_m) \\ 0 & \text{otherwise} \end{cases}, \tag{3.20}$$

where $\boldsymbol{u}'_m$ are the coordinates of the m-th pixel in $y^{(k)}(\boldsymbol{u})$ transformed to the coordinate grid of $x(\boldsymbol{u})$, and $\boldsymbol{v}_n$ are the coordinates of the n-th pixel in $x(\boldsymbol{u})$. This construction exploits the fact that the PSF is described by a narrow blur kernel and we set $W_{mn} = 0$ if $\boldsymbol{u}'_m$ and $\boldsymbol{v}_n$ does not affect each other.

One common assumption is to model the PSF by an isotropic Gaussian kernel $h(\boldsymbol{u}) = \exp(-\frac{1}{2}||\boldsymbol{u}||_2^2/(s^2\sigma_{\text{PSF}}^2))$ and to truncate $h(\boldsymbol{u})$ for $||\boldsymbol{u}||_2 > 3\sigma_{\text{PSF}}$ [Pick 07a] as depicted in Fig. 3.5. Notice that σ_{PSF} characterizes the PSF size in units of low-resolution pixels. For efficient storage, the joint system matrix $\boldsymbol{W}$ is assembled as a sparse matrix since most of the elements are zero due to the finite support of the blur kernel. The size of the matrix depends on the number of frames K, the image dimension M as well as the PSF radius N_{PSF}. In the asymptotic case, the number of non-zero elements in $\boldsymbol{W}$ is $\mathcal{O}(KMN_{\text{PSF}})$.

3.2.3 Discussion and Limitations of the Model

The benefit of the presented image formation model is that low-resolution data $\boldsymbol{y}$ can be simulated in an efficient way from a high-resolution image $\boldsymbol{x}$ by means of the system matrix $\boldsymbol{W}$, which is precomputed from motion and imaging parameters. This simulation only involves matrix-vector operations and is used to implement super-resolution via iterative energy minimization. For an efficient implementation, the matrix elements can be calculated in a parallel way [Wetz 13].

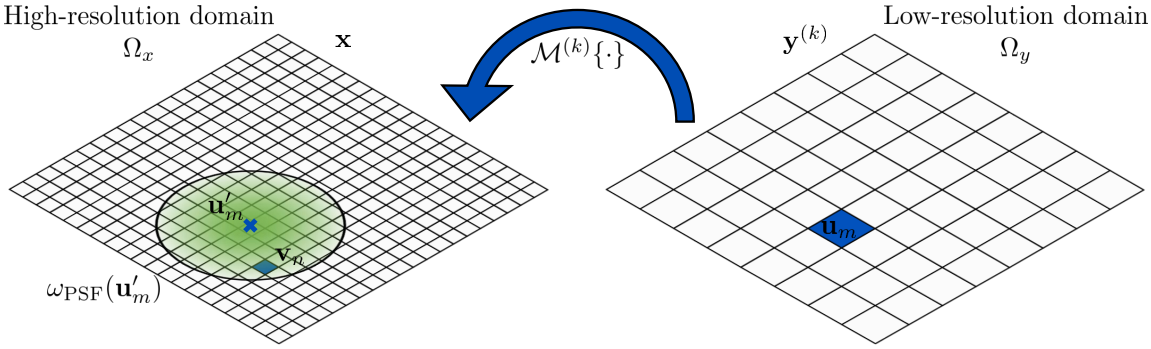


Figure 3.5: Construction of the system matrix $\boldsymbol{W}^{(k)}$ in an element-wise scheme. Each low-resolution pixel $\boldsymbol{u}_m$ is warped towards the high-resolution image $\boldsymbol{x}$ resulting in the transformed pixel $\boldsymbol{u}'_m$. The element W_{mn} is computed from $\boldsymbol{u}'_m$ and $\boldsymbol{v}_n$ assuming a radial symmetric PSF that is non-zero in the neighborhood $\omega_{\mathrm{PSF}}(\boldsymbol{u}'_m)$ but approaches zero otherwise.

It is important to note that the proposed discretization approximates the continuous model in Eq. (3.7) by assuming that the PSF blur kernel is the same in the reference frame and a subpixel warped frame. This is reasonable under subpixel motion that is approximately rigid. In prior work, other discretization schemes have been proposed, see e. g. [Pick 07a] and [Cape 04] as well as the references therein. These are more accurate under certain types of motion, e. g. affine motion, but are computationally more demanding.

There are also several practical limitations of the model that need to be considered. First, it is assumed that the characteristics of the imaging system in terms of the PSF are space invariant. In many practical applications, this assumption might be violated. Common examples are motion blur [Ma 15] or atmospheric blur that might be space variant. In order to model such effects, the construction of the system matrix needs to be implemented spatially adaptive to take a varying blur kernel into account. Another class of limitations is related to internal signal processing performed by a camera after capturing raw data. One issue is white balancing, which results in photometric variations over the low-resolution frames. In this chapter, such effects are not considered but the model can be extended to allow spatially and temporally varying photometric conditions, see Chapter 7. Such shortcomings of the model haven been quantitatively studied in [Pick 07a].

In the above derivation, we limited ourselves to single-channel images, where each pixel represents one discrete measurement. One interesting extension is to model the acquisition of color images in the RGB space. In today's low-cost cameras, color images are mosaiced since each pixel can only measure one spectral band according to a color-filter array (CFA) [Fars 06]. This CFA needs to be considered in color image super-resolution by extending the image formation model.

3.3 Bayesian Modeling of Super-Resolution

This section introduces multi-frame super-resolution from a Bayesian point of view. Let us describe the latent high-resolution image $\boldsymbol{x}$ along with the set of low-resolution observations $\boldsymbol{y}$ as random variables. The *observation model* that describes

the probability of observing a single frame $\boldsymbol{y}^{(k)}$ from the high-resolution image $\boldsymbol{x}$ is denoted by the probability density function (PDF) $p(\boldsymbol{y}^{(k)} \,|\, \boldsymbol{x})$. This conditional probability is defined in terms of the discrete image formation model in Eq. (3.17). If the observation noise ϵ is space invariant and follows a normal distribution with zero mean and covariance $\sigma^2_{\text{noise}}\boldsymbol{I}$, the observation model is given by:

$$\begin{aligned} p\left(\boldsymbol{y}^{(k)} \,|\, \boldsymbol{x}\right) &= \mathcal{N}\left(\boldsymbol{y}^{(k)} - \boldsymbol{W}^{(k)}\boldsymbol{x}; \boldsymbol{0}, \sigma^2_{\text{noise}}\boldsymbol{I}\right) \\ &:= \frac{1}{\sigma_{\text{noise}}\sqrt{2\pi}} \exp\left\{-\frac{\left(\boldsymbol{y}^{(k)} - \boldsymbol{W}^{(k)}\boldsymbol{x}\right)^\top \left(\boldsymbol{y}^{(k)} - \boldsymbol{W}^{(k)}\boldsymbol{x}\right)}{2\sigma^2_{\text{noise}}}\right\}, \end{aligned} \tag{3.21}$$

Notice that the observation model can be tailored to different noise characteristics. In Section 4.3, we present a model that accounts for space variant noise in the low-resolution observations.

Below, we review two commonly used approaches to infer the high-resolution image $\boldsymbol{x}$ from the set of low-resolution observations $\boldsymbol{y}$ under this distribution. Both approaches yield a formulation of super-resolution as unconstrained energy minimization and provide point estimates for the latent high-resolution image.

3.3.1 Maximum Likelihood Estimation

Let us assume that the sequence of low-resolution frames $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}$ are independent and identically distributed (i. i. d.) random variables. Then, one can derive the joint distribution to describe the probability of observing the low-resolution data $\boldsymbol{y}$ from the high-resolution image $\boldsymbol{x}$ according to the factorization:

$$p(\boldsymbol{y} \,|\, \boldsymbol{x}) = \prod_{k=1}^{K} \mathcal{N}\left(\boldsymbol{y}^{(k)} - \boldsymbol{W}^{(k)}\boldsymbol{x}; \boldsymbol{0}, \sigma^2_{\text{noise}}\boldsymbol{I}\right). \tag{3.22}$$

The objective of ML estimation is to infer a high-resolution image that best explains the set of low-resolution observations. If there is no prior knowledge about the high-resolution image available, it can be directly inferred from the point estimation:

$$\boldsymbol{x}_{\text{ML}} = \operatorname*{argmax}_{\boldsymbol{x}} p(\boldsymbol{y} \,|\, \boldsymbol{x}). \tag{3.23}$$

Taking the negative log-likelihood $L(\boldsymbol{x}) \propto -\log p(\boldsymbol{y} \,|\, \boldsymbol{x})$ of Eq. (3.23), this is equivalent to the unconstrained minimization problem:

$$\boldsymbol{x}_{\text{ML}} = \operatorname*{argmin}_{\boldsymbol{x}} L(\boldsymbol{x}), \tag{3.24}$$

where:

$$\begin{aligned} L(\boldsymbol{x}) &= \sum_{k=1}^{K} \left|\left| r\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right) \right|\right|_2^2 \\ &= \left|\left| r(\boldsymbol{x}, \boldsymbol{y}) \right|\right|_2^2, \end{aligned} \tag{3.25}$$

and $r(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}$ denotes the *residual error* of the estimate $\boldsymbol{x}$ w. r. t. the observations $\boldsymbol{y}$. This links the Gaussian observation model to least-square optimization.

The convex minimization problem in Eq. (3.24) can be solved in closed form. Unfortunately, this requires an inversion of the system matrix $\boldsymbol{W}$, which is computationally prohibitive for real-world problem sizes. For the purpose of a practical implementation, energy minimization is performed by means of iterative numerical optimization to avoid a direct inversion of the system matrix. Several approaches that are widely used in literature include steepest descent iterations with fixed [Elad 97, Li 10] or adaptive step size [Hard 97, Lee 03] as well as conjugate gradient (CG) based iteration schemes [Nguy 01b, Zibe 07]. In case of pure translational motion and a space invariant generative model, ML estimation can also be decomposed into a non-iterative interpolation and an iterative deblurring stage [Elad 01]. For details on the numerical optimization, we refer to Section 4.3.

3.3.2 Maximum A-Posteriori Estimation

In terms of MAP estimation, prior knowledge regarding the occurrence of high-resolution images is exploited instead of using a uniform prior. The motivation of this approach is that super-resolution is a highly ill-posed problem under practical conditions [Borm 04]. Thus, ML estimation that does not consider prior knowledge on the desired high-resolution image needs to be regularized to steer the reconstruction algorithm to a reasonable solution.

Figure 3.6 depicts this issue on a simulated dataset with known subpixel motion, where $K = 16$ low-resolution frames are obtained from a ground truth according to the proposed image formation model. This example considers low-resolution observations in the intensity range $[0, 1]$ that are corrupted by Gaussian noise at different standard deviations σ_{noise}. In the corresponding ML estimates, image noise in the input frames is severely amplified. Notice that in addition to the influence of image noise, super-resolution based on ML estimation is also ill-conditioned in case of uncertainties of model parameters [Pick 07a].

Let $p(\boldsymbol{x})$ be a prior distribution on the latent high-resolution image to model its appearance in Bayesian way. Similarly, let $p(\boldsymbol{y})$ be the distribution of the low-resolution frames. Then, using Bayes' rule, the posterior distribution $p(\boldsymbol{x} \,|\, \boldsymbol{y})$ is given by:

$$p(\boldsymbol{x} \,|\, \boldsymbol{y}) = \frac{p(\boldsymbol{y} \,|\, \boldsymbol{x}) \cdot p(\boldsymbol{x})}{p(\boldsymbol{y})} \propto p(\boldsymbol{y} \,|\, \boldsymbol{x}) \cdot p(\boldsymbol{x}) \,. \tag{3.26}$$

The goal of MAP estimation is to reconstruct a high-resolution image that maximizes the posterior $p(\boldsymbol{x} \,|\, \boldsymbol{y})$ according to:

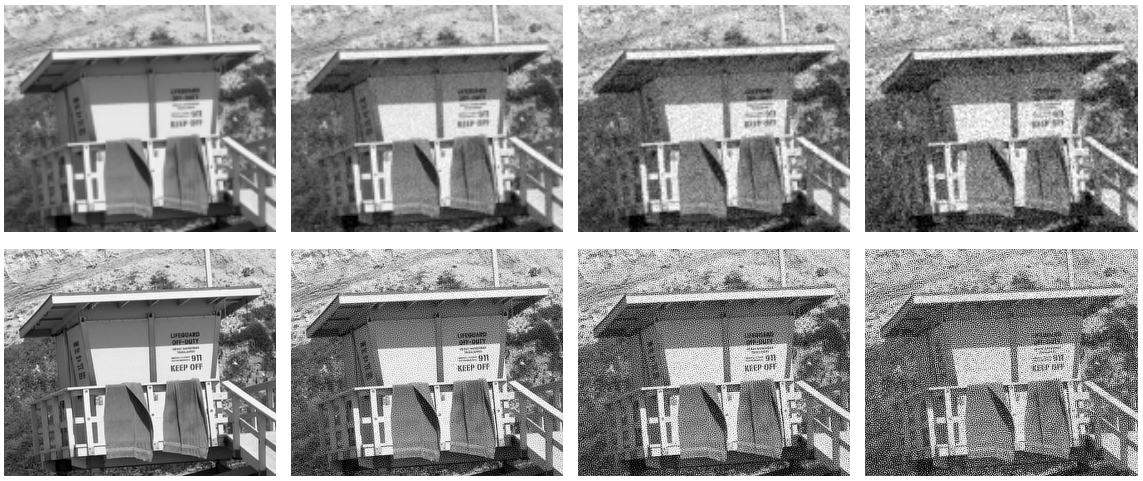
$$\boldsymbol{x}_{\text{MAP}} = \underset{\boldsymbol{x}}{\operatorname{argmax}} \left\{ p(\boldsymbol{y} \,|\, \boldsymbol{x}) \, p(\boldsymbol{x}) \right\} . \tag{3.27}$$

In order to derive this point estimation as an energy minimization problem, we use the negative log-likelihood associated with the posterior according to:

$$\boldsymbol{x}_{\text{MAP}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ L(\boldsymbol{x}) + \lambda R(\boldsymbol{x}) \right\} , \tag{3.28}$$

where $L(\boldsymbol{x})$ denotes the negative log-likelihood from ML estimation referred to as *data fidelity term*. $R(\boldsymbol{x}) \propto -\log p(\boldsymbol{x})$ denotes a *regularization term* with the regularization weight $\lambda \geq 0$ to weight this term relative to the data fidelity.

(a) Ground truth



(b) $\sigma_{\text{noise}} = 0$ **(c)** $\sigma_{\text{noise}} = 0.025$ **(d)** $\sigma_{\text{noise}} = 0.05$ **(e)** $\sigma_{\text{noise}} = 0.075$

Figure 3.6: Super-resolution using ML estimation on simulated data. First row: ground truth image used for this example. Second row: low-resolution frames simulated from the ground truth at different levels of Gaussian noise with standard deviation σ_{noise}. Third row: ML estimates using $K = 16$ low-resolution frames with magnification $s = 4$.

To define the prior distribution $p(\boldsymbol{x})$, various approaches emerged in literature. A wide class of these general-purpose models exploits smoothness, piecewise smoothness or sparsity of natural images in transform domains. We consider image priors that are expressed by the Boltzmann distribution [Bish 06]:

$$p(\boldsymbol{x}) = \frac{1}{Z(\sigma_{\text{prior}})} \exp\left\{-\frac{R(\boldsymbol{x})}{\sigma_{\text{prior}}}\right\}, \tag{3.29}$$

where $\sigma_{\text{prior}} > 0$ is a distribution scale parameter and $Z(\sigma_{\text{prior}})$ is the partition function used for normalization. Below, we review some of the most commonly used priors that are employed in this thesis. For the design of prior distributions that are tailored to the characteristics of natural images, we refer to Chapter 4.

Gaussian Prior. A Gaussian prior as the most basic and commonly used approach [Elad 97, Hard 97, Cape 03] considers a high-resolution image $\boldsymbol{x}$ as a spatially smooth signal. This should avoid noise amplification by ML estimation. In doing so, the prior $p(\boldsymbol{x})$ defined by Eq. (3.29) is parametrized by:

$$R_{\text{Gauss}}(\boldsymbol{x}) = ||\boldsymbol{Q}\boldsymbol{x}||_2^2, \tag{3.30}$$

where $\boldsymbol{Q} \in \mathbb{R}^{N \times N}$ is a circulant matrix to write a discrete convolution with a high-pass filter $\boldsymbol{q}$ as a matrix-vector product, i. e. $\boldsymbol{Q}\boldsymbol{x} \equiv \boldsymbol{x} \star \boldsymbol{q}$. Typical choices for $\boldsymbol{Q}$ are

(a) $\sigma_{\text{noise}} = 0$ **(b)** $\sigma_{\text{noise}} = 0.025$ **(c)** $\sigma_{\text{noise}} = 0.05$ **(d)** $\sigma_{\text{noise}} = 0.075$

Figure 3.7: MAP estimation on the simulated data in Fig. 3.6 with a Gaussian prior modeled by a discrete Laplacian [Hard 97] and a constant regularization weight ($\lambda = 0.05$). Super-resolved images are depicted at different noise standard deviations σ_{noise}.

the gradient determined by finite differences [Cape 03] or the discrete Laplacian [Hard 97, He 06]. This prior yields a Tikhonov regularized optimization problem also known as ridge regression [Hast 09]. The main benefits of this model are that it is feasible to use the prior for analytical computations and that it yields a convex regularization term that is easy to minimize by numerical optimization.

Figure 3.7 depicts the influence of the image prior on the simulated dataset in Fig. 3.6. In this example, the MAP estimation is based on a discrete Laplacian and a constant regularization weight ($\lambda = 0.05$). In contrast to ML estimation, noise amplification is reduced by regularization with the Gaussian prior. This stabilizes the reconstruction and improves the visual quality of super-resolved images.

Huber Prior. The major shortcoming of the Gaussian prior is that discontinuities in an image are penalized in the same manner as noise. This reduces the ability of edge reconstruction since sharp edges appear blurred due to the smoothness assumption. One approach to enhance edge reconstruction is to replace the L_2 norm of the Gaussian prior by a robust loss function to obtain a distribution with heavier tails. A common choice is to employ the Huber prior [Schu 96, Pick 07b]:

$$R_{\text{Huber}}(\boldsymbol{x}) = \sum_{i=1}^{N} \phi_{\text{Huber}}([\boldsymbol{Q}\boldsymbol{x}]_i)\,, \tag{3.31}$$

where $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ is a circulant matrix and $[\boldsymbol{z}]_i$ denotes the i-th element of the vector $\boldsymbol{z}$. The function $\phi_{\text{Huber}}(z)$ denotes the Huber loss applied element-wise to the high-pass filtered image $\boldsymbol{Q}\boldsymbol{x}$. In this work, we use the smooth approximation of the Huber loss that has continuous first- and second-order derivatives [Hart 04]:

$$\phi_{\text{Huber}}(z) = \delta_{\text{Huber}} \sqrt{1 + \left(\frac{z}{\delta_{\text{Huber}}}\right)^2} - \delta_{\text{Huber}}, \tag{3.32}$$

where δ_{Huber} is a scale parameter. This function behaves like the Gaussian prior for small z ($z \ll \delta_{\text{Huber}}$) and penalizes z quadratically. In case of large z ($z \gg \delta_{\text{Huber}}$), it is proportional to $|z|$. Hence, it features piecewise smooth regularization in order to enhance the reconstruction of discontinuities. Similar to the Gaussian prior, Eq. (3.31) yields a convex regularization term. However, in contrast to the Gaussian prior, it requires non-linear optimization techniques for energy minimization.

Total Variation. The Rudin, Osher and Fatemi (ROF) model [Rudi 92] also known as total variation (TV) has been originally introduced for image denoising. Later, it has also been employed for blind deconvolution [Chan 98] and super-resolution [Ng 07]. Unlike the Huber prior that provides piecewise smooth regularization, the TV prior explains an image as a piecewise constant signal.

The isotropic version of this prior introduced in [Rudi 92] is defined by:

$$R_{\mathrm{TV}}(\boldsymbol{x}) = \sum_{i=1}^{N} \sqrt{[\nabla_u \boldsymbol{x}]_i^2 + [\nabla_v \boldsymbol{x}]_i^2}, \tag{3.33}$$

where $\nabla_u \boldsymbol{x}$ and $\nabla_v \boldsymbol{x}$ denote the discrete image gradient in u- and v-direction, respectively. This convex regularization term exploits the sparsity of natural images in the gradient domain. Besides its application in image restoration problems, this prior has also great importance for regularization of ill-posed problems in the theory of compressed sensing [Dono 06, Cand 08].

Isotropic TV has the limitation that it considers the image gradient in horizontal and vertical direction only. For this reason, super-resolution is prone to staircasing artifacts in image regions with small gradient magnitudes. One common generalization of this approach is the use of bilateral total variation (BTV) [Fars 04b]. The BTV prior is inspired from bilateral filtering [Toma 98] and is given by:

$$R_{\mathrm{BTV}}(\boldsymbol{x}) = \sum_{m=-N_{\mathrm{BTV}}}^{N_{\mathrm{BTV}}} \sum_{n=-N_{\mathrm{BTV}}}^{N_{\mathrm{BTV}}} {\alpha_{\mathrm{BTV}}}^{|m|+|n|} \left|\left| \boldsymbol{x} - \boldsymbol{S}_u^m \boldsymbol{S}_v^n \boldsymbol{x} \right|\right|_1, \tag{3.34}$$

where $\boldsymbol{S}_u^m$ and $\boldsymbol{S}_v^n$ denote shifts of $\boldsymbol{x}$ by m pixel in u-direction and n pixel in v-direction, respectively. The shifts are performed in a $(2N_{\mathrm{BTV}}+1) \times (2N_{\mathrm{BTV}}+1)$ window, where $\alpha_{\mathrm{BTV}} \in]0,1]$ weights the difference between $\boldsymbol{x}$ and its shifted version according to the shift magnitude. This prior performs a multiscale analysis of the image gradient and yields convex regularization similar to isotropic TV.

3.4 Conclusion

This chapter introduced the computational framework that is utilized for the algorithm development in the remainder of this work. In the first part, a literature survey served as an overview regarding different paradigms of multi-frame super-resolution including frequency domain, interpolation-based as well as iterative spatial domain approaches. The algorithms proposed in this thesis are based on the iterative spatial domain formulation due to its flexibility in terms of the motion model and the ability to integrate prior knowledge. The second part covered the derivation of an image formation model to describe image acquisition in digital imaging from a mathematical viewpoint. This model was discretized in order to make it applicable for multi-frame super-resolution algorithms. Finally, super-resolution was formulated from a Bayesian perceptive as a statistical parameter estimation based on the discrete image formation model. For this purpose, ML and MAP estimation were discussed. These formulations state super-resolution as energy minimization problem that provides the basis for the computational methods introduced in the subsequent chapters.

CHAPTER 4

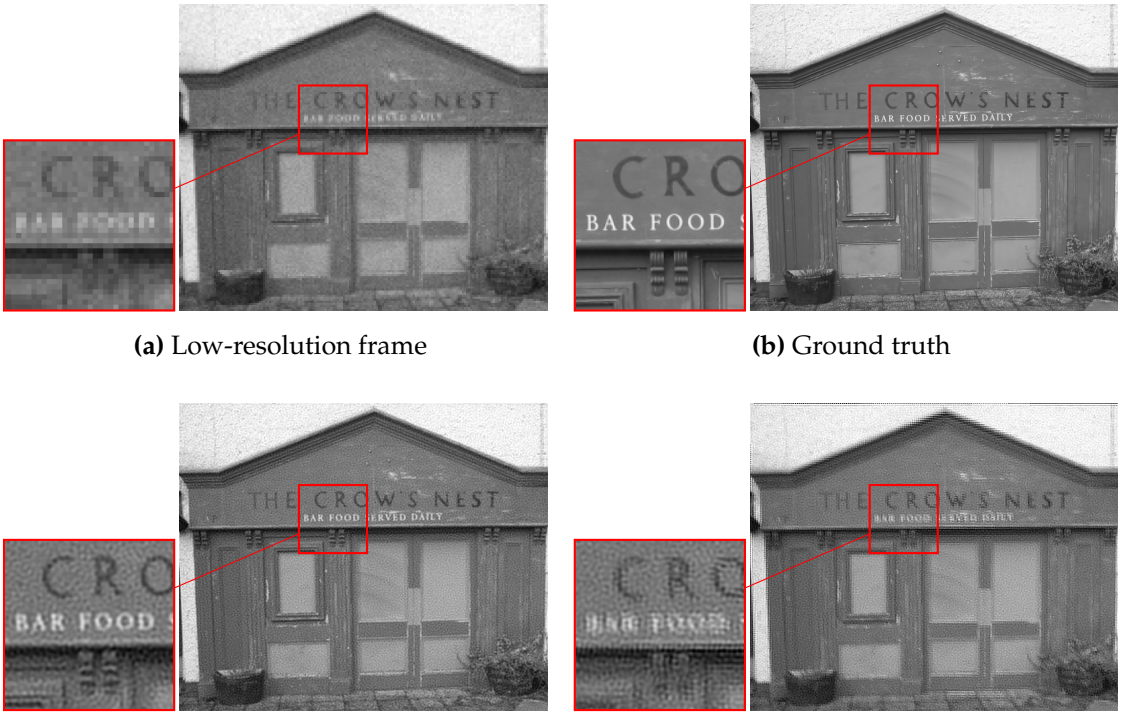
Robust Multi-Frame Super-Resolution with Sparse Regularization

The computational framework for multi-frame super-resolution as previously presented in Chapter 3 relies on a simplistic approximation of the true physics of image acquisition and accurate mathematical modeling of this process. For instance, it requires an accurate subpixel motion estimate and prior knowledge about the distribution of measurement noise. Uncertainty regarding these aspects limits the robustness of super-resolution in real-world applications. This chapter introduces a new algorithm for robust super-resolution imaging derived from a Bayesian point of view. The proposed method employs a confidence-aware observation model along with a sparse image prior and is implemented as iteratively re-weighted minimization. Unlike previous work, this approach features robust and edge preserving image reconstruction with small amount of parameter tuning, is flexible in terms of imaging models and computationally efficient.

Parts of this chapter have been originally published in [Kohl 16b] and have been later extended in [Berc 16].

4.1 Introduction

Robustness is one of the main design criteria for the development of multi-frame super-resolution algorithms. In this context, the term *robustness* refers to the ability of an algorithm to reconstruct a reasonable high-resolution image even in the presence of degenerated information employed in the reconstruction procedure. Conversely, an algorithm can be considered as not robust if it is severely affected by a small uncertainty of this information. In practice, such uncertainties are un-



(a) Low-resolution frame

(b) Ground truth

(c) Super-resolution with exact motion

(d) Super-resolution with inaccurate translation

Figure 4.1: Super-resolution under motion estimation uncertainty. (a) Simulated low-resolution frame. (b) High-resolution ground truth. (c) and (d) Super-resolved images (4× magnification) using an L_2 norm data fidelity term and Tikhonov regularization [Elad 97] with exact subpixel motion and inaccurate translations, respectively. The inaccurate motion estimate leads to ghosting artifacts due to the non-robust observation model.

avoidable and deteriorate super-resolution reconstruction. Let us discuss several aspects of practical relevance. These refer to motion estimation, image formation models, numerical optimization, and regularization.

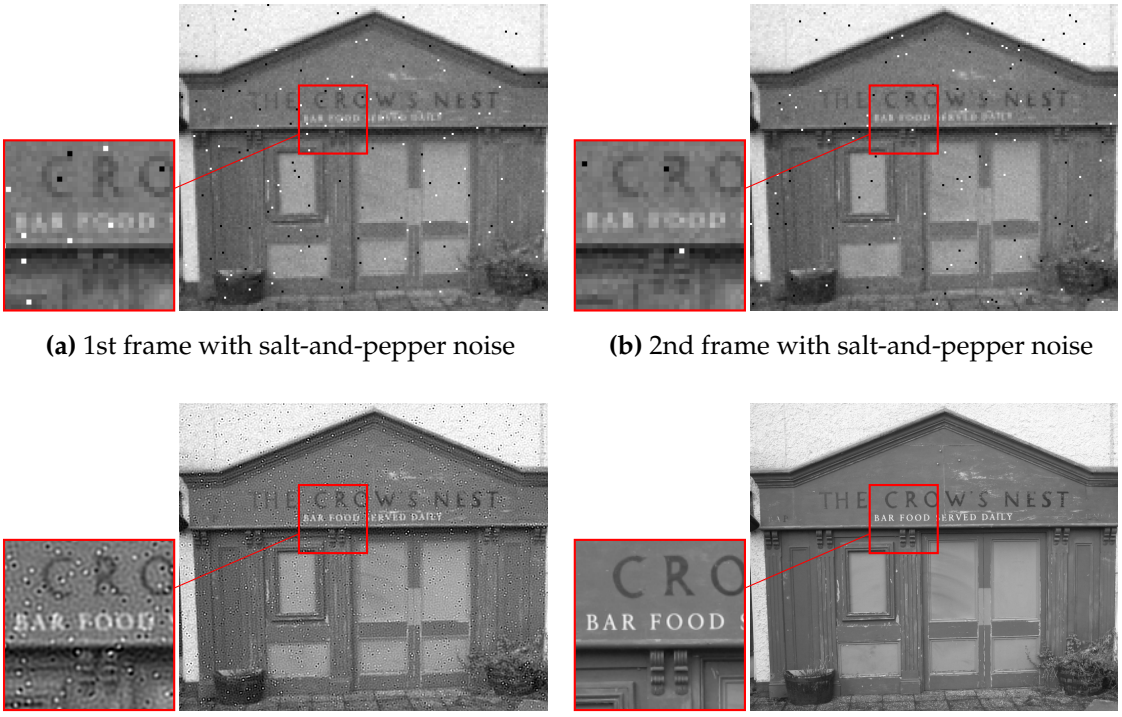
Motion Estimation Uncertainty. An important aspect for robust super-resolution is the uncertainty of subpixel motion. If this information is unknown, super-resolution requires an accurate motion estimate to provide reliable results. Unfortunately, this requirement is hard to fulfill using motion estimation on low-resolution images due to systematic artifacts like aliasing or blurring as well as random measurement noise [Vand 06b]. These issues cause uncertainties in the estimated subpixel motion. For this reason, motion estimation and super-resolution can be considered as a chicken-or-egg dilemma. In particular, this is the case if independently moving objects in a scene must be taken into account. Another challenging situation is non-rigid motion that causes ambiguities due to occlusions. In these cases, motion estimation may be affected by local outliers. In order to simplify motion estimation, some existing super-resolution algorithms are limited to simple parametric motion, e. g. globally rigid motion. However, parametric models are inappropriate in many applications, e. g. video upscaling [Kell 11].

The impact of motion estimation is demonstrated in Fig. 4.1, where $K = 16$ low-resolution frames (Fig. 4.1a) were generated from a ground truth (Fig. 4.1b) by simulating rigid motion and Gaussian noise. Super-resolution is based on an L_2 norm data fidelity and Tikhonov regularization using Laplacian filtering [Elad 97]. The super-resolved image using exact subpixel motion is depicted in Fig. 4.1c. Next, the translations $\boldsymbol{t} \in \mathbb{R}^2$ of every second frame were corrupted by uniform distributed errors $\boldsymbol{t}_\epsilon \in \mathbb{R}^2$ of random directions with $||\boldsymbol{t}_\epsilon||_2 = 1.0$. Super-resolution under this inaccurate motion is depicted in Fig. 4.1d. Notice that even under these small uncertainties, the reconstruction is severely affected by ghosting artifacts.

Model and Optimization Parameter Uncertainty. In addition to motion estimation, the uncertainties of parameters employed in the image formation model have a considerable impact on super-resolution. Compared to conditions in many real-world imaging setups, super-resolution usually approximates the true physics of image acquisition with simplified mathematical models. One aspect that is rarely considered is internal processing of image data in the camera system, e. g. image compression or white-balancing, see Section 3.2.3. Another issue is measurement noise that does not follow a simple space invariant normal distribution, e. g. due to invalid pixels related to impulse noise [Chan 05] or mixed noise [Xiao 11]. The influence of these aspects is demonstrated by corrupting low-resolution data by *salt-and-pepper* noise using a fraction of 0.5 % invalid pixels as depicted in Fig. 4.2a and Fig. 4.2b. Super-resolution on the corrupted low-resolution frames is not able to compensate for invalid pixels, see Fig. 4.2c.

Besides model parameters, there are also optimization parameters related to the formulation of super-resolution as energy minimization problem. One example are regularization weights that are selected prior to optimization. This is cumbersome as parameter selection is often performed off-line by trial-and-error or by automatic parameter selection schemes [Nguy 01a]. However, in both cases super-resolution is affected by an inappropriate selection and cannot compensate for the uncertainty of the parameters. This issue is demonstrated for the regularization weight λ in Fig. 4.3. In case of an underestimated regularization weight (Fig. 4.3a), super-resolution is affected by residual noise. As opposed to underestimation, an overestimated weight (Fig. 4.3c) results in oversmoothing. The optimal weight (Fig. 4.3b) results in a suitable tradeoff between residual noise and sharpness.

Ill-Posedness and Regularization. Super-resolution is known to be an ill-posed problem [Borm 04] and prior knowledge regarding the appearance of the images to be reconstructed is required to alleviate ill-posedness, see Section 3.3.2. This prior knowledge is leveraged by regularization techniques, where most parametric regularization terms are derived from Gaussian, Huber or TV priors. This has a crucial impact on the performance of super-resolution but most of these general-purpose priors are inadequate to model natural images [Bake 02]. For instance, discontinuities related to texture or edges are not explained appropriately due to the assumption of smooth or piecewise smooth images used to design these priors. This becomes crucial in presence of image noise as there is an inherent tradeoff between denoising and the preservation of edges or texture.



(a) 1st frame with salt-and-pepper noise

(b) 2nd frame with salt-and-pepper noise

(c) Super-resolution under salt-and-pepper noise

(d) Ground truth

Figure 4.2: Influence of invalid pixels to super-resolution. (a) and (b) Frames from Fig. 4.1 corrupted by salt-and-pepper noise. (c) and (d) Super-resolved image using an L_2 norm data fidelity term with Tikhonov regularization [Elad 97] and the ground truth. Notice that the non-robust observation model is unable to compensate for invalid pixels.

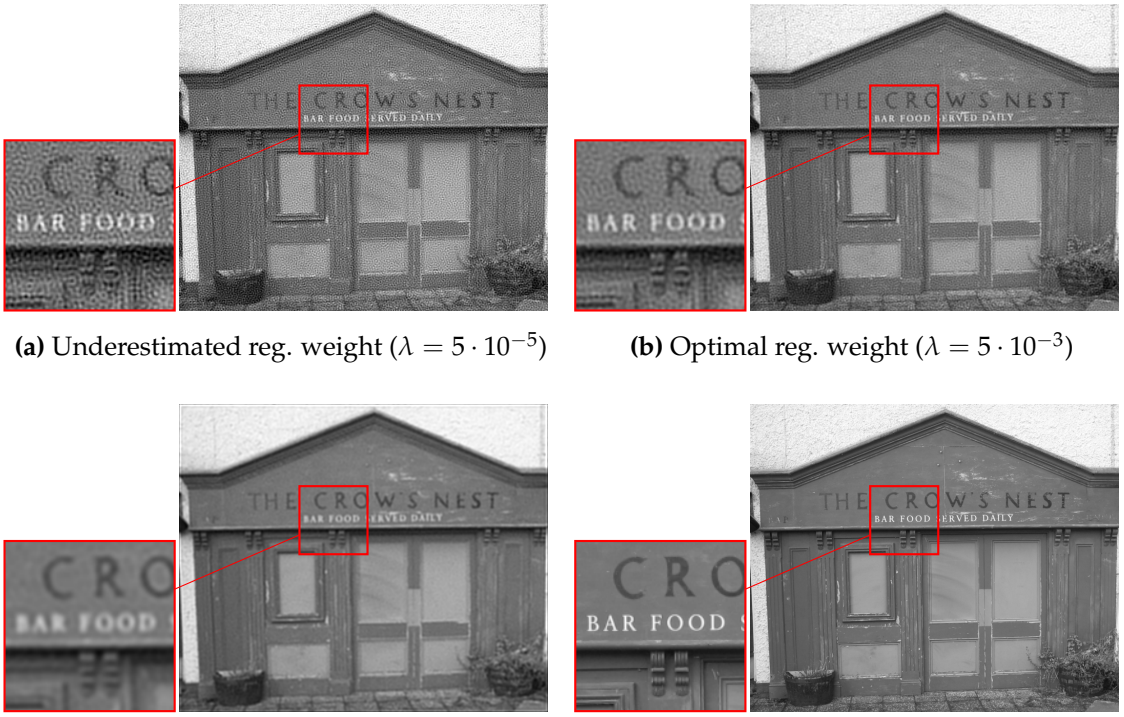
The remainder of this chapter is organized as follows. Section 4.2 presents a literature survey on related work in the area of robust super-resolution. Section 4.3 introduces a confidence-aware Bayesian model for robust super-resolution reconstruction. Section 4.4 introduces an iterative estimation scheme based on this model. Section 4.5 presents a comprehensive evaluation of this algorithm with comparisons to the state-of-the-art. Finally, Section 4.6 concludes this chapter.

4.2 Related Work

The algorithms most relevant to this work are based on Bayesian models. In particular, we are interested in the formulation of super-resolution as MAP estimation as well as related probabilistic methods. The focus in the design of robust algorithms in this area lies in outlier detection, optimization, and regularization, see Fig. 4.4.

Outlier Detection. The goal of outlier detection is to identify and downweight invalid observations termed *outliers*. According to Eq. (3.28), this is done by defining the data term:

$$L(\boldsymbol{x}) = \sum_{i=1}^{KM} \beta_i \left| \left[\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x} \right]_i \right|^p, \tag{4.1}$$



(a) Underestimated reg. weight ($\lambda = 5 \cdot 10^{-5}$)

(b) Optimal reg. weight ($\lambda = 5 \cdot 10^{-3}$)

(c) Overestimated reg. weight ($\lambda = 5 \cdot 10^{-1}$)

(d) Ground truth

Figure 4.3: Influence of the regularization weight λ to super-resolution. (a) - (d) Super-resolution using an L_2 norm data fidelity term and Tikhonov regularization [Elad 97] with an underestimated λ ($\lambda = 5 \cdot 10^{-5}$), an optimal λ ($\lambda = 5 \cdot 10^{-3}$), an overestimated λ ($\lambda = 5 \cdot 10^{-1}$) as well as the ground truth. Notice that an over- or underestimation lead to oversmoothing or noise amplification, respectively.

where $p \in [1,2]$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{KM})^\top$ denotes a confidence map to indicate valid observations referred to as *inliers*. The confidence β_i can be either a binary or a continuous variable, where $\beta_i = 1$ corresponds to an inlier y_i.

Deterministic approaches detect outliers in low-resolution observations. To this end, Zhao and Sawhney [Zhao 02] have proposed local image similarity assessment on displacement vector fields. In [Tana 10], Tanaka and Okutomi have proposed displacement estimation to construct confidence maps. This enables the detection of local outliers in optical flow. One drawback is that outlier detection does not exploit the presence of super-resolved data. Probabilistic strategies aggregate outlier detection and image reconstruction by expectation maximization (EM) [Harm 10, Cho 11]. These methods also focus on the detection of specific types of outliers, e. g. oversaturated pixels, by describing them in a probabilistic way.

In a different approach, Zomet et al. [Zome 01] have proposed robust gradient descent optimization. Instead of an explicit construction of the confidence map, outliers in the gradient descent update equations are filtered during iterative minimization. To remove outliers, median filtering of the gradient of the energy function is performed. This approach does not focus on specific types of outliers but has no proper theoretical justification in terms of convergence [Fars 04b].

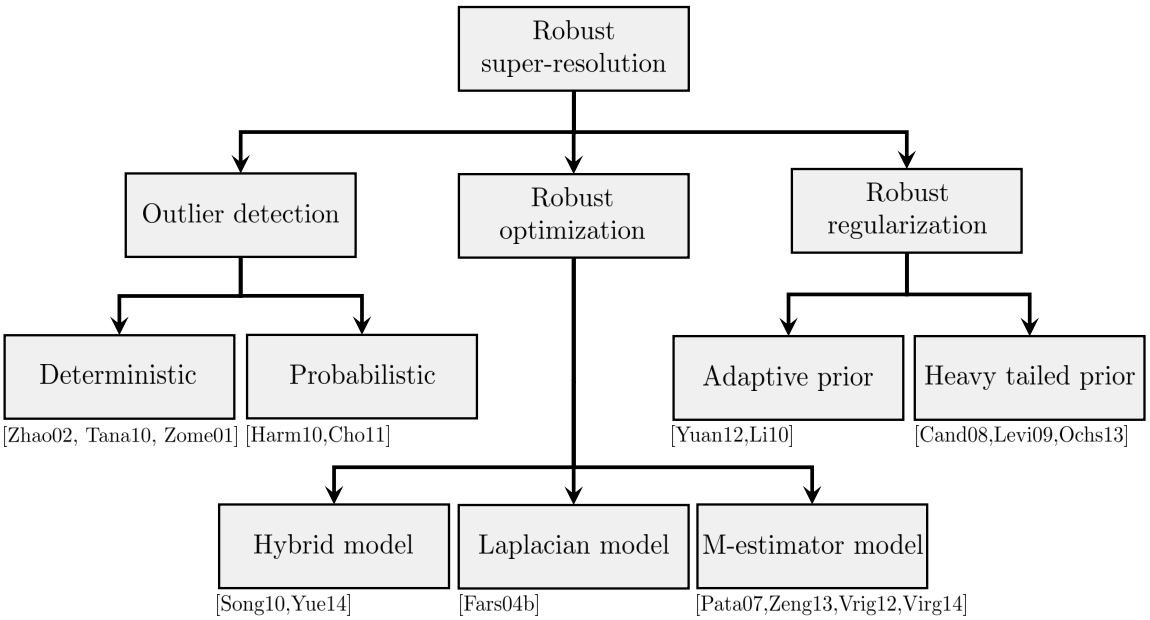


Figure 4.4: Overview of related work on robust super-resolution techniques.

Robust Optimization. In contrast to explicit outlier detection, outlier observations can be removed implicitly by defining the data fidelity as:

$$L(\boldsymbol{x}) = \sum_{i=1}^{KM} \phi_{\text{data}}([\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}]_i)\,, \tag{4.2}$$

where $\phi_{\text{data}}(\cdot)$ is a robust loss function. Robustness refers to the property that outliers are not penalized disproportionately as in case of the L_2 norm.

Farsiu et al. [Fars 04b] have introduced robust MAP estimation based on the L_1 norm, where the loss function in Eq. (4.2) is given by $\phi_{\text{data}}(z) = |z|$. This measure is statistically optimal in case of Laplacian noise. However, Laplacian noise is rarely the optimal model for inliers, where Gaussian noise is much more common. Inlier and outlier observations are also uniformly weighted and their influences are proportional to the magnitude of their residual errors. From a statistical point of view, this is not necessarily optimal. Besides the L_1 norm, *re-descending M-estimators* have been examined to design the loss function $\phi_{\text{data}}(z)$. The use of non-convex functions should further reduce the influence of outliers by rejecting them in the data fidelity term. In [Pata 07], Patanavijit and Jitapunkul have proposed the Lorentzian loss. Other widely used M-estimators are the Gaussian [Pham 08] or Tukey's biweight [Anas 09]. For the selection of the scale parameters of these functions, different adaptive schemes have been introduced [El Y 08a, El Y 08b]. Zeng and Yang [Zeng 13] have proposed an adaptive Huber function to consider a varying reliability of model parameters associated with the low-resolution frames. A similar approach is to employ hybrid error norms [Song 10, Yue 14], where $\phi_{\text{data}}(z)$ is an aggregation of the L_1 and the L_2 norm to combine their advantages.

One common issue of the aforementioned methods is that they rely on additional model parameters. For instance, regularization weights or scale parameters of M-estimators need to be specified. This task often requires user supervision or

ad-hoc methods based on empirical knowledge. Numerical methods for adaptive regularization weight selection have been developed by He and Kondi [He 06] and Vrigkas et al. [Vrig 12, Vrig 14]. However, these are based on Tikhonov regularization limiting the ability of edge reconstruction as discussed below.

Robust Regularization. Besides recognition-based image priors [Bake 02], most algorithms employ parametric prior distributions including Huber [Pick 07a] or TV models [Fars 04b, Ng 07]. These smoothness priors are characterized by convex regularization terms $R(\boldsymbol{x}) \propto -\log p(\boldsymbol{x})$ related to a distribution $p(\boldsymbol{x})$ and are not spatially adaptive. While this often leads to efficient algorithms, one inherent limitation is the ability to represent the characteristics of natural images in terms of sparsity. As we show in the derivation of the proposed prior, natural images are typically sparse and need to be represented by *heavy-tailed* distributions [Huan 99]. Such priors have been widely investigated for deblurring, where the Hyper-Laplacian distribution is a common choice [Levi 09, Kris 09, Kote 13]. In [Pata 07], Patanavijit and Jitapunkul have proposed non-convex Lorentzian-Laplacian regularization that implements a heavy-tailed prior for super-resolution.

Another class of priors aims at enhancing the parametric models to make them spatially adaptive. Yuan et al. [Yuan 12, Yuan 13] and Li et al. [Li 10] have presented spatially adaptive versions of TV and BTV, respectively. The idea is to decrease the impact of the regularization on discontinuities compared to the impact in homogenous regions. This may improve edge reconstruction compared to the unweighted counterparts of these priors. However, their benefit is highly dependent on additional feature extraction algorithms, e. g. the computation of second-order derivatives [Yuan 12] or entropy-based measures [Li 10].

Regularization of ill-posed problems has also been widely investigated in the theory of compressed sensing [Dono 06]. Here, sparse regularization is achieved by iteratively re-weighted L_1 norm optimization to approximate L_0 norm minimization. This scheme has been studied by Candes et al. [Cand 08] and Daubechies et al. [Daub 10] for sparse signal recovery, where it leads to sparser solutions compared to unweighted priors. This property makes these techniques attractive also for regularization in low-level vision problems [Ochs 13].

4.3 Bayesian Model for Robust Super-Resolution

This section introduces the mathematical model of the proposed super-resolution algorithm from a Bayesian perspective. This requires the definition of an observation model as well as a reasonable image prior. For both components, we propose space variant distributions to enhance space invariant modeling.

4.3.1 Space Variant Observation Model

In the most basic formulation of multi-frame super-resolution, the observation model $p(\boldsymbol{y} \mid \boldsymbol{x})$ is defined by a family of parametric distributions. For instance, one can employ a normal distribution [Elad 97] assuming additive Gaussian noise or a

Laplacian distribution [Fars 04b] assuming additive Laplacian noise in the image formation process, see Section 3.3. The main motivation behind this approach lies in its simplicity as noise can be fully described by a small number of parameters, e. g. its standard deviation, and the corresponding data fidelity term is convex. However, it has the shortcoming of being sensitive to outliers and cannot model spatially varying uncertainties, see Section 4.1.

We follow the assumption that the observation model can be reasonably described *locally* by a parametric distribution. Unlike Gaussian noise with fixed standard deviation for all observations, the proposed model employs a normal distribution with spatially varying standard deviation. This property is enforced by assigning pixel-wise confidence weights in spirit of outlier detection in Eq. (4.1). The observation model is given by the zero-mean weighted normal distribution $\mathcal{N}(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}; \boldsymbol{0}, \sigma^2_{\text{noise}}\boldsymbol{I}, \boldsymbol{\beta})$ defined by:

$$\begin{aligned} p(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\beta}) &= \mathcal{N}\left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}; \boldsymbol{0}, \sigma^2_{\text{noise}}\boldsymbol{I}, \boldsymbol{\beta}\right) \\ &:= \frac{1}{Z(\sigma_{\text{noise}}, \boldsymbol{\beta})} \exp\left\{-\frac{1}{2\sigma^2_{\text{noise}}}(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x})^\top \boldsymbol{B}(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x})\right\}, \end{aligned} \tag{4.3}$$

with normalization constant $Z(\sigma_{\text{noise}}, \boldsymbol{\beta})$, noise standard deviation σ_{noise}, and non-negative confidence weights $\boldsymbol{\beta} \in \mathbb{R}_0^{+KM}$ that are assembled to the diagonal matrix $\boldsymbol{B} = \operatorname{diag}(\beta_1, \ldots, \beta_{KM})$. In fact, assuming i. i. d. observations, Eq. (4.3) defines the normal distribution with spatially varying standard deviation:

$$p(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\beta}) \propto \prod_{m=1}^{KM} \exp\left\{-\frac{1}{2\sigma_m^2}\left[\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\right]_m^2\right\}, \tag{4.4}$$

where $\sigma_m = \sigma_{\text{noise}} / \sqrt{\beta_m}$ with $\beta_m \neq 0$ is the standard deviation at the m-th pixel.

This observation model shares several conceptual similarities with deterministic outlier detection. However, deterministic outlier detection is designed as a two-stage procedure, where confidence weights are determined on low-resolution data followed by super-resolution. This might lead to the selection of suboptimal confidence weights. In this chapter, we model the weights $\boldsymbol{\beta}$ as latent variables similar to probabilistic outlier detection [Harm 10, Cho 11] and estimate them simultaneously to the super-resolved image. This has the inherent advantage that we can gradually refine the confidence weights in an iterative algorithm.

4.3.2 Space Variant Image Prior

Similar to the observation model, most related works on image priors focused on parametric distributions that yield convex regularization terms in certain transform domains. In general, this transform domain is given by $\Omega_S \subset \mathbb{R}^{N_S}$ and a linear sparsifying transform of an image $\boldsymbol{x}$ is described by $S : \Omega_x \to \Omega_S$ with $S(\boldsymbol{x}) = \boldsymbol{S}\boldsymbol{x}$, where $\boldsymbol{S} \in \mathbb{R}^{N_S \times N}$ denotes the transform in matrix notation. The prior distribution exploits the sparse representation of an image under such transforms to regularize super-resolution reconstruction.

Next, different realization for $\boldsymbol{S}$ and its properties are compared. Based on these findings, a new sparsity-promoting prior for super-resolution is introduced.

Sparsity measure	Sparsifying transform $S(x)$				
	0th order		**1st order**		**2nd order**
	Identity	Roberts	Sobel	BTV	Laplacian
$s_{\text{Gini}}(z)$	0.26 ± 0.06	0.59 ± 0.06	0.59 ± 0.07	$\mathbf{0.68 \pm 0.05}$	0.61 ± 0.05
$s_{\text{Hoyer}}(z)$	0.09 ± 0.04	0.41 ± 0.08	0.41 ± 0.08	$\mathbf{0.53 \pm 0.06}$	0.43 ± 0.07
$s_{\text{Entropy}}(z)$	7.33 ± 0.34	5.46 ± 0.74	7.44 ± 0.76	$\mathbf{2.87 \pm 0.58}$	5.64 ± 0.69

Table 4.1: Sparsity of natural images from the LIVE database [Shei 16] in different transform domains. The sparsity is measured by mean $\pm$ standard deviation of the Gini index, the Hoyer index, and the entropy. The domains cover zeroth-order (identity), first-order (Roberts gradient, Sobel gradient, BTV) as well as second-order (Laplacian) transforms.

Analysis of Natural Image Statistics. The choice for the sparsifying transform $\boldsymbol{S}$ is based on an analysis of natural image statistics. This analysis studies different realizations for $\boldsymbol{S}$ based on high-pass filtering. These transforms include first-order methods based on the image gradient implemented by the Sobel operator, the Roberts operator as well as BTV ($L = 2$ and $\alpha_{\text{BTV}} = 0.5$). Moreover, the discrete Laplacian is examined as a second-order transform. For the sake of comparison, we also analyze the identity given by $\boldsymbol{S} = \boldsymbol{I}_{N\times N}$, which is considered as a zeroth-order transform. All transforms are compared by evaluating the sparsity of a transformed image $\boldsymbol{z} = \boldsymbol{S}\boldsymbol{x}$. In this context, sparsity refers to the amount of zero elements in the transform domain Ω_S.

To measure the sparsity of a transformed image $\boldsymbol{z}$ quantitatively, the Gini index $s_{\text{Gini}}(\boldsymbol{z})$, the Hoyer index $s_{\text{Hoyer}}(\boldsymbol{z})$ as well as the entropy $s_{\text{Entropy}}(\boldsymbol{z})$ are used [Hurl 09]. The higher $s_{\text{Gini}}(\boldsymbol{z})$ and $s_{\text{Hoyer}}(\boldsymbol{z})$ are, the higher the degree of sparsity of $\boldsymbol{z}$ obtained by the underlying transform. Conversely, a small entropy $s_{\text{Entropy}}(\boldsymbol{z})$ expresses higher sparsity. All measures are analyzed for 29 reference images of natural scenes that are available in the LIVE database [Shei 16], see Tab. 4.1. The implementation by the identity does not yield a sparse signal as the intensities itself do not follow a sparse distribution. Notice that the highest degree of sparsity is obtained by BTV indicating its efficiency to design sparse priors.

To prove the benefits of BTV to design the transform $\boldsymbol{S}$, we analyze the statistical distribution of $\boldsymbol{z}$. Figure 4.5 shows the discrete histogram assembled from all transformed samples z_i that are obtained from the 29 reference images using the BTV model. This demonstrates the clustering of the samples close to zero, which indicates sparsity of the transformed images. In order to model the histogram statistically, we employ the family of Hyper-Laplacian distributions [Kris 09]:

$$\begin{aligned} p(z_i) &= \mathcal{HL}\left(z_i; \mu_{\text{Prior}}, \sigma_{\text{prior}}, \nu\right) \\ &:= \frac{1}{Z(\sigma_{\text{prior}}, \nu)} \exp\left\{-\frac{1}{\nu}\left(\frac{|z_i - \mu_{\text{Prior}}|}{\sigma_{\text{prior}}}\right)^{\nu}\right\}, \end{aligned} \tag{4.5}$$

with shape parameter μ_{Prior}, scale parameters σ_{prior} and ν, and normalization constant $Z(\sigma_{\text{prior}}, \nu)$. This distribution is used to establish the image prior:

$$p(\boldsymbol{x}) = \prod_{i=1}^{N_S} \mathcal{HL}\left(z_i; \mu_{\text{Prior}}, \sigma_{\text{prior}}, \nu\right), \tag{4.6}$$

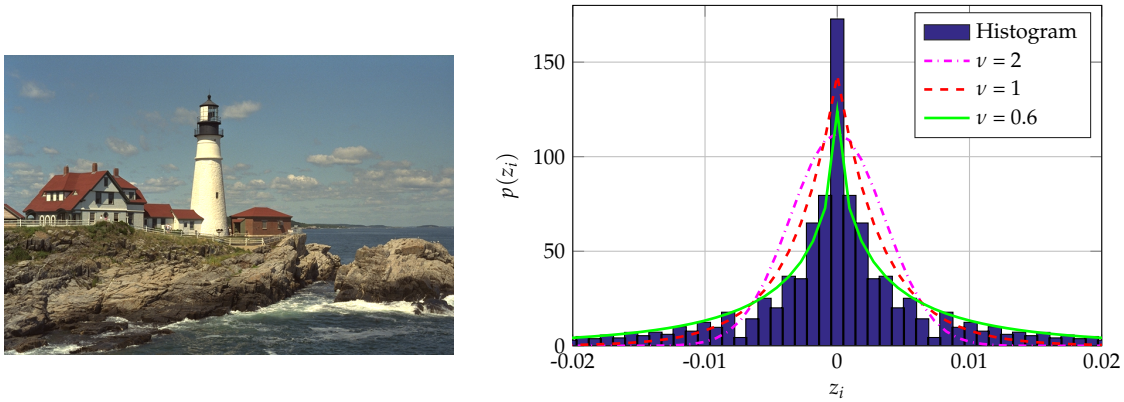


Figure 4.5: Analysis of the distribution $p(z_i)$ on 29 natural images [Shei 16] (left) using the BTV model. The discrete histogram (right) represents the empirical distribution of z_i in the reference images. The histogram is approximated by Hyper-Laplacian distributions using $\nu = 2$ (Gaussian), $\nu = 1$ (Laplacian), as well as $\nu = 0.6$ (heavy-tailed Hyper-Laplacian). Note the fit of the histogram tails for $\nu = 0.6$ compared to the Gaussian with $\nu = 2$.

where $z_1, \ldots, z_{N_S}$ are assumed to be i. i. d. variables. The Hyper-Laplacian is fitted to the discrete samples by ML estimation using $\nu = 2$ corresponding to a Gaussian $\mathcal{N}(z_i; \mu_{\text{Prior}}, \sigma^2_{\text{prior}})$, $\nu = 1$ corresponding to a Laplacian $\mathcal{L}(z_i; \mu_{\text{Prior}}, \sigma_{\text{prior}})$, as well as $\nu = 0.6$ corresponding to a heavy-tailed Hyper-Laplacian distribution.

Note that except for small z_i, Gaussian and Laplacian distributions provide poor fits to the statistical appearance of natural images. In Fig. 4.5, this is visible by the poor approximation of the histogram tails resulting in an inappropriate modeling of discontinuities. For $\nu < 1$, $p(z_i)$ follows a heavy-tailed distribution that is able to characterize the histogram tails in a reasonable way. This provides a better fit to the appearance of natural images than the Laplacian, which is consistent with recent findings in the area of natural scene statistics [Huan 99, Sriv 03]. For this reason, heavy-tailed priors became a common tool for image restoration [Levi 09, Kris 09, Kote 13] and compressed sensing [Cand 08, Daub 10].

Weighted Bilateral Total Variation. The design of the proposed image prior is motivated by these findings and exploits the sparsity of natural images in a transform domain Ω_S. However, instead of modeling the prior directly as a Hyper-Laplacian distribution, it is defined by the zero-mean weighted Laplacian distribution $\mathcal{L}(\boldsymbol{S}\boldsymbol{x}; \boldsymbol{0}, \sigma_{\text{prior}}\boldsymbol{I}, \boldsymbol{\alpha})$:

$$\begin{aligned} p(\boldsymbol{x} \,|\, \boldsymbol{\alpha}) &= \mathcal{L}\big(\boldsymbol{S}\boldsymbol{x}; \boldsymbol{0}, \sigma_{\text{prior}}\boldsymbol{I}, \boldsymbol{\alpha}\big) \\ &:= \frac{1}{Z(\sigma_{\text{prior}}, \boldsymbol{\alpha})} \exp\left\{ -\frac{||\boldsymbol{A}\boldsymbol{S}\boldsymbol{x}||_1}{\sigma_{\text{prior}}} \right\}, \end{aligned} \tag{4.7}$$

where σ_{prior} denotes a distribution scale parameter, $\boldsymbol{\alpha} \in \mathbb{R}_0^{+N_S}$ are confidence weights of the distribution in the transform domain assembled as the diagonal matrix $\boldsymbol{A} = \text{diag}\big(\alpha_1, \ldots, \alpha_{N_S}\big)$, and $Z(\sigma_{\text{prior}}, \boldsymbol{\alpha})$ is a normalization constant.

The regularization term associated with the distribution in Eq. (4.7) termed weighted bilateral total variation (WBTV) is based on the unweighted BTV in

Eq. (3.34) due to the performance of this approach to yield a sparse transform. Since ${\alpha_{\text{BTV}}}^{|m|+|n|} > 0$, we can reformulate the unweighted BTV according to:

$$\begin{aligned} R_{\text{BTV}}(\boldsymbol{x}) &= \sum_{m=-N_{\text{BTV}}}^{N_{\text{BTV}}} \sum_{n=-N_{\text{BTV}}}^{N_{\text{BTV}}} \left|\left| {\alpha_{\text{BTV}}}^{|m|+|n|} \left(\boldsymbol{I}_{N\times N} - \boldsymbol{S}_v^m \boldsymbol{S}_h^n\right) \boldsymbol{x} \right|\right|_1 \\ &= \sum_{m=-N_{\text{BTV}}}^{N_{\text{BTV}}} \sum_{n=-N_{\text{BTV}}}^{N_{\text{BTV}}} \left|\left| \boldsymbol{S}^{m,n} \boldsymbol{x} \right|\right|_1 = \left|\left| \boldsymbol{S} \boldsymbol{x} \right|\right|_1, \end{aligned} \tag{4.8}$$

where $\boldsymbol{S}^{m,n} = {\alpha_{\text{BTV}}}^{|m|+|n|}(\boldsymbol{I}_{N\times N} - \boldsymbol{S}_v^m \boldsymbol{S}_h^n) \in \mathbb{R}^{N\times N}$ denotes the transform associated with the shift (m,n). The overall transform $\boldsymbol{S} \in \mathbb{R}^{N_S \times N}$ with $N_S = (2N_{\text{BTV}}+1)^2 N$ for all shifts is assembled as:

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}^{-N_{\text{BTV}},-N_{\text{BTV}}} & \boldsymbol{S}^{-N_{\text{BTV}}+1,-N_{\text{BTV}}} & \dots & \boldsymbol{S}^{N_{\text{BTV}}-1,N_{\text{BTV}}} & \boldsymbol{S}^{N_{\text{BTV}},N_{\text{BTV}}} \end{pmatrix}^\top. \tag{4.9}$$

Then, WBTV regularization is conditioned on the weights $\boldsymbol{\alpha}$ according to:

$$R_{\text{WBTV}}(\boldsymbol{x} \,|\, \boldsymbol{\alpha}) := \left|\left| \boldsymbol{A}\boldsymbol{S}\boldsymbol{x} \right|\right|_1 = \sum_{m=-N_{\text{BTV}}}^{N_{\text{BTV}}} \sum_{n=-N_{\text{BTV}}}^{N_{\text{BTV}}} \sum_{i=1}^{N} \alpha_i^{m,n} \left| [\boldsymbol{S}^{m,n}\boldsymbol{x}]_i \right|, \tag{4.10}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^{-N_{\text{BTV}},-N_{\text{BTV}}}, \dots, \boldsymbol{\alpha}^{N_{\text{BTV}},N_{\text{BTV}}})^\top$ denotes the joint weight vector over all shifts and $\boldsymbol{\alpha}^{m,n} = (\alpha_1^{m,n}, \dots, \alpha_N^{m,n})^\top$ are weights associated with the shift (m,n).

This term allows us to locally adapt the prior $p(\boldsymbol{x} \,|\, \boldsymbol{\alpha})$ by controlling the weights $\boldsymbol{\alpha}$ similar to the locally adaptive BTV introduced by Li et al. [Li 10]. In particular, the impact of the regularization needs to be reduced on discontinuities compared to the behavior in flat regions. However, unlike [Li 10], the weights are handled as latent variables in the same way as those of the observation model. This avoids their explicit computation by means of feature detection in a preprocessing step.

4.3.3 Inference of the Model Confidence Weights

Our goal is to reconstruct the high-resolution image that best explains a set of low-resolution observations. If one knows the confidence weights employed in the Bayesian model, the high-resolution image could be inferred by the MAP framework presented in Section 3.3.2. However, if one does not know the weights, they need to be treated as latent variables in the estimation of the high-resolution image. For this purpose, three alternative approaches are examined.

Bayesian Marginalization. One approach is to marginalize over the latent variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Then, the high-resolution image is estimated from the marginal distribution. This Bayesian marginalization is formulated by:

$$\hat{\boldsymbol{x}} = \operatorname*{argmax}_{\boldsymbol{x}} \int_{\mathbb{R}^{KM}} \int_{\mathbb{R}^{N_S}} p(\boldsymbol{y}, \boldsymbol{\beta} \,|\, \boldsymbol{x})\, p(\boldsymbol{x}, \boldsymbol{\alpha})\, d\boldsymbol{\alpha} d\boldsymbol{\beta}, \tag{4.11}$$

where the integration is performed over all configurations of the confidence maps. Using the Bayes rule, marginalization over $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ yields:

$$\hat{\boldsymbol{x}} = \operatorname*{argmax}_{\boldsymbol{x}} \int_{\mathbb{R}_0^{+KM}} p(\boldsymbol{y} \,|\, \boldsymbol{\beta}, \boldsymbol{x})\, p(\boldsymbol{\beta}) \int_{\mathbb{R}_0^{+N_S}} p(\boldsymbol{x} \,|\, \boldsymbol{\alpha})\, p(\boldsymbol{\alpha})\, d\boldsymbol{\alpha} d\boldsymbol{\beta}, \tag{4.12}$$

where $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$ are the prior distributions assigned to the confidence weights.

Although this approach provides a theoretical basis for super-resolution under unknown confidence weights, there exist several practical limitations. One important restriction is that analytic marginalization is possible only for few relatively simple priors[1] or with simplistic approximations of the integration. An exact solution would require integration in a $KM + N_S$ dimensional space. This is computationally prohibitive for real-world applications, where the size of the parameter space lies in the range $KM + N_S \approx 10^6$. Another limitation is the parameter tuning that is required to define $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$ as these distributions would comprise additional hyperparameters.

Alternating MAP Estimation. As an alternative to marginalization, one can jointly estimate the super-resolved image and the confidence weights:

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \underset{\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}}{\operatorname{argmax}} \left\{ p(\boldsymbol{x} \,|\, \boldsymbol{\alpha}) \prod_{k=1}^{K} p\left(\boldsymbol{y}^{(k)} \,|\, \boldsymbol{x}, \boldsymbol{\beta}^{(k)}\right) p(\boldsymbol{\alpha})\, p(\boldsymbol{\beta}) \right\}, \tag{4.13}$$

where $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$ are priors for the confidence weights to obtain meaningful solutions of this underdetermined problem. Taking the negative log-likelihood of Eq. (4.13) leads to the joint energy minimization problem:

$$\begin{aligned}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \underset{\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}}{\operatorname{argmin}} \Bigg\{ & \sum_{i=1}^{KM} \beta_i \, |[\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}]_i|^2 + \lambda \sum_{i=1}^{N_S} \alpha_i \, |[\boldsymbol{S}\boldsymbol{x}]_i| \\ & + \log Z(\sigma_{\text{noise}}, \boldsymbol{\beta}) + \log Z(\sigma_{\text{prior}}, \boldsymbol{\alpha}) - \log p(\boldsymbol{\beta}) - \log p(\boldsymbol{\alpha}) \Bigg\}.\end{aligned} \tag{4.14}$$

This minimization problem can be solved by alternating MAP estimation for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{x}$. Hamada et al. [Hama 13] investigated this approach for a simplified model, where only confidence weights of a data fidelity term are taken into account. However, similar to Bayesian marginalization, the performance is highly dependent on the priors $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$ and only tractable for simplistic models.

Majorization-Minimization. Another approach to infer the confidence weights is to treat them as hidden information within an EM algorithm [Demp 77]. Related schemes have been successfully applied for probabilistic outlier removal [Cho 11]. However, similar to the aforementioned approaches, this concept requires a pure probabilistic formulation, i. e. an explicit definition of distributions $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$. These distributions are difficult to model for real-world problems, which limits the flexibility. For this reason, the proposed method is formulated as majorization-minimization (MM) algorithm [Hunt 04] as generalization of EM [Ba 14]. This does not require explicit modeling of $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$, and yields a computationally efficient approach for the inference of the confidence weights via *iteratively re-weighted minimization* [Cand 08, Scal 88].

[1] For instance, conjugate priors [Bish 06] to the data likelihood can be used. These priors enable an analytic calculation of the marginal distribution to infer latent hyperparameters [Oliv 09].

The iteratively re-weighted minimization framework comprises two steps, which results in a sequence of iterations $\{(\boldsymbol{x}^t, \boldsymbol{\alpha}^t, \boldsymbol{\beta}^t) : t = 1, \ldots, T\}$:

1. Given an estimate $\boldsymbol{x}^{t-1}$ for the latent high-resolution image, we first determine $\boldsymbol{\alpha}^t$ and $\boldsymbol{\beta}^t$ according to:

$$\boldsymbol{\alpha}^t = \alpha\left(\boldsymbol{x}^{t-1}\right), \tag{4.15}$$

$$\boldsymbol{\beta}^t = \beta\left(\boldsymbol{x}^{t-1}, \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}\right), \tag{4.16}$$

   where $\alpha : \mathbb{R}^{N_S} \to \mathbb{R}_0^{+N_S}$ and $\beta : \mathbb{R}^{KM} \to \mathbb{R}_0^{+KM}$ are *weighting functions* to compute confidence weights based on $\boldsymbol{x}^{t-1}$ and $\boldsymbol{y}$.

2. Given the weights $\boldsymbol{\alpha}^t$ and $\boldsymbol{\beta}^t$, we determine $\boldsymbol{x}^t$ according to the solution of the weighted minimization problem:

$$\boldsymbol{x}^t = \operatorname*{argmax}_{\boldsymbol{x}} \left\{ p\left(\boldsymbol{x} \,|\, \boldsymbol{\alpha}^t\right) \, p\left(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{\beta}^t\right) \right\}. \tag{4.17}$$

Both steps are alternated until convergence. For a detailed analysis of the relationship between this scheme and MM algorithms, we refer to Section 4.4.2.

4.4 Robust Super-Resolution Reconstruction

This section introduces a robust super-resolution algorithm based on iteratively re-weighted minimization. For the derivation of this algorithm, the basic computational steps for numerical optimization are outlined. Eventually, a theoretical study of the underlying Bayesian model and the proposed iteration scheme is provided by explicitly deriving this method as MM algorithm.

4.4.1 Iteratively Re-Weighted Minimization Algorithm

The general iteratively re-weighted minimization framework has some degrees of freedom that need to be adjusted for robust super-resolution. First, it utilizes weighting functions that need to be specified. Second, it assumes a fixed regularization weight λ that is adjusted prior to the iterative procedure. Hence, this approach is not adaptive regarding the characteristics of the low-resolution data.

The proposed algorithm is developed as adaptive iteration scheme. The weights $\boldsymbol{\alpha}^t$ and $\boldsymbol{\beta}^t$ are inferred by:

$$\boldsymbol{\alpha}^t = \alpha\left(\boldsymbol{x}^{t-1}, \sigma_{\text{noise}}^t\right), \tag{4.18}$$

$$\boldsymbol{\beta}^t = \beta\left(\boldsymbol{x}^{t-1}, \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}, \sigma_{\text{prior}}^t\right), \tag{4.19}$$

where $\alpha : \mathbb{R}^{N_S} \to \mathbb{R}_0^{+N_S}$ and $\beta : \mathbb{R}^{KM} \to \mathbb{R}_0^{+KM}$ are adaptive weighting functions that exploit σ_{noise}^t and σ_{prior}^t as scale parameters of the observation and the prior model. These scale parameters are adjusted at each iteration and characterize the data uncertainty in the underlying Bayesian model. To avoid manual parameter tuning, automatic hyperparameter estimation is used to determine the regularization weight λ^t at iteration t. Finally, the super-resolved image $\boldsymbol{x}^t$ is reconstructed using the confidence weights $\boldsymbol{\alpha}^t$ and $\boldsymbol{\beta}^t$ as well as the regularization weight λ^t.

Weight Estimation. To determine the confidence weights $\boldsymbol{\beta}^t$ of the observation model, the residual error $r(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}$ is analyzed. In this work, the weighting function in Eq. (4.18) is defined element-wise according to:

$$\beta(\boldsymbol{x}, \boldsymbol{y}, \sigma_{\text{noise}}) := \left(\beta_1(\boldsymbol{r}, \sigma_{\text{noise}}) \quad \dots \quad \beta_{KM}(\boldsymbol{r}, \sigma_{\text{noise}})\right)^\top \in \mathbb{R}_0^{+KM}, \tag{4.20}$$

where $\boldsymbol{r} = r(\boldsymbol{x}^{t-1}, \boldsymbol{y})$ denotes the residual error associated with the estimate $\boldsymbol{x}^{t-1}$ obtained at the previous iteration, and $\beta_i : \mathbb{R}^{KM} \to \mathbb{R}_0^+$ determines the weight for the i-th observation. The confidence weights are computed by considering frame-wise (global) outliers as well pixel-wise (local) outliers via the decomposition:

$$\beta_i(\boldsymbol{r}, \sigma_{\text{noise}}) := \underbrace{\beta_{i,\text{bias}}(\boldsymbol{r})}_{\text{frame-wise}} \cdot \underbrace{\beta_{i,\text{local}}(\boldsymbol{r}, \sigma_{\text{noise}})}_{\text{pixel-wise}}. \tag{4.21}$$

In order to detect outlier frames, we assume that the residual errors associated with the different frames need to be symmetric and zero-mean according to Eq. (4.3). Individual frames that violate this assumption are considered as outliers. Potential reasons for a violation of this assumption could be systematic errors like global photometric differences between the frames. We perform a bias detection [Zome 01] to identify such frames using the binary weighting function:

$$\beta_{i,\text{bias}}(\boldsymbol{r}) = \begin{cases} 1 & \text{if } \left|\text{median}\left(\boldsymbol{r}^{(k)}\right)\right| \leq c_{\text{bias}} \\ 0 & \text{otherwise} \end{cases}, \tag{4.22}$$

where $\boldsymbol{r}^{(k)}$ is the residual error of the k-th frame associated with the i-th observation, and $\text{median}(\cdot)$ denotes the sample median as robust estimator of the mean residual error [Zoub 12].

In addition to the detection of outlier frames, local outliers are detected pixel-wise using the bi-weight function:

$$\beta_{i,\text{local}}(\boldsymbol{r}, \sigma_{\text{noise}}) = \begin{cases} 1 & \text{if } |r_i| \leq c_{\text{local}}\sigma_{\text{noise}} \\ \frac{c_{\text{local}}\sigma_{\text{noise}}}{|r_i|} & \text{otherwise} \end{cases}, \tag{4.23}$$

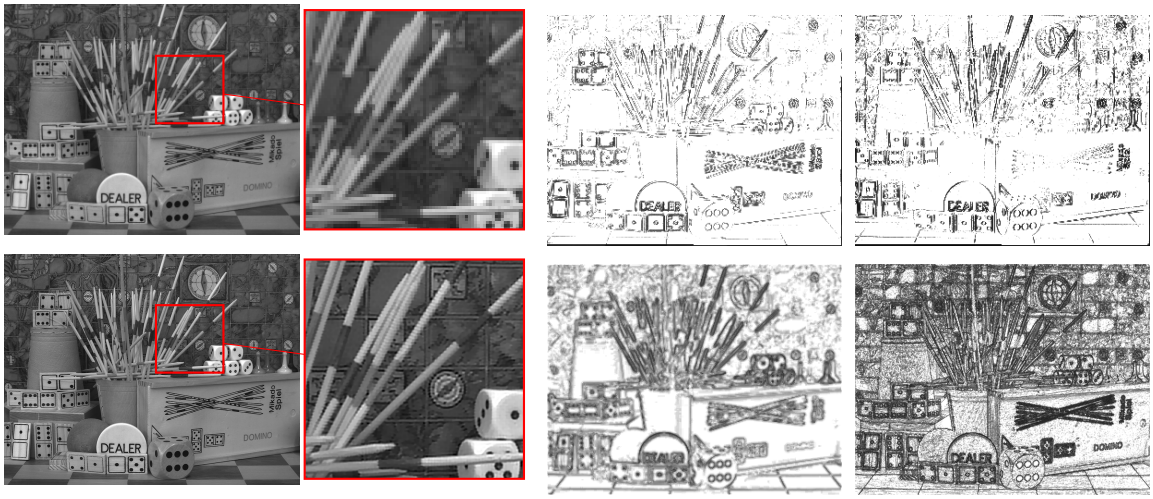
where σ_{noise} denotes an estimate of the standard deviation of the weighted normal distribution $\mathcal{N}(\boldsymbol{r}; \boldsymbol{0}, \sigma_{\text{noise}}^2 \boldsymbol{I}, \boldsymbol{\beta}^{t-1})$ and c_{local} is a tuning constant. Notice that a constant confidence is assigned to observations classified as inliers of a normal distribution, whereas outliers are weighted by their inverse residual errors. This function can downweight outliers related to non-Gaussian noise, e. g. impulsive noise, or locally inaccurate motion estimation.

The estimation of the prior weights $\boldsymbol{\alpha}^t$ in Eq. (4.19) follows a similar motivation and is done under the transform $\boldsymbol{z} = \boldsymbol{S}\boldsymbol{x}^{t-1}$ according to:

$$\alpha(\boldsymbol{x}, \sigma_{\text{prior}}) := \left(\alpha_1(\boldsymbol{z}, \sigma_{\text{prior}}) \quad \dots \quad \alpha_{N_S}(\boldsymbol{z}, \sigma_{\text{prior}})\right)^\top \in \mathbb{R}_0^{+N_S}. \tag{4.24}$$

The i-th weight is computed by the weighting function:

$$\alpha_i(\boldsymbol{z}, \sigma_{\text{prior}}) = \begin{cases} 1 & \text{if } |[Q(\boldsymbol{z})]_i| \leq c_{\text{prior}}\sigma_{\text{prior}} \\ p\frac{(c_{\text{prior}}\sigma_{\text{prior}})^{1-p}}{|[Q(\boldsymbol{z})]_i|^{1-p}} & \text{otherwise} \end{cases}, \tag{4.25}$$



(a) (b) (c)

Figure 4.6: Illustration of the proposed confidence weighting on an example image sequence. (a) Low-resolution image (top row) and super-resolved image with 4× magnification (bottom row) along with a zoom-in. (b) - (c) Gray-scale visualizations of the observation confidence weights (top row) and the prior weights (bottom row) after the 1st and the 10th iteration, respectively (bright regions denote higher weights). The observation weights identify outlier observations (e. g. due to inaccurate motion estimation) while the prior weights extract image structures for adaptive regularization.

where $p \in [0, 1]$ is referred to as *sparsity parameter*, σ_{prior} is an estimate of the scale parameter of the weighted Laplacian distribution $\mathcal{L}\left(\boldsymbol{z}; \boldsymbol{0}, \sigma_{\text{prior}} \boldsymbol{I}, \boldsymbol{\alpha}^{t-1}\right)$ and c_{prior} is a tuning constant. Note that in order to reduce the influence of isolated noisy pixels, these weights are inferred from a locally filtered version of the spatial information denoted by $Q(\boldsymbol{z})$. In this work, $Q(\cdot)$ is implemented by a 3×3 median filtering[2]. This scheme explains an image as a mixture of flat regions and discontinuities by assigning spatially adaptive weights. Accordingly, higher weights are assigned to flat regions while the influence of discontinuities is downweighted.

In Fig. 4.6, we illustrate the proposed weighting functions employed for iteratively re-weighted minimization. Figure 4.6 (top row) depicts the observation confidence weights associated with a single low-resolution frame affected by insufficient motion estimation. The weights are iteratively refined and model the low-resolution observations by mixed noise. In this example, mixed noise is related to the superposition of measurement noise and motion estimation uncertainty. Figure 4.6 (bottom row) depicts the prior weights in the domain of the super-resolved image. These weights are gradually refined over the iterations in order to make regularization spatially adaptive w. r. t. image structures. More specifically, lower weights are assigned to sharp edges to enhance their reconstruction.

Scale Parameter Estimation. The weighting functions in Eq. (4.23) and Eq. (4.25) require the knowledge of the scale parameters σ_{noise} and σ_{prior}, respectively. In

[2] In [Yuan 13], a related method has been proposed, where median filtering is used to extract edge information. This avoids the origination of false edges in spatially adaptive TV regularization.

order to avoid manual parameter tuning, both parameters are determined in an optimal way at each iteration. Given $\boldsymbol{x}^{t-1}$ and $\boldsymbol{\beta}^{t-1}$ obtained at the previous iteration and assuming a uniform prior $p(\sigma_{\text{noise}})$, we determine σ_{noise}^t via the ML estimator:

$$\sigma_{\text{noise}}^t = \underset{\sigma_{\text{noise}}}{\operatorname{argmax}}\, p\left(\boldsymbol{y} \,|\, \boldsymbol{x}^{t-1}, \boldsymbol{\beta}^{t-1}, \sigma_{\text{noise}}\right). \tag{4.26}$$

For outlier-insensitive estimation in Eq. (4.26), the scale parameter is computed from the median absolute deviation (MAD) [Scal 88] of $\boldsymbol{r}^{t-1} = \boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}^{t-1}$. In order to take different confidence weights associated with the low-resolution observations into account, the MAD is computed in a weighted version using the weighted median [Yin 96, Zhan 14b]:

$$\begin{aligned}\sigma_{\text{noise}}^t &= \sigma_0 \cdot \operatorname{mad}\left(\boldsymbol{r}^{t-1}, \boldsymbol{\beta}^{t-1}\right) \\ &= \sigma_0 \cdot \operatorname{median}\left(\begin{pmatrix} \left|r_1^{t-1} - \operatorname{median}\left(\boldsymbol{r}^{t-1}, \boldsymbol{\beta}^{t-1}\right)\right| \\ \vdots \\ \left|r_{KM}^{t-1} - \operatorname{median}\left(\boldsymbol{r}^{t-1}, \boldsymbol{\beta}^{t-1}\right)\right| \end{pmatrix}, \begin{pmatrix} \beta_1^{t-1} \\ \vdots \\ \beta_{KM}^{t-1} \end{pmatrix}\right),\end{aligned} \tag{4.27}$$

where we set $\sigma_0 = 1.4826$ to obtain a consistent estimate for the standard deviation of a normal distribution [Scal 88]. The quantities $\operatorname{mad}(\boldsymbol{r}, \boldsymbol{\beta})$ and $\operatorname{median}(\boldsymbol{r}, \boldsymbol{\beta})$ denote the weighted MAD and the weighted median of the residual error $\boldsymbol{r}$ under the confidence weights $\boldsymbol{\beta}$, respectively. The weighted median $\tilde{r} = \operatorname{median}(\boldsymbol{r}, \boldsymbol{\beta})$ generalizes the sample median and is defined as the point $\tilde{r}$, where the sum of the weights β_i associated with residuals r_i above and below $\tilde{r}$ fulfills:

$$\sum_{i:r_i<\tilde{r}} \beta_i < \frac{1}{2}\sum_i \beta_i \quad \text{and} \quad \sum_{i:r_i\geq\tilde{r}} \beta_i \leq \frac{1}{2}\sum_i \beta_i. \tag{4.28}$$

Similarly, the ML estimate for the scale parameter σ_{prior}^t is obtained from the distribution of $\boldsymbol{S}\boldsymbol{x}^{t-1}$. Given the weights $\boldsymbol{\alpha}^{t-1}$ determined at the previous iteration, the scale parameter at iteration t is determined by:

$$\sigma_{\text{prior}}^t = \sigma_0 \cdot \operatorname{mad}\left(Q\left(\boldsymbol{S}\boldsymbol{x}^{t-1}\right), \boldsymbol{\alpha}^{t-1}\right), \tag{4.29}$$

where $\sigma_0 = 1$ for the Laplacian distribution.

Hyperparameter Estimation. The selection of the regularization parameter λ has to deal with the following inherent tradeoff. On the one hand, if λ is underestimated, super-resolution is ill-conditioned and the reconstructed images are affected by residual noise. On the other hand, in case of an overestimate, the super-resolved images get blurred as illustrated in Fig. 4.3. In general, an optimal regularization weight is unknown and manual tuning based on trial-and-error procedures is time-consuming and error prone. In the proposed approach, an optimal λ also depends on the estimated confidence weights. Fully automatic approaches to select λ use, e. g. generalized cross validation (GCV) [Nguy 01a], the discrepancy principle [Wen 12], or Bayesian methods [Oliv 09, Baba 11]. Typically these methods deal with simplistic prior distributions, e. g. Gaussian priors [Vrig 14], or use

approximative schemes to determine a closed-form solution for the prior partition function [Oliv 09] to make parameter selection tractable.

In this work, a data-driven parameter selection that generalizes fairly well to different forms of the image prior is used. This approach is inspired by the work of Pickup et al. [Pick 07b] and is based on a two-fold cross validation like procedure that estimates the regularization parameter λ jointly with the super-resolved image. The advantage of this approach is that λ is adjusted at each iteration t as λ^t and the parameter selection exploits the model confidence weights as opposed to parameter selection prior to super-resolution. The key idea is to determine λ^t based on training observations such that it minimizes a cross validation error on a disjoint set of validation observations. For this purpose, the low-resolution observations $\boldsymbol{y}$ are decomposed into two disjoint subsets, where a fraction of δ, $0 < \delta < 1$ observations are used for parameter training and the remaining observations are hold back for validation. This is achieved by assembling a binary diagonal matrix $\boldsymbol{I}_\delta \in \{0,1\}^{KM \times KM}$, where the i-th element is $I_{\delta,i} = 1$ with probability δ to specify the training subset and $I_{\delta,i} = 0$ with probability $1 - \delta$ to specify the validation subset. Given a regularization weight λ, the super-resolved image reconstructed with this setting from the training observations is denoted by:

$$\boldsymbol{x}(\lambda) = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ (\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x})^\top \boldsymbol{I}_\delta \boldsymbol{B}^t (\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}) + \lambda \left|\left| \boldsymbol{A}^t \boldsymbol{S}\boldsymbol{x} \right|\right|_1 \right\}, \tag{4.30}$$

where $\boldsymbol{A}^t = \operatorname{diag}(\boldsymbol{\alpha}^t)$ and $\boldsymbol{B}^t = \operatorname{diag}(\boldsymbol{\beta}^t)$. The optimal weight λ^t is determined from the validation observations according to:

$$\lambda^t = \underset{\lambda}{\operatorname{argmin}}\, L_{\mathrm{cv}}\left(\lambda, \overline{\boldsymbol{I}_\delta}\right). \tag{4.31}$$

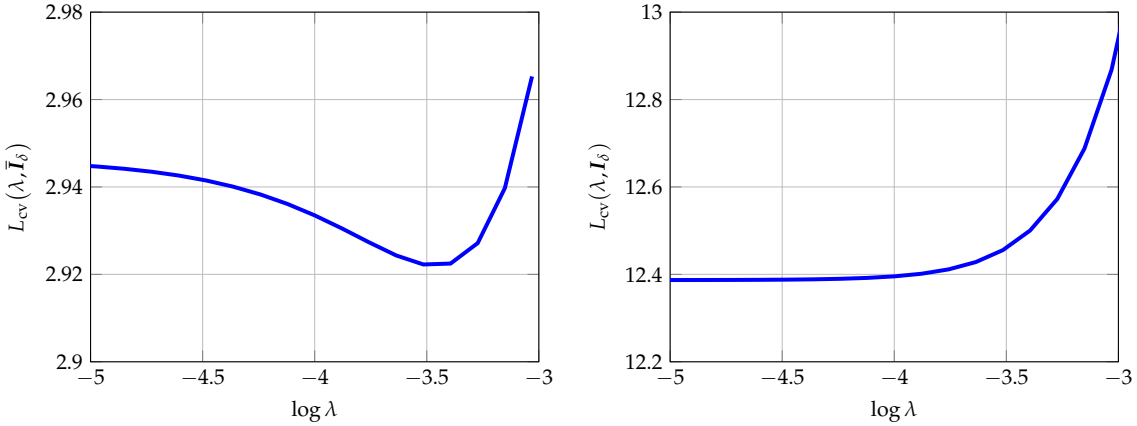
The cross validation error measures the fidelity of $\boldsymbol{x}(\lambda)$ on the validation set and is given by:

$$L_{\mathrm{cv}}\left(\lambda, \overline{\boldsymbol{I}_\delta}\right) = \left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}(\lambda)\right)^\top \overline{\boldsymbol{I}_\delta} \boldsymbol{B}^t \left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}(\lambda)\right), \tag{4.32}$$

where $\overline{\boldsymbol{I}_\delta}$ is obtained from $\boldsymbol{I}_\delta$ by flipping the diagonal elements. The behavior of this cross validation error is visualized in Fig. 4.7 in the log-transformed range $\log\lambda$. Unlike the error on the training observations $L_{\mathrm{cv}}(\lambda, \boldsymbol{I}_\delta)$ that is strictly monotonic increasing in Fig. 4.7b, the optimal regularization weight is the minimum of $L_{\mathrm{cv}}\left(\lambda, \overline{\boldsymbol{I}_\delta}\right)$ in Fig. 4.7a.

Notice that the minimization problem in Eq. (4.31) itself depends on an optimization problem. This makes the application of gradient-based optimization [Pick 07b] difficult as the gradient $\nabla L_{\mathrm{cv}}\left(\lambda, \overline{\boldsymbol{I}_\delta}\right)$ is not well-defined and its numerical approximation would be computationally expensive. To this end, the proposed parameter selection utilizes an adaptive grid search to solve Eq. (4.31). This search is performed in the domain of the log-transformed regularization weight $\log\lambda$ instead of using the linear space for λ. For the first iteration of the proposed algorithm, λ^1 is selected as the global minimum of $L_{\mathrm{cv}}\left(\lambda, \overline{\boldsymbol{I}_\delta}\right)$. For this task, a grid search is performed over the initial search range $[\log\lambda_l, \log\lambda_u]$. In the subsequent iterations ($t > 1$), λ^{t-1} is used as initial guess to refine it to λ^t with search range:

$$\left[\log\lambda^{t-1} - \frac{1}{t},\ \log\lambda^{t-1} + \frac{1}{t}\right]. \tag{4.33}$$



(a) Error on the validation observations

(b) Error on the training observations

Figure 4.7: Behavior of the cross validation error for the selection of an optimal regularization weight. (a) Cross validation error $L_{\text{cv}}(\lambda, \overline{\boldsymbol{I}_\delta})$ in the log-transformed range $\log \lambda$ on the validation observations. (b) Error $L_{\text{cv}}(\lambda, \boldsymbol{I}_\delta)$ on the training observations. Note that $L_{\text{cv}}(\lambda, \boldsymbol{I}_\delta)$ is strictly monotonic increasing while $L_{\text{cv}}(\lambda, \overline{\boldsymbol{I}_\delta})$ has a unique minimum.

The number of iterations is adaptively adjusted at each iteration of iteratively reweighted minimization. For $t = 1$, it is initialized by $T_{\text{cv}}^1 = T_{\text{cv}}$. Then, it is gradually reduced to $T_{\text{cv}}^t = \lceil 0.5 \cdot T_{\text{cv}}^{t-1} \rceil$. This enables a parameter selection of moderate computational effort while avoiding the limitation of a gradient-based search.

Image Reconstruction. Once the confidence weights $\boldsymbol{\alpha}^t$ and $\boldsymbol{\beta}^t$ as well as the regularization weight λ^t are determined, the super-resolved image $\boldsymbol{x}^t$ is estimated via the weighted energy minimization problem:

$$\boldsymbol{x}^t = \underset{\boldsymbol{x}}{\operatorname{argmin}}\, F^t(\boldsymbol{x}). \tag{4.34}$$

The energy function that is minimized at iteration t is given by:

$$F^t(\boldsymbol{x}) = (\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x})^\top \boldsymbol{B}^t (\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}) + \lambda^t \left|\left|\boldsymbol{A}^t \boldsymbol{S}\boldsymbol{x}\right|\right|_1, \tag{4.35}$$

where $\boldsymbol{A}^t = \operatorname{diag}(\boldsymbol{\alpha}^t)$ and $\boldsymbol{B}^t = \operatorname{diag}(\boldsymbol{\beta}^t)$. This convex and unconstrained minimization problem provides an MAP estimate for $\boldsymbol{x}^t$ under the given parameters and is numerically solved by means of gradient-based techniques. In this thesis, we employ scaled conjugate gradient (SCG) iterations [Nabn 02] to solve for $\boldsymbol{x}^t$ and to enhance the rate of convergence compared to steepest descent schemes. SCG iterations seek a stationary point:

$$\nabla_{\boldsymbol{x}} F^t(\boldsymbol{x}) = -2\boldsymbol{B}^t \boldsymbol{W}^\top (\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}) + \lambda^t \boldsymbol{A}^t \boldsymbol{S}^\top \operatorname{sign}(\boldsymbol{A}^t \boldsymbol{S}\boldsymbol{x}) \stackrel{!}{=} \boldsymbol{0}. \tag{4.36}$$

Numerical optimization requires a smooth and continuous differentiable regularization term to facilitate gradient-based iterations. Therefore, the regularization term is approximated by the convex Charbonnier function [Char 94]:

$$\phi_{\text{Char}}(\boldsymbol{z}) := \sum_{i=1}^{N_S} \sqrt{z_i^2 + \tau}. \tag{4.37}$$

For small τ ($\tau = 10^{-4}$), this provides a reasonable approximation of the L_1 norm while avoiding the non-differentiability. Consequently, the gradient of the regularization term is given by:

$$\boldsymbol{A}^t\boldsymbol{S}^\top \operatorname{sign}\left(\boldsymbol{A}^t\boldsymbol{S}\boldsymbol{x}\right) \approx \boldsymbol{A}^t\boldsymbol{S}^\top \cdot \psi_{\text{Char}}(\boldsymbol{A}^t\boldsymbol{S}\boldsymbol{x}), \tag{4.38}$$

where $\psi_{\text{Char}}(z) = \nabla_z \phi_{\text{Char}}(z)$ is the gradient of the Charbonnier function:

$$\psi_{\text{Char}}(\boldsymbol{z}) = \begin{pmatrix} \psi_{\text{Char}}(z_1) & \psi_{\text{Char}}(z_2) & \dots & \psi_{\text{Char}}(z_{N_S}) \end{pmatrix}^\top \tag{4.39}$$

$$\psi_{\text{Char}}(z_i) = z_i \left(\sqrt{z_i^2 + \tau} \right)^{-1}. \tag{4.40}$$

Coarse-to-Fine Optimization. Although re-weighted minimization according to Eq. (4.34) is convex, it is important to note that the overall optimization problem solved by iteratively re-weighted minimization is non-convex. Intuitively, this is caused by the fact that the convergence is affected by the initialization of the confidence weights and thus it may converge to different local minimums. For this reason, iteratively re-weighted minimization is implemented in a *coarse-to-fine* scheme as shown in Algorithm 4.1. In this approach, the initial confidence weights are set to $\boldsymbol{\alpha}^0 = \boldsymbol{1}$ and $\boldsymbol{\beta}^0 = \boldsymbol{1}$, where $\boldsymbol{1}$ is an all-one vector. The super-resolved image $\boldsymbol{x}^0$ is initialized by the temporal median of the motion compensated low-resolution frames that severs as an outlier-insensitive initial guess [Fars 03]. The magnification factor is initialized by a lower value than the desired magnification starting with $s^1 = 1$. Then, it is gradually increased by Δs per iteration such that $s^t = s^{t-1} + \Delta s$ until the desired magnification is reached. In order to solve Eq. (4.34), $\boldsymbol{x}^{t-1}$ is propagated as initial guess to determine $\boldsymbol{x}^t$.

We perform a maximum number of T_{irwsr} iterations in the outer optimization loop, a maximum number of T_{scg} iterations for SCG in the inner loop, and an initial number of T_{cv} iterations for cross validation based parameter selection. As a termination criterion we use the absolute difference between $\boldsymbol{x}^t$ and $\boldsymbol{x}^{t-1}$ according to:

$$\max_{i=1,\dots,N} \left(\left| x_i^{t-1} - x_i^t \right| \right) < \eta, \tag{4.41}$$

where η denotes the termination tolerance.

This approach has two benefits compared to single-scale optimization. First, it reduces the risk of getting stuck in local minimums as the non-convexity of the energy function in a lower dimensional space is less crucial. Second, it reduces the computational costs as more iterations of the computational demanding hyper-parameter estimation are done more efficiently for smaller magnification factors.

4.4.2 Algorithm Analysis

In this section, we analyze Algorithm 4.1 regarding the following aspects. First and foremost, we discuss the relationship of iteratively re-weighted minimization to MM algorithms. This links the proposed weighted optimization to the solution of a non-convex energy minimization problem. Afterwards, based on this relationship to the MM theory, we prove the convergence of the underlying iteration scheme.

Algorithm 4.1 Super-resolution using iteratively re-weighted minimization

Input: Initial guess for $\boldsymbol{x}^0$ (high-resolution image), $\boldsymbol{\alpha}^0$ and $\boldsymbol{\beta}^0$ (confidence weights), s^0 (magnification factor), and $[\log \lambda_l, \log \lambda_u]$ (regularization weight search range)
Output: Final high-resolution image $\boldsymbol{x}$, confidence weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and regularization weight λ

while Convergence criterion in Eq. (4.41) not fulfilled and $t \leq T_{\text{irwsr}}$ **do**
  Select magnification factor $s^t = \min(s^{t-1} + \Delta s, s)$
  Propagate $\boldsymbol{x}^{t-1}$ in coarse-to-fine scheme using magnification factor s^t
  Compute scale parameters σ^t_{noise} and σ^t_{prior} according to Eq. (4.27) – (4.29)
  Compute confidence weights $\boldsymbol{\alpha}^t$ and $\boldsymbol{\beta}^t$ according to Eq. (4.22) – (4.25)
  Compute regularization weight λ^t according to Eq. (4.31) with T^t_{cv} iterations
  $t_{\text{scg}} \leftarrow 1$
  while Convergence criterion in Eq. (4.41) not fulfilled and $t_{\text{scg}} \leq T_{\text{scg}}$ **do**
    Update $\boldsymbol{x}^t$ by SCG iteration for Eq. (4.36)
    $t_{\text{scg}} \leftarrow t_{\text{scg}} + 1$
  end while
  $t \leftarrow t + 1$
end while

Relationship to Majorization-Minimization Algorithms. The proposed super-resolution method can be considered as MM algorithm, see Section 4.3. The basic notion of this class of algorithms is to replace the direct minimization of a difficult – potentially non-convex function – with the minimization of a surrogate function. Compared to the original non-convex function, this surrogate function is easier to optimize. A surrogate function that can be employed in this context is referred to as *majorizing function* [Hunt 04] and is defined as follows.

Definition 4.1 (Majorizing function). *Let $F(\boldsymbol{x})$ be a real-valued function. Then, the real-valued function $\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1})$ is called a majorizing function for $F(\boldsymbol{x})$ at $\boldsymbol{x}^{t-1} \in \mathbb{R}^N$ if:*

1. *$\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1}) \geq F(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^N$, and*
2. *$\tilde{F}(\boldsymbol{x}^{t-1}, \boldsymbol{x}^{t-1}) = F(\boldsymbol{x}^{t-1})$.*

Let us now consider the robust and sparse reconstruction given by the minimum of the non-convex energy function:

$$F(\boldsymbol{x}) = \sum_{i=1}^{KM} \phi_{\text{Huber}}([\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}]_i) + \lambda \sum_{i=1}^{N_S} \phi_p([\boldsymbol{S}\boldsymbol{x}]_i)\,, \tag{4.42}$$

where the data fidelity is given by the Huber loss with scale parameter σ_{noise}:

$$\phi_{\text{Huber}}(z) = \begin{cases} z^2 & \text{if } |z| \leq \sigma_{\text{noise}} \\ 2\sigma_{\text{noise}}|z| - \sigma^2_{\text{noise}}, & \text{otherwise.} \end{cases}, \tag{4.43}$$

and the regularization term is defined by the mixed L_1/L_p norm with $p \in [0,1]$ and scale parameter σ_{prior}:

$$\phi_p(z) = \begin{cases} |z| & \text{if } |z| \leq \sigma_{\text{prior}} \\ \sigma^{1-p}_{\text{prior}}|z|^p & \text{otherwise,} \end{cases}. \tag{4.44}$$

This function comprises an outlier-insensitive data fidelity term and sparse regularization. Moreover, let us define the convex energy:

$$\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1}) = F^t(\boldsymbol{x}, \boldsymbol{x}^{t-1}) + \sum_{i=1}^{KM} \rho\left(\left[\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}^{t-1}\right]_i\right) + \lambda \sum_{i=1}^{N_S} \tau\left(\left[\boldsymbol{S}\boldsymbol{x}^{t-1}\right]_i\right), \tag{4.45}$$

where:

$$\rho(z) = \begin{cases} 0 & \text{if } z \leq \sigma_{\text{noise}} \\ \sigma_{\text{noise}}^2 \left(\frac{z}{\sigma_{\text{noise}}} - 1\right) & \text{otherwise} \end{cases}, \tag{4.46}$$

$$\tau(z) = \begin{cases} 0 & \text{if } |z| \leq \sigma_{\text{prior}} \\ (1-p)\sigma_{\text{prior}}^{1-p}|z|^p & \text{otherwise} \end{cases}, \tag{4.47}$$

and $F^t(\boldsymbol{x}, \boldsymbol{x}^{t-1})$ is the energy function in Eq. (4.34) as optimized by Algorithm 4.1 with regularization weight λ. Notice that $\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1})$ and $F^t(\boldsymbol{x}, \boldsymbol{x}^{t-1})$ are equal up to the non-negative terms $\rho(\cdot)$ and $\tau(\cdot)$ that are independent of $\boldsymbol{x}$. Thus, $\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1})$ is an upper bound for $F^t(\boldsymbol{x}, \boldsymbol{x}^{t-1})$ and the minimizer of these functions w. r. t. $\boldsymbol{x}$ are equivalent.

The relation of iteratively re-weighted minimization to MM algorithms is established by the following theorem.

Theorem 4.1. *The convex energy function $\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1})$ in Eq. (4.45) is a majorizing function for the non-convex energy function $F(\boldsymbol{x})$ in Eq. (4.42) at $\boldsymbol{x} = \boldsymbol{x}^{t-1}$.*

Proof. The proof of this theorem is given in Appendix A.2.1. □

If the scale parameters σ_{noise} and σ_{prior} as well as the regularization weight λ are assumed to be constant, the proposed algorithm can be considered as an MM algorithm to minimize the non-convex energy in Eq. (4.42). The basic principle of this scheme is to successively construct majorizing functions $\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1})$ at $\boldsymbol{x}^{t-1}$ to obtain a refined estimate $\boldsymbol{x}^t$, see Fig. 4.8. Thus, direct optimization of a non-convex energy function is casted to a sequence of weighted but convex optimizations. This relationship also clarifies the properties of the proposed algorithm regarding robustness as minimization of the confidence-aware observation model is related to minimizing the Huber loss. Similarly, minimization based on the WBTV prior is related to minimizing the sparsity-promoting L_1/L_p regularization term.

Convergence Analysis. Based on this relationship, we establish a convergence proof of iteratively re-weighted minimization. In order to study the convergence, let $\boldsymbol{x}^0$ be the initial guess and σ_{noise}, σ_{prior} as well as λ constant parameters over the iterations. Then, the objective value $F(\boldsymbol{x})$ in Eq. (4.42) converges within a finite number of iterations, which is stated by the following theorem.

Theorem 4.2. *Let $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^T$ be an iteration sequence obtained by iteratively re-weighted minimization. Then, for all $t = 2, \ldots, T$ there exists a strict positive $\underline{\beta}$ such that:*

$$F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^t) \geq \underline{\beta} \left|\left|\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t\right|\right|_2^2. \tag{4.48}$$

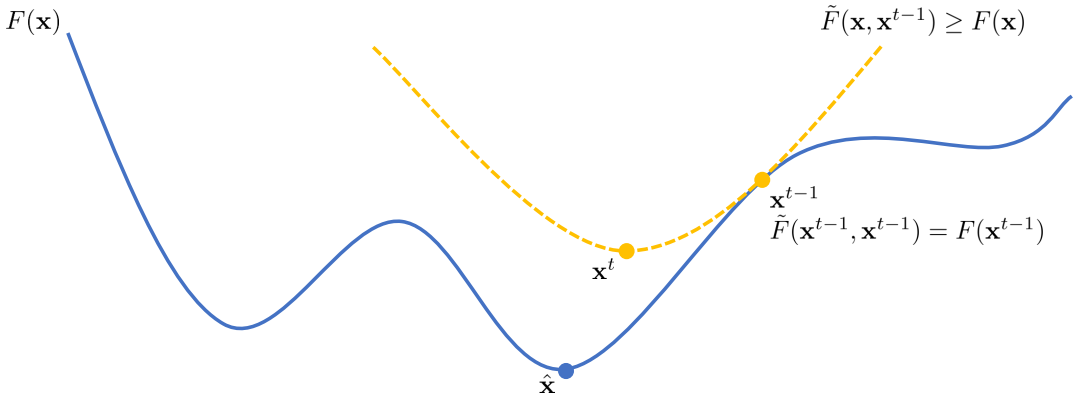


Figure 4.8: Illustration of the MM principle. The minimization of the non-convex function $F(\boldsymbol{x})$ is casted to the iterative minimization of convex majorizing functions $\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1})$.

Proof. The proof of this theorem is given in Appendix A.2.2. □

This theorem states that the objective value $F(\boldsymbol{x}^t)$ is monotonically decreasing. Since $F(\boldsymbol{x})$ is a lower-bounded function, $F(\boldsymbol{x}^t)$ converges to an extreme value, and so does Algorithm 4.1. For a detailed experimental convergence study with adaptive scale and regularization parameters, we refer to Section 4.5.3.

4.5 Experiments and Results

The experimental results reported in this section provide a comparison of the proposed method to several state-of-the-art algorithms as well as an in-depth analysis of its numerical properties. This includes quantitative evaluations on simulated data and qualitative assessment of super-resolution on real images. We focus on super-resolution under challenging conditions in real-world applications, including motion estimation uncertainty or image noise with space variant properties.

4.5.1 Experiments on Simulated Data

To enable quantitative evaluations, simulated low-resolution data with a known ground truth for the desired high-resolution data was used. The ground truth images were projected by the image formation model to obtain their low-resolution counterparts. This mapping was described by rigid motion with uniform distributed translation $\boldsymbol{t} = (t_u, t_v)^\top$, $t_u, t_v \in [-3, 3]$ and rotation angle $\varphi \in [-1, 1]$, a Gaussian PSF ($\sigma_{\text{PSF}} = 0.5$), and subsampling according to the desired magnification factor s. Each frame was corrupted by a superimposition of intensity-dependent Poisson noise, Gaussian noise with standard deviation σ_{noise}, and salt-and-pepper noise at level ν_{noise} that specifies the amount of invalid pixels. For each ground truth, the simulation was performed ten times and all results were averaged over these randomized realizations. Grayscale converted reference images were taken from the LIVE database [Shei 16] consisting of color photographs of natural scenes. The peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) [Wang 04b] were used to compare super-resolution to a ground truth.

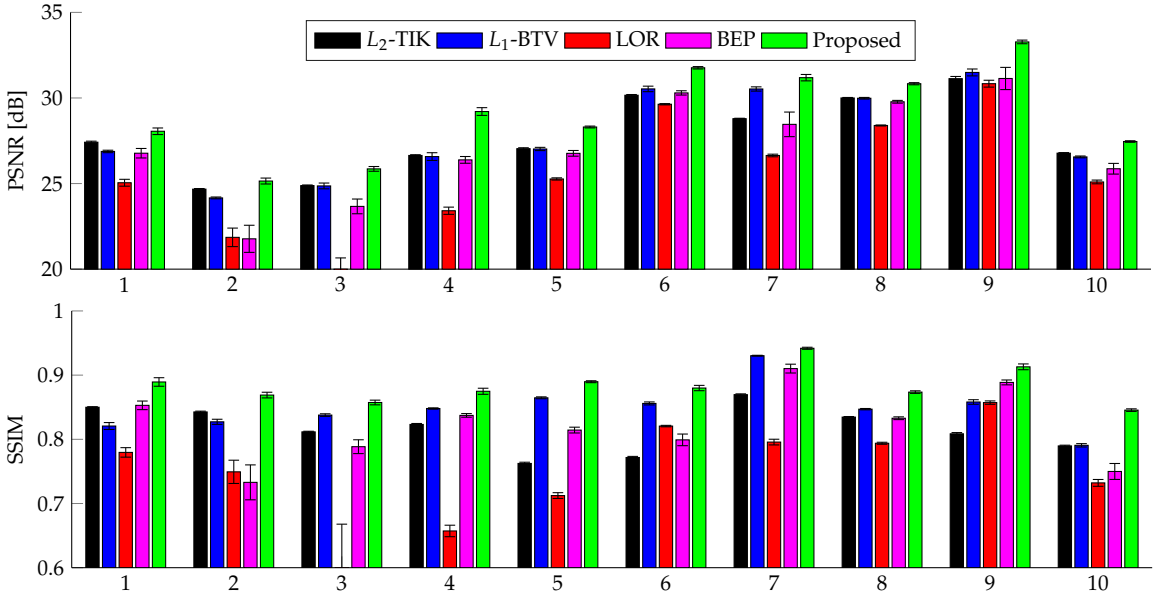
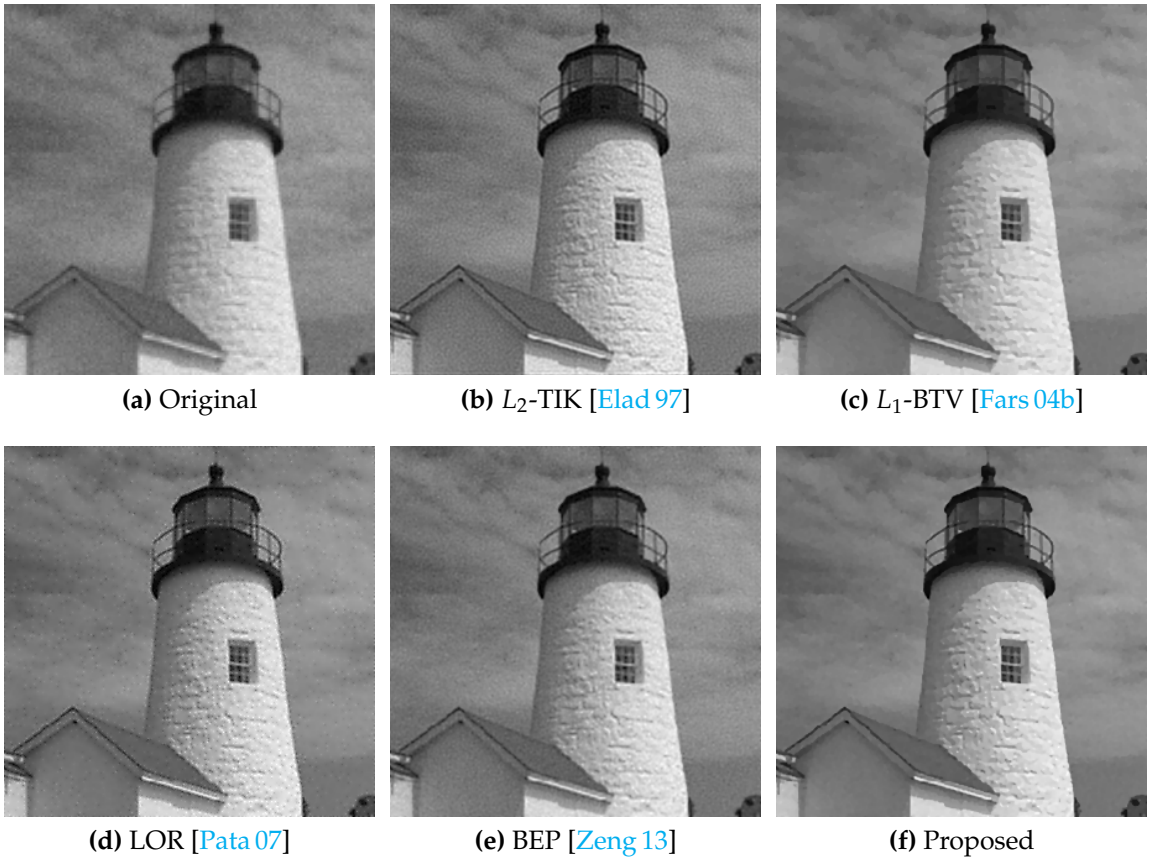


Figure 4.9: Mean ± standard deviation of the PSNR and SSIM achieved by the competing super-resolution algorithms under exact subpixel motion estimation. The benchmark includes ten simulated datasets with ten randomly generated image sequences per dataset.

Iteratively re-weighted minimization was used with $T_{\text{irwsr}} = 10$, $T_{\text{scg}} = 10$, $T_{\text{cv}} = 20$, and termination tolerance $\eta = 0.001$. The WBTV parameters were set to $N_{\text{BTV}} = 2$ and $\alpha_{\text{BTV}} = 0.7$ with sparsity parameter $p = 0.5$. The tuning constants of the underlying weighting functions were set to $c_{\text{bias}} = 0.02$ and $c_{\text{local}} = c_{\text{prior}} = 2.0$ for images given in the intensity range $[0, 1]$ according to [Kohl 16b].

The proposed approach was compared to several related spatial domain reconstruction algorithms, namely L_2 norm minimization coupled with Tikhonov regularization (L_2-TIK) [Elad 97], L_1 norm minimization coupled with BTV regularization (L_1-BTV) [Fars 04b], Lorentzian M-estimator based super-resolution (LOR) [Pata 07], and adaptive super-resolution with bilateral edge preserving regularization (BEP) [Zeng 13]. For a fair evaluation of these algorithms, their regularization weights were selected for each dataset individually using a grid search on a training sequence and maximization of the PSNR. Notice that the proposed algorithm does not require off-line regularization parameter selections.

Effect of Image Noise. Figure 4.9 depicts a benchmark of the competing super-resolution algorithms in a baseline experiment by utilizing the exact subpixel motion from the simulation process. In this experiment, ten datasets with sequences of $K = 8$ low-resolution frames were employed for super-resolution with magnification $s = 2$. Each frame was degraded by a fixed level of Gaussian noise ($\sigma_{\text{noise}} = 0.02$). Note that the proposed algorithm consistently outperformed the state-of-the-art in terms of both measures. In comparison to L_1-BTV, the PSNR and SSIM measures were improved by 1.2 decibel (dB) and 0.03, respectively. A qualitative comparison is shown in Fig. 4.10, where the proposed algorithm achieved decent results in terms of the reconstruction of image structures while the competing methods were prone to oversmoothing or residual noise.



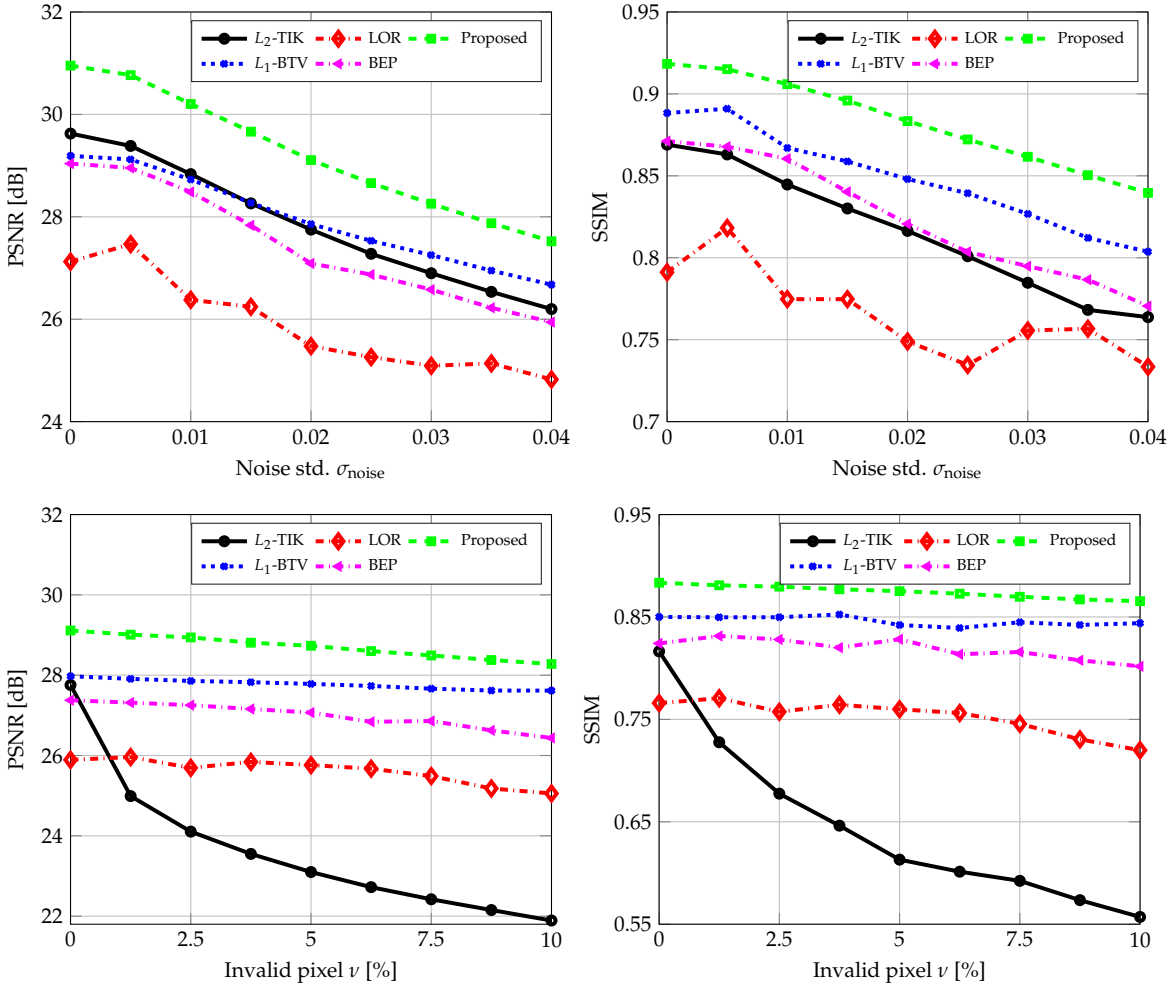
(a) Original **(b)** L_2-TIK [Elad 97] **(c)** L_1-BTV [Fars 04b]

(d) LOR [Pata 07] **(e)** BEP [Zeng 13] **(f)** Proposed

Figure 4.10: Super-resolution ($K = 8$ frames, magnification $s = 2$) with exact subpixel motion and additive Gaussian noise ($\sigma_{\text{noise}} = 0.02$) on the simulated *lighthouse* dataset with a comparison of different combinations of observation models and prior distributions.

The influence of image noise was investigated by varying the levels of Gaussian noise ($\sigma_{\text{noise}} \in [0, 0.04]$) and salt-and-pepper noise ($\nu_{\text{noise}} \in [0, 0.15]$). The averaged PSNR and SSIM measures over ten realization of these experiments are plotted in Fig. 4.11. In the presence of invalid pixels, the L_2-TIK method failed to reconstruct reliable high-resolution data as this approach does not compensate for outliers. The different robust models (L_1-BTV [Fars 04b], LOR [Pata 07], BEP [Zeng 13], and the proposed method) were less sensitive. Moreover, the proposed method quantitatively outperformed the competing robust models for both noise types. See Fig. 4.12 for a comparison on example data with salt-and-pepper noise.

Effect of Motion Estimation Uncertainty. The effect of motion estimation uncertainty was studied by simulating deviations between the true and the actual motion model. For this purpose, for two out of eight frames, an isotropic scaling factor to simulate camera zoom was considered such that the motion associated with these frames deviated from the rigid motion model. Motion estimation was performed using the enhanced correlation coefficient (ECC) optimization framework proposed by Evangelidis and Psarakis [Evan 08] assuming rigid motion. Hence, the frames affected by scaling can be considered as outliers.

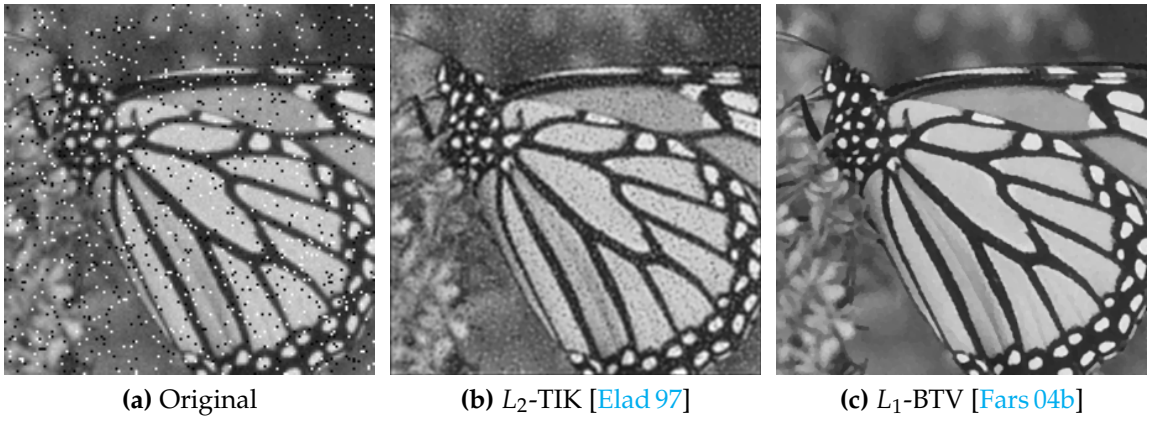


(a) PSNR vs. noise level **(b)** SSIM vs. noise level

Figure 4.11: PSNR and SSIM of super-resolution with image noise. Top row: performance of the competing algorithms under Gaussian noise of varying standard deviations σ_{noise}. Bottom row: influence of salt-and-pepper noise at different levels ν_{noise}.

Figure 4.13 depicts a benchmark of super-resolution on ten simulated datasets in this situation, where the scaling factor followed a normal distribution $\mathcal{N}(c; 1, \sigma_c^2)$ with standard deviation $\sigma_c = 0.05$. Notice that the L_2-TIK method was prone to inaccurate motion estimation while the different robust methods were less sensitive. In this benchmark, the proposed method outperformed the state-of-the-art on most of the datasets. Compared to L_1-BTV, the PSNR and SSIM measures were enhanced by 0.7 dB and 0.04, respectively. See Fig. 4.14 for a qualitative comparison among the competing algorithms. The effect of motion estimation uncertainty is visible by ghosting artifacts that were avoided by the proposed method.

Effect of Photometric Variations. The image formation models widely used in literature ignore several effects of digital imaging, see Section 3.2.3. This includes varying photometric conditions during image acquisition caused by camera white balancing or time variant lighting conditions. If varying photometric conditions



(a) Original **(b)** L_2-TIK [Elad 97] **(c)** L_1-BTV [Fars 04b]

(d) LOR [Pata 07] **(e)** BEP [Zeng 13] **(f)** Proposed

Figure 4.12: Super-resolution ($K = 8$ frames, magnification $s = 2$) on the simulated *monarch* dataset with mixed Gaussian noise ($\sigma_{\text{noise}} = 0.02$) and salt-and-pepper noise with a comparison of different combinations of observation models and prior distributions.

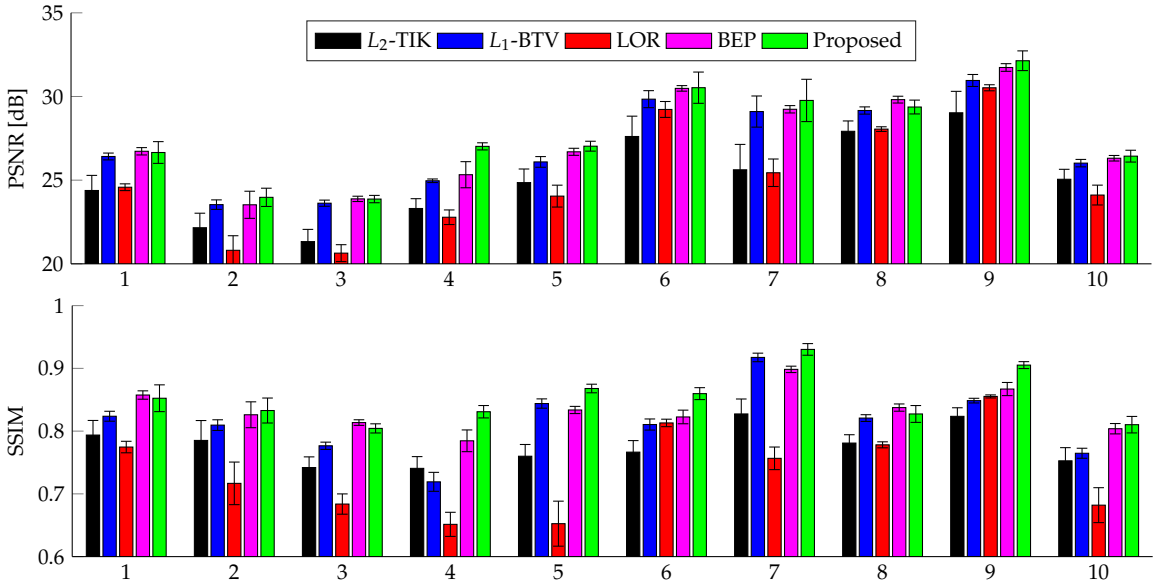


Figure 4.13: Mean $\pm$ standard deviation of the PSNR and SSIM measures in Fig. 4.9 in the presence of inaccurate motion estimation.

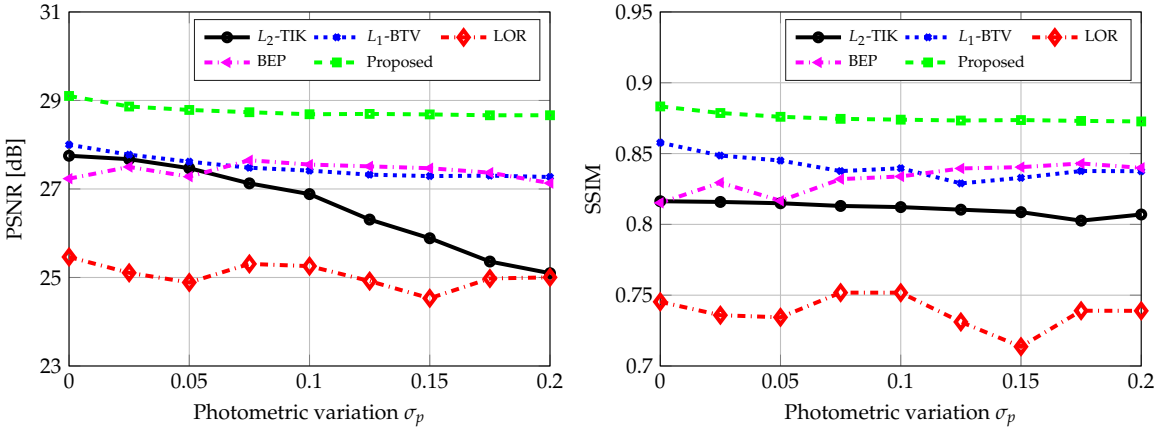
(a) Original **(b)** L_2-TIK [Elad 97] **(c)** L_1-BTV [Fars 04b]

(d) LOR [Pata 07] **(e)** BEP [Zeng 13] **(f)** Proposed

Figure 4.14: Super-resolution ($K = 8$ frames, magnification $s = 2$) on the simulated *cemetery* dataset in the presence of inaccurate motion estimation. The uncertainty of motion parameters led to ghosting artifacts in case of a non-robust observation model (L_2-TIK [Elad 97]), while robust models (L_1-BTV [Fars 04b], LOR [Pata 07], BEP [Zeng 13], and the proposed method) compensated for this uncertainty.

should be taken into account, the image formation model needs to be extended and photometric registration has to be employed to estimate photometric parameters [Cape 03, Cape 04]. However, photometric registration might be error prone and its uncertainty leads to outliers in super-resolution reconstruction. To evaluate the impact of photometric variations, an original low-resolution frame $\boldsymbol{y}^{(k)}$ is corrupted according to $\boldsymbol{z}^{(k)} = \gamma_m^{(k)} \boldsymbol{y}^{(k)} + \gamma_a^{(k)} \boldsymbol{1}$ to obtain a distorted frame $\boldsymbol{z}^{(k)}$ [Cape 04]. For two randomly selected frames, photometric variations were simulated by choosing uniform distributed parameters in $[-\frac{1}{2}\sigma_p, +\frac{1}{2}\sigma_p]$ for $\gamma_a^{(k)}$ and in $[1 - \frac{1}{2}\sigma_p, 1 + \frac{1}{2}\sigma_p]$ for $\gamma_m^{(k)}$, where σ_p reflects the parameter uncertainty.

Figure 4.15 depicts the impact of photometric variations at different levels σ_p. In this situation, the L_2-TIK approach was affected by an intensity bias as captured by the PSNR. The robust algorithms were less sensitive to photometric variations. In particular, the proposed method consistently achieved the highest quality measures since photometric variations were successfully compensated by bias detection. See Fig. 4.16 for a visual comparison of this behavior on the *lighthouse* dataset. Here, photometric variations in the input frames caused an intensity bias that is apparent in the L_2-TIK reconstruction but compensated by the proposed method.



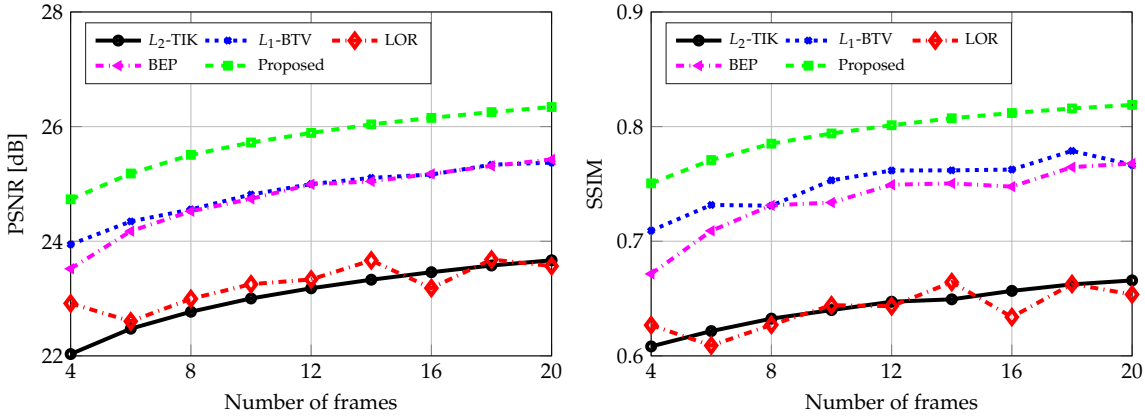
(a) PSNR vs. photometric variations **(b)** SSIM vs. photometric variations

Figure 4.15: PSNR and SSIM of super-resolution at different levels of photometric variations. All photometric variations were simulated by uniform distributed global contrast and brightness changes with standard deviation σ_p relative to a reference image.

(a) Original **(b)** L_2-TIK [Elad 97] **(c)** L_1-BTV [Fars 04b]

(d) LOR [Pata 07] **(e)** BEP [Zeng 13] **(f)** Proposed

Figure 4.16: Super-resolution ($K = 8$ frames, magnification $s = 2$) on the *lighthouse* dataset with photometric variations ($\sigma_p = 0.15$). Photometric variations in input frames led to an intensity bias under non-robust models (L_2-TIK [Elad 97]), while robust models (L_1-BTV [Fars 04b], LOR [Pata 07], BEP [Zeng 13], and the proposed method) were less sensitive.



(a) PSNR vs. sequence length **(b)** SSIM vs. sequence length

Figure 4.17: PSNR and SSIM of super-resolution for different numbers of input frames.

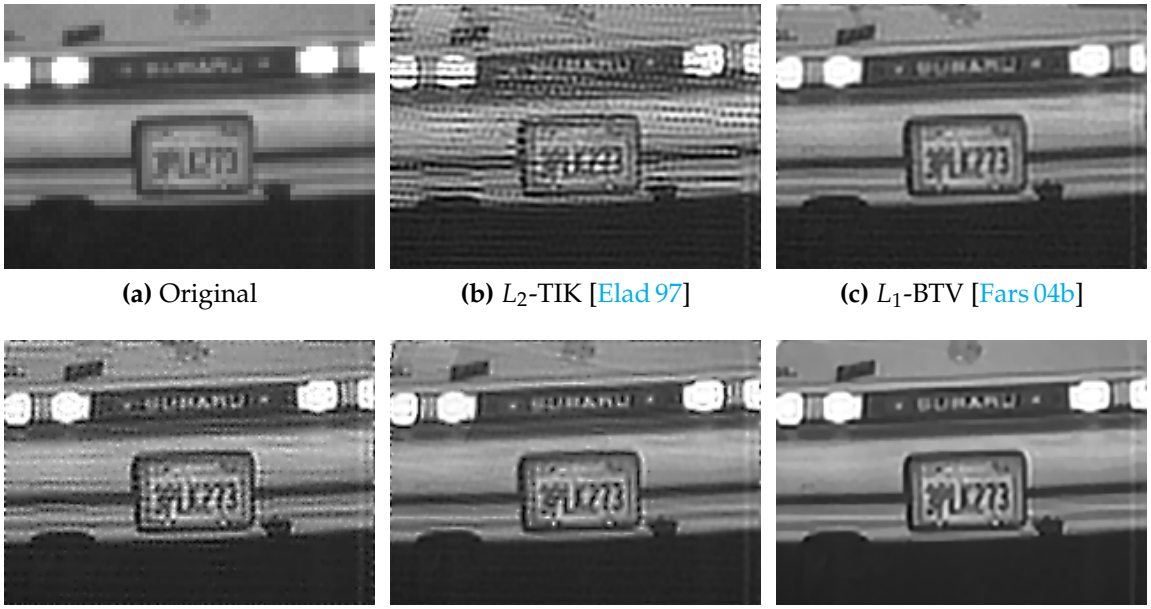
Effect of the Sequence Length. One relevant parameter for super-resolution is the number of low-resolution input frames. This parameter was investigated for the magnification factor $s = 3$ as larger magnifications typically require more input frames. Throughout this experiment, the fraction of invalid pixels was set to $\nu_{\text{noise}} = 0.01$ to simulate outliers and the exact subpixel motion was utilized.

In Fig. 4.17, we depict the quality measures versus the number of low-resolution input frames. As expected, a larger number of frames resulted in more accurate reconstructions indicated by an increasing PSNR and SSIM. However, even in the case of long input sequences, the performance of L_2-TIK was limited due to the presence of outliers. In comparison to the competing algorithms, the proposed method performed best in terms of both quality measures. Notice that the proposed method with $K = 8$ provided competitive results to L_1-BTV and BEP with $K = 20$ frames. Hence, it is more economical regarding the number of input frames. This study also considered the important use case of underdetermined super-resolution, which is the case for $K < s^2$. Even in this challenging situation that appeared for $K < 9$, the proposed algorithm provided reliable reconstructions w. r. t. the ground truth and outperformed the state-of-the-art.

4.5.2 Experiments on Real Data

The proposed method was qualitatively evaluated on real image data in two different applications. First, experiments with natural images were conducted. These are challenging due to the uncertainty of subpixel motion estimation. Second, experimental results in the field of 3-D range imaging are presented. Here, super-resolved range data was reconstructed from low-resolution range images captured with a Time-of-Flight (ToF) sensor that is affected by space variant noise.

Evaluation on Natural Images. For the experiments on natural scenes, image sequences with different types of subpixel motion were used. Figure 4.18 compares the different super-resolution algorithms on the *car* sequence taken from the MDSP database [Fars 16]. This experiment aims at super-resolving a license plate



(a) Original **(b)** L_2-TIK [Elad 97] **(c)** L_1-BTV [Fars 04b]

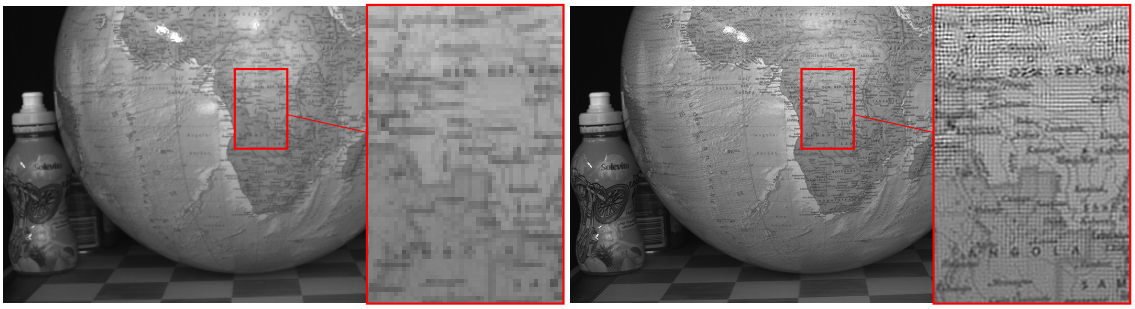
(d) LOR [Pata 07] **(e)** BEP [Zeng 13] **(f)** Proposed

Figure 4.18: Super-resolution on the *car* dataset to identify a license plate ($K = 12$ frames, magnification $s = 3$). The subpixel motion is related to out-of-plane movements of the car. Figure reused from [Kohl 16b] with the publisher's permission ©2016 IEEE.

using $K = 12$ frames. Super-resolution was applied with magnification $s = 3$ and a Gaussian PSF ($\sigma_{\text{PSF}} = 0.4$). The subpixel motion across these frames followed an affine model with a substantial amount of scaling related to out-of-plane movements of the car. The motion estimation for this sequence was performed by ECC optimization [Evan 08]. Note that large car movements made motion estimation difficult and resulted in outliers due to misregistrations for individual frames. Consequently, L_2-TIK was affected by ghosting artifacts, while the robust algorithms were less sensitive. In terms of the recovery of the license plate, the proposed method provided an artifact-free and sharp reconstruction.

Figure 4.19 depicts super-resolution on the *globe* sequence [Kohl 17] acquired with a Basler acA2000-50gm CMOS camera. For this experiment, $K = 17$ low-resolution frames captured by 4×4 hardware binning on the sensor array relative to the maximum pixel resolution were used. Super-resolution was performed with magnification $s = 4$ and a Gaussian PSF ($\sigma_{\text{PSF}} = 0.4$). The subpixel motion was related to a superposition of rigid camera movements and an independent rotation of the globe. In order to handle this non-rigid model, the variational optical flow algorithm proposed by Liu [Liu 09] was employed for motion estimation. Notice that optical flow computation was error-prone due to occlusions that were caused by large rotations of the globe. Such outliers resulted in artifacts on the globe surface in the L_2-TIK reconstruction. The proposed method was robust against these outliers and achieved a decent recovery of text on the globe surface.

Evaluation on Range Images. For the experiments in range imaging, a PDM CamCube 3.0 ToF camera was used to measure the 3-D scene in Fig. 4.20. Range



(a) Original **(b)** L_2-TIK [Elad 97]

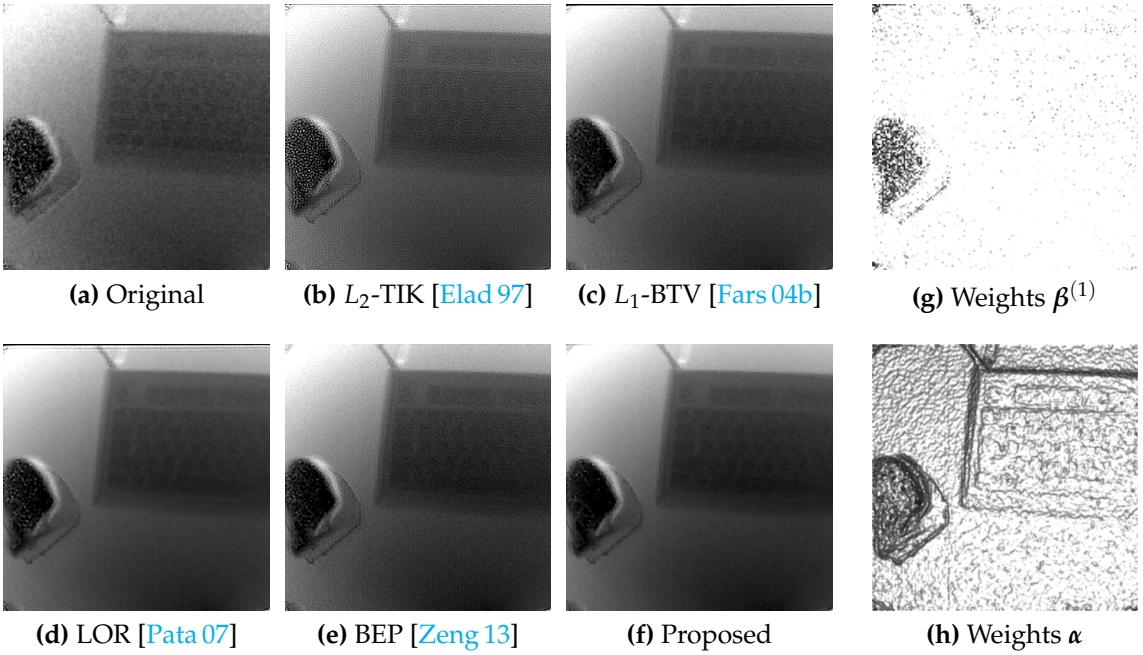
(c) L_1-BTV [Fars 04b] **(d)** LOR [Pata 07]

(e) BEP [Zeng 13] **(f)** Proposed

Figure 4.19: Super-resolution on the *globe* dataset ($K = 17$ frames, magnification $s = 4$). The subpixel motion is a mixture of camera movements and a rotation of the globe.

data was acquired with 200 × 200 px at a frame rate of 30 Hz and super-resolution was applied to sets of range images. In addition to the low spatial resolution, the reliability of the ToF sensor was affected by intensity-dependent errors on the black surface of the punch. This resulted in space variant noise, i. e. larger uncertainties of range data on the punch surface compared to regions with brighter illumination, which is a common issue in ToF imaging. Super-resolution was applied with $K = 16$ frames, a Gaussian PSF ($\sigma_{\text{PSF}} = 0.5$) and magnification $s = 3$. Motion estimation was performed by ECC optimization with an affine model.

In this example, the proposed method achieved the best behavior under space variant noise as shown by the reconstruction of flat surfaces and object edges. Range data with lower confidence, i. e. measurements affected by higher noise levels, was successfully determined as a by-product of iteratively re-weighted minimization. This is visible in the visualization of the observation confidence map



(a) Original **(b)** L_2-TIK [Elad 97] **(c)** L_1-BTV [Fars 04b] **(g)** Weights $\boldsymbol{\beta}^{(1)}$

(d) LOR [Pata 07] **(e)** BEP [Zeng 13] **(f)** Proposed **(h)** Weights $\boldsymbol{\alpha}$

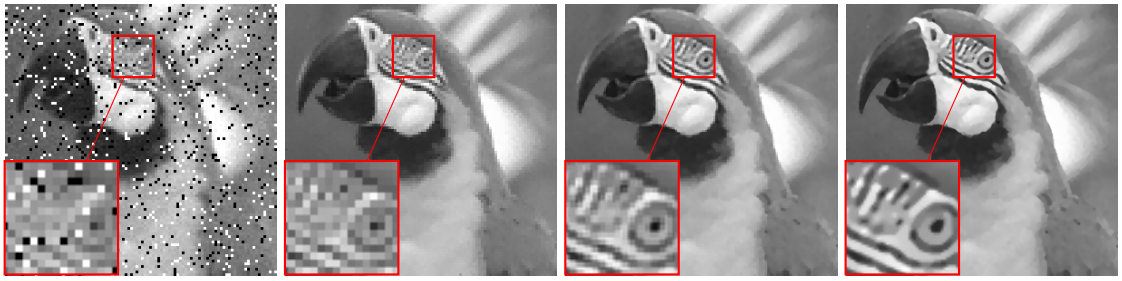
Figure 4.20: Super-resolution for ToF range images in the presence of space variant noise ($K = 16$ frames, magnification $s = 3$). (b) - (f) Super-resolved images obtained by the competing algorithms. (g) - (h) Observation weights $\boldsymbol{\beta}^{(1)}$ associated with the first frame and prior weights $\boldsymbol{\alpha}$ visualized in grayscale (brighter regions denote higher weights). Figure reused from [Kohl 16b] with the publisher's permission ©2016 IEEE.

$\boldsymbol{\beta}^{(1)}$ associated with the first frame. Here, lower weights were assigned to surfaces affected by intensity-dependent noise. Similarly, the confidence map $\boldsymbol{\alpha}$ of WBTV steers the regularization to improve the reconstruction of depth discontinuities.

4.5.3 Convergence and Parameter Sensitivity

To confirm the convergence of iteratively re-weighted minimization experimentally, the parameter estimates provided by the proposed algorithm were traced over the iterations. The convergence is studied on simulated images under a mixture of Gaussian noise ($\sigma_{\text{noise}} = 0.02$) and salt-and-pepper noise ($\nu_{\text{noise}} \in [0, 0.1]$). To evaluate the sensitivity of the algorithm regarding the initial guess, the proposed initialization computed by the motion-compensated temporal median was compared to an initialization computed by bicubic upsampling of a single frame. Super-resolved images at different iterations with magnification factor $s = 3$ using $K = 12$ frames and the temporal median as initial guess are shown in Fig. 4.21.

Figure 4.22 depicts the average PSNR and SSIM measures of the super-resolved images at different iterations and different amounts of invalid pixels over ten random realizations of the experiment. Independently of the noise level and the initial guess, iteratively re-weighted minimization converged within the first five iterations. This also appeared in case of a large amount of outliers and confirms the convergence of the iteration scheme. In addition, the behavior of the adaptive



(a) Original **(b)** Iteration 1 **(c)** Iteration 2 **(d)** Iteration 10

Figure 4.21: Illustration of the convergence of iteratively re-weighted minimization. The example depicts super-resolved images ($K = 12$ frames, magnification $s = 3$) at different iterations for the *parrots* dataset. The low-resolution input images are affected by a mixture of Gaussian noise ($\sigma_{\text{noise}} = 0.02$) and salt-and-pepper noise ($\nu_{\text{noise}} = 0.1$).

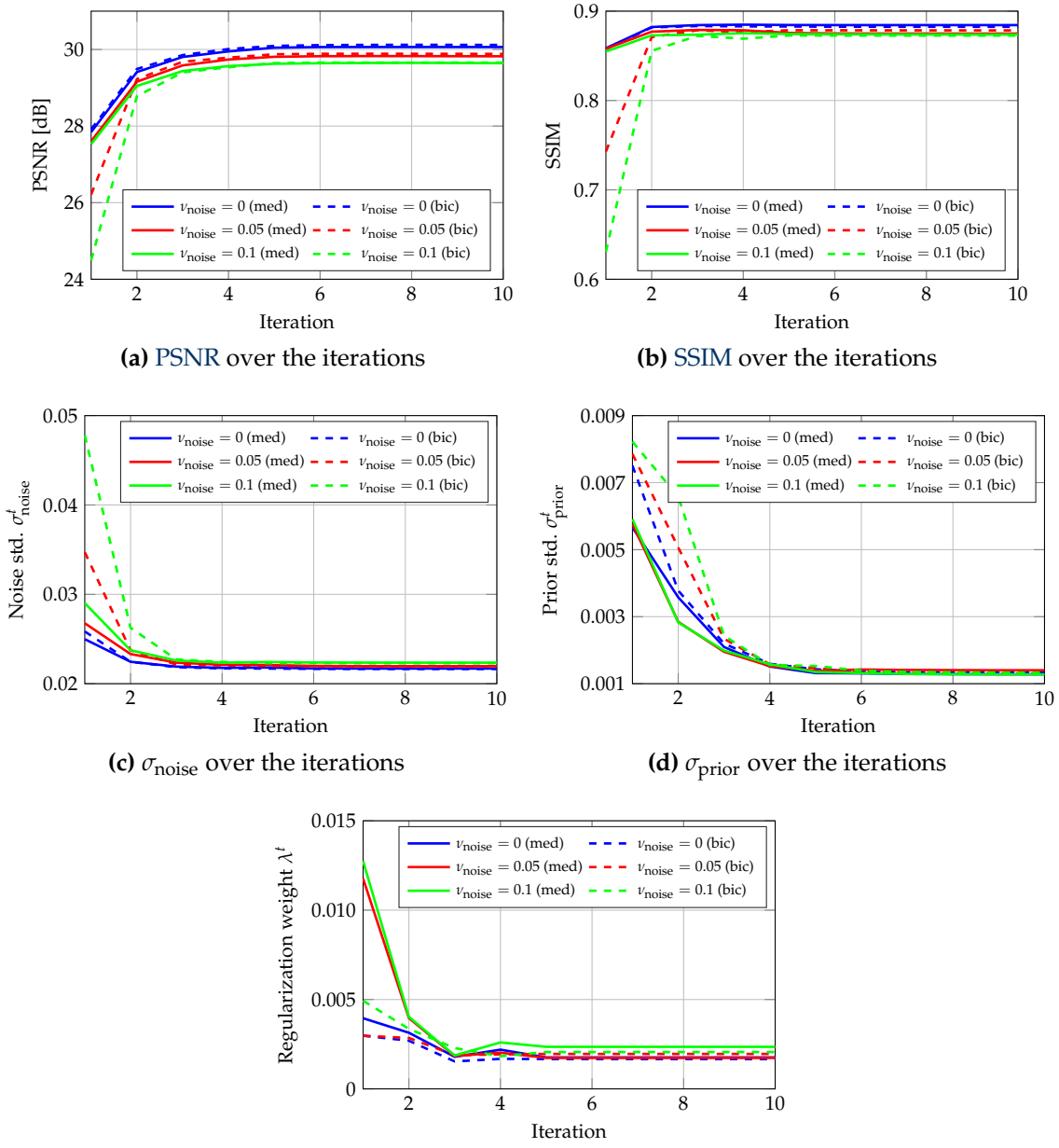
scale and regularization parameter estimation is depicted. Similar to the latent high-resolution image, these estimates converged within a few iterations.

One relevant parameter of iteratively re-weighted minimization is the sparsity parameter p of the underlying prior weighting function. This parameter controls how strong sparsity is enforced and $p < 1$ implements a heavy-tailed prior distribution, see Section 4.4.2. The impact of this parameter is studied at different noise levels. To this end, low-resolution images were corrupted by a mixture of Gaussian noise ($\sigma_{\text{noise}} \in [0, 0.04]$) and salt-and-pepper noise ($\nu_{\text{noise}} = 0.01$). Figure 4.23 compares BTV ($p = 1.0$) to the proposed WBTV ($p < 1$) on an example dataset. Notice that WBTV regularization contributed to an improved reconstruction of fine textures. Furthermore, it was less sensitive to staircasing in homogenous image regions. The means and the standard deviations of the PSNR and SSIM measures for ten random realizations of this experiment are plotted in Fig. 4.24 for different noise levels. In these experiments, $p < 1$ enhanced the accuracy of super-resolution due to the edge-aware reconstruction compared to the unweighted BTV prior. The contributions of the sparse prior were more substantial for larger noise levels. In this work, p is chosen in the range $[0.3, 0.8]$ as a too small p decreases the numerical stability of weight computation and increases the degree of non-convexity of the underlying optimization problem. Conversely, a too large p limits the benefit of the WBTV prior. In summary, $p = 0.5$ is a reasonable choice for natural images and generalizes fairly well to scenes with different content.

4.5.4 Computational Complexity

This section reports the computational complexity of the proposed algorithm in terms of computation time as well as the number of energy function evaluations for numerical optimization. In Tab. 4.2, we compare these performance characteristics on image sequences of different sizes[3]. To this end, the *car* and the *globe* sequences (see Section 4.5.2) were used. In both experiments, L_2-TIK converged

[3]These experiments were performed on an Intel Xeon CPU E5-1630v4 with 3.7 GHz and 64 GB RAM using a non-parallelized MATLAB implementation.



(a) PSNR over the iterations

(b) SSIM over the iterations

(c) σ_{noise} over the iterations

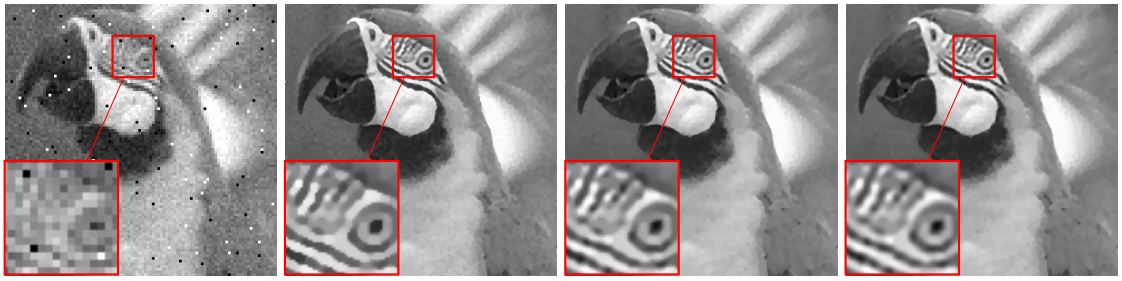
(d) σ_{prior} over the iterations

(e) λ over the iterations

Figure 4.22: Convergence analysis of iteratively re-weighted minimization. (a) - (b) PSNR and SSIM depicted for different amounts of invalid pixels on the dataset in Fig. 4.21. The iterations were initialized by the temporal median (med) and bicubic upsampling of a single frame (bic). (c) - (e) adaptive estimates of σ_{noise}, σ_{prior}, and λ.

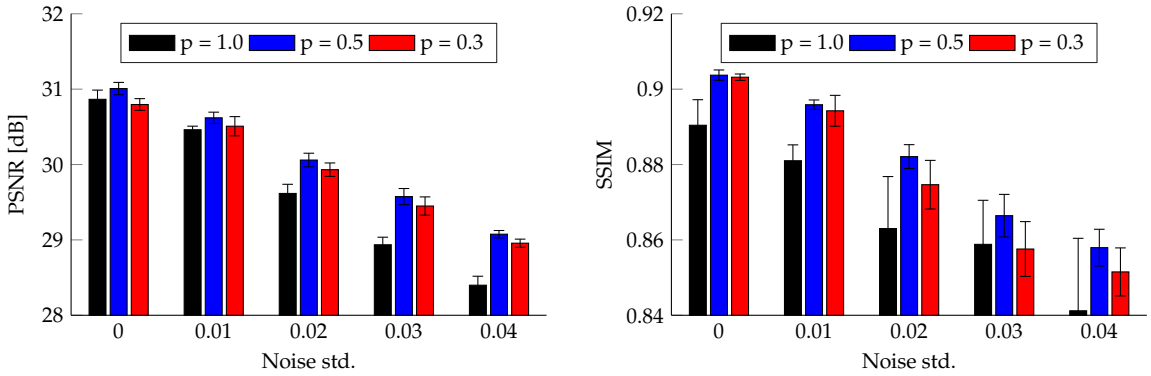
quite fast and required the lowest computation time while the different robust algorithms required more iterations.

To examine the complexity of the different computational stages, iteratively re-weighted minimization was evaluated with the proposed hyperparameter selection (adaptive λ) and with a bypass of this stage (constant λ). Here, the adaptive algorithm increased the computation time compared to the competing methods that do not provide an automatic parameter selection. Notice that a bypass of



(a) Original **(b)** BTV ($p = 1.0$) **(c)** WBTV ($p = 0.5$) **(d)** WBTV ($p = 0.3$)

Figure 4.23: Impact of the sparsity parameter p of the proposed prior weighting function on the *parrots* dataset. The low-resolution frames are affected by a mixture of Gaussian noise ($\sigma_{\text{noise}} = 0.04$) and salt-and-pepper noise ($\nu_{\text{noise}} = 0.01$). This example compares the BTV prior ($p = 1.0$) to the proposed WBTV prior using different settings of the sparsity parameter ($p = 0.3$ and $p = 0.5$). Notice the recovery of the texture for $p < 0.5$.



(a) PSNR vs. noise level σ_{noise} **(b)** SSIM vs. noise level σ_{noise}

Figure 4.24: Impact of the sparsity parameter p of the prior weighting function to the performance of iteratively re-weighted minimization. The parameter sensitivity was assessed on the dataset in Fig. 4.23 at different Gaussian noise levels. The proposed choice $p = 0.5$ led to superior results for all noise levels.

this stage considerably reduced the complexity. In this case, the computation time was comparable to those of the state-of-the-art. Moreover, the proposed coarse-to-fine optimization was compared to a single-scale implementation. Note that by-passing the coarse-to-fine optimization increased the computation time, which reveals the benefit of the proposed iteration scheme.

4.6 Conclusion

This chapter introduced a robust super-resolution algorithm from a Bayesian perspective. This approach is based on adaptive confidence weighting to define space variant observation and prior distributions. The confidence weights are treated as latent variables in the Bayesian model and are inferred simultaneously to the super-resolved image in an adaptive scheme by means of iteratively re-weighted

Table 4.2: Computational complexity of iteratively re-weighted minimization and several state-of-the-art algorithms. This analysis includes the computation times and the number of energy function evaluations for numerical optimization. The L_2-TIK method is considered as baseline and the numbers in brackets denote the relative increase of the computation time and the number of function evaluations compared to L_2-TIK. Iteratively re-weighted minimization was evaluated with (w/) and without (w/o) coarse-to-fine optimization as well as with adaptive regularization and with constant regularization weight.

Super-resolution algorithm	*Globe* sequence (510×270, $K = 17$ frames)		*Car* sequence (70×50, $K = 12$ frames)	
	Time [s]	# Fun. eval.	Time [s]	# Fun. eval.
State-of-the-art				
L_2-TIK [Elad 97]	174 ($\times$1.0)	50 ($\times$1.0)	2 ($\times$1.0)	48 ($\times$1.0)
L_1-BTV [Fars 04b]	383 ($\times$2.2)	50 ($\times$ 1.0)	4 ($\times$2.0)	50 ($\times$ 1.0)
LOR [Pata 07]	323 ($\times$1.9)	50 ($\times$1.0)	4 ($\times$2.0)	50 ($\times$ 1.0)
BEP [Zeng 13]	2291 ($\times$13.2)	50 ($\times$1.0)	18 ($\times$9.0)	50 ($\times$ 1.0)
Proposed				
w/ coarse-to-fine (adapt. λ)	1198 ($\times$6.9)	75 ($\times$ 1.5)	15 ($\times$7.5)	82 ($\times$1.7)
w/ coarse-to-fine (const. λ)	914 ($\times$5.3)	50 ($\times$ 1.0)	8 ($\times$4)	50 ($\times$1.0)
w/o coarse-to-fine (adapt. λ)	3098 ($\times$17.8)	80 ($\times$1.6)	26 ($\times$13.0)	90 ($\times$1.9)

minimization. Mathematically, this technique can be derived as an MM algorithm. Iteratively re-weighted minimization combines the advantages of robustness regarding outliers in image formation with sparse regularization to enhance the reconstruction of edges and texture. As an additional merit, it does not require an extensive manual parameter tuning and provides an automatic parameter selection in a computationally efficient way.

In a baseline benchmark with mixed Gaussian and Poisson noise, iteratively re-weighted minimization achieved average gains of 1.2 dB and 0.03 in terms of PSNR and SSIM over related robust algorithms. In a benchmark that considered inaccurate motion estimation, it improved the PSNR and SSIM by 0.7 dB and 0.04, respectively. The iteration scheme showed fast convergence and converged to stationary points within five iterations regardless of the initialization and the fraction of outliers in the image formation.

Notice that throughout this chapter, subpixel motion is initially estimated on low-resolution frames prior to super-resolution. However, the proposed framework is extensible to treat subpixel motion as hidden information. In [Berc 16], iteratively re-weighted minimization has been formulated via confidence-aware Levenberg-Marquardt optimization [Marq 63]. This enables joint motion estimation and super-resolution to enhance the accuracy of an initial motion estimate.

Part II

Multi-Sensor Super-Resolution for Hybrid Imaging

CHAPTER 5

Multi-Sensor Super-Resolution using Guidance Images

The algorithms investigated in Part I of this work provide super-resolution of a single modality only. Part II of this thesis examines the extension of super-resolving images of one modality in the presence of a complementary modality. The key idea of this approach termed *multi-sensor* super-resolution is to steer image reconstruction by guidance images. For this purpose, we present a computational framework that employs guidance data to enhance motion estimation and regularization compared to conventional algorithms dealing with a single modality only. Moreover, we present an outlier detection scheme for multi-sensor super-resolution as an extension of this framework. As an important application of practical relevance, the proposed method is evaluated in the field of *hybrid range imaging*. In this application, high-resolution photometric data is used as guidance to super-resolve low-resolution 3-D range data. The presented experimental evaluation reveals that multi-sensor super-resolution outperforms conventional reconstruction algorithms that work solely on range data.

This methodology has been introduced by Köhler et al. [Kohl 13b, Kohl 14b, Kohl 15b] and later presented by Haase [Haas 16] for interventional imaging.

5.1 Introduction

Over the past decades, the vast majority of super-resolution algorithms has been designed to handle image data of a single modality. As these traditional approaches exploit information acquired with a single sensor, they are referred to as *single-sensor* super-resolution. While this concept has the benefit of great flexibility regarding its applicability in different imaging systems, it suffers from the inherent drawback that only information present in low-resolution data is utilized. For

instance, motion estimation as one of the most essential prerequisites of super-resolution needs to be carried out on low-resolution images, which might be error prone in practical applications. In addition, all algorithmic stages of these approaches are formulated for a single modality, e. g. to design image priors in Bayesian methods. If super-resolution is applied in *hybrid imaging* as the major goal of this chapter, this basic concept essentially ignores the presence of additional information captured by different modalities. In this context, hybrid imaging refers to a class of techniques that combines a set of complementary modalities in a common system by means of *sensor data fusion*.

Hybrid Imaging Technologies. Let us first review several popular hybrid imaging technologies that have emerged in literature along with their basic characteristics. All of these techniques have in common that the involved modalities are complementary in terms of their properties. Hence, their fusion enables a comprehensive representation of the underlying scene.

Some of the most popular hybrid imaging systems have been developed for healthcare. This includes the fusion of functional nuclear imaging such as positron emission tomography (PET) with structural imaging modalities such as computed tomography (CT) or magnetic resonance imaging (MRI) that are widely used in radiology. Combinations of these technologies have been engineered in PET/CT [Beye 00] or PET/MRI [Jude 08] scanners. For instance in this context, structural imaging features the acquisition of anatomical information with high spatial resolution while nuclear imaging provides an acquisition of functional processes in lower resolution. Hybrid imaging has also been proposed for 3-D range imaging as the primary application in this chapter. In this field, measurements of 3-D surface information can be gained by means of active sensor technologies such as ToF [Kolb 10] or structured light [Scha 03]. In addition to the surface information encoded by range images that are acquired with these technologies, other optical techniques are used to capture photometric information of the same scene simultaneously [Han 13]. The modalities are complementary, as photometric data provides high-resolution color and texture information while active range sensors acquire the corresponding surface information.

One common observation in most of these systems is that some modalities feature a high spatial resolution while others are available in lower resolution. In practice, these gaps of the sensor characteristics are caused by technological or economic reasons. This initiates the development of novel super-resolution algorithms that exploit multiple modalities and their complementary natures in order to enhance the traditional single-sensor approaches.

Multi-Sensor Super-Resolution. The target of the proposed super-resolution approach for hybrid imaging is to identify one modality that is available at high spatial resolution as a *guidance* modality. Accordingly, in contrast to the conventional single-sensor approaches, this chapter shows how resolution enhancement for one modality can be steered by such guidance images. Guidance images are exploited in various ways including 1) motion estimation, 2) spatially adaptive regularization, and 3) outlier detection as vital parts of multi-sensor super-resolution. This

method is driven by the hypothesis that guidance images of high quality in terms of their spatial resolution and signal-to-noise ratio (SNR) hold the potential to enhance super-resolution of another modality.

The remainder of this chapter is organized as follows. Section 5.2 provides a literature survey on related super-resolution and filtering techniques. In Section 5.3, we introduce a multi-sensor framework that employes guidance data for motion estimation and spatially adaptive regularization. Afterwards, Section 5.4 extends this method by outlier detection that is driven by guidance data. Section 5.5 studies the application of this framework in hybrid 3-D range imaging, where low-resolution range images are super-resolved under the guidance of high-resolution photometric data. Finally, Section 5.6 presents a summary of this chapter.

5.2 Related Work

Compared to the great number of algorithms for single-sensor resolution enhancement, there are only a few approaches that deal with the problem of multi-sensor super-resolution. Prior work in this field typically addresses specific imaging setups. An early method has been introduced in the pioneering work of Zomet and Peleg [Zome 02]. This method addresses super-resolution of multi-channel images and is driven by the strategy that super-resolution of one of the channels can be guided by the remaining channels. For this purpose, it exploits statistical redundancies across the channels to derive an observation model with a virtual prediction error. Instead of minimizing the residual error as done in single-sensor algorithms, this virtual prediction error is minimized. Super-resolution for color and infrared data have been considered as example applications but the experiments are limited to single-image upsampling. One limitation in comparison to the approach presented in this chapter is that it does not consider motion estimation as an integral part of super-resolution reconstruction.

Methods that are conceptually closely related to the approach of Zomet and Peleg [Zome 02] include local image filters. Some of the well known techniques are guided upsampling [He 13, He 10] or joint bilateral upsampling [Kopf 07] to upsample an image under the guidance of a second one. Here, range imaging is one important application, where range data upsampling is guided by high-resolution color images. More recently, these filters have been extended by various approaches, e. g. non-local means regularization [Park 11], anisotropic total generalized variation regularization [Fers 13] or photometric and range co-sparse analysis [Kiec 13]. However, despite their success, these approaches were designed for single-image upsampling and use image formation models of limited flexibility compared to the models proposed for multi-frame super-resolution. In particular, effects of the camera PSF are seldom modeled by these local filters. For a generalization of such filters towards multi-frame super-resolution, we refer to Chapter 6.

Another mentionable approach that employs the concept of guidance images has been proposed by Kennedy et al. [Kenn 07] for PET/CT scanners in medicine. This method super-resolves PET scans by using anatomical information gained from CT data. Even if this approach is conceptually interesting, it is limited to PET resolution enhancement and does not generalize to other imaging setups.

5.3 Multi-Sensor Super-Resolution Framework

In this section, we present the basis of the proposed multi-sensor super-resolution framework that processes low-resolution images under the guidance of a complementary modality. The main novelty of this methodology is two-fold. First, a filter-based technique is presented that uses high-resolution guidance data to obtain a reliable motion estimate for super-resolution. In addition, a regularization technique is introduced that exploits high-resolution guidance images to adaptively regularize super-resolution on a second modality. Both techniques have the goal to improve robustness and accuracy of the framework compared to a single-sensor approach that does not exploit guidance data.

5.3.1 Framework Overview

The proposed framework aims at reconstructing a high-resolution image $\boldsymbol{x}$ from a set of low-resolution frames $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}$ termed *input images*. For each input image $\boldsymbol{y}^{(k)}$, there exists a corresponding frame $\boldsymbol{z}^{(k)}$ that is acquired with another modality and termed *guidance image*[1]. Each guidance image $\boldsymbol{z}^{(k)}$ is encoded as a $L_u \times L_v$ image. In fact, the pixel resolution $L = L_u L_v$ can be much higher than those of the associated input image $\boldsymbol{y}^{(k)}$ given by $M = M_u M_v$ to take advantage of the guidance data.

We assume that each pair $(\boldsymbol{y}^{(k)}, \boldsymbol{z}^{(k)})$ is aligned to each other by means of sensor data fusion. This alignment is described by the pixel-wise mapping:

$$\boldsymbol{u}_y = \Phi(\boldsymbol{u}_z)\,, \tag{5.1}$$

where $\boldsymbol{u}_y \in \Omega_y$ denotes a pixel in an input image and $\boldsymbol{u}_z \in \Omega_z$ denotes the corresponding pixel position in the guidance image that encodes the same position in the captured scene but with a different modality. Note that the mapping $\Phi : \Omega_z \to \Omega_y$ needs not be bijective. In particular, one important situation is the case of a surjective mapping. In this situation, a set of pixel coordinates in a guidance image maps to the same pixel in the corresponding input image, which appears if the pixel resolution of the guidance data is higher than those of the input images. In the most common setup, a set of pixels $\boldsymbol{v}_z \in \omega_{yz}(\boldsymbol{u}_z)$ is mapped to the same position $\boldsymbol{u}_y$, see Fig. 5.1. Conversely, $\Phi^{-1}(\boldsymbol{u}_y)$ denotes a set of pixels in the guidance image that are associated with $\boldsymbol{u}_y$ under the assumption that the mapping is invertible[2]. If the underlying mapping is applied to fuse guidance and input images, both domains are aligned up a scale factor to preserve their pixel resolutions. For the sake of convenience, we limit ourselves to static systems, i. e. the sensors involved in the system have a fixed relative orientation. Hence, the mapping is constant over all frames. In this case, sensor data fusion can be either achieved by a software-based calibration involving image registration or can be directly implemented by the imaging system.

[1] Each guidance image $\boldsymbol{z}^{(k)}$ is synchronized in time with the respective input image $\boldsymbol{y}^{(k)}$.

[2] Notice that depending on the implementation of the sensor data fusion, the mapping might not be invertible and occlusions needs to be considered. For instance, this is the case in stereo vision setups used for hybrid range imaging [Kohl 15b].

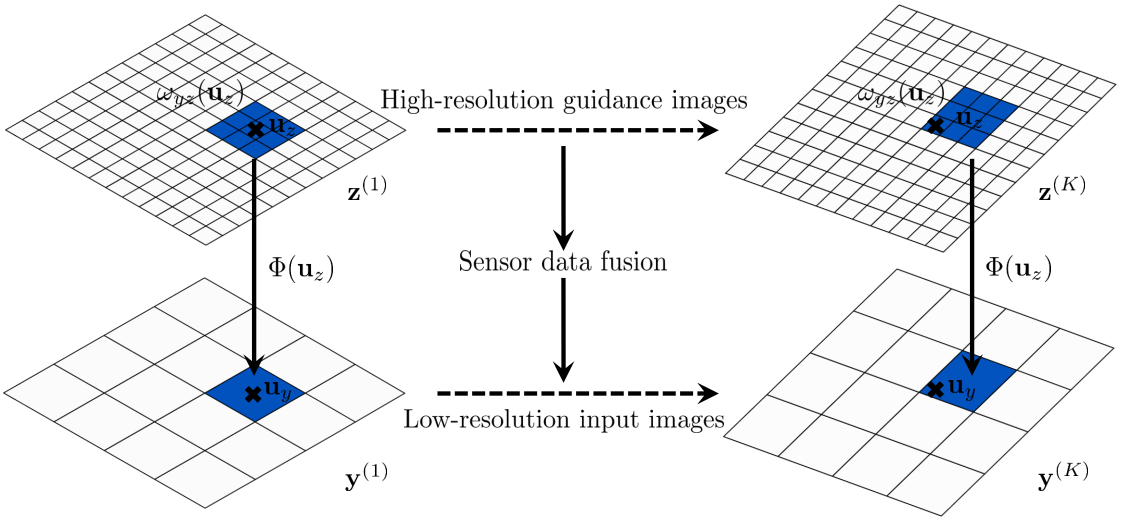


Figure 5.1: Illustration of sensor data fusion between low-resolution input images and the associated high-resolution guidance data for multi-sensor super-resolution. Each pair $(\boldsymbol{y}^{(k)}, \boldsymbol{z}^{(k)})$ is geometrically aligned up to a scale factor. In the proposed framework, the local neighborhood $\omega_{yz}(\boldsymbol{u}_z)$ centered at the pixel position $\boldsymbol{u}_z$ in a guidance image is mapped to the pixel position $\boldsymbol{u}_y$ in an input image according to the mapping $\Phi(\boldsymbol{u}_z)$. We assume a fixed mapping $\Phi(\boldsymbol{u}_z)$ over the sequence of input and guidance images.

Once the guidance data $\boldsymbol{z}$ is fused with the low-resolution input data $\boldsymbol{y}$, the proposed framework builds on the MAP estimator for the high-resolution image $\hat{\boldsymbol{x}}$ according to:

$$\boldsymbol{x}_{\mathrm{MAP}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ L_{\mathrm{MSR}}(\boldsymbol{x}, \boldsymbol{z}) + \lambda R_{\mathrm{MSR}}(\boldsymbol{x}, \boldsymbol{z}) \right\}. \tag{5.2}$$

The guidance data $\boldsymbol{z}$ is involved in two components. In terms of the observation model defined by the data fidelity term $L_{\mathrm{MSR}}(\boldsymbol{x}, \boldsymbol{z})$, guidance images are used to estimate subpixel motion. This avoids a direct motion estimation on low-resolution frames with the goal to enhance the accuracy of super-resolution. In terms of the image prior, a spatially adaptive regularization term $R_{\mathrm{MSR}}(\boldsymbol{x}, \boldsymbol{z})$ weighted by $\lambda \geq 0$ is used. This term exploits both, a super-resolved image $\boldsymbol{x}$ as well as the guidance data $\boldsymbol{z}$ with the goal of taking advantage of structural correlation across both modalities. Both novelties of the multi-sensor framework over the single-sensor counterpart are introduced in the following subsections.

5.3.2 Motion Estimation using Guidance Images

The proposed framework explicitly employs displacement vector fields as the most flexible approach to model subpixel motion. For this purpose, motion estimation is realized by means of optical flow computation [Liu 09]. In [Zhao 02], Zhao and Sawhney suggested that reconstruction-based super-resolution is feasible with this kind of motion estimation under the prerequisite of small noise in the estimated flow. However, the accuracy of optical flow is limited by noise, blur or aliasing present in low-resolution data. For this reason, many attempts have been made to treat image reconstruction and optical flow estimation in a joint framework, e. g.

by probabilistic methods [Fran 07]. This has the goal to compensate for inaccurate optical flow.

To meet the requirement regarding precise optical flow and to circumvent its direct estimation on low-resolution data, the proposed motion estimation is driven by guidance images and implemented as computationally efficient local filtering of dense displacement vector fields. In this *filter-based* approach, we determine displacements fields $m_z^{(k)}(\boldsymbol{u}_z)$ of each guidance image $\boldsymbol{z}^{(k)}$ relative to a fixed reference frame $\boldsymbol{z}^{(r)}$ with $r \neq k$. In order to obtain the associated displacements in the domain of the input images, we take advantage of the sensor data fusion with the guidance data. For the realization of motion estimation, we assume that motion present in input data is also encoded by the guidance data if both modalities are co-aligned. Intuitively this means that both sensors involved in this setup „see" the same scene and describe the same motion.

Based on sensor data fusion, a displacement vector field $m_y(\boldsymbol{u}_y)$ for a single input frame $\boldsymbol{y}$ relative to the reference frame is obtained from $m_z(\boldsymbol{u}_z)$ estimated on guidance images as depicted in Fig. 5.2. This process consists of two steps:

1. The displacement field $m_z(\boldsymbol{u}_z)$ that is given in terms of pixel units in the domain of the guidance images is first rescaled element-wise to:

$$\tilde{m}_z(\boldsymbol{u}_z) = \begin{pmatrix} \frac{M_u}{L_u} \cdot m_{z,u}(\boldsymbol{u}_z) \\ \frac{M_v}{L_v} \cdot m_{z,v}(\boldsymbol{u}_z) \end{pmatrix}. \tag{5.3}$$

   Thus, the displacements $\tilde{m}_z(\boldsymbol{u}_z)$ are defined in units of low-resolution pixels.

2. The displacements on the input frames are determined from the intermediate displacements obtained in the first stage by resampling described by the filter operation $m_y(\boldsymbol{u}) = \Delta\{\tilde{m}_z(\boldsymbol{u})\}$. The filtering of the displacement field is performed element-wise according to:

$$\begin{aligned} m_y(\boldsymbol{u}_y) &= \Delta\{\tilde{m}_z(\boldsymbol{u}_z)\} \\ &:= \begin{pmatrix} \Delta_{\omega_{yz}(\boldsymbol{u}_z),u}\{\tilde{m}_{z,u}(\boldsymbol{u}_z)\} \\ \Delta_{\omega_{yz}(\boldsymbol{u}_z),v}\{\tilde{m}_{z,v}(\boldsymbol{u}_z)\} \end{pmatrix}, \end{aligned} \tag{5.4}$$

   where $\omega_{yz}(\boldsymbol{u}_z)$ denotes the set of pixels in a local window centered at position $\boldsymbol{u}_z$ that corresponds to a single pixel position $\boldsymbol{u}_y$ in the low-resolution input image, see Fig. 5.2.

This filter-based technique enables noise suppression to deal with single erroneously estimated displacements. In addition to noise suppression, it needs to preserve motion discontinuities in the original displacement fields. To this end, the resampling used in the second stage of the proposed technique is formulated as local median filtering:

$$\Delta_{\omega_{yz}(\boldsymbol{u}_z),i}\{\tilde{m}_{z,i}(\boldsymbol{u}_z)\} = \text{median}_{\boldsymbol{v}_z \in \omega_{yz}(\boldsymbol{u}_z)}(\tilde{m}_{z,i}(\boldsymbol{v}_z)), \tag{5.5}$$

where $\text{median}_{\boldsymbol{v} \in \omega_{yz}(\boldsymbol{u})}(\cdot)$ computes the median of the displacements estimated in the local neighborhood $\omega_{yz}(\boldsymbol{u})$ in the coordinate direction $i \in \{u, v\}$.

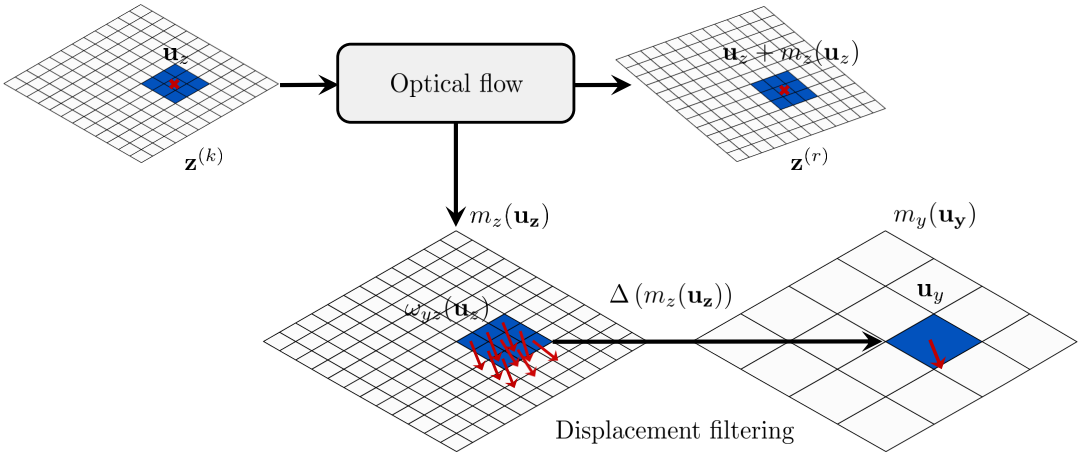


Figure 5.2: Flowchart of filter-based motion estimation using guidance images. First, the displacement field $m_z(\boldsymbol{u}_z)$ is gained by optical flow estimation of the frame $\boldsymbol{z}^{(k)}$ towards the reference frame $\boldsymbol{z}^{(r)}$. Then, the displacement field $m_y(\boldsymbol{u}_y)$ is determined from $m_z(\boldsymbol{u}_z)$ using patch-wise filtering with the neighborhood $\omega_{yz}(\boldsymbol{u}_z)$.

5.3.3 Spatially Adaptive Regularization using Guidance Images

Spatially adaptive regularization exploits the fact that input and guidance images capture the same structural content but are encoded by different modalities. In particular, we make use of the assumption that there exist correlations between both image types due to common structures. While this is of course not completely true over the entire image, one can assume correlations in terms of a set of structural features, e. g. areas or edges. This idea takes the same line as related concepts that became a standard in hybrid image processing, e. g. color-guided range upsampling [Park 11, Fers 13, Kiec 13] or CT-guided PET reconstruction [Kenn 07].

In this chapter, correlations among modalities are modeled by a weighting function $\alpha : \Omega_x \times \Omega_z \to \mathbb{R}_0^{+N}$ that exploits the image $\boldsymbol{x}$ and a corresponding guidance image $\boldsymbol{z}$. This function is used to control the regularization term:

$$R_{\mathrm{MSR}}(\boldsymbol{x}, \boldsymbol{z}) = \alpha\,(\boldsymbol{x}, \boldsymbol{z})^{\top}\, \phi_{\mathrm{MSR}}(\boldsymbol{Q}\boldsymbol{x})\,, \tag{5.6}$$

where $\phi_{\mathrm{MSR}}(\cdot)$ denotes a loss function applied on a high-pass filtered version of $\boldsymbol{x}$ to penalize residual noise[3]. This is done using a discrete filter modeled by the circulant matrix $\boldsymbol{Q} \in \mathbb{R}^{N \times N}$. Similar to the sparse regularization proposed in Chapter 4, the adaptive weights $\alpha(\boldsymbol{x}, \boldsymbol{z})$ are controlled in such a way that the regularizer does not penalize discontinuities in $\boldsymbol{x}$. In contrast to Chapter 4, the selection of the weights is steered by guidance data. For this purpose, the guidance images are used to extract discontinuities that are considered as relevant structural features. This process is driven by edge detection on the guidance image $\boldsymbol{z}$ to obtain an edge map $\tau : \Omega_z \to \{0, 1\}^L$, where $\tau(\boldsymbol{u}) = 1$ indicates an edge at position $\boldsymbol{u}$.

The adaptive weights are first computed in the domain of the guidance data as:

$$\tilde{\alpha}\,(\tilde{\boldsymbol{x}}, \boldsymbol{z}) = \begin{pmatrix} \tilde{\alpha}_1\,(\tilde{\boldsymbol{x}}, \boldsymbol{z}) & \tilde{\alpha}_2\,(\tilde{\boldsymbol{x}}, \boldsymbol{z}) & \dots & \tilde{\alpha}_L\,(\tilde{\boldsymbol{x}}, \boldsymbol{z}) \end{pmatrix}^{\top} \in \mathbb{R}_0^{+L}, \tag{5.7}$$

[3] Since the proposed framework is flexible regarding the choice of the loss function $\phi_{\mathrm{MSR}}(\cdot)$, we tailor the regularization to the characteristics of specific applications. For instance, the Huber loss can be used to implement piecewise-smooth regularization, see [Kohl 13b].

where $\tilde{\boldsymbol{x}}$ is the image $\boldsymbol{x}$ resampled to the size of the guidance image $\boldsymbol{z}$ using bicubic interpolation. The weight at the i-th pixel position $\boldsymbol{u}_i$ is computed by:

$$\tilde{\alpha}_i\left(\tilde{\boldsymbol{x}}, \boldsymbol{z}\right) = \begin{cases} \exp\left\{-\frac{\rho(\tilde{\boldsymbol{x}}, \boldsymbol{z}, \omega_{xz}(\boldsymbol{u}_i))}{\tau_0}\right\} & \text{if } \tau(\boldsymbol{u}_i) = 1 \\ 1 & \text{otherwise} \end{cases}, \tag{5.8}$$

where $\rho(\tilde{\boldsymbol{x}}, \boldsymbol{z}, \omega_{xz}(\boldsymbol{u}_i))$ reflects the degree of correlation computed by a similarity measure between the image $\tilde{\boldsymbol{x}}$ and guidance data $\boldsymbol{z}$ in a $N_{xz} \times N_{xz}$ local neighborhood $\omega_{xz}(\boldsymbol{u}_i)$ centered at $\boldsymbol{u}_i$. The parameter τ_0 denotes a contrast factor to map the local image similarity at positions corresponding to an edge to a weight $\tilde{\alpha}_i(\tilde{\boldsymbol{x}}, \boldsymbol{z}) \in [0, 1]$. Finally, the weights $\alpha(\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}_0^{+N}$ for the regularization term in Eq. (5.6) are obtained by bicubic interpolation of $\tilde{\alpha}(\tilde{\boldsymbol{x}}, \boldsymbol{z}) \in \mathbb{R}_0^{+L}$ to the domain of super-resolved data. This reduces the impact of regularization for image regions associated with edges in the guidance data according to the similarity measure.

In Eq. (5.8), the similarity measure $\rho(\tilde{\boldsymbol{x}}, \boldsymbol{z}, \omega_{xz}(\boldsymbol{u}))$ indicates how reasonable the assumption of correlations in terms of discontinuities actually is. In particular, it needs to downweight the impact of guidance data if this assumption does not hold true for certain image regions. Multi-modal measures are utilized to analyze these similarities. This can be achieved by cross correlation as used in mutual-structure filters [Shen 15] or by information theoretic measures. In this work, the similarity is assessed by local mutual information (LMI) that has been also successfully applied in related fields of image enhancement like adaptive TV denoising [Guo 08]. The LMI is computed following the definition of Pluim et al. [Plui 03]:

$$\rho_{\mathrm{mi}}(\tilde{\boldsymbol{x}}, \boldsymbol{z}, \omega_{xz}(\boldsymbol{u})) = \sum_{\tilde{x}_i, z_i \in \omega_{xz}(\boldsymbol{u})} p(\tilde{x}_i, z_i) \log\left(\frac{p(\tilde{x}_i, z_i)}{p(\tilde{x}_i)\, p(z_i)}\right), \tag{5.9}$$

where $p(\tilde{x}_i, z_i)$ is an estimate of the joint PDF of the samples extracted from the local neighborhood $\omega_{xz}(\boldsymbol{u})$ in the images $\tilde{\boldsymbol{x}}$ and $\boldsymbol{z}$. Similarly, $p(\tilde{x}_i)$ and $p(z_i)$ are the corresponding marginals. To define $\rho(\tilde{\boldsymbol{x}}, \boldsymbol{z}, \omega_{xz}(\boldsymbol{u}))$, we use the normalized LMI:

$$\rho(\tilde{\boldsymbol{x}}, \boldsymbol{z}, \omega_{xz}(\boldsymbol{u})) = -\frac{\rho_{\mathrm{mi}}(\tilde{\boldsymbol{x}}, \boldsymbol{z}, \omega_{xz}(\boldsymbol{u}))}{\sum_{\tilde{x}_i, z_i \in \omega_{xz}(\boldsymbol{u})} p(\tilde{x}_i, z_i) \log p(\tilde{x}_i, z_i)}. \tag{5.10}$$

The size of $\omega_{xz}(\boldsymbol{u})$ in Eq. (5.10) is adjusted to balance between too many empty bins in the joint histogram $p(\tilde{x}_i, z_i)$ and a too low resolution of LMI.

5.3.4 Numerical Optimization

In this basic approach to multi-sensor super-resolution, we solve Eq. (5.2) with a Gaussian observation model. This leads to the data fidelity term:

$$L_{\mathrm{MSR}}(\boldsymbol{x}, \boldsymbol{z}) = \sum_{k=1}^{K} \left|\left|\boldsymbol{y}^{(k)} - \boldsymbol{W}_z^{(k)} \boldsymbol{x}\right|\right|_2^2, \tag{5.11}$$

where the system matrix $\boldsymbol{W}_z^{(k)}$ is parametrized with the displacement fields obtained from the optical flow on the guidance images, see Section 5.3.2. The regularization term is driven by the adaptive weights $\boldsymbol{\alpha}$, see Section 5.3.3.

Algorithm 5.1 Two-stage multi-sensor super-resolution

Input: Initial guess for high-resolution image $\boldsymbol{x}$
Output: Final high-resolution image $\boldsymbol{x}$ and spatially adaptive regularization weights $\boldsymbol{\alpha}$

for $k = 1, \ldots, K$ **do**
    Compute optical flow $m_z^{(k)}(\boldsymbol{u}_z)$ of $\boldsymbol{z}^{(k)}$ towards the reference $\boldsymbol{z}^{(r)}$
    Compute $m_y^{(k)}(\boldsymbol{u}_z)$ by local filtering of $m_z^{(k)}(\boldsymbol{u}_z)$ according to Eq. (5.4)
end for
while SCG convergence criteria not fulfilled **do**
    Update $\boldsymbol{x}$ by SCG iteration for Eq. (5.2) using uniform weights $\boldsymbol{\alpha} = \boldsymbol{1}$
end while
Set $\boldsymbol{z}$ to the temporal motion-compensated median of $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(K)}$
Compute spatially adaptive weights $\boldsymbol{\alpha}$ according to Eq. (5.8)
while SCG convergence criteria not fulfilled **do**
    Update $\boldsymbol{x}$ by SCG iteration for Eq. (5.2) using the adaptive weights $\boldsymbol{\alpha}$
end while

The optimization of Eq. (5.2) is performed in an alternating fashion to jointly estimate the super-resolved image and the weights $\boldsymbol{\alpha}$. We limit this scheme to two stages corresponding to two optimization loops as outlined in Algorithm 5.1.

In the first stage, we reconstruct an intermediate solution for the super-resolved image using SCG iterations [Nabn 02] given uniform weights in the regularization term. The initial guess for these iterations is obtained by bicubic interpolation of the reference input image. In the second stage, we compute the spatially adaptive weights $\boldsymbol{\alpha}$ using the super-resolved image of the first stage and the respective guidance image. Instead of using a single image, the spatially adaptive weights are obtained from the motion-compensated temporal median of the sequence $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(K)}$ to enhance the accuracy of edge detection. Finally, given the weights $\boldsymbol{\alpha}$, the image reconstructed in the first stage is iteratively refined by SCG and spatially adaptive regularization.

5.4 Outlier Detection for Robust Multi-Sensor Super-Resolution

The framework presented in the previous section is based on several idealizing assumptions limiting its robustness in real-world applications. First and foremost, it neglects outliers in the optical flow estimated on guidance data. While noise in the displacement fields can be compensated by the proposed filter-based technique, motion estimation is prone to occlusions or inaccurate flows in textureless regions. In addition, we derived Algorithm 5.1 under a Gaussian observation model that neglects outliers in low-resolution data.

In order to deal with the aforementioned issues, this section presents an outlier detection scheme as extension of the two-stage framework. This approach is divided into two separate detection schemes applied on guidance and input data. These techniques yield confidence maps associated with the input data and their displacement fields, which are combined for robust super-resolution.

5.4.1 Outlier Detection on Guidance Images

Similar to the filter-based technique in Section 5.3.2, the detection of outliers in terms of motion estimation is driven by guidance images. The proposed outlier detection is inspired by the image similarity based method of Zhao and Sawhney [Zhao 02] but adopted in such a way that it is applicable on guidance images instead of using the low-resolution input frames directly. This approach determines the reliability of the estimated displacement fields that affects the robustness of super-resolution.

For outlier detection, the reference frame in the domain of the guidance images denoted as $z^{(r)}$ is warped towards each of the remaining frames $z^{(k)}, k \neq r$ according to the estimated displacements. Then, we assess the consistency of each target frame $z^{(k)}$ relative to the warped reference $\tilde{z}^{(k)}$ as shown in Fig. 5.3 for one of these pairs. Similar to the approach in [Zhao 02], this consistency is assessed by the normalized cross correlation (NCC) that is used as a local similarity measure. The local NCC at the i-th pixel position $\boldsymbol{u}_i$ in the guidance images is computed for the local window $\omega_{yz}(\boldsymbol{u}_i)$ centered at $\boldsymbol{u}_i$ according to:

$$\rho_{\mathrm{ncc}}\left(\boldsymbol{z}^{(k)}, \tilde{\boldsymbol{z}}^{(k)}, \omega_{yz}(\boldsymbol{u}_i)\right) = \frac{\sum\limits_{\boldsymbol{v}_j \in \omega_{yz}(\boldsymbol{u}_i)} \left(z_j^{(k)} - \mu_i\right)\left(\tilde{z}_j^{(k)} - \tilde{\mu}_i\right)}{\sqrt{\sum\limits_{\boldsymbol{v}_j \in \omega_{yz}(\boldsymbol{u}_i)} \left(z_j^{(k)} - \mu_i\right)^2 \sum\limits_{\boldsymbol{v}_j \in \omega_{yz}(\boldsymbol{u}_i)} \left(\tilde{z}_j^{(k)} - \tilde{\mu}_i\right)^2}}, \tag{5.12}$$

where $z_j^{(k)}$ and $\tilde{z}_j^{(k)}$ are the elements in $\boldsymbol{z}^{(k)}$ and $\tilde{\boldsymbol{z}}^{(k)}$ at the position $\boldsymbol{v}_j \in \omega_{yz}(\boldsymbol{u}_i)$, and μ_i and $\tilde{\mu}_i$ are the local means of $\boldsymbol{z}^{(k)}$ and $\tilde{\boldsymbol{z}}^{(k)}$ in $\omega_{yz}(\boldsymbol{u}_i)$, respectively. In order to transform this local similarity to the domain of the low-resolution images, it is computed by means of patch-wise processing as shown in Fig. 5.3.

The local similarity measures the fidelity of the image warping according to the optical flow and is used for outlier detection. In the proposed detection scheme, the local NCC computed in the range $[-1, +1]$ is used to determine the confidence weight:

$$\beta_{z,i}\left(\boldsymbol{z}^{(k)}\right) = \begin{cases} \frac{1}{2}\rho_{\mathrm{ncc}}\left(\boldsymbol{z}^{(k)}, \tilde{\boldsymbol{z}}^{(k)}, \omega_{yz}(\boldsymbol{u}_i)\right) + \frac{1}{2} & \text{if } \rho_{\mathrm{ncc}}\left(\boldsymbol{z}^{(k)}, \tilde{\boldsymbol{z}}^{(k)}, \omega_{yz}(\boldsymbol{u}_i)\right) \geq \rho_0 \\ 0 & \text{otherwise} \end{cases}, \tag{5.13}$$

where $\rho_0 \in [-1, +1]$ is a fixed threshold to classify observations associated with a poor local similarity as an outlier. For the remaining observations that are not classified as outliers, a higher local image similarity due to more accurate optical flow estimation indicates a higher confidence. The confidence map for the k-th frame is assembled as:

$$\beta_z\left(\boldsymbol{z}^{(k)}\right) = \left(\beta_{z,1}\left(\boldsymbol{z}^{(k)}\right) \quad \beta_{z,2}\left(\boldsymbol{z}^{(k)}\right) \quad \ldots \quad \beta_{z,M}\left(\boldsymbol{z}^{(k)}\right)\right)^\top. \tag{5.14}$$

Then, the joint confidence map for the entire image sequence is constructed as:

$$\beta_z\left(\boldsymbol{z}\right) = \left(\beta_z\left(\boldsymbol{z}^{(1)}\right) \quad \beta_z\left(\boldsymbol{z}^{(2)}\right) \quad \ldots \quad \beta_z\left(\boldsymbol{z}^{(K)}\right)\right)^\top. \tag{5.15}$$

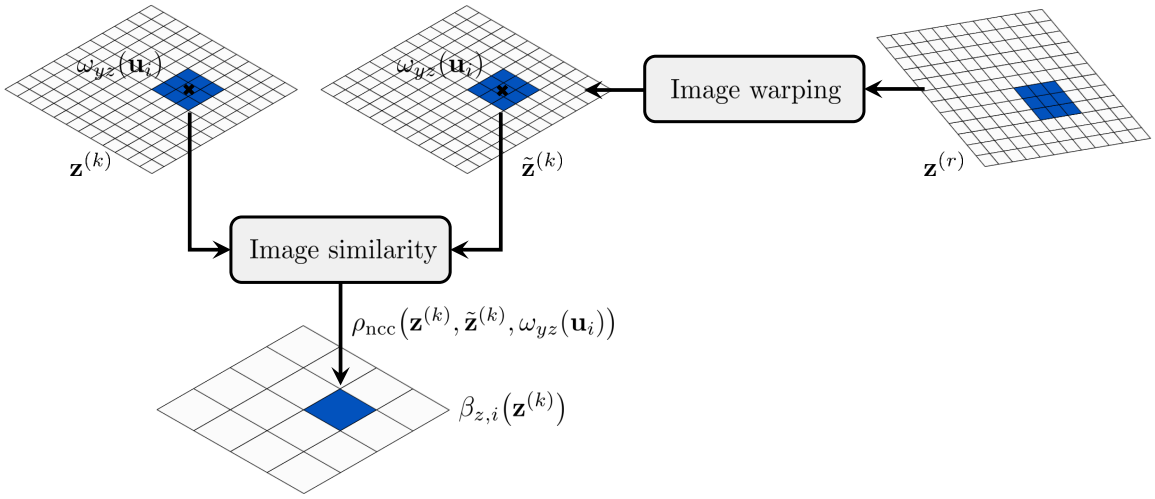


Figure 5.3: Illustration of outlier detection on guidance images. The local image similarity between the k-th frame $\boldsymbol{z}^{(k)}$ and the warped reference $\tilde{\boldsymbol{z}}^{(k)}$ according to the estimated optical flow is used to calculate the confidence weight $\beta_{z,i}(\boldsymbol{z}^{(k)})$. The resulting confidence map $\beta_z\left(\boldsymbol{z}^{(k)}\right)$ is constructed by patch-wise processing with the neighborhood $\omega_{yz}(\boldsymbol{u}_i)$ and defined in the domain of the input images.

5.4.2 Outlier Detection on Input Images

In addition to outlier detection in the estimated displacement fields based on the guidance images, the low-resolution input frames themselves are assessed to remove outliers. This takes outliers due to non-Gaussian noise into account that cannot be detected on the guidance images. This outlier detection is formulated in an implicit way in accordance to the algorithm presented in Section 4.4.

Let $\boldsymbol{x}$ be an estimate for the super-resolved image. Then, we define the confidence weight associated with the i-th pixel in the k-th frame $\boldsymbol{y}^{(k)}$ according to:

$$\beta_{y,i}\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right) = \beta_{\mathrm{bias},i}\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right) \beta_{\mathrm{local},i}\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right), \tag{5.16}$$

where the weighting functions are given by:

$$\beta_{\mathrm{bias},i}\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right) = \begin{cases} 1 & \text{if } \left|\mathrm{median}\left(\boldsymbol{y}^{(k)} - \boldsymbol{W}_z^{(k)}\boldsymbol{x}\right)\right| \leq c_{\mathrm{bias}} \\ 0 & \text{otherwise} \end{cases}, \tag{5.17}$$

$$\beta_{\mathrm{local},i}\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right) = \begin{cases} 1 & \text{if } \left|\left[\boldsymbol{y}^{(k)} - \boldsymbol{W}_z^{(k)}\boldsymbol{x}\right]_i\right| \leq c_{\mathrm{local}}\sigma_{\mathrm{noise}} \\ \frac{c_{\mathrm{local}}\sigma_{\mathrm{noise}}}{\left|\left[\boldsymbol{y}^{(k)} - \boldsymbol{W}_z^{(k)}\boldsymbol{x}\right]_i\right|} & \text{otherwise} \end{cases}, \tag{5.18}$$

with noise level σ_{noise} and tuning constants c_{bias} and c_{local} to detect biased frames along with local outliers as shown in Section 4.4.1. The confidence map for the k-th low-resolution frame is assembled according to:

$$\beta_y\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right) = \left(\beta_{y,1}\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right) \quad \beta_{y,2}\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right) \quad \ldots \quad \beta_{y,M}\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right)\right)^\top. \tag{5.19}$$

Then, the joint confidence map for all low-resolution observations is given by:

$$\beta_y\left(\boldsymbol{x}, \boldsymbol{y}\right) = \left(\beta_y\left(\boldsymbol{x}, \boldsymbol{y}^{(1)}\right) \quad \beta_y\left(\boldsymbol{x}, \boldsymbol{y}^{(2)}\right) \quad \ldots \quad \beta_y\left(\boldsymbol{x}, \boldsymbol{y}^{(K)}\right)\right)^\top. \tag{5.20}$$

5.4.3 Numerical Optimization

For an outlier-aware numerical optimization, super-resolution is performed by means of iteratively re-weighted minimization. We estimate the super-resolved image iteratively via a sequence of weighted minimization problems:

$$\boldsymbol{x}^t = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ (\boldsymbol{y} - \boldsymbol{W}_z \boldsymbol{x})^\top \boldsymbol{B}^t (\boldsymbol{y} - \boldsymbol{W}_z \boldsymbol{x}) + \lambda R_{\mathrm{MSR}}(\boldsymbol{x}, \boldsymbol{z}) \right\}, \tag{5.21}$$

where $R_{\mathrm{MSR}}(\boldsymbol{x}, \boldsymbol{z})$ denotes the spatially adaptive regularization term defined in Eq. (5.6) with constant regularization weight[4] $\lambda \geq 0$. In this weighted minimization, $\boldsymbol{B}^t$ denotes the confidence map at iteration t. In order to take outliers in the displacement fields and the low-resolution observations into account, $\boldsymbol{B}^t$ is constructed according to:

$$\boldsymbol{B}^t = \operatorname{diag}\Big(\beta_z(\boldsymbol{z}) \odot \beta_y\left(\boldsymbol{x}^{t-1}, \boldsymbol{y}\right) \Big), \tag{5.22}$$

where $\odot$ is the Hadamard (element-wise) product and $\beta_z(\boldsymbol{z})$ is the confidence map determined from the local image similarity according to Eq. (5.15). Notice that these confidence weights can be pre-computed and kept fixed over the iterations. The confidence map $\beta_y(\boldsymbol{x}^{t-1}, \boldsymbol{y})$ is updated dynamically according to Eq. (5.20) based on the estimate $\boldsymbol{x}^{t-1}$ obtained at the previous iteration. In this dynamic weighting scheme, we use an adaptive estimate of the noise standard deviation σ_{noise} at each iteration according to the MAD rule, see Section 4.4.1.

This iterative procedure is initialized by the super-resolved image $\boldsymbol{x}^0$ and the corresponding adaptive weights $\boldsymbol{\alpha}$ for the regularization term. These initializations are obtained by the two-stage approach introduced in Algorithm 5.1. Afterwards, the super-resolved image is gradually refined using SCG iterations. The overall optimization scheme is outlined in Algorithm 5.2.

5.5 Application to Hybrid Range Imaging

This section validates the proposed multi-sensor framework in the field of hybrid range imaging that is considered as example application. In this area, we aim at reconstructing high-resolution surface information from sequences of low-resolution range images to overcome resolution limitations of current range sensors. Prior work in this field approached this task solely on range images as shown by Schuon et al. [Schu 08, Schu 09] or Bhavsar and Rajagopalan [Bhav 12]. Unlike these single-sensor approaches, we employ high-resolution color images as a guidance to super-resolve low-resolution range data.

We first introduce a tailor-made image formation model for range imaging. Subsequently, this model is adopted in the multi-sensor framework to formulate range super-resolution reconstruction. Eventually, we present a proof-of-concept evaluation for ToF imaging based on simulated datasets. For a thorough experimental evaluation on real image data within the scope of interventional medical imaging, we refer to Chapter 8.

[4] An adaptive version of this algorithm, where the regularization weight λ is re-computed per iteration can be developed using the cross validation scheme presented in Section 4.4.1.

Algorithm 5.2 Robust multi-sensor super-resolution using outlier detection

Input: Initial guess for high-resolution image $\boldsymbol{x}^0$ and adaptive regularization weights $\boldsymbol{\alpha}$
Output: Final high-resolution image $\boldsymbol{x}$ and joint confidence map $\boldsymbol{B}$

for $k = 1 \dots K$ **do**
    Construct confidence map $\beta_z(\boldsymbol{z}^{(k)})$ according to Eq. (5.14)
end for
Construct confidence map $\beta_z(\boldsymbol{z})$ for all guidance images according to Eq. (5.15)
$t \leftarrow 1$
while Convergence criterion not fulfilled **do**
    Update confidence map $\beta_y(\boldsymbol{x}^{t-1}, \boldsymbol{y})$ according to Eq. (5.20)
    Update joint confidence map $\boldsymbol{B}^t$ according to Eq. (5.22)
    while SCG convergence criterion not fulfilled **do**
        Update $\boldsymbol{x}^t$ by SCG iteration for Eq. (5.21) using adaptive weights $\boldsymbol{\alpha}$
    end while
    $t \leftarrow t + 1$
end while

5.5.1 Image Formation Model for Range Imaging

In terms of the image formation model, we need to adopt the model presented in Chapter 3 to describe the formation of low-resolution range images from high-resolution ones. Here, one crucial aspect is the formulation of the motion model. Following the concept of multi-sensor super-resolution, the motion associated with the k-th range image $\boldsymbol{y}^{(k)}$ is first estimated by means of optical flow from the color image $\boldsymbol{z}^{(k)}$ and subsequently projected to the domain of the range data. Then, the motion on the range images encoded by dense displacement fields should ideally represent the motion that appears in the 3-D space.

However, under a general type of camera motion, the actual motion that appears in range images cannot be described solely based on 2-D displacement fields. One prominent example is camera motion orthogonal to the measured surface and the sensor image plane, which is referred to as *out-of-plane* motion. In this case, the motion in color images and hence the estimated displacements on range images provide only a 2-D view on the actual scene motion. For this reason, it does not explain the out-of-plane component appropriately. A similar situation appears if there is a tilting of a surface across two frames. Notice that related range super-resolution techniques [Schu 08, Schu 09, Bhav 12] ignore this limitation of the motion model. However, neglecting this aspect is only reasonable for situations that allow an accurate description of the actual motion by 2-D displacement fields, whereas more general types of motion lead to a bias in super-resolution reconstruction [Kohl 15b]. Therefore, we extend the image formation model to better explain the actual motion.

Formulation of the Model. To enhance the modeling of scene motion and to take out-of-plane movements into account, a transformation of the measured range values is used in addition to 2-D displacement fields. For the derivation of the motion model, let $\boldsymbol{y}^{(r)}$ be the reference range image and $\boldsymbol{y}^{(k)}$ be the frame that needs to be explained under this model. The most simplest but non-trivial transformation be-

tween $\boldsymbol{y}^{(r)}$ and $\boldsymbol{y}^{(k)}$ as a better approximation to the actual 3-D motion is given by:

$$\boldsymbol{y}^{(k)} = \gamma_m^{(k)} \mathcal{M}\{\boldsymbol{y}^{(r)}\} + \gamma_a^{(k)}, \tag{5.23}$$

where the motion operator $\mathcal{M}\{\boldsymbol{y}^{(r)}\}$ describes the subpixel motion in the domain of the range data, and $\gamma_m^{(k)} \in \mathbb{R}$ and $\gamma_a^{(k)} \in \mathbb{R}$ are called the *range correction* parameters. While the former describes motion on the image plane, the latter account for more general types of 3-D motion. The additive parameter $\gamma_a^{(k)}$ describes a global shift of the range values and can roughly explain out-of-plane motion. Similarly, the multiplicative parameter $\gamma_m^{(k)}$ describes a shearing of the range values.

The range correction parameters are used to formulate the image formation model:

$$\boldsymbol{y}^{(k)} = \gamma_m^{(k)} \boldsymbol{W}_z^{(k)} \boldsymbol{x} + \gamma_a^{(k)} + \epsilon^{(k)}, \tag{5.24}$$

which describes the formation of the k-th range image from the high-resolution range data $\boldsymbol{x}$ according to the system matrix $\boldsymbol{W}_z^{(k)}$ and the observation noise $\epsilon^{(k)}$. The system matrix is defined by the motion information on the image plane given by displacement vector fields as well as the underlying sampling model.

Model Parameter Estimation. The range correction parameters need to be estimated from the low-resolution range images $\boldsymbol{y}^{(k)}$, $k \neq r$ relative to the reference frame $\boldsymbol{y}^{(r)}$. From a conceptual point of view, this is equivalent to a photometric registration of intensity images acquired under varying photometric conditions as shown in the work of Capel and Zisserman [Cape 03]. Let $(y_i, \tilde{y}_i)$ be a pair of range values that are obtained from $\boldsymbol{y}^{(r)}$ and $\tilde{\boldsymbol{y}}^{(k)}$, where $\tilde{\boldsymbol{y}}^{(k)}$ is the k-th frame $\boldsymbol{y}^{(k)}$ warped towards the reference frame according to the estimated displacement fields. Then, the range correction parameters can be determined by pair-wise registration. The registration associated with the k-th frame is formulated as the line fitting problem:

$$(\hat{\gamma}_m, \hat{\gamma}_a) = \underset{\gamma_m, \gamma_a}{\operatorname{argmin}} \sum_{i=1}^{M} \phi_{\text{range}}(\tilde{y}_i - \gamma_m y_i - \gamma_a), \tag{5.25}$$

where $\phi_{\text{range}}(r)$ denotes a loss function applied to the residual errors for M pairs of range values. A simulated example is depicted in Fig. 5.4.

Estimating the range correction parameters is challenging due to random measurement noise, systematic errors in the range data, or outliers in optical flow. To deal with these issues, the range correction needs to be performed by robust parameter estimation and is applied on a median filtered version of original range values. In this work, Eq. (5.25) is solved by probabilistic optimization using the M-estimator sample consensus (MSAC) algorithm [Torr 00]. The loss function that is used for MSAC is given by the truncated least-squares term:

$$\phi_{\text{range}}(r) = \min\left(\delta_{\text{msac}}^2, r^2\right). \tag{5.26}$$

This loss function assigns a constant penalty δ_{msac}^2 to residual errors that exceed the threshold δ_{msac}^2, which are referred to as outliers. Residual errors that fall below this threshold are penalized quadratically and are considered as inliers. In this

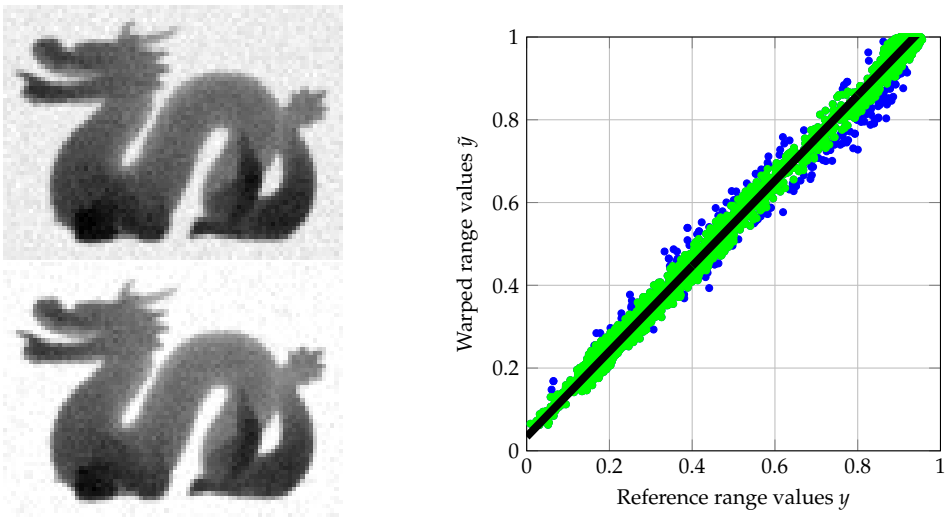


Figure 5.4: Range correction for a pair of range images in presence of out-of-plane motion. Left: Reference frame $\boldsymbol{y}^{(r)}$ and k-th frame $\boldsymbol{y}^{(k)}$, $k \neq r$ warped towards the reference according to the displacement field on the image plane. Right: Scatter plot of the range values $(y_i, \tilde{y}_i)$ drawn from $\boldsymbol{y}^{(r)}$ and $\boldsymbol{y}^{(k)}$ shown in blue. Inliers detected by MSAC ($\sigma_{\text{noise}} = 0.025$, $T_{\text{msac}} = 200$) are marked in green and the fitted model is visualized by the black line.

work, the threshold is adaptively set to $\delta_{\text{msac}} = 1.96\sigma_{\text{noise}}$ to achieve a correct classification of 95 % of the true inliers under the assumption that the range values are affected by zero-mean Gaussian noise with standard deviation σ_{noise}.

For a probabilistic optimization of Eq. (5.25), the initial parameter values for MSAC are set to $\gamma_m = 1$ and $\gamma_a = 0$, and only parameter settings that result in lower objective values are accepted within estimation to avoid unreliable solutions. Then, at each iteration, two pairs of range values $(y^1, \tilde{y}^1)$ and $(y^2, \tilde{y}^2)$ are randomly drawn from the images $\boldsymbol{y}^{(r)}$ and $\tilde{\boldsymbol{y}}^{(k)}$. From these pairs, (γ_m, γ_a) is computed in closed form. Accordingly, the pairs $(y_i, \tilde{y}_i)$ for $i = 1, \ldots, M$ are classified either as inliers or outliers in accordance to the objective value, which yields an inlier set associated with the current iteration. This procedure is repeated for T_{msac} iterations to detect the optimal inlier set $\mathcal{Y}_{\text{min}}$ that leads to a minimum objective value. Finally, the inlier set $\mathcal{Y}_{\text{min}}$ is used to gain $(\hat{\gamma}_m, \hat{\gamma}_a)$ by linear least-squares estimation, see Fig. 5.4. Algorithm 5.3 summarizes the overall procedure to determine the range correction parameters for one pair of range images.

5.5.2 Range Super-Resolution Reconstruction

Let us next adopt Algorithm 5.1 and Algorithm 5.2 to the desired application. This requires custom observation and prior models.

The observation model for range data is described by a weighted normal distribution to account for space variant noise characteristics of current range sensors. Hence, we employ the confidence-aware data fidelity term:

$$L_{\text{MSR}}(\boldsymbol{x}) = \sum_{i=1}^{K} \left(\boldsymbol{y}^{(k)} - \gamma_m^{(k)} \boldsymbol{W}_z^{(k)} \boldsymbol{x} - \gamma_a^{(k)} \right)^{\top} \boldsymbol{B}^{(k)} \left(\boldsymbol{y}^{(k)} - \gamma_m^{(k)} \boldsymbol{W}_z^{(k)} \boldsymbol{x} - \gamma_a^{(k)} \right), \quad (5.27)$$

Algorithm 5.3 M-estimator sample consensus (MSAC) based range correction

Input: Pair of range images $\boldsymbol{y}^{(r)}$ (reference) and $\boldsymbol{y}^{(k)}$ (template)
Output: Range correction parameters $\hat{\gamma}_m$ and $\hat{\gamma}_a$

Determine $\tilde{\boldsymbol{y}}^{(k)}$ by warping $\boldsymbol{y}^{(k)}$ towards the reference $\boldsymbol{y}^{(r)}$
Initialize $\gamma_m \leftarrow 1$, $\gamma_a \leftarrow 0$, and $\phi_{\min} \leftarrow \sum_{i=1}^{M} \phi(\tilde{y}_i - \gamma_m y_i - \gamma_a)$
for $t = 1, \ldots, T_{\text{msac}}$ **do**
    Draw randomly selected pairs $(y^1, \tilde{y}^1)$ and $(y^2, \tilde{y}^2)$ from $\boldsymbol{y}^{(r)}$ and $\tilde{\boldsymbol{y}}^{(k)}$
    Estimate range correction parameters (γ_m, γ_a) from $(y^1, \tilde{y}^1)$ and $(y^2, \tilde{y}^2)$
    Initialize objective value $\phi^t \leftarrow 0$ and inlier set $\mathcal{Y}^t \leftarrow \{\}$
    for $i = 1, \ldots, M$ **do**
        Determine residual error $r_i = \tilde{y}_i - \gamma_m y_i - \gamma_a$
        if $r_i^2 < \delta_{\text{msac}}^2$ **then**
            Update inlier set according to $\mathcal{Y}^t \leftarrow \mathcal{Y}^t \cup \{(y_i, \tilde{y}_i)\}$
        end if
        Update objective value according to $\phi^t \leftarrow \phi^t + \min(\delta_{\text{msac}}^2, r_i^2)$
    end for
    if $\phi^t < \phi_{\min}$ **then**
        $\phi_{\min} \leftarrow \phi^t$ and $\mathcal{Y}_{\min} \leftarrow \mathcal{Y}^t$
    end if
end for
Determine $\hat{\gamma}_m$ and $\hat{\gamma}_a$ by least-squares estimation on the optimal inlier set $\mathcal{Y}_{\min}$

which is derived from the image formation model in Eq. (5.24). $\boldsymbol{B}^{(k)}$ denotes the joint confidence map of the k-th frame constructed from the range and color data. In the two-stage approach according to Algorithm 5.1, this confidence map is set to $\boldsymbol{B}^{(k)} = \boldsymbol{I}$. For robust super-resolution according to Algorithm 5.2, the confidence map is computed dynamically.

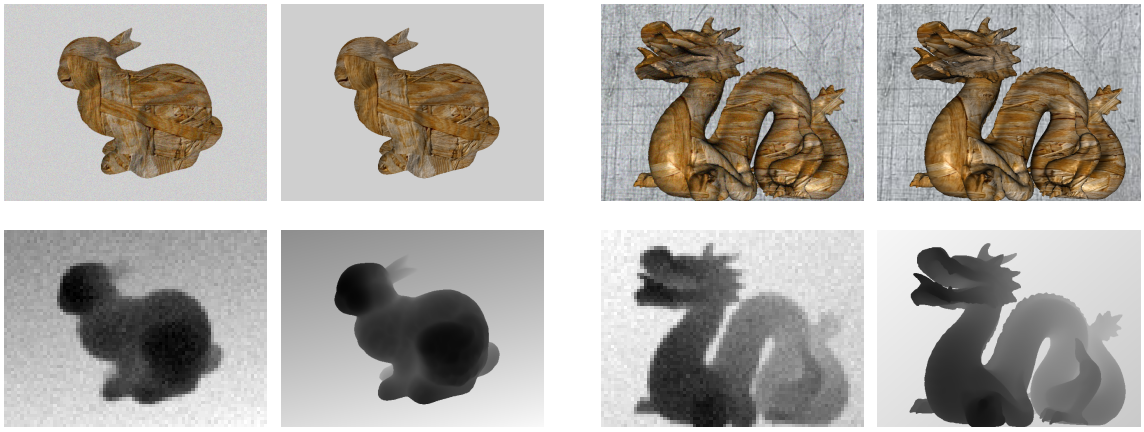
The image prior is formulated via piecewise smooth regularization to describe the appearance of smooth surfaces and depth discontinuities captured in range data. This type of regularization can be achieved by the Huber prior, see Section 3.3.2. Combined with the proposed spatially adaptive scheme, the regularization term for range super-resolution is defined as:

$$R_{\text{MSR}}(\boldsymbol{x}, \boldsymbol{z}) = \sum_{i=1}^{N} \alpha_i(\boldsymbol{x}, \boldsymbol{z}) \cdot \left(\delta_{\text{Huber}} \sqrt{1 + \left(\frac{[\boldsymbol{Q}\boldsymbol{x}]_i}{\delta_{\text{Huber}}} \right)^2} - \delta_{\text{Huber}} \right), \tag{5.28}$$

where δ_{Huber} denotes the Huber threshold parameter. $\boldsymbol{Q} \in \mathbb{R}^{N \times N}$ denotes the filter kernel of a discrete Laplacian expressed as a circulant matrix to exploit the curvature of range data for regularization. We define this circulant matrix according to:

$$\boldsymbol{Q}\boldsymbol{x} \equiv \frac{1}{4} \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \star \boldsymbol{X}, \tag{5.29}$$

where $\boldsymbol{X} \in \mathbb{R}^{M_u \times M_v}$ is a representation of $\boldsymbol{x} \in \mathbb{R}^{M_u M_v}$ in matrix notation and $\star$ denotes the discrete 2-D convolution.



(a) Simulated data **(b)** Ground truth **(c)** Simulated data **(d)** Ground truth

Figure 5.5: Simulated range and color data along with the ground truth data obtained from the artificial *Stanford Bunny* and the *Dragon* scenes.

5.5.3 Experiments and Results

Let us now present an experimental evaluation of the proposed multi-sensor super-resolution techniques in hybrid range imaging. The goal of this study is two-fold. On the one hand, we aim at comparing multi-sensor super-resolution to the conventional single-sensor approach as implemented by state-of-the-art algorithms [Schu 08, Schu 09, Bhav 12]. On the other hand, the influence of the different multi-sensor techniques including motion estimation, spatially adaptive regularization and outlier detection to the performance of the proposed framework is studied.

In order to conduct a quantitative evaluation, we limited ourselves to experiments on artificial datasets with known ground truth. Experiments on real range data corrupted with systematic errors within the scope of a medical application are presented in Chapter 8. This simulation addresses the conditions of commercially available ToF sensors that are characterized by a low spatial resolution compared to color sensors. Figure 5.5 depicts the *Stanford Bunny* and the *Dragon* scenes that were taken form the Stanford 3-D Scanning Repository[5] for this study. Geometrically aligned range and color images were obtained from 3-D mesh representations of these scenes using the Range Imaging Toolkit (RITK) [Wasz 11a]. The ground truth data was captured in a pixel resolution of 640×480 px and is available online[6]. For the simulation of a realistic color sensor, the color images were encoded in a pixel resolution of 640×480 px but blurred according to a Gaussian PSF ($\sigma_{\text{PSF}} = 0.5$) and disturbed by zero-mean Gaussian noise ($\sigma_{\text{noise}} = 0.002$). The corresponding range images were simulated in a pixel resolution of 80×60 px with a Gaussian PSF ($\sigma_{\text{PSF}} = 0.5$) and additive Gaussian noise ($\sigma_{\text{noise}} = 0.025$). This setup was used to generate four datasets showing the artificial scenes from different perspectives and with different textures. The displacements across the frames of these image sequences were related to rigid camera movements.

[5] http://graphics.stanford.edu/data/3Dscanrep

[6] https://www5.cs.fau.de/research/data/multi-sensor-super-resolution-datasets

(a) Direct **(b)** Filter-based **(c)** Confidence map

Figure 5.6: Comparison of motion estimation strategies to obtain displacement fields on range images. (a) Displacement field magnitudes obtained by the *direct* approach using optical flow estimation on range images. (b) Displacement field magnitudes obtained by the proposed filter-based approach that exploits color images as guidance (bright regions denote higher magnitudes). (c) Confidence map computed on the color data for the displacement field in (b) (bright regions denote higher weights).

In the following experiments, the algorithms summarized in Tab. 5.1 along with their model parameters were analyzed. All reconstruction algorithms are based on a Huber prior with $\delta_{\text{Huber}} = 5 \cdot 10^{-4}$ and $\lambda = 0.08$. To evaluate the impact of the techniques proposed in this chapter, the algorithms differ in the way of how color images are exploited. Single-sensor super-resolution (SSR) that works solely on the range data is considered as the baseline. Multi-sensor super-resolution (MSR) utilizes color images in Algorithm 5.1 for filter-based motion estimation using uniform weights for regularization. The adaptive multi-sensor super-resolution (AMSR) uses the color images also for spatially adaptive regularization. Adaptive multi-sensor super-resolution with outlier detection (AMSR-OD) augments AMSR with the outlier detection scheme in Algorithm 5.2.

Direct vs. Filter-Based Motion Estimation. Let us first present a comparison of the different strategies for motion estimation as a prerequisite for super-resolution. In these experiments, the computation of displacement fields was performed by the variational optical flow algorithm introduced by Liu [Liu 09]. In the single-sensor approach, optical flow was obtained directly on the range images, whereas the multi-sensor approaches used the proposed filter-based technique on color images. A qualitative comparison among these strategies is shown in Fig. 5.6. While direct motion estimation (Fig. 5.6a) was error prone and resulted in noisy displacement fields, the filter-based technique (Fig. 5.6b) accurately recorded camera motion. As shown in the experiments reported below, this substantially affects the accuracy of super-resolved range information.

In the context of motion estimation, the proposed outlier detection that is driven by color images provides a confidence map associated with the estimated displacement fields (Fig. 5.6c). This can further enhance the robustness of super-resolution compared to a reconstruction without proper outlier detection.

Single-Sensor vs. Multi-Sensor Super-Resolution. In Fig. 5.7, we compare the different reconstruction algorithms on the *Dragon-1* dataset using $K = 25$ frames

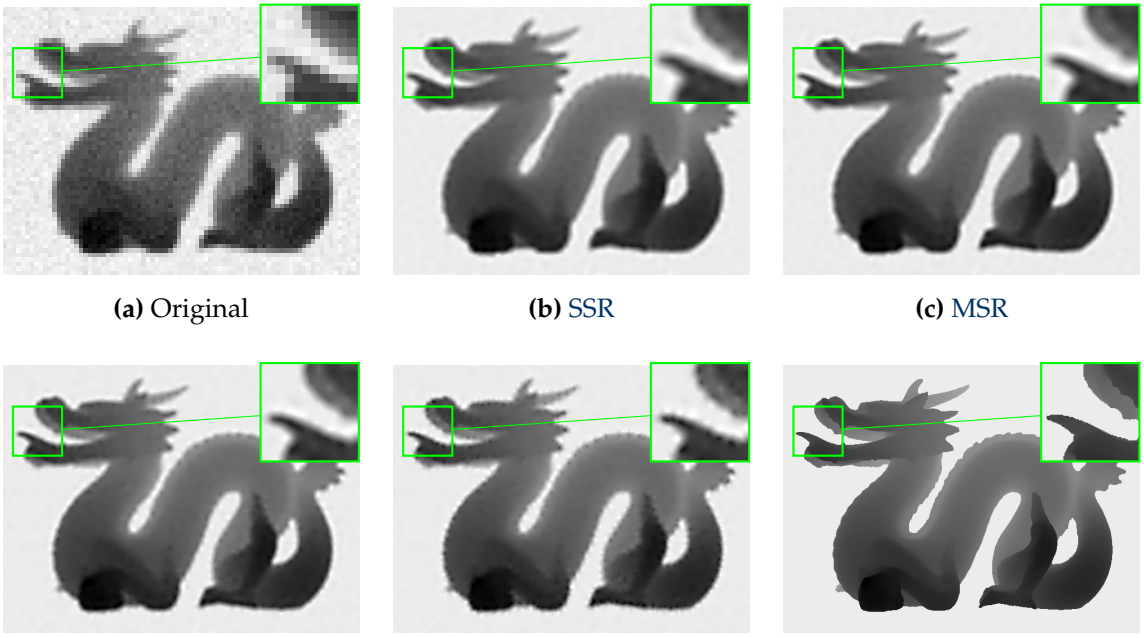
Table 5.1: Overview of the reconstruction approaches that are examined for their use in hybrid range imaging along with their optimization parameters. The multi-sensor framework is employed in different versions (MSR, AMSR, AMSR-OD) using filter-based motion estimation, spatially adaptive regularization, and outlier detection for range data based on color images. The single-sensor algorithm (SSR) that is not guided by color images is considered as the baseline.

Reconstruction algorithm	**Algorithm properties**		
	Motion estimation	**Adaptive regularization**	**Outlier detection**
Single-sensor super-resolution (SSR)	direct	✗	✗
Multi-sensor super-resolution (MSR)	filter-based	✗	✗
Adaptive multi-sensor super-resolution (AMSR)	filter-based	✓ $\tau_0 = 0.06$, $N_{xz} = 5$	✗
Adaptive multi-sensor super-resolution with outlier detection (AMSR-OD)	filter-based	✓ $\tau_0 = 0.06$, $N_{xz} = 5$	✓ $\rho_0 = 0.5$

and a magnification factor $s = 4$. In comparison to SSR that is affected by inaccurate motion estimation, the different multi-sensor approaches (MSR, AMSR, AMSR-OD) improved the accuracy of range information. More specifically, the reconstruction algorithms that implement spatially adaptive regularization enhanced the reconstruction of depth discontinuities compared to the non-adaptive methods.

These properties are confirmed by a quantitative comparison based on the PSNR and SSIM measures of super-resolved range data relative to the ground truth. Figure 5.8 summarizes the statistics of both measures on four datasets, where each dataset comprises 15 randomly generated image sequences. This reveals that the different multi-sensor approaches outperformed the single-sensor approach in terms of both measures. Among the different multi-sensor algorithms, the highest accuracy was obtained by AMSR-OD that uses color images for motion estimation, spatially adaptive regularization and outlier detection. In comparison to the SSR reconstruction, AMSR-OD improved the mean PSNR and SSIM by 0.9 dB and 0.02, respectively.

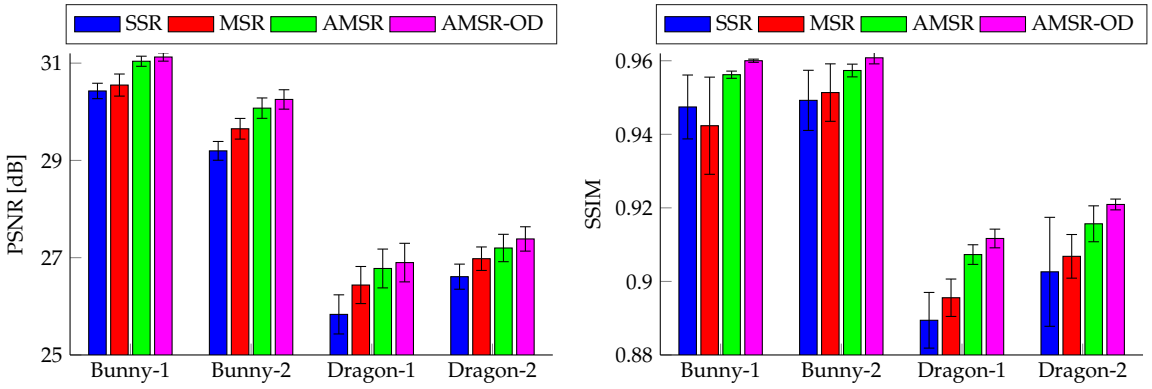
Influence of the Model Parameters. Figure 5.9 reports the behavior of the competing methods on the *Dragon-2* dataset regarding the choice of the regularization weight λ on a logarithmic scaled axis. It is worth noting that the different multi-sensor algorithms considerably outperformed the single-sensor algorithm regardless of the choice of the regularization weight over several orders of magnitude ($-2 \leq \log \lambda \leq 0.5$). In the case of an overestimation of this parameter ($\log \lambda \geq 0.5$), which resulted in oversmoothing of the range data, the different approaches showed a similar behavior.



(a) Original **(b)** SSR **(c)** MSR

(d) AMSR **(e)** AMSR-OD **(f)** Ground truth

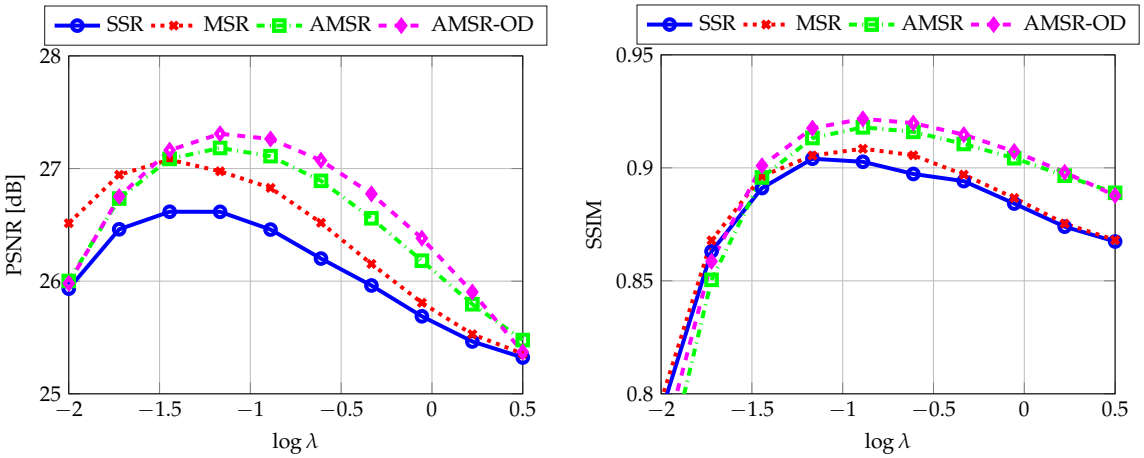
Figure 5.7: Comparison of original, low-resolution range data, single-sensor super-resolution (SSR) and the different multi-sensor approaches (MSR, AMSR and AMSR-OD) to the ground truth range data on the *Dragon-1* dataset. The reconstructions were obtained from $K = 25$ frames with magnification factor $s = 4$.



(a) PSNR on four datasets **(b)** SSIM on four datasets

Figure 5.8: Mean $\pm$ standard deviation of the PSNR and SSIM measures on four datasets obtained from the artificial *Stanford Bunny* and *Dragon* scenes. Both measures were evaluated for 15 randomly generated image sequences per dataset.

Another relevant model parameter is the contrast factor τ_0 that is used for spatially adaptive regularization. Figure 5.10 depicts the influence of this parameter on the *Dragon-2* dataset. If the contrast factor was overestimated ($\log \tau_0 \geq -1.0$), one can observe that the adaptive approaches (AMSR and AMSR-OD) behave like the non-adaptive algorithms (SSR and MSR) that can be considered as the baseline. In case of an underestimation ($\log \tau_0 \leq -1.5$), the adaptive approaches were prone



(a) PSNR vs. regularization weight λ **(b)** SSIM vs. regularization weight λ

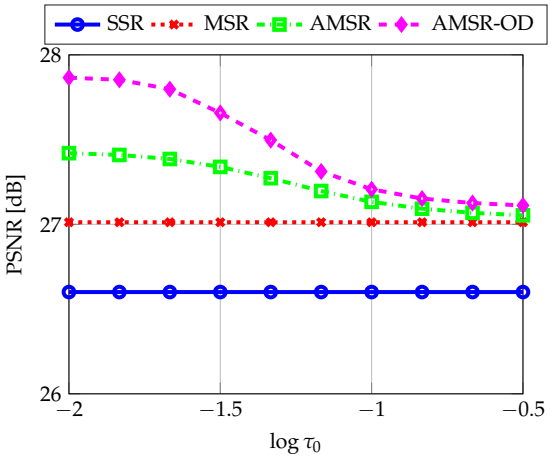
Figure 5.9: Parameter sensitivity study for the regularization weight λ. The PSNR and SSIM measures at each parameter setting are averaged over 15 random realizations of the experiment on the *Dragon-2* dataset. The classical single-sensor approach (SSR) is compared to the different proposed multi-sensor approaches (MSR, AMSR, AMSR-OD).

to texture copying artifacts as texture in the color images was erroneously transferred to the super-resolved range data. This behavior was captured quantitatively by the SSIM. Notice that this texture copying is comparable to related color-guided range filtering and upsampling techniques [Park 11, Kiec 13, Fers 13] but is controllable by limiting the contrast factor to a reasonable range ($-1.5 \leq \log \tau_0 \leq -1.0$).

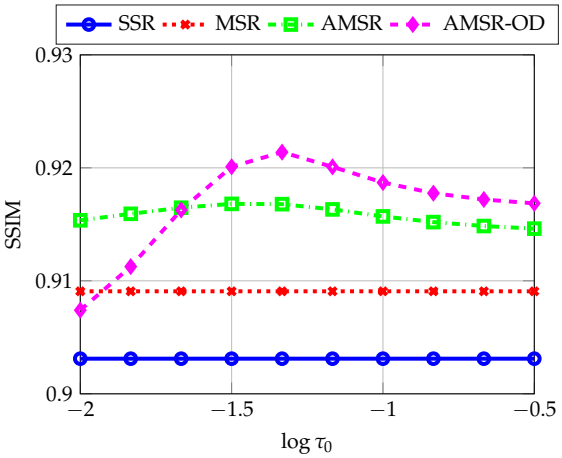
5.6 Conclusion

This chapter studied super-resolution for a single modality under the guidance of a complementary modality. In contrast to the algorithms presented in the first part of this thesis, the proposed multi-sensor framework takes advantage of additional guidance data. This aims at enhancing accuracy and robustness of super-resolution reconstruction. The computational stages that are steered by guidance images include motion estimation, spatially adaptive regularization as well as outlier detection. These concepts yield two algorithms: In the two-stage algorithm, a filter-based technique to obtain displacement fields in the domain of low-resolution data from guidance images and spatially adaptive regularization steered by the guidance images is utilized for image reconstruction. In the iteratively re-weighted minimization algorithm, confidence maps for outlier detection are constructed by exploiting guidance data.

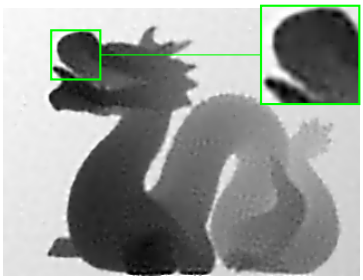
In order to prove the benefit of multi-sensor super-resolution over the single-sensor counterpart, hybrid range imaging was considered as example application. The goal was to super-resolve range data acquired with low-cost sensors under the guidance of color images fused with the range data. Multi-sensor super-resolution was tailored to range imaging by extending the underlying image for-



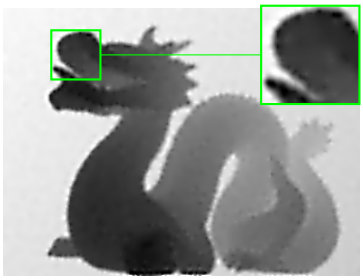
(a) PSNR vs. contrast factor τ_0



(b) SSIM vs. contrast factor τ_0



(c) $\log \tau_0 = -2.0$



(d) $\log \tau_0 = -1.3$

(e) $\log \tau_0 = -0.5$

Figure 5.10: Parameter sensitivity study for the contrast factor τ_0 of spatially adaptive regularization. Top row: PSNR and SSIM for different τ_0. Both measures are averaged over 15 random realizations of the experiment on the *Dragon-2* dataset. The non-adaptive algorithms (SSR and MSR) are the baseline and are compared to the adaptive algorithms (AMSR and AMSR-OD). Bottom row: AMSR-OD for an underestimated, an optimal, and an overestimated τ_0, respectively. Note the texture copying artifacts in case of a too low τ_0 ($\log \tau_0 = -2.0$). For an appropriate parameter setting ($\log \tau_0 = -1.3$), spatially adaptive regularization shows good tradeoffs between unwanted texture copying and the reconstruction of depth discontinuities.

mation model. We demonstrated in a simulation study that the proposed multi-sensor approach is able to take advantage of color images. The combination of the multi-sensor techniques provided superior surface reconstructions compared to the single-sensor approach and led to improvements of the PSNR and SSIM of 0.9 dB and 0.02, respectively.

CHAPTER 6

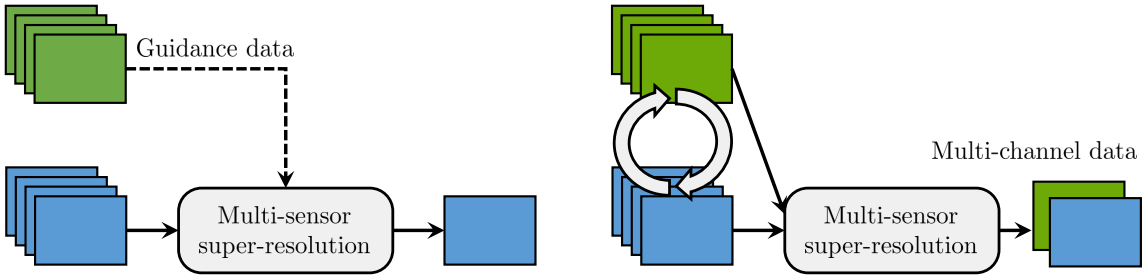
Multi-Sensor Super-Resolution using Locally Linear Regression

This chapter introduces a generalization of multi-sensor super-resolution that is applicable for an arbitrary number of modalities in hybrid imaging but does not require reliable guidance information. For this problem formulation, the modalities are represented by *multi-channel* images. This framework builds on a Bayesian model that features a novel image prior to exploit sparsity of the channels in a transform domain as well as locally linear regressions across them. Super-resolution is then derived via joint estimation of high-resolution image channels along with latent hyperparameters of the Bayesian model. In order to solve this non-convex optimization problem efficiently, an alternating minimization algorithm is developed. The proposed methodology is validated for resolution enhancement in various applications in computer vision, including color- and multispectral imaging as well as 3-D range imaging.

Parts of this chapter have been originally published in [Ghes 14] and [Kohl 15c].

6.1 Introduction

In Chapter 5, a first multi-sensor super-resolution method for hybrid imaging has been presented. In principle, this concept relies on two fundamental prerequisites. First and foremost, the underlying reconstruction algorithm super-resolves only a single modality. In addition, it relies on the existence of a reliable guidance modality that is used to steer super-resolution. Despite the advantages over a single-sensor formulation, these requirements limit the use of this approach for



(a) Guidance image framework (Chapter 5) **(b)** Multi-channel framework (Chapter 6)

Figure 6.1: Comparison of the different approaches to multi-sensor super-resolution. The guidance image framework in (a) uses static guidance information of a single modality to steer super-resolution for another modality. The multi-channel framework in (b) exploits dependencies among $C \geq 2$ modalities (image channels) to jointly super-resolve them.

many target applications. Some prominent examples are color or multispectral imaging, where current systems feature the acquisition of multiple spectral bands ranging from three to several hundreds. Since it is inadequate to super-resolve single spectral bands only, this initiated the development of color [Goto 04, Fars 06] and multispectral [Akgu 05, Ague 06, Zhan 12a] super-resolution techniques. The second shortcoming is that the required guidance data might not be available or it requires unjustifiable efforts to provide them. This situation appears in range imaging with devices that lack reliable high-resolution color sensors to gain guidance data [Ghes 14]. This is connected with the feature extraction from the guidance, e. g. in terms of spatially adaptive regularization [Kohl 15b] or outlier detection [Kohl 14b], that deteriorates in case of insufficient image quality.

In this chapter, we sacrifice the concept of guidance images to circumvent the aforementioned limitations and approach multi-sensor super-resolution from a generalized perspective. For this purpose, a set of modalities is represented by *multi-channel* images in the underlying mathematical framework. In contrast to processing the individual channels one after another, super-resolution is performed jointly for the entire set of channels. The basic assumption of this approach is the existence of *inter-channel dependencies*, such as geometrical structures that are visible in multiple channels. A well known example are color images that exhibit a high degree of correlation among their spectral bands as widely studied in color image processing [Gala 91, Kats 93, Schu 95]. In the context of multi-sensor super-resolution, inter-channel dependencies can be exploited in a Bayesian formulation as prior knowledge for the reconstruction of high-resolution multi-channel images. We capture these dependencies by a novel locally linear regression (LLR) image prior that is flexible with regard to the number of channels and does not rely on additional guidance information. This prior steers super-resolution *dynamically* as opposed to the static, feature-based techniques in Chapter 5. Based on this Bayesian model, we derive the simultaneous estimation of the unknown high-resolution channels along with latent prior hyperparameters as a joint energy minimization problem that is solved by confidence-aware optimization. In Fig. 6.1, we illustrate this methodology in comparison to the guidance image framework.

The remainder of this chapter is structured as follows. Section 6.2 provides a literature review regarding related methods. Section 6.3 introduces a Bayesian model of multi-channel images that is used in Section 6.4 to formulate multi-sensor super-resolution via joint energy minimization. Section 6.5 presents an in-depth analysis of this model and theoretical comparisons to related methods. In Section 6.6, we report an experimental evaluation by studying multiple target applications including color, multispectral and range imaging along with comparisons to the state-of-the-art in these domains. Finally, Section 6.7 draws a conclusion.

6.2 Related Work

The proposed approach to multi-sensor super-resolution exploits mutual dependencies among image channels to jointly super-resolve them. In particular, as shown in this chapter, we are interested in modeling statistical dependencies across the channels as a prior distribution for Bayesian estimation. Below, we provide a survey on similar concepts that have been successfully applied in related areas.

Color Imaging. In color imaging, dependencies among spectral bands have been widely investigated. Here, most commercially available cameras acquire red (R), green (G) and blue (B) spectral bands that form the RGB space. However, due to economic reasons, the sensor array is usually equipped with a CFA and is made sensitive to a single color per pixel. The interpolation of full RGB measurements referred to as *demosaicing* [Kimm 99] can be considered as some sort of super-resolution. To avoid inconsistencies among interpolated color channels, inter-channel dependencies are exploited for demosaicing, which can be done via color ratios [Kimm 99] or color correlation terms [Kere 99].

Later, correlations among color bands have been studied for various tasks such as denoising [Kere 98], deconvolution [Moli 03, Vega 06], and sparse representation of color images [Mair 08], among others. In the area of image restoration, Ono and Yamada [Ono 16] have proposed local color nuclear norm regularization based on the color-line property [Omer 04], which states that color bands in local image regions are linearly dependent. In [Fatt 14], Fattal employed this property for color image dehazing. Moreover, demosaicing of color images has been augmented with the notion of multi-frame super-resolution as proposed in the work of Gotoh and Okutomi [Goto 04]. This approach considers dependencies between color channels by a transformation of the highly correlated RGB space into luminance and chrominance components that are modeled by different prior distributions. In [Fars 06], Farsiu et al. proposed a related method based on an inter-channel regularization in the RGB space. This regularization enforces consistency in terms of locations and orientations of edges captured in the different channels. Such techniques avoid color artifacts, e. g. color bleeding, and serve as a strong prior for image super-resolution.

Multi- and Hyperspectral Imaging. In the area of multi- and hyperspectral image processing, different attempts have been made at extending color image restora-

tion to a larger number of spectral bands. In [Akgu 05], Akgun et al. proposed a generalized image formation model for hyperspectral images. Similar to color image restoration, such methods benefit from incorporating dependencies among the spectral bands to their underlying model. In the method of Zhang et al. [Zhan 12a], statistical dependencies are considered by applying a principal component analysis (PCA) on the original spectral bands. Then, super-resolution is performed on PCA compressed hyperspectral data. A different notion is to employ correlations with a high spatial but a low spectral resolution image [Ague 06, Akht 15, Lana 15] to steer hyperspectral super-resolution. Notice that this concept is closely related to guidance image based super-resolution in Chapter 5.

Joint and Mutual Structure Filtering. *Joint* image filters process a single input image driven by a guidance image. In this area, the guidance is assumed to be static following the same line of thought as guidance image based super-resolution. Local filters related to this concept include joint bilateral [Kopf 07], guided [He 13, Hore 14] or weighted median filtering [Ma 13, Zhan 14b], see Section 5.2. Such filters can also be learned from example data using convolutional neural networks [Li 16]. Global filters that are formulated via regularized energy minimization have been developed for range image upsampling [Park 11, Fers 13, Kiec 13] as well as cross-field image restoration [Yan 13]. These approaches impose properties of the filter output based on the given guidance image using implicit regularization terms. Contrary to the method proposed in this chapter, joint filtering has the common prerequisite of high-quality guidance data similar to guided image based super-resolution. Another limitation is that joint filtering is prone to erroneously transfer image structures to the filter output that are only present in the guidance image.

In contrast to the aforementioned techniques, Shen et al. [Shen 15] introduced *mutual structure* filtering to simultaneously filter input and guidance data with consideration of structural inconsistencies. This inconsistency-aware and dynamic formulation alleviates the erroneous structure transfer compared to filters with pure static guidance. In [Ham 15], Ham et al. proposed with a similar motivation static and dynamic guided filtering that combines regularization gained by static guidance data and the filter input. Although these methods improve flexibility and robustness of joint filtering, they ignore reasonable models of the image formation process and address denoising problems rather than super-resolution. Contrary to the proposed super-resolution approach, these filters are developed as single-image methods without considering multi-frame processing.

6.3 Bayesian Modeling of Multi-Channel Images

In this chapter, an unknown, high-resolution multi-channel image is represented as the composite of C disjoint channels denoted by $\boldsymbol{x} = (\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_C^\top)^\top$, where each channel $\boldsymbol{x}_i$, $i = 1, \ldots, C$ is represented by a $N_i \times 1$ vector. For the sake of notational brevity, we limit ourselves to channels with consistent dimensions, i. e. $N_i = N$ for all $i = 1, \ldots, C$. Each high-resolution channel $\boldsymbol{x}_i$ is related to a sequence of K

low-resolution frames $\boldsymbol{y}_i = (\boldsymbol{y}_i^{(1)}, \ldots, \boldsymbol{y}_i^{(K)})^\top$. Here, $\boldsymbol{y}_i^{(k)}$ is the k-th frame associated with the i-th channel and is represented by a $M_i \times 1$ vector, where we again assume the same dimension in all channels, i. e. $M_i = M$ for all $i = 1, \ldots, C$. Furthermore, we denote the composite of C channels by $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_C^\top)^\top$ that represents the entire set of low-resolution observations as $(C \cdot K \cdot M) \times 1$ vector.

Let us consider the low-resolution observations $\boldsymbol{y}$ along with the unknown multi-channel image $\boldsymbol{x}$ as random variables. We aim at determining the joint posterior probability over all channels:

$$\begin{aligned} p(\boldsymbol{x}, \Phi \,|\, \boldsymbol{y}) &= p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_C, \Phi \,|\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_C) \\ &= \frac{p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_C \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_C) \cdot p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_C \,|\, \Phi) \cdot p(\Phi)}{p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_C)}, \end{aligned} \tag{6.1}$$

where $p(\boldsymbol{y} \,|\, \boldsymbol{x}) = p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_C \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_C)$ is the conditional probability of obtaining the entire set of observations $\boldsymbol{y}$ from the unknown multi-channel image $\boldsymbol{x}$ and $p(\boldsymbol{x} \,|\, \Phi) = p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_C \,|\, \Phi)$ is the prior probability for $\boldsymbol{x}$. In Eq. (6.1), Φ are latent hyperparameters of the imaging process with the assigned distribution $p(\Phi)$.

This section proceeds with the definition of these distributions in a hierarchical way as follows. First, the observation model $p(\boldsymbol{y} \,|\, \boldsymbol{x})$ is developed. Accordingly, a prior distribution $p(\boldsymbol{x} \,|\, \Phi)$ is assigned to the multi-channel image $\boldsymbol{x}$ to model its statistical appearance. Eventually, we introduce a prior distribution $p(\Phi)$ that is employed for an inference of the hyperparameters Φ.

6.3.1 Multi-Channel Observation Model

In order to calculate the posterior probability in Eq. (6.1), the following assumptions are made to derive the conditional probability $p(\boldsymbol{y} \,|\, \boldsymbol{x})$ that represents the observation model: 1) The low-resolution channels $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_C$ are mutually independent assuming statistically independent noise among the channels, and 2) the formation of a low-resolution channel $\boldsymbol{y}_i$ depends only on the corresponding high-resolution channel $\boldsymbol{x}_i$ but is independent on the remaining channels, see Fig. 6.2. Hence, the conditional probability $p(\boldsymbol{y} \,|\, \boldsymbol{x})$ can be factorized according to:

$$p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_C \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_C) = \prod_{i=1}^{C} p(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i) = \prod_{i=1}^{C} \prod_{k=1}^{K} p\left(\boldsymbol{y}_i^{(k)} \,|\, \boldsymbol{x}_i\right), \tag{6.2}$$

where $p(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i)$ is the joint conditional probability of observing the frames of the low-resolution channel $\boldsymbol{y}_i$ from the high-resolution channel $\boldsymbol{x}_i$ and $p(\boldsymbol{y}_i^{(k)} \,|\, \boldsymbol{x}_i)$ is the conditional probability of observing a single frame $\boldsymbol{y}_i^{(k)}$. The formation of the low-resolution observations $\boldsymbol{y}_i$ from the high-resolution channel $\boldsymbol{x}_i$ is described by:

$$\boldsymbol{y}_i = \boldsymbol{W}_i \boldsymbol{x}_i + \epsilon_i = \begin{pmatrix} \boldsymbol{D}\boldsymbol{H}_i \boldsymbol{M}_i^{(1)} \boldsymbol{x}_i \\ \vdots \\ \boldsymbol{D}\boldsymbol{H}_i \boldsymbol{M}_i^{(K)} \boldsymbol{x}_i \end{pmatrix} + \begin{pmatrix} \epsilon_i^{(1)} \\ \vdots \\ \epsilon_i^{(K)} \end{pmatrix}, \tag{6.3}$$

where $\boldsymbol{W}_i$ is the system matrix and ϵ_i is additive measurement noise for this channel. The system matrix $\boldsymbol{W}_i$ comprises subsampling modeled by $\boldsymbol{D}$, which is assumed to be constant over the channels. The circulant matrix $\boldsymbol{H}_i$ denotes space

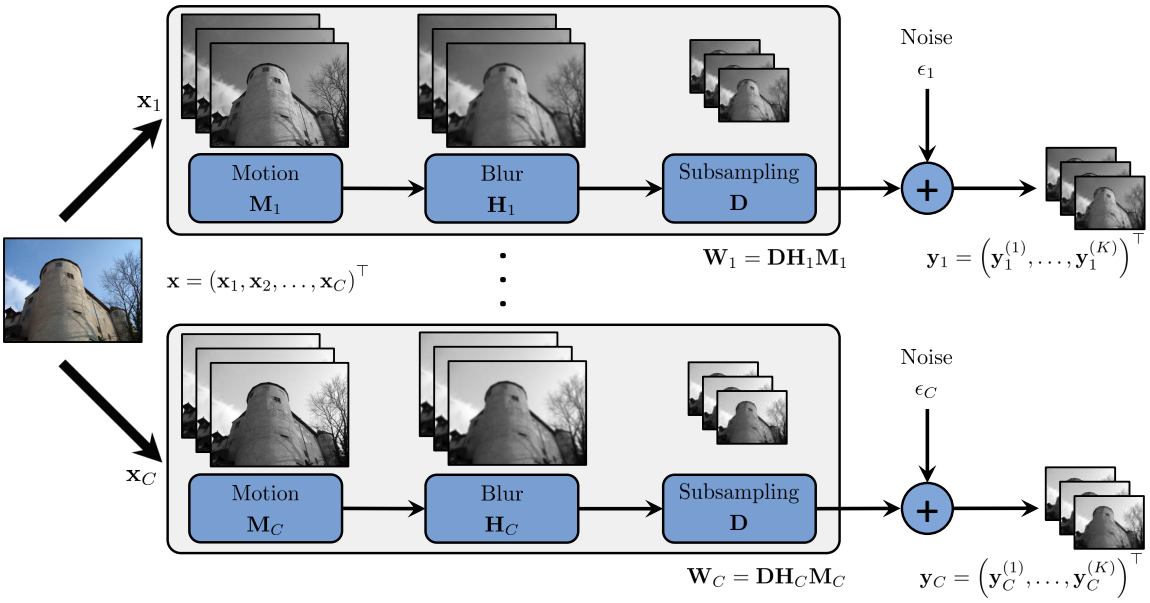


Figure 6.2: Formation of low-resolution multi-channel observations $\boldsymbol{y}$ encoded in C channels and K frames from the high-resolution multi-channel image $\boldsymbol{x}$.

invariant blur associated with the PSF of the i-th channel that might be varying over the channels. $\boldsymbol{M}_i^{(k)}$ models subpixel motion relative to the reference coordinate grid for the k-th frame associated with the i-th channel.

Based on the factorization in Eq. (6.2) and the image formation in Eq. (6.3), the observation model is given by the distribution:

$$p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_C \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_C) \propto \exp\left\{ -\sum_{i=1}^{C} \sum_{m=1}^{KM} \phi_{\boldsymbol{y}_i}\left([\boldsymbol{y}_i - \boldsymbol{W}_i \boldsymbol{x}_i]_m\right) \right\}, \tag{6.4}$$

where $\phi_{\boldsymbol{y}_i} : \mathbb{R} \to \mathbb{R}_0^+$ denotes a loss function to model the noise distribution for the i-th channel. In spirit of the weighted normal distribution proposed in Chapter 4 and to make the model robust to space variant noise and outliers, we define Eq. (6.4) based on the Huber loss:

$$\phi_{\boldsymbol{y}_i}(z) = \begin{cases} z^2 & \text{if } |z| \leq \sigma_{\text{noise},i} \\ 2\sigma_{\text{noise},i}|z| - \sigma_{\text{noise},i}^2 & \text{otherwise} \end{cases}, \tag{6.5}$$

where $\sigma_{\text{noise},i}$ denotes the distribution scale parameter that characterizes the noise level associated with the i-th channel.

6.3.2 Multi-Channel Image Prior Model

Let us now define the image prior $p(\boldsymbol{x} \,|\, \Phi)$ employed in the posterior probability in Eq. (6.1). In general, this prior needs to be factorized according to:

$$\begin{aligned} p(\boldsymbol{x} \,|\, \Phi) &= p(\boldsymbol{x}_1 \,|\, \Phi)\, p(\boldsymbol{x}_2 \,|\, \boldsymbol{x}_1, \Phi) \ldots p(\boldsymbol{x}_C \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{C-1}, \Phi) \\ &= p(\boldsymbol{x}_1 \,|\, \Phi) \prod_{i=2}^{C} p(\boldsymbol{x}_i \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}, \Phi)\,. \end{aligned} \tag{6.6}$$

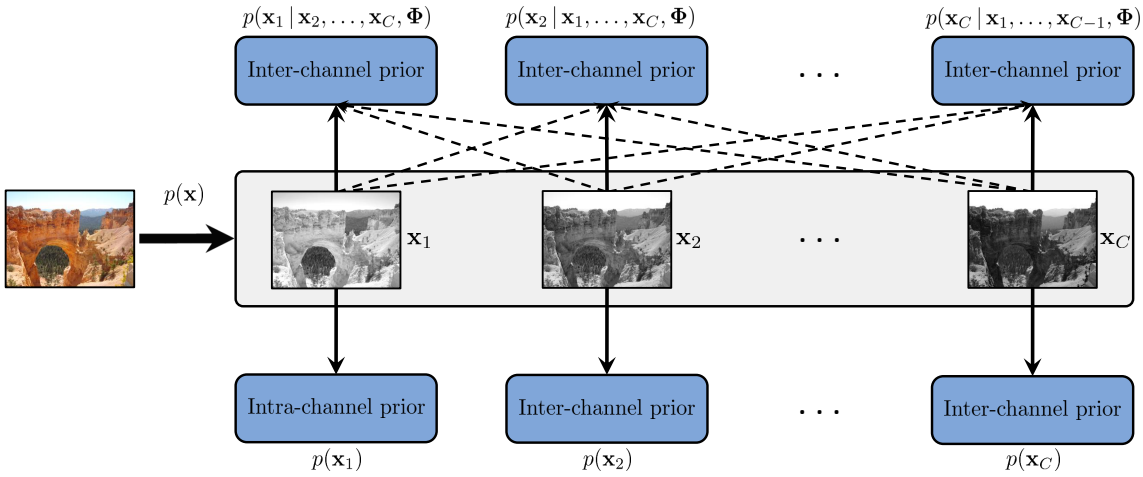


Figure 6.3: Schematic representation of the prior distribution $p(\boldsymbol{x})$. The proposed model consists of an inter-channel prior to describe statistical dependencies among pairs of channels and an intra-channel prior to describe the appearance of the individual channels.

Contrary to the observation model, this factorization considers dependencies among the high-resolution channels as the key notion of the Bayesian formulation.

The prior distribution $p(\boldsymbol{x} \,|\, \Phi)$ considers two complementary aspects. On the one hand, it models the statistical appearance of each individual channel $\boldsymbol{x}_i$, which is related to an *intra*-channel prior. On the other hand, it accounts for statistical dependencies of each channel $\boldsymbol{x}_i$ relative to all other channels $\boldsymbol{x}_j$, $i \neq j$ that is considered by an *inter*-channel prior. Using a pair-wise approach to consider these dependencies, the joint distribution $p(\boldsymbol{x} \,|\, \Phi)$ in Eq. (6.6) is given by:

$$p(\boldsymbol{x} \,|\, \Phi) = \prod_{i=1}^{C} p(\boldsymbol{x}_i \,|\, \mathcal{X}_i, \Phi)\,, \tag{6.7}$$

and the prior distribution $p(\boldsymbol{x}_i \,|\, \mathcal{X}_i, \Phi)$ associated with the channel $\boldsymbol{x}_i$ is written as:

$$p(\boldsymbol{x}_i \,|\, \mathcal{X}_i, \Phi) \propto \exp\left\{ -\lambda_i R_{\text{intra}}(\boldsymbol{x}_i) - \sum_{j=1, j\neq i}^{C} \mu_{ij} R_{\text{inter}}\left(\boldsymbol{x}_i, \boldsymbol{x}_j, \Phi_{ij}\right) \right\}, \tag{6.8}$$

where $\mathcal{X}_i = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_C\}$. In Eq. (6.8), $R_{\text{intra}}(\boldsymbol{x}_i)$ denotes the regularization term for the intra-channel prior distribution associated with $\boldsymbol{x}_i$. Similarly, $R_{\text{inter}}\left(\boldsymbol{x}_i, \boldsymbol{x}_j, \Phi_{ij}\right)$ denotes the regularization term of the inter-channel prior for the channel pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ parametrized by a set of hyperparameters Φ_{ij} as shown below. The regularization weight $\lambda_i \geq 0$ denotes the contribution of the intra-channel prior to model the statistical appearance of the multi-channel image solely based on the individual channels. The regularization weight $\mu_{ij} \geq 0$ denotes the contribution of the inter-channel prior between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Hence, it considers statistical dependencies according to Eq. (6.6), whereas in case of $\mu_{ij} = \mu_{ji} = 0$ the channels $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are treated as independent. We call the prior distribution *symmetric*, if the regularization weights fulfill the property $\mu_{ij} = \mu_{ji}$ for all $i, j = 1, \ldots, C$.

Intra-Channel Prior. Following the state-of-the-art in image restoration for single-channel images [Kris 09], the intra-channel prior needs to account for the sparsity of the individual channels in a certain transform domain. The corresponding regularization term adopts WBTV [Kohl 16b] as previously introduced in Chapter 4 and is given by:

$$R_{\text{intra}}(\boldsymbol{x}_i) := \sum_{n=1}^{N_S} \phi_{\boldsymbol{x}_i}([\boldsymbol{S}\boldsymbol{x}_i]_n)\,, \tag{6.9}$$

where $\boldsymbol{S} \in \mathbb{R}^{N_S \times N}$ with $N_S = (2N_{\text{BTV}}+1)^2 N$ denotes the linear sparsifying transform:

$$\boldsymbol{S} = \begin{pmatrix} \alpha_{\text{BTV}}^{|-N_{\text{BTV}}|+|-N_{\text{BTV}}|} \left(\boldsymbol{I}_{N\times N} - \boldsymbol{S}_{\text{v}}^{-N_{\text{BTV}}} \boldsymbol{S}_{\text{h}}^{-N_{\text{BTV}}}\right) \\ \vdots \\ \alpha_{\text{BTV}}^{|+N_{\text{BTV}}|+|+N_{\text{BTV}}|} \left(\boldsymbol{I}_{N\times N} - \boldsymbol{S}_{\text{v}}^{+N_{\text{BTV}}} \boldsymbol{S}_{\text{h}}^{+N_{\text{BTV}}}\right) \end{pmatrix}. \tag{6.10}$$

Here, $\alpha_{\text{BTV}} \in]0,1]$ is the BTV weighting factor, $N_{\text{BTV}} \geq 1$ is the BTV window size, and $\boldsymbol{S}_{\text{v}}^m$ and $\boldsymbol{S}_{\text{h}}^n$ model vertical and horizontal shifts of $\boldsymbol{x}_i$ by m and n pixels, respectively. The loss function $\phi_{\boldsymbol{x}_i} : \mathbb{R} \to \mathbb{R}_0^+$ in Eq. (6.9) is given by the mixed L_1/L_p norm:

$$\phi_{\boldsymbol{x}_i}(z) = \begin{cases} |z| & \text{if } |z| \leq \sigma_{\text{prior},i} \\ \sigma_{\text{prior},i}^{1-p_i} \cdot |z|^{p_i} & \text{otherwise} \end{cases}, \tag{6.11}$$

where $\sigma_{\text{prior},i}$ and $p_i \in [0,1]$ are the prior distribution scale parameter and the sparsity parameter for the channel $\boldsymbol{x}_i$, respectively.

Inter-Channel Prior. In Eq. (6.8), the inter-channel prior accounts for pair-wise statistical dependencies among the image channels. For this purpose, we follow the assumption that such dependencies can be described *locally* by means of linear regressions on a patch-wise basis. Let $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ be a pair of disjoint image channels. Then, the regression of $\boldsymbol{x}_i$ towards $\boldsymbol{x}_j$ at the n-th pixel position is given by:

$$x_{j,m} = C_{ij,n} x_{i,m} + b_{ij,n}, \qquad \text{for all } m \in \omega_{\text{LLR}}(n), \tag{6.12}$$

where the parameters $C_{ij,n}$ and $b_{ij,n}$ denote the local regression coefficients for a $(2N_{\text{LLR}}+1) \times (2N_{\text{LLR}}+1)$ image patch $\omega_{\text{LLR}}(n)$ centered at the n-th pixel. These coefficients are assumed to be constant for all pixel positions $m \in \omega_{\text{LLR}}(n)$, see Fig. 6.4. Based on this patch-wise relationship, the inter-channel prior for the channel $\boldsymbol{x}_i$ is defined via the fidelity of a regression towards each of the remaining channels $\boldsymbol{x}_j$, $j \neq i$ over all local patches. This fidelity is stated by the LLR model:

$$R_{\text{inter}}(\boldsymbol{x}_i, \boldsymbol{x}_j, \Phi_{ij}) := \sum_{n=1}^{N} \phi_{\boldsymbol{x}_{ij}}\left(\left[\boldsymbol{C}_{ij}\boldsymbol{x}_i + \boldsymbol{b}_{ij} - \boldsymbol{x}_j\right]_n\right), \tag{6.13}$$

where $\boldsymbol{C}_{ij} = \text{diag}(C_{ij,1}, \ldots, C_{ij,N}) \in \mathbb{R}^{N\times N}$ and $\boldsymbol{b}_{ij} = (b_{ij,1}, \ldots, b_{ij,N})^\top \in \mathbb{R}^N$ are regression coefficients over the entire image assembled from the pixel-wise coefficients $C_{ij,n}$ and $b_{ij,n}$ in Eq. (6.12), respectively. We denote by $\Phi_{ij} = \{\boldsymbol{C}_{ij}, \boldsymbol{b}_{ij}\}$ the set of coefficients that are treated as hyperparameters of the prior distribution. In Eq. (6.13), $\phi_{\boldsymbol{x}_{ij}} : \mathbb{R} \to \mathbb{R}_0^+$ denotes a loss function to measure the regression fidelity.

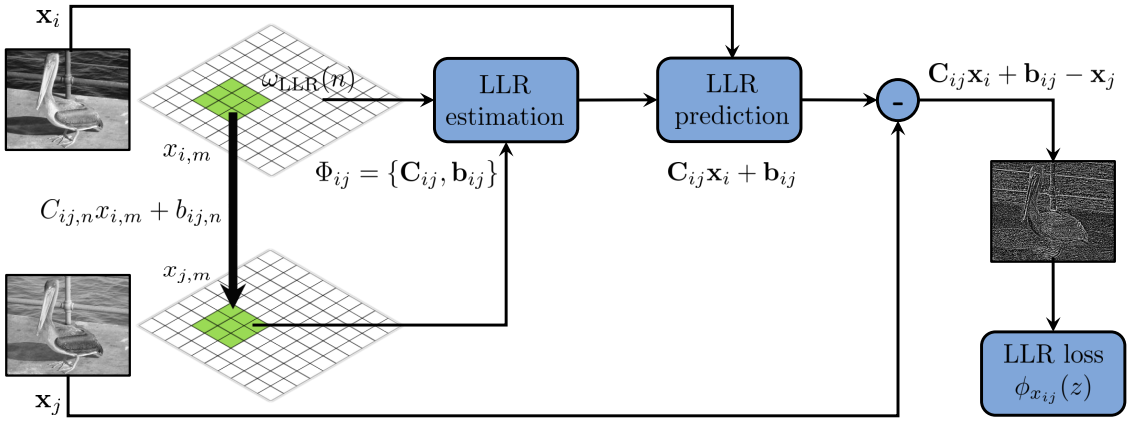


Figure 6.4: Schematic representation of the locally linear regression (LLR) model of the channel $\boldsymbol{x}_i$ towards the channel $\boldsymbol{x}_j$ depicted for pairs of color channels. We establish the LLR model for image patches $\omega_{\text{LLR}}(n)$ and define the corresponding prior distribution based on the regression residual error and an outlier-insensitive loss function.

In order to tolerate outliers regarding the linear regression assumption, we define the LLR model according to Tukey's biweight function [Meer 91]:

$$\phi_{\boldsymbol{x}_{ij}}(z) = \begin{cases} \frac{1}{6}\sigma^2_{\text{LLR},ij}\left(1 - \left(1 - \frac{z^2}{\sigma^2_{\text{LLR},ij}}\right)^3\right) & \text{if } |z| \leq \sigma_{\text{LLR},ij} \\ \frac{1}{6}\sigma^2_{\text{LLR},ij} & \text{otherwise} \end{cases}, \tag{6.14}$$

where $\sigma_{\text{LLR},ij}$ is the distribution scale parameter for the channels $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

It is worth noting that similar regression models have been proposed previously for multi-sensor super-resolution [Zome 01, Ghes 14] as well as joint filtering [He 13, Shen 15]. In particular, the regression in Eq. (6.13) generalizes the concept of guided filtering as proposed by He et al. [He 13]. However, the key novelty is that the proposed prior is applicable to an arbitrary number of image channels and represents a Bayesian interpretation of mutual dependencies, while guided filtering considers only dependencies of a filter input relative to a static guidance. Moreover, since Eq. (6.13) is formulated via an outlier-insensitive loss function, it is spatially adaptive and features robustness regarding image regions that violate the linear regression assumption. We elaborate on these properties with comparisons to related methods in more detail in Section 6.5.

Prior on the Hyperparameters. The LLR prior distribution relies on knowledge regarding the regression coefficients but in general these parameters are unknown. For this reason, they are treated as latent hyperparameters and in order to estimate them, we need to assign a meaningful prior distribution $p(\Phi)$. Let us assume that the coefficients associated with the different channels are mutually independent random variables. Then, the joint prior distribution $p(\Phi) = p(\Phi_{11}, \ldots, \Phi_{CC})$ with $\Phi_{ij} = \{\boldsymbol{C}_{ij}, \boldsymbol{b}_{ij}\}$ can be factorized to:

$$p(\Phi) = \prod_{i=1}^{C} \prod_{j=1, j\neq i}^{C} p(\boldsymbol{C}_{ij})\, p(\boldsymbol{b}_{ij})\,. \tag{6.15}$$

In this work, $p(\boldsymbol{b}_{ij})$ is assumed to be a uniform distribution. In addition, $p(\boldsymbol{C}_{ij})$ is adopted from ridge regression [He 13] and is given by the normal distribution:

$$p(\boldsymbol{C}_{ij}) \propto \exp\left\{-\epsilon_{ij}\left|\left|\boldsymbol{C}_{ij}\right|\right|_F^2\right\}, \tag{6.16}$$

where $||\cdot||_F$ is the Frobenius norm that is given by $||\boldsymbol{C}_{ij}||_F = ||\mathrm{diag}(\boldsymbol{C}_{ij})||_2$ for the diagonal matrix $\boldsymbol{C}_{ij}$ and $\epsilon_{ij} \geq 0$ denotes a hyperparameter regularization weight for the channels $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Intuitively, this prior distribution penalizes large coefficients $\boldsymbol{C}_{ij}$. As shown in Section 6.4, the benefit of this prior is that the regression coefficients can be estimated in closed-form within the proposed algorithm[1].

6.4 Bayesian Multi-Channel Super-Resolution

This section aims at the development of parameter estimation techniques based on the proposed Bayesian model of multi-channel images. More specifically, we are interested in obtaining point estimates of the high-resolution channels $\boldsymbol{x}$ given the low-resolution, multi-channel observations $\boldsymbol{y}$. For this purpose, two approaches can be distinguished.

6.4.1 Sequential Maximum A-Posteriori Estimation

Let us first discuss the *sequential* estimation of the unknown high-resolution image channels, which serves as a baseline approach in this chapter. Therefore, let the inter-channel prior $p(\boldsymbol{x}_i \,|\, \boldsymbol{x}_j, \Phi_{ij})$ for each pair of channels be a uniform distribution. Accordingly, statistical dependencies among the channels can be ignored and the MAP estimate for the high-resolution image $\boldsymbol{x}$ is given by:

$$\boldsymbol{x}_{\mathrm{MAP}} = \underset{\boldsymbol{x}}{\mathrm{argmax}} \prod_{i=1}^{C} p(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i) \prod_{i=1}^{C} p(\boldsymbol{x}_i)\,. \tag{6.17}$$

Based on this simplifying assumption, the unknown high-resolution channels $\boldsymbol{x}_i$ for $i = 1, \ldots, C$ are reconstructed independently by minimizing the negative log-likelihood of Eq. (6.17):

$$\boldsymbol{x}_{i,\mathrm{MAP}} = \underset{\boldsymbol{x}_i}{\mathrm{argmin}} \left\{L(\boldsymbol{x}_i) + \lambda_i R_{\mathrm{intra}}(\boldsymbol{x}_i)\right\}. \tag{6.18}$$

This approach is based solely on the observation model related to the data fidelity term $L(\boldsymbol{x}_i) \propto -\log p(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i)$ and the intra-channel regularization term $R_{\mathrm{intra}}(\boldsymbol{x}_i)$ associated with the considered channel. The advantage of sequential estimation is its conceptual simplicity, as Eq. (6.18) can be solved straightforwardly using super-resolution for single-channel images. However, its limitation is that statistical dependencies are ignored, which might cause inconsistencies among the super-resolved image channels.

[1] This prior leads to a ridge regression problem to determine the regression coefficients [He 13].

6.4.2 Joint Maximum A-Posteriori Estimation

Our aim is to jointly estimate the image channels under consideration of inter-channel dependencies. Using the joint posterior distribution in Eq. (6.1), the goal is to determine the high-resolution image $\boldsymbol{x}$ along with the latent regression coefficients Φ by the joint MAP estimation:

$$(\boldsymbol{x}_{\mathrm{MAP}}, \Phi_{\mathrm{MAP}}) = \operatorname*{argmax}_{\boldsymbol{x},\Phi} p(\boldsymbol{y} \,|\, \boldsymbol{x})\, p(\boldsymbol{x} \,|\, \Phi)\, p(\Phi)\,. \tag{6.19}$$

Notice that Eq. (6.19) is non-convex due to the non-convexity of the regularization terms related to the prior distribution. Moreover, the dimension of the parameter space $(\boldsymbol{x}, \Phi)$ is $CN + C(C-1)N = C^2N$ and the number of observations is CKM. Thus, the dimension of the parameter space grows quadratically as a function of the number of channels and joint MAP estimation is underdetermined if $KM < CN$. These properties make iterative optimization challenging and a joint numerical optimization would be computationally prohibitive due to the high dimensionality of the parameter space.

For an efficient numerical solution, we alternatively solve Eq. (6.19) w. r. t. the high-resolution image $\boldsymbol{x}$ and the regression coefficients Φ while keeping the other parameters fixed. Starting at an initial guess $(\boldsymbol{x}^0, \Phi^0)$, this leads to a sequence of estimates $(\boldsymbol{x}^t, \Phi^t)$ according to the iteration scheme:

$$\Phi^t = \operatorname*{argmax}_{\Phi} p\left(\boldsymbol{x}^{t-1} \,|\, \Phi\right) p(\Phi)\,, \tag{6.20}$$

$$\boldsymbol{x}^t = \operatorname*{argmax}_{\boldsymbol{x}} p(\boldsymbol{y} \,|\, \boldsymbol{x})\, p\left(\boldsymbol{x} \,|\, \Phi^t\right), \tag{6.21}$$

for $t = 1, \ldots, T_{\mathrm{am}}$. Let us now derive efficient solutions for both substeps.

Estimation of the Regression Coefficients. Given the estimate $\boldsymbol{x}^{t-1}$ for the image channels at iteration $t-1$, the latent regression coefficients are obtained by minimizing the negative log-likelihood of Eq. (6.20):

$$\Phi^t = \operatorname*{argmin}_{\Phi} \sum_{i=1}^{C} \sum_{j=1, j\neq i}^{C} F_{ij}^t(\boldsymbol{C}_{ij}, \boldsymbol{b}_{ij}). \tag{6.22}$$

This energy minimization is separable and the regression coefficients among the i-th and j-th channel are obtained by optimizing the negative log-likelihood of the corresponding prior distributions. In order to solve Eq. (6.22) efficiently, an upper bound of this log-likelihood term is minimized. For this purpose, the non-convex biweight loss function $\phi_{x_{ij}}(z)$ is rewritten by means of an MM algorithm [Hunt 04], see Appendix A.3.1. This leads to the confidence-aware energy function:

$$F_{ij}^t(\boldsymbol{C}_{ij}, \boldsymbol{b}_{ij}) = \left(\boldsymbol{C}_{ij}\boldsymbol{x}_i^{t-1} + \boldsymbol{b}_{ij} - \boldsymbol{x}_j^{t-1}\right)^\top \boldsymbol{K}_{ij}^t \left(\boldsymbol{C}_{ij}\boldsymbol{x}_i^{t-1} + \boldsymbol{b}_{ij} - \boldsymbol{x}_j^{t-1}\right) + \epsilon_{ij} \left|\left|\boldsymbol{C}_{ij}\right|\right|_F^2. \tag{6.23}$$

The confidence weights $\boldsymbol{K}_{ij}^t$ used in this convex energy function at iteration t are assembled as the diagonal matrix:

$$\boldsymbol{K}_{ij}^t = \operatorname{diag}\left(\kappa_1\left(\boldsymbol{C}_{ij}^{t-1}, \boldsymbol{b}_{ij}^{t-1}\right) \quad \kappa_2\left(\boldsymbol{C}_{ij}^{t-1}, \boldsymbol{b}_{ij}^{t-1}\right) \quad \ldots \quad \kappa_N\left(\boldsymbol{C}_{ij}^{t-1}, \boldsymbol{b}_{ij}^{t-1}\right)\right), \tag{6.24}$$

and the weighting function to obtain the k-th weight is given by:

$$\kappa_k(\boldsymbol{C}_{ij}, \boldsymbol{b}_{ij}) = \begin{cases} \left(1 - \left(\frac{r_{ij,k}(\boldsymbol{C}_{ij}, \boldsymbol{b}_{ij})}{c_{\mathrm{LLR}} \cdot \sigma^t_{\mathrm{LLR},ij}}\right)^2\right)^2 & \text{if } \left|r_{ij,k}(\boldsymbol{C}_{ij}, \boldsymbol{b}_{ij})\right| \leq c_{\mathrm{LLR}} \sigma^t_{\mathrm{LLR},ij} \\ 0 & \text{otherwise} \end{cases}, \tag{6.25}$$

$$r_{ij}(\boldsymbol{C}_{ij}, \boldsymbol{b}_{ij}) = Q_{\omega_{\mathrm{LLR}}}\left(\boldsymbol{C}_{ij}\boldsymbol{x}_i^{t-1} + \boldsymbol{b}_{ij} - \boldsymbol{x}_j^{t-1}\right), \tag{6.26}$$

where $r_{ij}(\boldsymbol{C}_{ij}, \boldsymbol{b}_{ij})$ denotes a filtered version of the regression residual error among the given channels. Notice that in order to reduce the influence of isolated pixels with large regression error and to avoid the origination of pseudo-structures by the inter-channel prior, $Q_{\omega_{\mathrm{LLR}}}(\cdot)$ is implemented as median filter with window size $(2N_{\mathrm{LLR}} + 1) \times (2N_{\mathrm{LLR}} + 1)$. The tuning constant c_{LLR} for Tukey's biweight loss is set to be $c_{\mathrm{LLR}} = 4.6851$ to achieve a 95 % asymptotic efficiency under a normal distribution of the regression error [Meer 91]. The unknown distribution scale parameter $\sigma^t_{\mathrm{LLR},ij}$ is adaptively updated at each iteration based on the weighted MAD rule under the weights $\boldsymbol{K}_{ij}^{t-1}$ according to:

$$\sigma^t_{\mathrm{LLR},ij} = \sigma_0 \cdot \mathrm{mad}\left(r_{ij}(\boldsymbol{C}_{ij}^{t-1}, \boldsymbol{b}_{ij}^{t-1}), \boldsymbol{K}_{ij}^{t-1}\right), \tag{6.27}$$

where $\sigma_0 = 1.4826$ to obtain a consistent estimate under a normal distribution for the regression inliers [Scal 88].

The minimization of the energy function in Eq. (6.23) needs to consider overlapping image patches to establish linear regressions. However, the regression coefficients are defined to be constant within each patch according to the definition of the LLR model, see Section 6.3.2. It is worth noting that this constraint avoids the trivial solution for the regression coefficients ($\boldsymbol{C}_{ij}^t = \boldsymbol{0}$ and $\boldsymbol{b}_{ij}^t = -\boldsymbol{x}_i^{t-1}$). In order to consider this constraint, we utilize the separability of Eq. (6.23) and estimate local regression coefficients associated with the image $\omega_{\mathrm{LLR}}(k)$ centered at the k-th pixel position in the image channels $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ according to:

$$(\tilde{C}^t_{ij,k}, \tilde{b}^t_{ij,k}) = \underset{C_{ij,k}, b_{ij,k}}{\mathrm{argmin}} \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} \left(C_{ij,k} x_{i,l}^{t-1} + b_{ij,k} - x_{j,l}^{t-1}\right)^2 + \epsilon_{ij} C_{ij,k}^2 \ , \tag{6.28}$$

where the confidence weights are computed by $\kappa_{ij,l} = \kappa_l(\boldsymbol{C}_{ij}^{t-1}, \boldsymbol{b}_{ij}^{t-1})$ according to Eq. (6.25). This ridge regression problem is equivalent to confidence-aware guided filtering [Hore 14]. Hence, the local coefficients are computed in closed-form:

$$\tilde{C}^t_{ij,k} = \frac{\mathrm{E}_{\omega_{\mathrm{LLR}}(k)}\left(\boldsymbol{x}_i^{t-1} \odot \boldsymbol{x}_j^{t-1}, \boldsymbol{K}_{ij}^t\right) - \mathrm{E}_{\omega_{\mathrm{LLR}}(k)}\left(\boldsymbol{x}_i^{t-1}, \boldsymbol{K}_{ij}^t\right) \cdot \mathrm{E}_{\omega_{\mathrm{LLR}}(k)}\left(\boldsymbol{x}_j^{t-1}, \boldsymbol{K}_{ij}^t\right)}{\mathrm{E}_{\omega_{\mathrm{LLR}}(k)}\left(\boldsymbol{x}_i^{t-1} \odot \boldsymbol{x}_i^{t-1}, \boldsymbol{K}_{ij}^t\right) + \epsilon_{ij}}, \tag{6.29}$$

$$\tilde{b}^t_{ij,k} = \mathrm{E}_{\omega_{\mathrm{LLR}}(k)}\left(\boldsymbol{x}_j^{t-1}, \boldsymbol{K}_{ij}^t\right) - \tilde{C}^t_{ij} \cdot \mathrm{E}_{\omega_{\mathrm{LLR}}(k)}\left(\boldsymbol{x}_i^{t-1}, \boldsymbol{K}_{ij}^t\right), \tag{6.30}$$

where $\mathrm{E}_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{z}, \boldsymbol{K})$ denotes the weighted mean in the image patch $\omega_{\mathrm{LLR}}(k)$ centered at the k-th pixel in $\boldsymbol{z}$ under the confidence weights $\boldsymbol{K}$, see Appendix A.3.2.

The regression coefficients over the entire image are computed by averaging the local coefficients $\tilde{\boldsymbol{C}}_{ij}$ and $\tilde{\boldsymbol{b}}_{ij}$ corresponding to overlapping image patches following related strategies in image filtering [Hore 14, He 13]. Thus, the regression coefficients for the image channels $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are obtained by:

$$\boldsymbol{C}_{ij}^t = \operatorname{diag}\left(\mathrm{E}_{\omega_{\mathrm{LLR}}(1)}\left(\tilde{\boldsymbol{C}}_{ij}^t, \boldsymbol{K}_{ij}^t\right) \quad \dots \quad \mathrm{E}_{\omega_{\mathrm{LLR}}(N)}\left(\tilde{\boldsymbol{C}}_{ij}^t, \boldsymbol{K}_{ij}^t\right)\right), \tag{6.31}$$

$$\boldsymbol{b}_{ij}^t = \left(\mathrm{E}_{\omega_{\mathrm{LLR}}(1)}\left(\tilde{\boldsymbol{b}}_{ij}^t, \boldsymbol{K}_{ij}^t\right) \quad \dots \quad \mathrm{E}_{\omega_{\mathrm{LLR}}(N)}\left(\tilde{\boldsymbol{b}}_{ij}^t, \boldsymbol{K}_{ij}^t\right)\right)^\top. \tag{6.32}$$

It is important to note that this calculation of the regression coefficients for a single pair of image channels can be implemented with a time complexity of $\mathcal{O}(N)$ using box filtering, see Section 6.5.2. This considerably accelerates the hyperparameter estimation compared to the use of gradient-based optimization.

Estimation of the Image Channels. Given the estimate Φ^t for the regression coefficients, the high-resolution image $\boldsymbol{x}^t$ is estimated by jointly minimizing the negative log-likelihood of Eq. (6.21) w. r. t. the individual channels $\boldsymbol{x}_1, \dots, \boldsymbol{x}_C$. That is, we obtain $\boldsymbol{x}^t$ as the solution of the energy minimization problem:

$$\boldsymbol{x}^t = \operatorname*{argmin}_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_C} F^t(\boldsymbol{x}). \tag{6.33}$$

Similar to the estimation of the regression coefficients, the optimization of this non-convex log-likelihood term is rewritten as an MM algorithm [Hunt 04]. This leads to a weighted but convex minimization problem, see Section 4.4.2. Thus, the joint energy function is given by:

$$\begin{aligned} F^t(\boldsymbol{x}) = &\sum_{i=1}^{C} \left(\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{x}_i\right)^\top \boldsymbol{B}_i^t \left(\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{x}_i\right) + \lambda_i \left|\left|\boldsymbol{A}_i^t\boldsymbol{x}_i\right|\right|_1 \\ &+ \sum_{j=1, j\neq i}^{C} \mu_{ij} \left(\boldsymbol{C}_{ij}^t\boldsymbol{x}_i + \boldsymbol{b}_{ij}^t - \boldsymbol{x}_j\right)^\top \boldsymbol{K}_{ij}^t \left(\boldsymbol{C}_{ij}^t\boldsymbol{x}_i + \boldsymbol{b}_{ij}^t - \boldsymbol{x}_j\right), \end{aligned} \tag{6.34}$$

where the confidence weights for the i-th channel at iteration t are assembled as the diagonal matrices:

$$\boldsymbol{A}_i^t = \operatorname{diag}\left(\alpha_{i,1}\left(\boldsymbol{x}^{t-1}\right) \quad \alpha_{i,2}\left(\boldsymbol{x}^{t-1}\right) \quad \dots \quad \alpha_{i,N_S}\left(\boldsymbol{x}^{t-1}\right)\right), \tag{6.35}$$

$$\boldsymbol{B}_i^t = \operatorname{diag}\left(\beta_{i,1}\left(\boldsymbol{x}^{t-1}\right) \quad \beta_{i,2}\left(\boldsymbol{x}^{t-1}\right) \quad \dots \quad \beta_{i,KM}\left(\boldsymbol{x}^{t-1}\right)\right), \tag{6.36}$$

using the weighting functions $\alpha_{i,k}(\boldsymbol{x}^{t-1}) \equiv \alpha_k(\boldsymbol{x}_i^{t-1})$ and $\beta_{i,k}(\boldsymbol{x}^{t-1}) \equiv \beta_k(\boldsymbol{x}_i^{t-1})$ with adaptive scale parameter selection as previously introduced in Section 4.4.1, see Eqs. (4.21) - (4.25). Note that both weighting functions can be applied channel-wise to determine the corresponding confidence weights and to majorize the negative log-likelihood of Eq. (6.19).

Numerical optimization of Eq. (6.33) is performed by means of SCG iterations starting from $\boldsymbol{x}^{t-1}$ obtained at the previous iteration. This gradient-based iteration scheme seeks a stationary point:

$$\nabla_{\boldsymbol{x}} F^t(\boldsymbol{x}) = \left(\frac{\partial F^t}{\partial \boldsymbol{x}_1}(\boldsymbol{x}) \quad \frac{\partial F^t}{\partial \boldsymbol{x}_2}(\boldsymbol{x}) \quad \dots \quad \frac{\partial F^t}{\partial \boldsymbol{x}_C}(\boldsymbol{x})\right)^\top = \boldsymbol{0}, \tag{6.37}$$

where the gradient of the joint energy function w. r. t. the k-th channel, $k = 1, \ldots, C$ is computed in closed-form:

$$\begin{aligned}\frac{\partial F^t}{\partial \boldsymbol{x}_k} = &-2\boldsymbol{B}_k^t \boldsymbol{W}_k^\top \left(\boldsymbol{y}_k - \boldsymbol{W}_k \boldsymbol{x}_k\right) + \lambda_k \cdot \boldsymbol{A}_k^t \boldsymbol{S}^\top \operatorname{sign}\left(\boldsymbol{A}_k^t \boldsymbol{S} \boldsymbol{x}_k\right) \\ &+ \sum_{j=1, j\neq k}^{C} 2\mu_{kj} \cdot \boldsymbol{K}_{kj}^t \boldsymbol{C}_{kj}^t \left(\boldsymbol{C}_{kj}^t \boldsymbol{x}_k + \boldsymbol{b}_{kj}^t - \boldsymbol{x}_j\right) - \sum_{i=1, i\neq k}^{C} 2\mu_{ik} \cdot \boldsymbol{K}_{ik}^t \left(\boldsymbol{C}_{ik}^t \boldsymbol{x}_i + \boldsymbol{b}_{ik}^t - \boldsymbol{x}_k\right).\end{aligned} \tag{6.38}$$

To facilitate gradient-based optimization, the derivatives of non-smooth L_1 norm terms are approximated by the Charbonnier function, i. e. $\operatorname{sign}(z) \approx z/(\sqrt{z^2 + \tau})$, where τ is a small constant ($\tau = 10^{-4}$) to ensure differentiability at $z = 0$ [Char 94].

Overall Optimization Algorithm. We divide the proposed alternating minimization scheme in an outer and two inner optimization loops, see Algorithm 6.1. In order to provide an initialization for the iterations, a sequential estimation of the high-resolution image channels is performed. This can be done by minimizing Eq. (6.33) without inter-channel prior (i. e. $\mu_{ij} = 0$ for all $i, j = 1, \ldots, C$), which is equivalent to a channel-wise iteratively re-weighted reconstruction as previously introduced in Chapter 4 using constant regularization weights λ_i. Subsequently, the regression coefficients are computed pair-wise based on Eqs. (6.29),(6.30) while the high-resolution channels are estimated jointly by SCG iterations for Eq. (6.33).

In total, we perform a maximum number of T_{am} iterations for alternating minimization and a maximum number of T_{scg} iterations for SCG to estimate the high-resolution image channels. As a termination criterion we choose the maximum absolute difference among consecutive iterations:

$$\max_{i=1,\ldots,C} \left(\max_{k=1,\ldots,N} \left(\left| x_{i,k}^t - x_{i,k}^{t-1} \right| \right) \right) < \eta, \tag{6.39}$$

where η denotes the termination tolerance.

6.5 Model and Algorithm Analysis

In this section, we present an in-depth analysis of the proposed Bayesian model along with the joint MAP estimation approach. In particular, this covers a study regarding the adaptivity of the prior distribution as well as the computational complexity and the convergence of the algorithm. Eventually, a theoretical comparison of the Bayesian model to related state-of-the-art methods is presented.

6.5.1 Adaptivity of the Regression Model

Unlike many of the closely related joint filters [Kopf 07, Zhan 14b, He 13], the proposed inter-channel prior is robust against inconsistent structures among channels. Prominent examples for this issue appear in range imaging, where texture

Algorithm 6.1 Multi-sensor super-resolution using locally linear regression (LLR)

Input: Initial guess for high-resolution multi-channel image $\boldsymbol{x}^0$
Output: Final high-resolution multi-channel image $\boldsymbol{x}$ with LLR coefficients $\boldsymbol{C}_{ij}$, $\boldsymbol{b}_{ij}$ and $\boldsymbol{K}_{ij}$

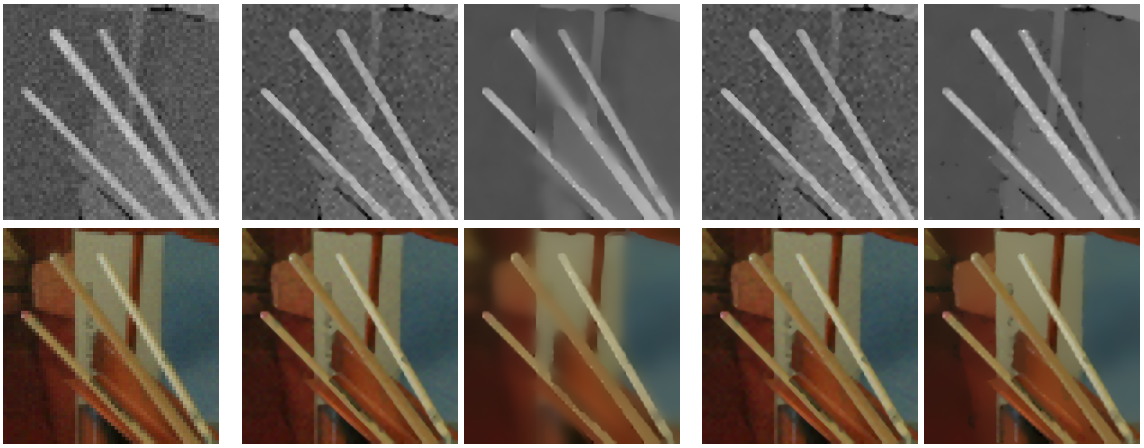
$t \leftarrow 1$
while Convergence criterion in Eq. (6.39) not fulfilled and $t \leq T_{\text{am}}$ **do**
  for $i = 1, \ldots, C$ **do**
    for $j = 1, \ldots, C$ **do**
      Estimate confidence weights $\boldsymbol{K}_{ij}^t$ for LLR prior by Eq. (6.25)
      Estimate LLR coefficients $\boldsymbol{C}_{ij}^t$ and $\boldsymbol{b}_{ij}^t$ by Eqs. (6.29),(6.30)
    end for
  end for
  $t_{\text{scg}} \leftarrow 1$
  while Convergence criterion in Eq. (6.39) not fulfilled and $t_{\text{scg}} \leq T_{\text{scg}}$ **do**
    Update high-resolution channels $\boldsymbol{x}^t$ by SCG iteration for Eq. (6.33)
    $t_{\text{scg}} \leftarrow t_{\text{scg}} + 1$
  end while
  $t \leftarrow t + 1$
end while

does not necessarily coincide with surface information, or in multispectral restoration with structural inconsistencies among the spectral bands [Shen 15]. One essential property of the inter-channel prior is its adaptivity regarding these inconsistencies to avoid the erroneous transfer of structures from original channels to complementary reconstructed ones.

Figure 6.5 investigates this adaptivity in the context of joint upsampling of single range and color images. In this simulated data example, the inter-channel prior is adopted in two different versions. On the one hand, we analyze the proposed regression based on Tukey's biweight loss that tolerates outliers related to inconsistent structures and is referred to as adaptive LLR. On the other hand, Tukey's biweight is replaced by the L_2 norm, which leads to a simplified model that is referred to as non-adaptive LLR. The inter-channel regularization weight μ controls the impact of the regressions among range and color data to the upsampled channels[2]. Figure 6.5b and Fig. 6.5c demonstrates that the non-adaptive model is prone to texture-copying artifacts in case of an overestimated regularization weight μ. Moreover, structural inconsistencies cause an oversmoothing of the color data. The adaptive regression depicted in Fig. 6.5d and Fig. 6.5e features higher robustness against unwanted texture-copying. This behavior is quantitatively analyzed in Fig. 6.6 (left) by the PSNR of upsampled range and color data over a wide range of parameter settings. Here, the adaptive model features a lower sensitivity regarding an overestimation of the regularization weight. Note that the non-adaptive model might converge to a solution that is even inferior to a simple sequential upsampling, which is considered as the baseline.

In addition, Fig. 6.6 (right) depicts the parameter sensitivity analysis regarding the hyperparameter regularization weight ϵ. Notice that the adaptive model is

[2] In this analysis, we limit ourselves to uniform inter-channel weights μ and LLR hyperparameter regularization weights ϵ for the regression between all pairs of channels.



(a) Original **(b)** Non-adaptive ($\mu = 0.01$) **(c)** Non-adaptive ($\mu = 1.00$) **(d)** Adaptive ($\mu = 0.01$) **(e)** Adaptive ($\mu = 1.00$)

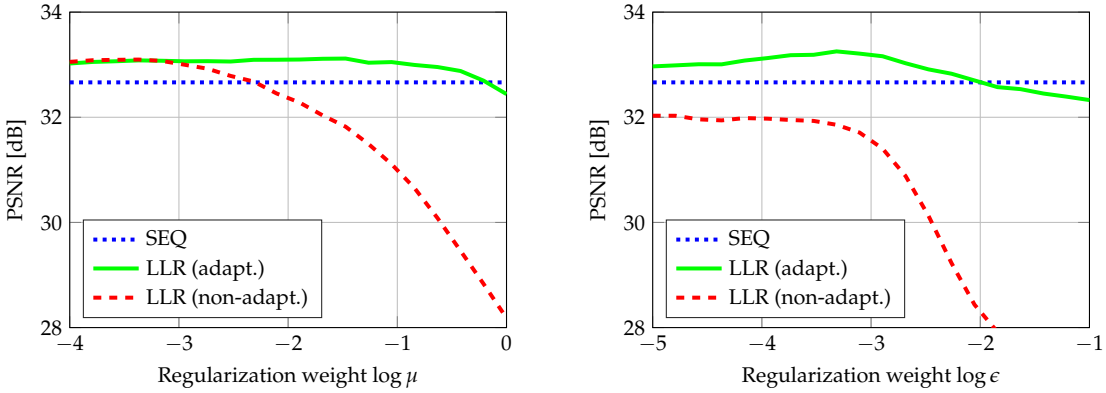
Figure 6.5: Analysis of the LLR prior for joint upsampling of range and color data ($4\times$ upsampling). (a) Simulated range and color data. (b) and (c) Upsampled range and color data using a non-adaptive version of the LLR prior based on the L_2 norm with different inter-channel regularization weights μ. (d) and (e) Upsampled range and color data with the proposed adaptive LLR prior based on Tukey's biweight loss. Notice the texture-copying artifacts and the oversmoothing caused by the non-adaptive version of the prior.

stable over a wide range of parameter settings and outperforms the non-adaptive counterpart by a large margin.

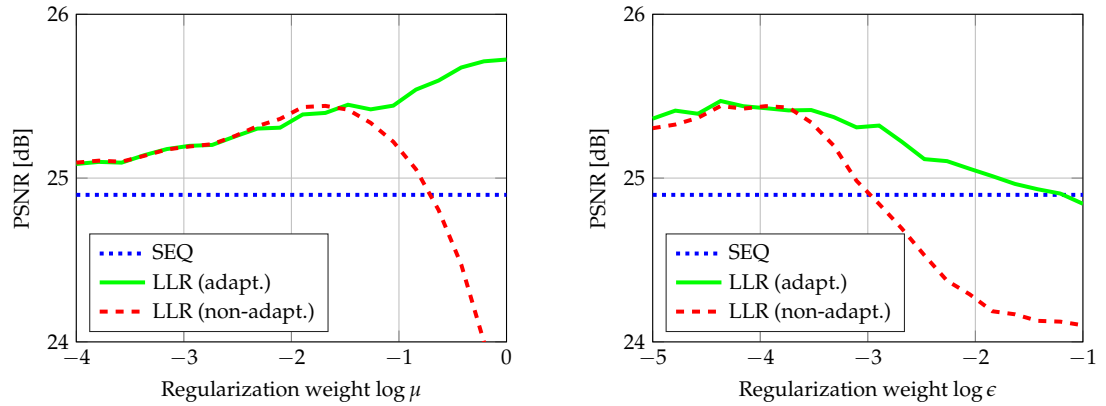
6.5.2 Computational Complexity and Convergence

The time complexity of Algorithm 6.1 is related to two phases. First, given C channels, the regression coefficients for $C(C-1)$ pairs of channels are computed according to Eqs. (6.29),(6.30) at each iteration. This involves element-wise vector products as well as confidence-aware box filtering. Given N pixels per channel, this can be implemented with time complexity $\mathcal{O}(N)$ for each pair by means of integral images and is independent on the regression patch size [Crow 84, He 13]. Hence, the regression coefficient estimation has $\mathcal{O}\left(C^2N\right)$ time complexity. Second, the estimation of the high-resolution channels using SCG requires the computation of the joint energy function and its gradient based on Eqs. (6.34) - (6.37). Given K frames of size M pixels, this can be implemented by sparse matrix-vector products as well as element-wise vector products and has $\mathcal{O}\left(CKM + C^2N\right)$ time complexity at each inner iteration. For T_{scg} iterations for SCG in the inner optimization loop, the overall time complexity of a single alternating minimization iteration in Algorithm 6.1 is $\mathcal{O}\left(C^2N + T_{\text{scg}}(CKM + C^2N)\right)$.

In addition to the complexity, let us also investigate the convergence of Algorithm 6.1 experimentally. In Fig. 6.7, the convergence is analyzed for the joint range and color upsampling example in Fig. 6.5. This depicts the progress of the PSNR of the range data as well as the sum of absolute differences between the channels for successive iterations. Furthermore, we show a comparison regarding the influence of the initial guess to the solution of the underlying non-convex energy



(a) Sensitivity analysis using the PSNR on color data



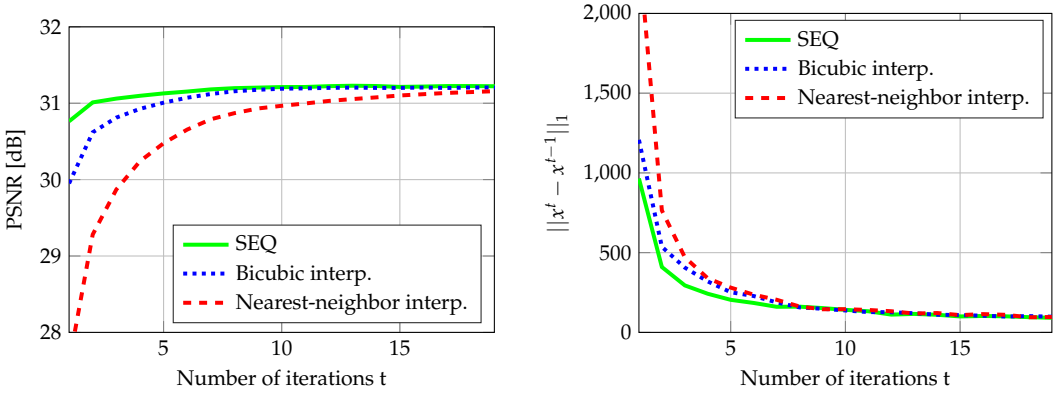
(b) Sensitivity analysis using the PSNR on range data

Figure 6.6: Sensitivity analysis of the LLR prior for joint range and color upsampling including a comparison of a non-adaptive version of the prior using the L_2 norm to the adaptive one based on Tukey's biweight loss. Sequential upsampling of the channels (SEQ) is considered as the baseline. (a) Sensitivity regarding the inter-channel regularization weight μ. (b) Sensitivity regarding the hyperparameter regularization weight ϵ. Notice that adaptive LLR features a higher stability over a wider range of parameter settings.

minimization problem. For this purpose, let us compare the proposed initialization provided by a sequential reconstruction without inter-channel prior to bicubic and nearest-neighbor interpolations. While the former provides a more accurate starting point, the latter are easy to compute. Albeit different initializations are used, alternating minimization converges to comparable solutions. We found that $T_{\text{am}} = 10$ iterations for alternating minimization with $T_{\text{scg}} = 10$ iterations for SCG are typically sufficient for convergence.

6.5.3 Connection to Related Methods

The Bayesian model introduced in Section 6.3 features a combination of two complementary paradigms of image filtering and restoration. On the one hand, this includes local filtering (e. g. [Kopf 07, Zhan 14b]), where the goal is to obtain a filter output image from an input by considering relationships between both in



(a) PSNR over successive iterations **(b)** Absolute differences among iterations

Figure 6.7: Convergence analysis of alternating minimization for multi-channel super-resolution. We compared different initializations including a sequential reconstruction of the channels (SEQ) as well as channel-wise bicubic and nearest-neighbor interpolation.

local patches. Global methods (e. g. [Ham 15, Shen 15]) on the other hand aim at an implicit reconstruction of a filter output by optimizing global energy functions. For the sake of notational brevity, let us study the relationship to these paradigms for $C = 2$ channels. Here, a single iteration of alternating minimization aims at optimizing the joint energy function:

$$\begin{aligned} F(\boldsymbol{x}_1, \boldsymbol{x}_2, \Phi) = & \underbrace{\phi_y(\boldsymbol{y}_1 - \boldsymbol{W}_1\boldsymbol{x}_1) + \phi_y(\boldsymbol{y}_2 - \boldsymbol{W}_2\boldsymbol{x}_2) + \lambda_1\phi_x(\boldsymbol{x}_1) + \lambda_2\phi_x(\boldsymbol{x}_2)}_{\text{global term } F_{\text{global}}(\boldsymbol{x}_1, \boldsymbol{x}_2)} \\ & + \underbrace{\mu_{12}\phi_{x_{12}}(\boldsymbol{C}_{12}\boldsymbol{x}_1 + \boldsymbol{b}_{12} - \boldsymbol{x}_2) + \mu_{21}\phi_{x_{21}}(\boldsymbol{C}_{21}\boldsymbol{x}_2 + \boldsymbol{b}_{21} - \boldsymbol{x}_1)}_{\text{local term } F_{\text{local}}(\boldsymbol{x}_1, \boldsymbol{x}_2, \Phi)} \\ & + \underbrace{\epsilon_{12}||\text{diag}(\boldsymbol{C}_{12})||_2^2 + \epsilon_{21}||\text{diag}(\boldsymbol{C}_{21})||_2^2}_{\text{hyperparameter regularization}}. \end{aligned} \tag{6.40}$$

The observation model and the intra-channel prior are related to the global energy $F_{\text{global}}(\boldsymbol{x}_1, \boldsymbol{x}_2)$, while the inter-channel prior forms the local energy $F_{\text{local}}(\boldsymbol{x}_1, \boldsymbol{x}_2, \Phi)$. This local term appears as an additional regularizer in this inverse problem. Similar formulations appear in two closely related filtering techniques.

Relation to Guided Filtering. The mixed local/global formulation in Eq. (6.40) provides a generalization of the well known guided filter [He 13] as a prominent example for a local operator. For this consideration, let $\lambda_1 = \lambda_2 = 0$ and let us drop the global data fidelity term. Moreover, let $\mu_{21} = \epsilon_{21} = 0$. Then, the optimization of Eq. (6.40) can be simplified to the minimization of the energy function:

$$F_{\text{GF}}(\boldsymbol{x}_1, \boldsymbol{x}_2, \Phi) = \phi_{x_{12}}(\boldsymbol{C}_{12}\boldsymbol{x}_1 + \boldsymbol{b}_{12} - \boldsymbol{x}_2) + \epsilon_{12}||\text{diag}(\boldsymbol{C}_{12})||_2^2. \tag{6.41}$$

If we keep the channel $\boldsymbol{x}_1$ fixed, optimizing $F_{\text{GF}}(\boldsymbol{x}_1, \boldsymbol{x}_2, \Phi)$ is a generalized version of guided filtering for $\boldsymbol{x}_2$ under the guidance of $\boldsymbol{x}_1$. Compared to conventional guided filtering, the proposed algorithm provides several valuable extensions. First and foremost, the local model adopts the robust loss function $\phi_{x_{ij}}(z)$

to establish linear regressions, while guided filtering relies on simple least-squares estimation making it prone to outliers. This extension is sensible in case of inconsistent structures, see Section 6.5.1. In addition, the proposed formulation couples local filtering with a global model. This offers the flexibility to model the image formation for multi-frame super-resolution in contrast to a purely local filtering that might exhibit halo artifacts [He 13]. Similar relations applies in comparison to other local filters, e. g. joint bilateral filtering [Kopf 07].

Relation to Mutual Structure Filtering. Eq. (6.40) also provides a generalization of mutual structure filtering as proposed by Shen et al. [Shen 15]. To outline this relationship, let us assume that the system matrices of the individual channels are given by the identity, i. e. $\boldsymbol{W}_1 = \boldsymbol{W}_2 = \boldsymbol{I}$. In addition, let us drop the intra-channel prior, i. e. $\lambda_1 = \lambda_2 = 0$. Then, Eq. (6.40) can be simplified to:

$$\begin{aligned} F_{\mathrm{MS}}(\boldsymbol{x}_1, \boldsymbol{x}_2, \Phi) = \; & \phi_y(\boldsymbol{y}_1 - \boldsymbol{x}_1) + \phi_y(\boldsymbol{y}_2 - \boldsymbol{x}_2) \\ & + \mu_{12}\phi_{x_{12}}(\boldsymbol{C}_{12}\boldsymbol{x}_1 + \boldsymbol{b}_{12} - \boldsymbol{x}_2) + \mu_{21}\phi_{x_{21}}(\boldsymbol{C}_{21}\boldsymbol{x}_2 + \boldsymbol{b}_{21} - \boldsymbol{x}_1) \\ & + \epsilon_{12}\big|\big|\mathrm{diag}(\boldsymbol{C}_{12})\big|\big|_2^2 + \epsilon_{21}\big|\big|\mathrm{diag}(\boldsymbol{C}_{21})\big|\big|_2^2. \end{aligned} \tag{6.42}$$

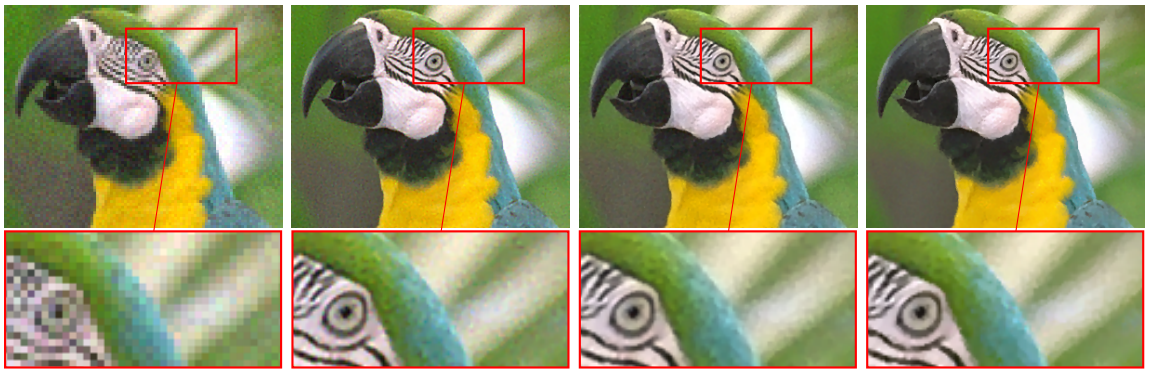
Joint minimization of $F_{\mathrm{MS}}(\boldsymbol{x}_1, \boldsymbol{x}_2, \Phi)$ w. r. t. the regression coefficients Φ and the channels $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is conceptually equivalent to mutual structure filtering. However, notice that the algorithm in [Shen 15] considers a denoising problem ($\boldsymbol{W}_1 = \boldsymbol{W}_2 = \boldsymbol{I}$) and employs the L_2 norm for $\phi_x(z)$ and $\phi_{x_{ij}}(z)$. These simplifications make joint minimization efficient to compute, since closed-form solutions can be derived for the latent image channels and the regression coefficients. In contrast to this approach, the proposed algorithm enables multi-frame super-resolution.

6.6 Experiments and Results

This section presents a detailed experimental evaluation of the proposed multi-channel super-resolution. Several applications that are of great interest in computer vision are being investigated. The main focus of this study lies on resolution enhancement for color and 3-D range images as two classical applications. In addition, further experiments including multispectral image upsampling as well as joint segmentation and super-resolution are presented.

6.6.1 Applications in Color Imaging

In terms of color image resolution enhancement, we aim at simultaneously super-resolving color channels in the RGB space. For this purpose, the LLR model is adopted to exploit dependencies among the spectral bands that are common in case of natural images [Omer 04]. The proposed algorithm was compared to the following approaches to color super-resolution. As a baseline approach, a sequential super-resolution of the color channels that served as initial guess for multi-channel super-resolution was evaluated. This is conceptually equivalent to the algorithm introduced by Köhler et al. [Kohl 16b] for single-channel images. In



(a) Original **(b)** Sequential **(c)** IDP [Fars 06] **(d)** LLR

Figure 6.8: Color super-resolution ($K = 10$ frames, magnification $s = 3$) on simulated data with a comparison of a sequential reconstruction of the color channels to a multi-channel reconstruction using IDP regularization [Fars 06] and the LLR prior.

addition, multi-channel super-resolution was evaluated using the inter-color dependency penalty (IDP) proposed by Farsiu et al. [Fars 06] as an alternative to the proposed prior[3]. This regularization term penalizes mismatches in terms of the location and orientation of edges among the color channels. In the sequel, experiments on simulated and real color image sequences are presented.

Simulated Data. In order to conduct a quantitative evaluation, simulated color image sequences ($K = 10$ frames) with randomly generated rigid motion were generated from the LIVE Database [Shei 16]. This simulation comprises a Gaussian PSF ($\sigma_{\text{PSF}} = 0.5$), subsampling ($s = 3$) as well as additive Gaussian noise ($\sigma_{\text{noise}} = 0.03$) for each color channel. Throughout these experiments, the exact subpixel motion was used for super-resolution to explicitly study the impact of the different prior models. The inter-channel regularization weights μ_{ij} as well as the hyperparameter weights ϵ_{ij} were defined in a symmetric way ($\mu_{ij} = \mu_{ji}$ and $\epsilon_{ij} = \epsilon_{ji}$). The intra-channel prior parameters were set to $\lambda_i = 4 \cdot 10^{-3}$ with BTV window size $N_{\text{BTV}} = 1$ and weighting factor $\alpha_{\text{BTV}} = 0.5$ for each channel. The inter-channel parameters were set to $\mu_{ij} = 0.5$ and $\epsilon_{ij} = 10^{-4}$ with LLR patch size $N_{\text{LLR}} = 3$ for all pairs of channels.

Figure 6.8 compares sequential super-resolution of the individual color channels to multi-channel super-resolution. In contrast to the sequential approach, both multi-channel methods avoided inconsistencies between the super-resolved color channels. This resulted in lower noise levels in homogeneous areas while the sequential approach caused color artifacts in these regions. In a quantitative comparison on 20 simulated datasets, the proposed method achieved the highest PSNR and SSIM measures, see Fig. 6.9. On average, compared to the sequential and the IDP super-resolution, the proposed algorithm improved the PSNR (SSIM) by 1.5 dB (0.04) and 0.5 dB (0.01), respectively.

[3]Chrominance and luminance regularization as well as demosaicing as proposed in [Fars 06] were omitted to exclusively evaluate the influence of the inter-channel regularization.

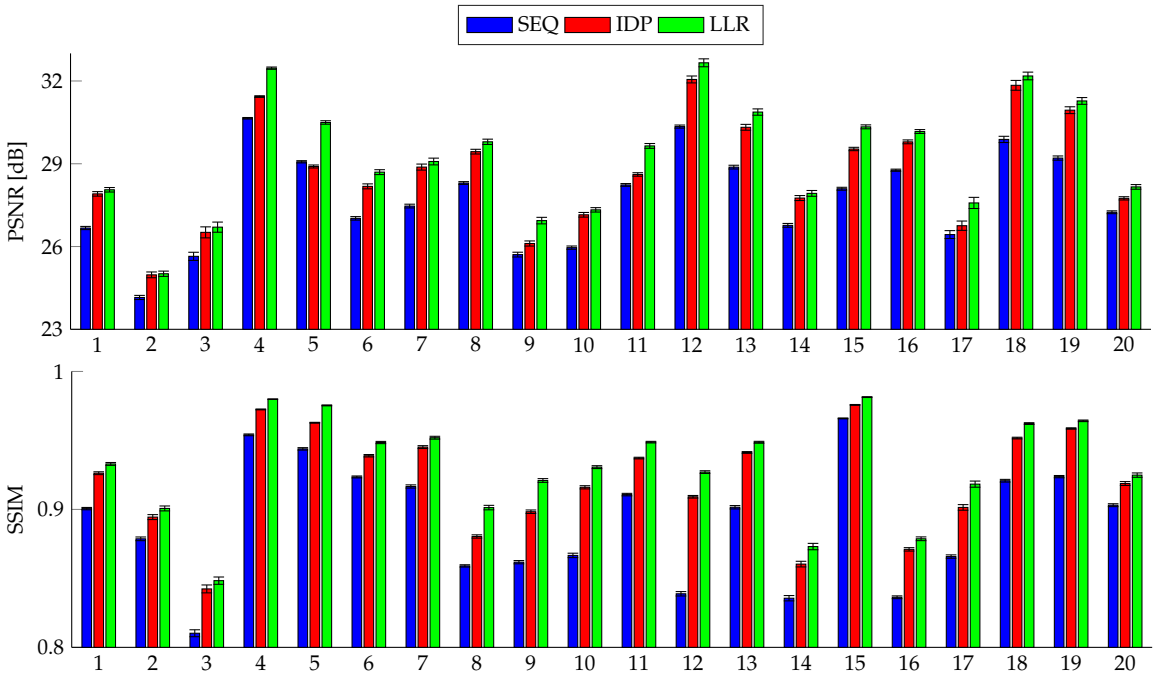


Figure 6.9: Mean ± standard deviation of the PSNR and the SSIM of 20 simulated color image datasets obtained from the LIVE database [Shei 16]. For each dataset, 15 randomly simulated image sequences were generated to determine the quality measure statistics. This benchmark compares sequential super-resolution of the color channels (SEQ) to multi-channel super-resolution using IDP regularization [Fars 06] and the LLR prior.

Real Data. Figure 6.10 shows a qualitative comparison of the competing methods on the *Bookcase* sequence ($K = 30$ frames) taken from the MDSP benchmark dataset [Fars 16]. Each pixel in these color images reflects full RGB information to make additional demosaicing unnecessary. Super-resolution was conducted with magnification $s = 3$ and a Gaussian PSF ($\sigma_{\text{PSF}} = 0.4$). The unknown subpixel motion was described by an affine model and estimated by ECC optimization [Evan 08]. Throughout these experiments, we set the intra-channel prior parameters to $\lambda_i = 5 \cdot 10^{-3}$, $N_{\text{BTV}} = 1$ and $\alpha_{\text{BTV}} = 0.5$ for each channel. The inter-channel parameters were set to $\mu_{ij} = 20$, $\epsilon_{ij} = 10^{-4}$ and $N_{\text{LLR}} = 1$ for all pairs of channels.

Similar to the previous experiments, multi-channel super-resolution gets rid of color artifacts that appeared in the super-resolved color channels obtained by the sequential approach. Examples regarding these artifacts are jagged edges and color bleeding as shown in the highlighted region that contains text. These artifacts were better compensated by multi-channel super-resolution using the LLR prior.

6.6.2 Applications in Range Imaging

In Chapter 5, super-resolution for 3-D range data was investigated by exploiting color images as a static guidance. The multi-channel technique presented in this chapter is applicable to similar setups but does not rely on the existence of accurate guidance data. The following experiments consider single-image upsampling as well as multi-frame super-resolution in the context of range imaging.

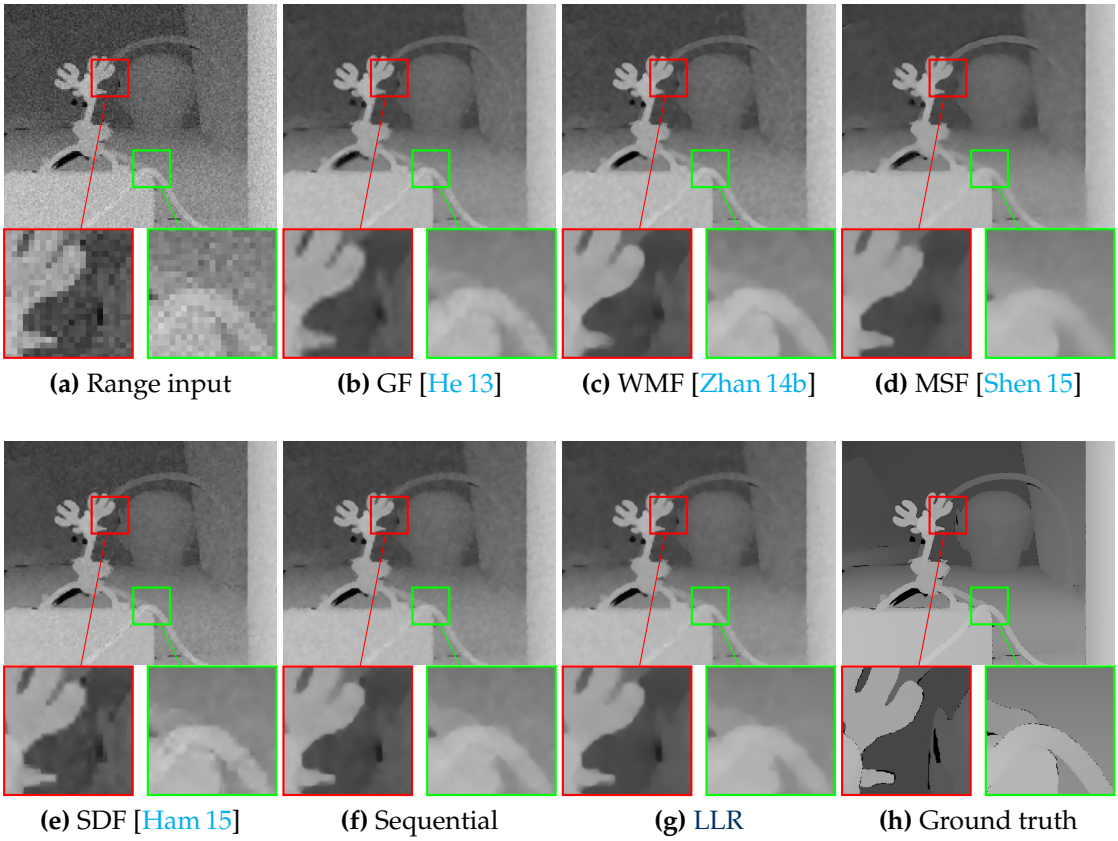
(a) Original **(b)** Sequential **(c)** IDP [Fars 06] **(d)** LLR

Figure 6.10: Color super-resolution ($K = 30$ frames, magnification $s = 3$) on the *Bookcase* sequence [Fars 16] with a comparison of sequential color channel reconstruction to multi-channel reconstructions using IDP regularization [Fars 06] and the LLR prior.

Joint RGB-D Upsampling. In the domain of single image joint RGB-D upsampling, we aim at simultaneously upsampling $C = 4$ channels (RGB color plus depth) from their low-resolution counterparts. This is a highly underdetermined reconstruction problem but exploiting mutual dependencies among range and color data serves as a strong prior to alleviate this issue.

For the sake of a quantitative evaluation, artificial RGB-D images were obtained from the ground truth color and disparity data of the Middlebury Stereo Datasets [Hirs 07, Scha 07]. The formation of low-resolution images considered the conditions of low-cost ToF sensors and comprises a Gaussian PSF ($\sigma_{\text{PSF}} = 0.3$) and subsampling ($s = 8$) relative to the ground truth. These degradations were jointly applied to all channels to consider the absence of reliable guidance information. In addition, color and range channels were corrupted by additive Gaussian noise with standard deviations 0.02 and 0.04, respectively. We adopted the LLR prior with a symmetric distribution for joint multi-channel upsampling. The inter-channel parameters were set to $\mu_{ij} = 1.5$, $\epsilon_{ij} = 10^{-4}$ and $N_{\text{LLR}} = 5$ for all pairs of channels. For the intra-channel prior a BTV window size of $N_{\text{BTV}} = 2$ with weighting factor $\alpha_{\text{BTV}} = 0.5$ was used. The intra-channel regularization weights were set to $\lambda_i = 5 \cdot 10^{-4}$ for the color channels ($i \in \{1, 2, 3\}$) and $\lambda_i = 10^{-2}$ for the range channel ($i = 4$) to reflect the noise levels of both modalities.

Figure 6.11 depicts a comparison of the proposed joint multi-channel upsampling against various state-of-the-art filters for color guided range upsampling using publicly available reference implementations. This includes the guided filter (GF) [He 13], the weighted median filter (WMF) [Zhan 14b], the mutual structure filter (MSF) [Shen 15] as well as the static and dynamic guidance filter (SDF) [Ham 15]. Among them, GF and WMF employ color data as a static guidance that was obtained by bicubic upsampling of the input color image. As a consequence, these methods gained a limited quality of the upsampled range data due to the absence of reliable color data. In addition to a pure static guidance, MSF and SDF also exploit range data regularization similar to the proposed model. Although these approaches were less sensitive to erroneous texture-copying since structural



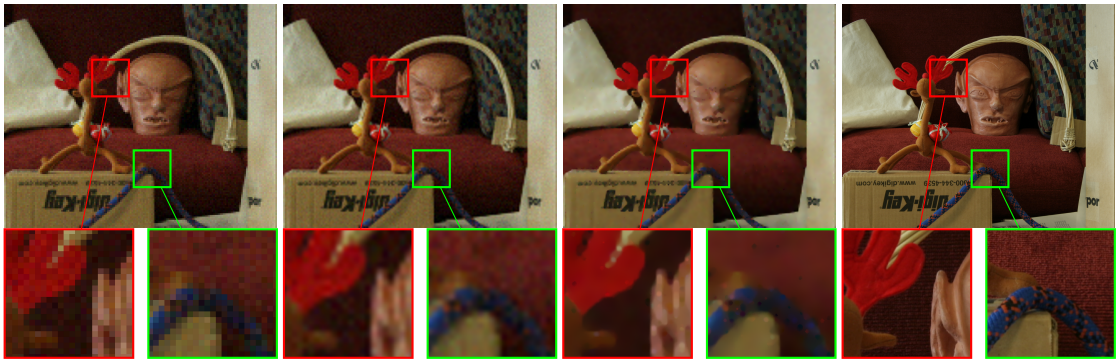
(a) Range input **(b)** GF [He 13] **(c)** WMF [Zhan 14b] **(d)** MSF [Shen 15]

(e) SDF [Ham 15] **(f)** Sequential **(g)** LLR **(h)** Ground truth

Figure 6.11: Single image RGB-D upsampling (magnification $s = 8$) on the Middlebury Stereo Datasets [Hirs 07, Scha 07] with visual comparison of the upsampled range images. (a) Low-resolution range image, (b) guided filter [He 13], (c) weighted median filter [Zhan 14b], (d) mutual structure filter [Shen 15], (e) static and dynamic guidance filter [Ham 15], (f) sequential upsampling of range and color channels without inter-channel prior, (g) multi-channel upsampling using the LLR prior, (h) ground truth range data.

inconsistencies could be considered, they do not incorporate an appropriate image formation model. Thus, they were inherently limited for upsampling of structures that were lost in low-resolution range *and* color data. Opposed to the state-of-the-art, the proposed method simultaneously upsampled all channels under a reasonable image formation model and profited from enhanced color data for the task of range upsampling. This is depicted in Fig. 6.12 showing the original and bicubic interpolation of the color channels in comparison to the color image reconstructed as a by-product of joint multi-channel upsampling.

In Tab. 6.1, a benchmark on various Middlebury datasets is summarized. For fair comparisons, the parameters of the different joint image filters were adjusted to each dataset individually to optimize the PSNR, while the proposed joint multi-channel upsampling was employed with the aforementioned default parameters. In these experiments, joint multi-channel upsampling based on LLR considerably outperformed a simple sequential upsampling of the channels. Moreover, the joint upsampling considerably improved the accuracy of range information compared to the state-of-the-art filters on most of the evaluated datasets.



(a) Color input **(b)** Bicubic interp. **(c)** LLR **(d)** Ground truth

Figure 6.12: Single image RGB-D upsampling on the Middlebury Stereo Datasets [Hirs 07, Scha 07] with visual comparison of the upsampled color data. (a) Low-resolution image, (b) bicubic upsampling, (c) upsampling using the LLR prior, (d) ground truth color image.

Table 6.1: Quantitative evaluation of joint RGB-D image upsampling on the Middlebury Stereo Datasets [Hirs 07, Scha 07] (8× upsampling). We compared the PSNR of upsampled range data using different joint image filters (GF [He 13], WMF [Zhan 14b], MSF [Shen 15], SDF [Ham 15]) to multi-channel upsampling using a sequential (channel-wise) approach as well as the proposed LLR prior. All joint filters used the bicubic upsampled color images as static guidance. For each dataset, the best and the second best results are highlighted.

	Art	*Books*	*Dolls*	*Laundry*	*Moebius*	*Reindeer*
Interpolation						
Nearest-neighbor	26.04	26.44	26.90	26.93	26.80	26.75
Bicubic	27.82	27.99	28.56	28.62	28.42	28.46
Joint filters						
GF [He 13]	30.57	32.69	34.46	33.25	33.48	32.55
WMF [Zhan 14b]	**30.81**$^{)2}$	**32.93**$^{)2}$	34.32	33.47	**33.55**$^{)2}$	32.94
MSF [Shen 15]	30.43	**33.14**$^{)1}$	**34.50**$^{)2}$	**33.49**$^{)2}$	33.50	**33.10**$^{)2}$
SDF [Ham 15]	30.43	31.80	33.23	33.33	32.85	32.09
Multi-channel						
Sequential	30.71	32.36	33.98	33.30	33.23	32.96
LLR	**30.93**$^{)1}$	32.75	**34.56**$^{)1}$	**33.81**$^{)1}$	**33.72**$^{)1}$	**33.42**$^{)1}$

Photogeometric Super-Resolution. Contrary to related joint filters, the proposed model can be directly applied to gain resolution enhancement of range and color data from a multi-frame perspective referred to as *photogeometric super-resolution* [Ghes 14]. We investigate this methodology for current ToF cameras that provide geometric information by 3-D range images along with photometric information encoded by amplitude images at the same pixel resolution at a video frame rate. Figure 6.13 and Fig. 6.14 depict super-resolution on range and amplitude data of

two datasets captured with a Mesa SR-4000 camera[4]. Both channels were acquired at a resolution of 176×144 px and subpixel motion across the sequences was induced by camera translations. Motion estimation was implemented via variational optical flow [Liu 09] on the amplitude images. We used the LLR prior for $C = 2$ channels with symmetric distribution. The inter-channel parameters were set to $\mu_{ij} = 10^2$, $\epsilon_{ij} = 10^{-4}$ and $N_{\mathrm{LLR}} = 1$ for all pairs of channels. The intra-channel parameters were set to $\lambda_i = 5 \cdot 10^{-4}$ for the amplitude data ($i = 1$) and $\lambda_i = 2 \cdot 10^{-3}$ for the range data ($i = 2$) with $N_{\mathrm{BTV}} = 1$ and $\alpha_{\mathrm{BTV}} = 0.5$ for both channels.

We compared the proposed technique against adaptive multi-sensor super-resolution (AMSR) as presented in Section 5.5. AMSR was applied to both channels separately using the amplitude data as static guidance for its feature-based regularizer. In addition, we evaluated sequential super-resolution of both channels under the proposed model without inter-channel prior.

Photogeometric super-resolution simultaneously enhanced geometric and photometric information. In comparison to a sequential reconstruction of both channels, exploiting mutual dependencies between range and amplitude data boosted range super-resolution even further. The LLR prior could also better capture such dependencies compared to AMSR that used low-resolution amplitude data directly to steer the underlying regularization technique. Notice that in this setup, amplitude images do not necessarily meet the quality requirements for a static guidance. This improvement of the range regularization resulted in superior reconstructions of depth discontinuities and surfaces.

6.6.3 Further Applications

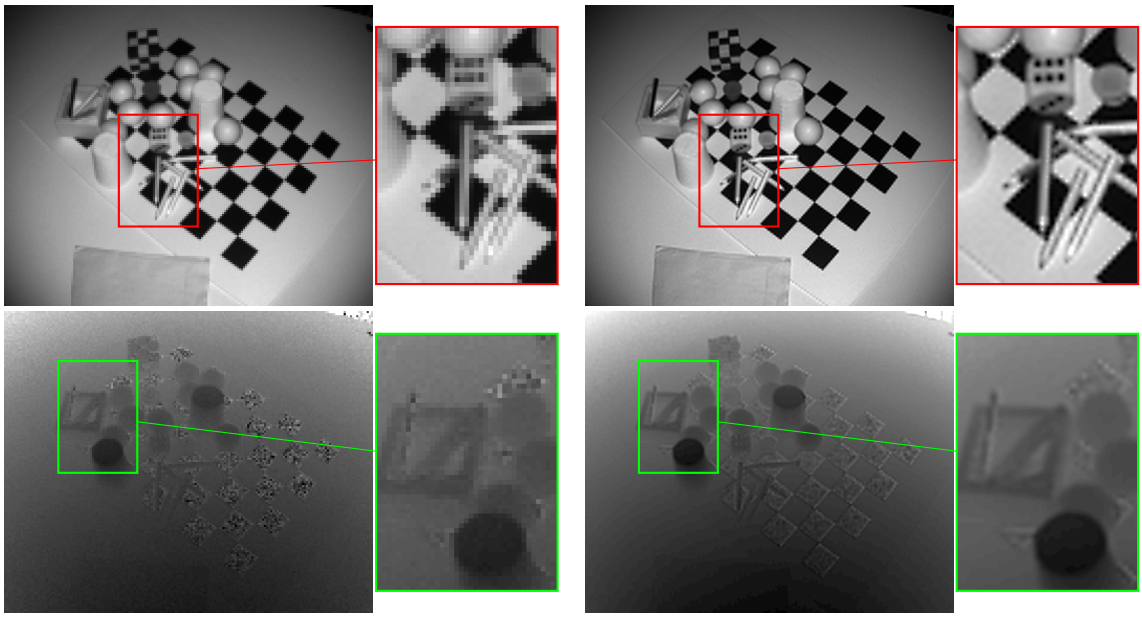
The proposed methodology facilitates numerous vision tasks beyond classical resolution enhancement in color and range imaging. Below, we briefly investigate two further example applications of practical relevance.

Multispectral Image Upsampling. The proposed multi-channel model generalizes to multispectral imaging for $C \gg 3$ channels in a straightforward way as extension of color image processing. Let us consider the application of single-image upsampling, where we target at estimating high-resolution multispectral images from single low-resolution ones[5]. In this highly underdetermined reconstruction problem, we exploit the high degree of correlations among the spectral bands of multispectral images. Figure 6.15 depicts multispectral upsampling on an example image taken from the Harvard dataset [Chak 11]. The multispectral data consists of $C = 31$ bands that correspond to central wavelengths between 420 and 720 nm. Here, we depict a false-color visualization using a self organizing map [Jord 14] (top row) along with a single spectral band centered at 600 nm (bottom row).

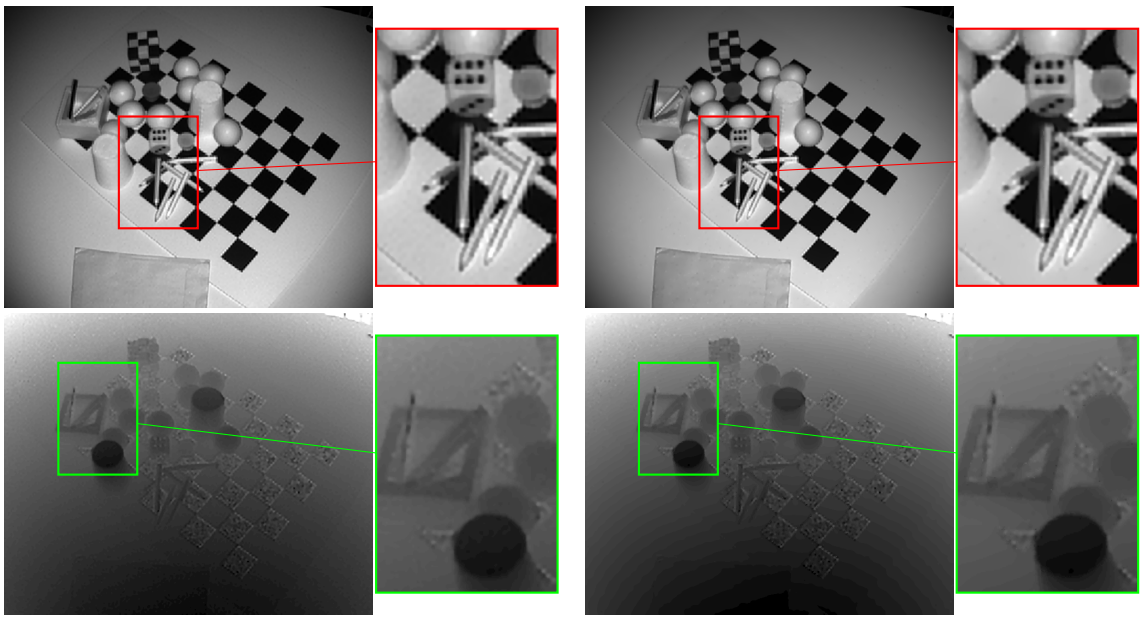
For this application, we employed the LLR prior with a symmetric distribution. The inter-channel parameters were set to $\mu_{ij} = 5 \cdot 10^{-3}$, $\epsilon_{ij} = 10^{-4}$ and $N_{\mathrm{LLR}} = 3$

[4]The data acquisition for this study was done in collaboration with Peter Fürsattel at Metrilus GmbH, Erlangen, Germany.

[5]The reconstruction algorithm can easily be extended to multi-frame reconstruction using motion estimation techniques tailored for multi- and hyperspectral data, see e. g. [Zhan 12a].

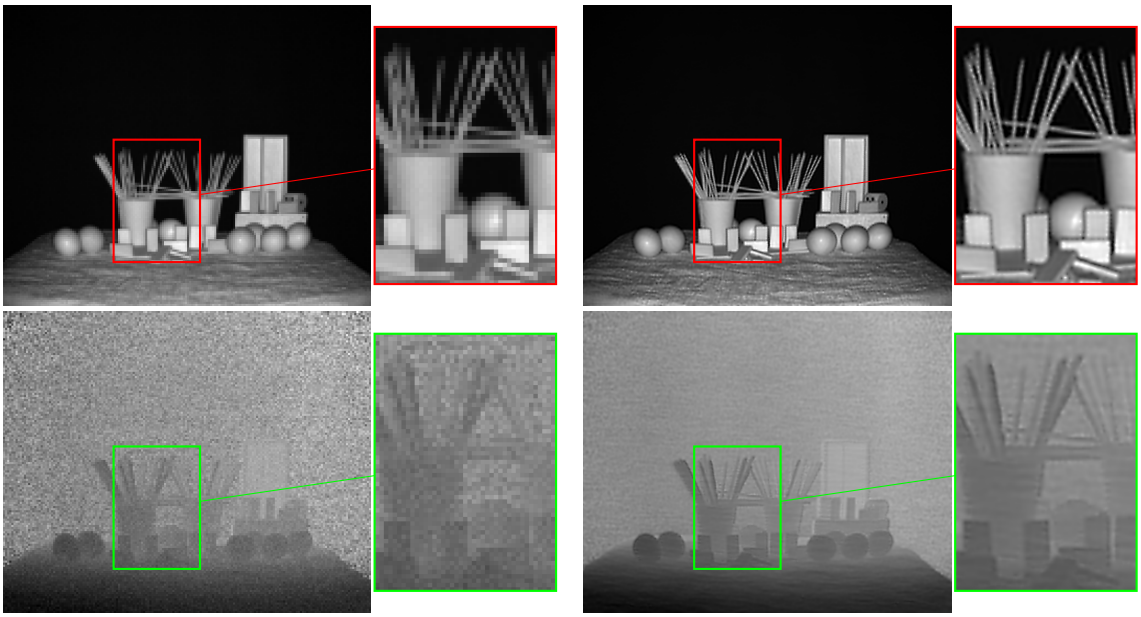


(a) Original **(b)** AMSR



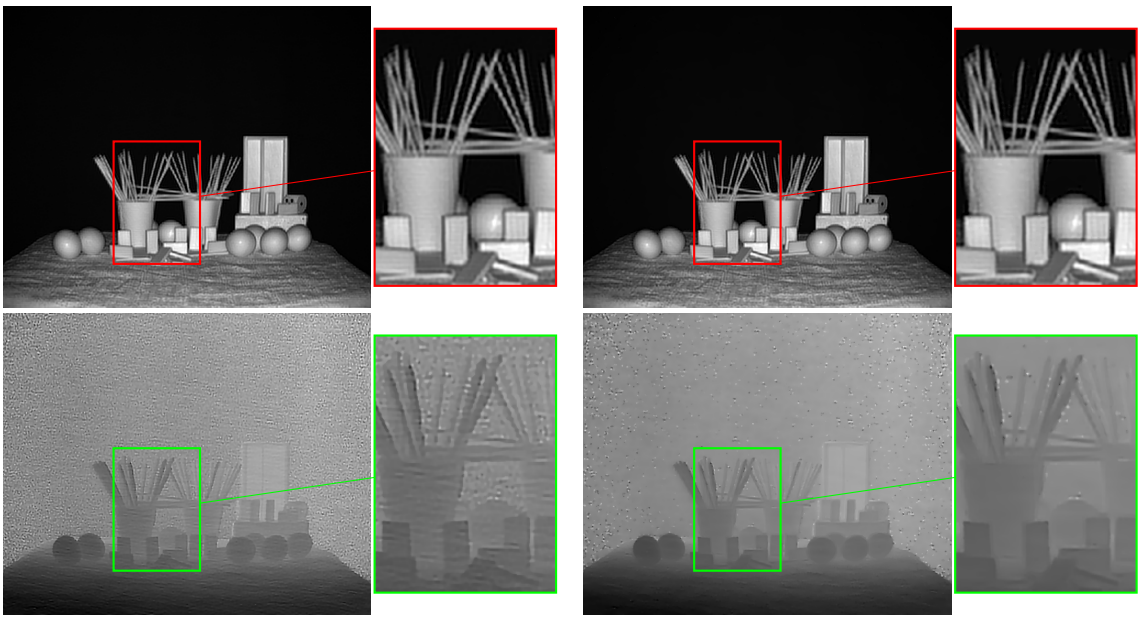
(c) Sequential **(d)** LLR

Figure 6.13: Photogeometric super-resolution on amplitude and range data ($K = 16$ frames, magnification $s = 4$) of the *checkerboard* dataset. The image data was captured with a Mesa SR-4000 ToF camera. (a) Original amplitude (first row) and range data (second row), (b) super-resolved data using the adaptive multi-sensor super-resolution (AMSR) presented in Section 5.5, (c) super-resolved data using sequential processing of both channels and (d) super-resolved data gained by multi-channel processing using the LLR prior.



(a) Original

(b) AMSR



(c) Sequential

(d) LLR

Figure 6.14: Photogeometric super-resolution on amplitude and range data ($K = 21$ frames, magnification $s = 4$) of the *games* dataset. The image data was captured with a Mesa SR-4000 ToF camera. (a) Original amplitude (first row) and range data (second row), (b) super-resolved data using the adaptive multi-sensor super-resolution (AMSR) presented in Section 5.5, (c) super-resolved data using sequential processing of both channels and (d) super-resolved data gained by multi-channel processing using the LLR prior.

(a) Original **(b)** Sequential **(c)** IDP [Fars 06] **(d)** LLR

Figure 6.15: Multispectral image upsampling ($C = 31$ spectral bands, magnification $s = 2$). Top row: False-color visualization for the original image (a), sequential upsampling of the different channels without inter-channel prior (b), multi-channel upsampling using IDP regularization [Fars 06] (c), and multi-channel upsampling using the LLR prior (d). Bottom row: Single spectral band at wavelength 600 nm. Notice that IDP produced unwanted structure copying artifacts in the depicted image region.

for all pairs of channels. The corresponding intra-channel parameters were set to $\lambda_i = 5 \cdot 10^{-4}$, $N_{\mathrm{BTV}} = 1$ and $\alpha_{\mathrm{BTV}} = 0.5$ for all channels. For the sake of comparison to the LLR prior, sequential upsampling of the different channels without inter-channel prior as well as multi-channel upsampling using IDP regularization [Fars 06] extended to an arbitrary number of channels was applied.

In comparison to sequential upsampling, both multi-channel approaches led to a decrease of color artifacts. Similar to color super-resolution, these color artifacts appeared as residual noise due to disregarding mutual dependencies. However, notice that in this example the assumption of strong mutual dependencies was violated for certain pairs of channels due to inconsistent structures. In Fig. 6.15, this assumption was violated for the image region that contains text. This resulted in erroneously copied structure from other, original channels to the upsampled ones in case of IDP regularization. The proposed outlier-insensitive prior based

(a) Original **(b)** Sequential binarization & SR **(c)** Joint binarization & SR

Figure 6.16: Joint binarization and super-resolution on two-tone images ($K = 20$ frames, magnification $s = 3$). (a) - (b) Single grayscale image and the decoupled use of super-resolution (top row) along with the corresponding binarizations modeled as probability maps (bottom row). (c) Joint binarization and super-resolution using the LLR prior. It is worth noting that the joint approach better removed ringing and compression artifacts.

on Tukey's biweight loss features higher robustness against inconsistencies and avoided these structure copying artifacts.

Joint Segmentation and Super-Resolution. One recent trend in image processing is to couple image analysis with retrospective image enhancement via joint frameworks [Shen 07, Lela 15]. In [Kohl 15d], segmentation driven deblurring has been proposed to boost both subtasks compared to their decoupled usage. Following a similar notion, the multi-channel methodology can be adopted to joint segmentation and super-resolution. We can employ consistency among images and a corresponding segmentation as a strong prior to leverage resolution enhancement. This assumption is reasonable for text document images that feature a correlation among intensity values and the appearance of text and symbols.

We demonstrate this concept in Fig. 6.16 for binarization on compressed two-tone frames of the *Adyoron* sequence [Fars 16]. Motion estimation across the sequence ($K = 20$ frames) was performed by ECC optimization [Evan 08] with an

affine model. We employed the LLR prior with symmetric distribution for $C = 2$ channels with $\mu_{ij} = 25$, $\epsilon_{ij} = 10^{-4}$ and $N_{\text{LLR}} = 3$. The intra-channel parameters were set to $\lambda_i = 10^{-3}$ for the grayscale image ($i = 1$) and $\lambda_i = 10^{-1}$ for the binarization ($i = 2$) with $N_{\text{BTV}} = 2$ and $\alpha_{\text{BTV}} = 0.5$ for both channels.

Figure 6.16a depicts a single grayscale image along with its binarization using the intensity-based soft-clustering proposed in [Kohl 15d] for text images. We represent a binarization of an image $\boldsymbol{x} \in \mathbb{R}^N$ as a probability map $\boldsymbol{s} \in [0,1]^N$, where $s_i = 0$ indicates a dark structure, e. g. font, at the i-th pixel. Note that the binarization failed to detect text and symbols accurately on aliased low-resolution data. For the sake of comparison, Fig. 6.16b depicts super-resolution on the grayscale images followed by the binarization in a sequential manner. This served as initialization for the joint approach in Fig. 6.16c that considered grayscale data and its binarization as coupled channels via the LLR prior. Notice that the joint approach better compensated ringing and compression artifacts and reconstructed both channels simultaneously at a super-resolved scale. This can serve as a more reliable basis for subsequent image analysis tasks like text detection and recognition [Espa 11] or writer identification [Chri 15] to name a few.

6.7 Conclusion

This chapter proposed a novel approach to multi-sensor super-resolution that is applicable to a variety of current computer vision applications. As the core idea, we introduced a Bayesian model that accounts for the image formation process of multi-channel images as well as a LLR prior distribution to consider mutual dependencies among different image channels. Subsequently, we developed Bayesian parameter estimation techniques based on this model, where we proposed a joint multi-channel reconstruction algorithm to take inter-channel dependencies into consideration. We presented a thorough analysis of this model including comparisons to related methods, where we discussed connections to several popular image filters. Eventually, we studied several target applications, where the proposed method generalized fairly well and outperformed the state-of-the-art. In color image super-resolution as a classical application, LLR improved the PSNR by 1.5 dB and the SSIM by 0.04 compared to conventional channel-wise super-resolution. Unlike related work, LLR can handle single- and multi-frame resolution enhancement in a unified framework and does neither rely on guidance information nor on a fixed number of image channels. We also presented potential applications beyond classical super-resolution including multispectral image upsampling as well as joint segmentation and super-resolution on two-tone images.

Albeit its flexible and unified formulation, this approach can be further tailored to specific domains. One extension is to optionally augment the prior distribution by a static guidance as proposed for related image filters [Ham 15]. This might further boost super-resolution for applications, where such guidance data is available. Another extension includes the acceleration of the algorithm. In this chapter, we proposed a brute-force approach to exploit mutual dependencies among all pairs of image channels but a suitable dimensionality reduction could enhance the efficiency in case of a very large number of channels.

Part III

Super-Resolution in Medical Imaging

CHAPTER 7

Applications in Retinal Fundus Video Imaging

Over the past years, super-resolution has found its entry into various fields of medical imaging and has been examined for diagnostic or interventional workflows, e. g. in radiology [Gree 08, Robi 10]. This chapter presents new applications of super-resolution in retinal video imaging that has recently emerged as a branch of today's ophthalmic imaging technologies. As the primary contribution, a tailored method to reconstruct high-resolution retinal fundus images from video sequences taken from the human eye background is presented. Super-resolution exploits natural human eye movements that occur during an examination and cause subpixel motion across video frames. As an additional contribution, a novel method to assess noise and sharpness characteristics in fundus images in a fully automatic manner is presented. This quality measure is employed in a new automatic hyperparameter selection scheme for super-resolution reconstruction.

The proposed super-resolution framework has been originally published in [Kohl 14a] and image quality assessment has been first published in [Kohl 13a].

7.1 Introduction and Medical Background

In today's ophthalmology, retinal *fundus imaging* is one of the most frequently used techniques to gain information about the human eye background non-invasively [Patt 06]. The primary scopes of this structural imaging technology are documentation [Abra 10] and computer-aided screening of eye diseases such as diabetic retinopathy [Abra 15] or glaucoma [Bock 10, Kohl 15a], among others. This is mainly because of the cost efficiency and availability of fundus cameras compared to other modalities like optical coherence tomography (OCT) [Huan 91]. The most common approach in clinical practice are single-shot techniques that provide high-resolution color photographs of the human retina. Consequently, they provide static information of the retina. Another trend that has recently emerged is

fundus video imaging to gain dynamic measurements. The main motivation behind the use of video capable cameras is to measure fast temporal changes of the retina, e. g. the cardiac cycle [Torn 15]. Compared to single-shot imaging, video imaging features a high temporal resolution but single frames are limited in terms of their spatial resolution and SNR, see Fig. 7.1. This is mainly caused by technological aspects, e. g. increased light exposure times, but also related to economic reasons as the use of mobile and cost-effective hardware is desirable [Hohe 15].

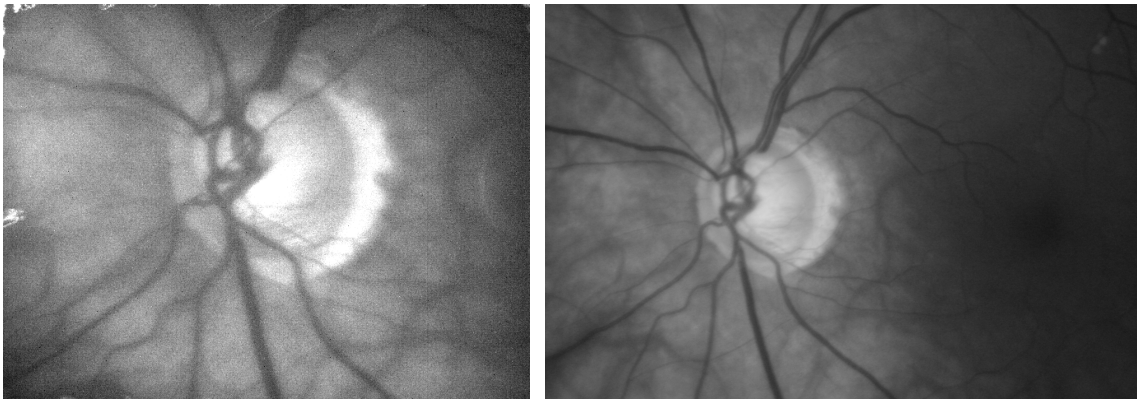
Image enhancement and restoration is an emerging field of research in retinal imaging in order to enhance the diagnostic usability of fundus images. One approach are temporal denoising schemes. In [Kohl 12], Köhler et al. have proposed adaptive temporal averaging using a registration based compensation of natural eye movements that appear during video imaging. This approach recovers single denoised images from a set of noisy frames. Another direction of prior work are blind deconvolution algorithms with the goal to recover a sharp retinal image from a blurred one. These methods can be applied in a multi-frame scheme to image pairs acquired in longitudinal examinations as shown by Marrugo et al. [Marr 11c]. However, one limitation of blind deconvolution is that it does not enhance the spatial resolution in terms of pixel sampling.

This chapter investigates a different direction and aims at improving the diagnostic usability of retinal images by multi-frame super-resolution. The basic idea behind this approach is to utilize natural eye movements [Rayn 98, Duch 07] during video imaging as a cue for super-resolution reconstruction. In [Muri 11], Murillo et al. proposed this idea to enhance scanning laser ophthalmoscopy. In parallel to the work presented in this thesis, single- and multi-frame resolution enhancement for fundus imaging have also been studied by Thapa et al. [Thap 14]. In general, super-resolution applied to retinal images is challenging due to optical aberrations caused by the optics of the human eye. Other common issues are photometric distortions or specular reflections that lead to oversaturations. These effects are related to the external illumination and the imaging through a small pupil, which makes homogeneous illuminations of the retina difficult to achieve. Moreover, fundus video data might be affected by a poor SNR as the light exposure during the examination needs to be low to avoid impairments of the patient. Therefore, robust estimation techniques are required to deal with these conditions.

The remainder of this chapter is structured as follows. Section 7.2 develops a domain-specific image formation model that provides the basis of super-resolution in fundus video imaging. Section 7.3 presents a super-resolution algorithm that employs a novel hyperparameter selection scheme termed *quality self-assessment* to jointly estimate a super-resolved image along with optimal regularization parameters. Section 7.4 presents an experimental evaluation of the proposed framework in fundus video imaging. Finally, Section 7.5 draws a conclusion.

7.2 Image Formation Model for Retinal Imaging

In order to reconstruct high-resolution retinal images, we exploit sequences of low-resolution video frames. For the sake of notational brevity, we limit ourselves in this chapter to monochromatic video data and hence a single frame represents the



(a) (b)

Figure 7.1: Single-shot versus video techniques in retinal imaging. (a) Single frame (640×480 px, 15° field of view (FOV)) acquired from a glaucoma patient using the video camera system developed by Tornow et al. [Torn 15]. (b) Single photograph (1944×1296 px, 22.5° FOV) captured from the same patient with a commercially available Kowa nonmyd fundus camera. For fair comparison to monochromatic video data, the green color channel of the Kowa image is depicted.

luminance of a retinal image. For super-resolution in a Bayesian framework, we need to formulate an image formation model that relates the individual frames to the unknown high-resolution image. In the following subsections, we adopt the model previously presented in Section 3.2 to the conditions in retinal imaging.

7.2.1 Derivation of the Motion Model

One of the most important aspects of the proposed model is the fact that subpixel motion across video frames that are captured during an examination is related movements of the eye relative to a fundus camera. According to Rayner [Rayn 98], such eye movements can be categorized into pursuits, vergence, vestibular motion, and saccades that differ in their causes, amplitudes, and speeds. Pursuit and vergence motion are related to continuously fixating moving targets and nearby, static targets, respectively. Vestibular motion is caused by small motion of the head or the entire body. Saccades refer to abrupt motion with high velocity that occur during the fixation to a new target [Duch 07]. In addition to these movements, notice that the human eye is never completely still when fixating a target due to natural eye motion caused by tremors, drifts and microsaccades [Rayn 98, Duch 07]. Such movements occur randomly and have small amplitudes. As they are unavoidable during an examination, there is no need to induce camera motion by means of mechanical components to enable super-resolution reconstruction.

To describe eye movements mathematically, we follow the model developed by Can et al. [Can 02] that approximates the human retina as a spherical surface. Let us consider two views of the retina with eye movements among them. These views are associated with two video frames $\boldsymbol{x}^{(r)}$ and $\boldsymbol{x}^{(k)}$ captured with one single camera, where $\boldsymbol{x}^{(r)}$ denotes the reference frame. Furthermore, let $\boldsymbol{U}^{(r)} \in \mathbb{R}^3$ and $\boldsymbol{U}^{(k)} \in \mathbb{R}^3$ be the coordinates of a single point on the retina in the 3-D space transformed to

the local camera coordinate systems associated with these frames. According to [Can 02], we obtain $\boldsymbol{U}^{(k)} = (U, V, \pi(U,V))^\top$ by the quadratic surface model:

$$\pi(U,V) = \pi_1 U^2 + \pi_2 V^2 + \pi_3 UV + \pi_4 U + \pi_5 V + \pi_6, \tag{7.1}$$

where the parameter set $\pi = \{\pi_1, \ldots, \pi_6\}$ describes the retina shape. Since the curvature of the retina is negligible compared to its distance to the camera center, we can assume a weak-perspective camera model [Hart 04] that is represented by the projection matrix $\boldsymbol{P} \in \mathbb{R}^{3\times4}$ for both views. Then, $\boldsymbol{U}^{(k)}$ and $\boldsymbol{U}^{(r)}$ are mapped to $\boldsymbol{u}^{(k)} = \boldsymbol{P}\boldsymbol{U}^{(k)}$ and $\boldsymbol{u}^{(r)} = \boldsymbol{P}\boldsymbol{U}^{(r)}$ on the image plane in the two frames. As eye motion can be considered as rigid [Duch 07], the relationship between $\boldsymbol{U}^{(k)}$ and $\boldsymbol{U}^{(r)}$ is given by $\boldsymbol{U}^{(r)} = \boldsymbol{R}\boldsymbol{U}^{(k)} + \boldsymbol{t}$, where $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ and $\boldsymbol{t} \in \mathbb{R}^3$ denote a rotation matrix and a translation vector, respectively. This leads to the relationship between the 3-D point $\boldsymbol{U}^{(r)} = (U', V', \pi(U',V'))^\top$ and its 2-D projection $\boldsymbol{u}^{(k)}$ according to:

$$U' = R_{11} \cdot \frac{s_0 u - c_u}{f_u} + R_{12} \cdot \frac{s_0 v - c_v}{f_v} + R_{13} \cdot \zeta(u,v) + t_u, \tag{7.2}$$

$$V' = R_{21} \cdot \frac{s_0 u - c_u}{f_u} + R_{22} \cdot \frac{s_0 v - c_v}{f_v} + R_{23} \cdot \zeta(u,v) + t_v, \tag{7.3}$$

where f_u and f_v are the focal lengths, $(c_u, c_v)^\top$ is the camera center, and s_0 is a scaling parameter of the weak-perspective camera. $\zeta(u,v)$ is the quadratic equation:

$$\zeta(u,v) = \zeta_1 u^2 + \zeta_2 v^2 + \zeta_3 uv + \zeta_4 u + \zeta_5 v + \zeta_6, \tag{7.4}$$

where the parameters $\zeta = \{\zeta_i, \ldots, \zeta_6\}$ depend on the shape parameters π in Eq. (7.1) and the projection matrix $\boldsymbol{P}$. Based on Eqs. (7.2),(7.3), we can establish the relationship between the 2-D points $\boldsymbol{u}^{(r)} = (u', v')^\top$ and $\boldsymbol{u}^{(k)} = (u, v)^\top$ by the quadratic image-to-image transformation:

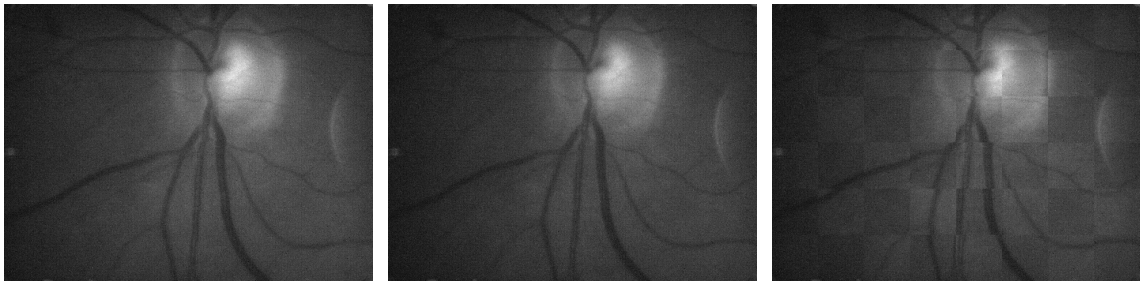
$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_6 \\ \theta_7 & \theta_8 & \theta_9 & \theta_{10} & \theta_{11} & \theta_{12} \end{pmatrix} \begin{pmatrix} u^2 & v^2 & uv & u & v & 1 \end{pmatrix}^\top, \tag{7.5}$$

where the transformation parameters θ_i depend on the shape parameters π, the projection matrix $\boldsymbol{P}$, as well as the eye movement characterized by $\boldsymbol{R}$ and $\boldsymbol{t}$.

In contrast to the derivation in [Can 02] for general types of rigid eye motion, the proposed model considers video imaging with high frame rates. For super-resolution, we exploit miniature movements when fixating static targets over short time intervals, i. e. tremors and microsaccades [Rayn 98], but neglect movements with larger amplitudes over longer intervals. As this motion is small compared to the FOV (see Fig. 7.2), we assume that $\boldsymbol{t} \approx \boldsymbol{0}$ and $\boldsymbol{R} \approx \boldsymbol{I}$. Therefore, we trade the accuracy of the quadratic transformation in Eq. (7.5) against improved robustness of parameter estimation using the affine transformation [Fang 06]:

$$\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_4 & \theta_5 & \theta_6 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}. \tag{7.6}$$

This homography relates eye motion to translation, rotation, scaling and shearing on the image plane. These effects are described by six degrees of freedom given by $\Theta = \{\theta_1, \ldots, \theta_6\}$ (see Section 3.2.2) as opposed to the twelve degrees of freedom in Eq. (7.5). Notice that related eye motion models are based on globally rigid homographies [Kola 15, Kola 16], which is a specialization of the affine model.



(a) Frame 1 **(b)** Frame 12 **(c)** Checkerboard overlay

Figure 7.2: Illustration of natural, miniature eye movements in retinal fundus video imaging. (a) - (b) Two frames (frames 1 and 12) with eye movements captured at a frame rate of 25 Hz. (c) Checkerboard visualization for both frames. Notice that miniature movements can be perceived in the checkerboard visualization but are small compared to the FOV.

7.2.2 Derivation of the Photometric Model

In practice, intensity changes among fundus video frames are not exclusively related to eye motion. Another effect that needs to be modeled are photometric variations, which can be distinguished in spatial and temporal ones, see Fig. 7.3. Spatial variations are caused by illumination inhomogeneities within a single image [Marr 11a]. Such variations occur due to the curved shape of the retina, which makes it difficult to achieve homogenous illumination conditions at the image center and in peripheral regions. Furthermore, the illumination depends on the eye anatomy and the presence of diseases. Temporal variations are caused by brightness changes across multiple frames, which can be caused by eye movements or pulsatile changes [Torn 15]. Since both types of variations are unavoidable in general, photometric registration is required to compensate for them. This can be achieved by retrospective correction methods [Kola 11, Zhen 12] that are commonly applied as a preprocessing step in retinal image restoration [Marr 11c].

In this work, photometric variations are corrected jointly to super-resolution instead of correcting the input frames in a preprocessing step. The proposed model extends the global photometric model introduced by Capel and Zisserman [Cape 03]. This model describes the relation between an image $\boldsymbol{x}^{(k)}$ with photometric variations and a reference image $\boldsymbol{x}$ according to:

$$\boldsymbol{x}^{(k)} = \boldsymbol{\gamma}_m^{(k)} \odot \boldsymbol{x} + \gamma_a^{(k)} \boldsymbol{1}. \tag{7.7}$$

$\boldsymbol{\gamma}_m \in \mathbb{R}^N$ denotes a *bias field* associated with $\boldsymbol{x} \in \mathbb{R}^N$ to consider uneven illumination of the retina relative to the reference, which is formulated pixel-wise in a multiplicative model. $\gamma_a \in \mathbb{R}$ denotes a global brightness offset to describe intensity variations over time. Note that in this formulation, the set of photometric parameters $\Gamma = \{\boldsymbol{\gamma}_m, \gamma_a\}$ is defined in the domain of high-resolution data.

7.2.3 Joint Photogeometric and Sampling Model

The formation of low-resolution video from a high-resolution image is described by combining the motion model in Eq. (7.6) with the photometric one in Eq. (7.7).

(a) Frame 1 **(b)** Frame 18 **(c)** Checkerboard overlay

Figure 7.3: Illustration of spatial and temporal photometric variations in fundus video imaging. (a) - (b) Two frames (frames 1 and 18) with eye movements and varying photometric conditions. (c) Corresponding checkerboard visualization. Notice the temporal brightness changes and the illumination inhomogeneities within both frames.

To describe this process in a physically appropriate manner, it is assumed that the geometric transformation related to eye movements takes place as the first operation, followed by the photometric transformation and the sampling onto the domain of the low-resolution data. We limit ourselves to space and time invariant modeling of the sampling process as a reasonable assumption for retinal images captured over a short time period with a high frame rate and a small FOV. Hence, the formation of the k-th low-resolution frame is described by:

$$\boldsymbol{y}^{(k)} = \boldsymbol{D}\boldsymbol{H}\left(\gamma_m^{(k)} \odot \left(\boldsymbol{M}^{(k)}\boldsymbol{x}\right) + \gamma_a^{(k)}\boldsymbol{1}\right) + \epsilon^{(k)}, \tag{7.8}$$

where $\boldsymbol{M}^{(k)}$ denotes the subpixel motion related to eye movements, $\gamma_m^{(k)}$ and $\gamma_a^{(k)}$ are the photometric parameters, and $\epsilon^{(k)}$ denotes additive noise for the k-th frame. $\boldsymbol{D}$ and $\boldsymbol{H}$ are time invariant and describe subsampling and a LSI blur kernel. The latter approximates the superposition of the camera PSF and unavoidable optical aberrations in the human eye [Marr 11c].

In Eq. (7.8), photometric variations are modeled as atmospheric effects in the domain of high-resolution data. However, under spatially smooth variations as a common assumption of retrospective illumination correction [Hou 06], we can simplify Eq. (7.8) using the approximations[1]:

$$\boldsymbol{D}\boldsymbol{H}\left(\gamma_m^{(k)} \odot \boldsymbol{M}^{(k)}\boldsymbol{x}\right) \approx \boldsymbol{D}\gamma_m^{(k)} \odot \boldsymbol{D}\boldsymbol{H}\boldsymbol{M}^{(k)}\boldsymbol{x}, \tag{7.9}$$

$$\gamma_a^{(k)}\boldsymbol{D}\boldsymbol{H}\boldsymbol{1} \approx \gamma_a^{(k)}\boldsymbol{D}\boldsymbol{1}. \tag{7.10}$$

Based on these approximations, Eq. (7.8) can be rewritten to explain the formation of low-resolution frames from a high-resolution image according to:

$$\boldsymbol{y}^{(k)} = \gamma_m^{(k)} \odot \boldsymbol{W}^{(k)}\boldsymbol{x} + \gamma_a^{(k)}\boldsymbol{1} + \epsilon^{(k)}, \tag{7.11}$$

where the system matrix $\boldsymbol{W}^{(k)}$ is assembled element-wise from the blur kernel and motion parameters (see Eq. (3.20)). Notice that in Eq. (7.11) the photometric parameters $\gamma_m^{(k)}$ and $\gamma_a^{(k)}$ are defined in the domain of low-resolution frames.

[1]Similar approximations regarding the photometric parameters have also been proposed for related image formation models in the field of blind deconvolution [Marr 11c].

7.3 Super-Resolution with Quality Self-Assessment

The proposed framework aims at reconstructing an eye movement compensated high-resolution fundus image from multiple low-resolution video frames while simultaneously compensating photometric variations. The algorithm developed in this section is divided into a registration and a reconstruction stage as follows:

1. Given a sequence of low-resolution frames, a *photogeometric registration* is accomplished to estimate the latent eye motion as well as the photometric parameters that describe the image formation process according to Eq. (7.11).

2. Given the estimate of the photogeometric model, a high-resolution image is reconstructed by MAP estimation from the low-resolution frames.

First, this section presents photogeometric registration for fundus videos employed in the initial registration stage. Subsequently, an iterative optimization scheme is developed referred to as super-resolution with quality self-assessment to jointly estimate latent Bayesian hyperparameters along with the high-resolution image. Eventually, a tailored quality measure for a fully automatic assessment of image noise and sharpness within quality self-assessment is proposed.

7.3.1 Photogeometric Registration Algorithm

Photogeometric registration is accomplished in two steps. First, the photometric parameters that are related to spatial and temporal illumination variations are estimated from the observed, low-resolution frames. Second, the geometric parameters that describe eye movements are determined under consideration of the photometric parameters.

Photometric Registration. Given the set of low-resolution frames $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}$, the multiplicative bias fields that describe spatial photometric variations are estimated for each frame separately. Following state-of-the-art retrospective illumination correction techniques [Hou 06], a bias field is assumed to be a spatially smooth signal. In this chapter, a parametric approach is employed that represents a bias field $\boldsymbol{\gamma}_m^{(k)}$ by the superposition of spatially smooth basis functions as a B-spline surface [Kola 11]. Hence, $\boldsymbol{\gamma}_m^{(k)}$ is computed by B-spline fitting of the intensities in the corresponding low-resolution frame $\boldsymbol{y}^{(k)}$, which implicitly enforces the smoothness condition.

The brightness offsets $\gamma_a^{(k)}$ that describe temporal photometric variations are determined by pair-wise registration. This is done in a robust manner by computing the median temporal brightness of each frame $\boldsymbol{y}^{(k)}$ relative to the reference $\boldsymbol{y}^{(r)}$ under the estimated bias fields for both frames according to:

$$\gamma_a^{(k)} = \operatorname{median}\left(\left(\boldsymbol{\gamma}_m^{(k)}\right)^{-1} \odot \boldsymbol{y}^{(k)}\right) - \operatorname{median}\left(\left(\boldsymbol{\gamma}_m^{(r)}\right)^{-1} \odot \boldsymbol{y}^{(r)}\right), \tag{7.12}$$

where $(\boldsymbol{\gamma}_m^{(k)})^{-1}$ and $(\boldsymbol{\gamma}_m^{(r)})^{-1}$ are the pixel-wise inverted bias fields of the k-th frame and the reference frame, respectively.

Geometric Registration. Once the photometric parameters are determined, the geometric parameters that describe eye movements by the affine image-to-image homography in (7.6) are obtained by pair-wise registration. For the k-th low-resolution frame $\boldsymbol{y}^{(k)}$, these motion parameters are determined as the solution of the intensity-based registration problem:

$$\Theta^{(k)} = \underset{\Theta}{\operatorname{argmax}}\, \rho\left(\mathcal{M}_\Theta\left\{\left(\gamma_m^{(k)}\right)^{-1} \odot \boldsymbol{y}^{(k)} - \gamma_a^{(k)}\mathbf{1}\right\}, \left(\gamma_m^{(r)}\right)^{-1} \odot \boldsymbol{y}^{(r)}\right), \tag{7.13}$$

where $\mathcal{M}_\Theta\{\cdot\}$ denotes image warping towards the reference frame according to the motion parameters Θ and $\rho : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_0^+$ is an image similarity measure.

It is worth noting that Eq. (7.13) compensates for photometric variations using the photometric parameters in the similarity measure. In order to enhance the robustness of the geometric registration regarding residual variations, the normalized cross correlation is used for the similarity measure. This pair-wise registration is implemented by ECC optimization [Evan 08], which iteratively solves Eq. (7.13) for the motion parameters in a coarse-to-fine scheme.

7.3.2 Super-Resolution Reconstruction Algorithm

Given the low-resolution observations $\boldsymbol{y}$ and the photogeometric registration parameters $\{\Theta^{(k)}, \Gamma^{(k)}\}_{k=1}^K$, the high-resolution image $\boldsymbol{x}$ is inferred according to the MAP estimation:

$$\boldsymbol{x}_{\mathrm{MAP}} = \underset{\boldsymbol{x}}{\operatorname{argmax}}\, p\left(\boldsymbol{y} \,|\, \boldsymbol{x}, \{\Theta^{(k)}, \Gamma^{(k)}\}_{k=1}^K\right) p(\boldsymbol{x}), \tag{7.14}$$

where $p(\boldsymbol{y} \,|\, \boldsymbol{x}, \{\Theta^{(k)}, \Gamma^{(k)}\}_{k=1}^K)$ denotes the observation model related to Eq. (7.11). For the prior distribution, we assign the exponential form $p(\boldsymbol{x}) \propto \exp(-\lambda R(\boldsymbol{x}))$ with regularization term $R(\boldsymbol{x})$ and regularization weight $\lambda \geq 0$.

Notice that MAP estimation in Eq. (7.14) requires prior knowledge regarding the regularization weight λ. However, in the desired application, the choice of this parameter strongly depends on the imaging conditions. Thus, there might be a considerable variance regarding the optimal parameter for video data of different subjects, which makes its off-line selection on training data difficult. To this end, the regularization weight is treated as a latent hyperparameter. In contrast to *data* driven hyperparameter selection (see Section 4.4), the regularization weight is inferred in a *quality* driven way termed image quality self-assessment. This optimization scheme employs a quality measure that quantifies image noise and sharpness. This enables a fully automatic parameter selection and provides an objective quality measure for super-resolution as a by-product of the optimization algorithm.

The estimations of the high-resolution image and the optimal regularization weight are treated as two coupled subproblems. Our goal is to infer the regularization weight λ according to:

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmax}}\, Q\big(\boldsymbol{x}(\lambda)\big), \tag{7.15}$$

Algorithm 7.1 Super-resolution with image quality self-assessment

Input: Initial guess for image $\boldsymbol{x}$ and regularization weight search range $[\log \lambda_l, \log \lambda_u]$
Output: Final high-resolution image $\boldsymbol{x}$ with optimal regularization weight $\hat{\lambda}$

$\lambda \leftarrow \lambda_l$ and $Q_{\max} \leftarrow 0$
while $\lambda \leq \lambda_u$ **do**
  for $t = 1, \ldots, T_{\text{scg}}$ **do**
    Update $\boldsymbol{x}$ by SCG iteration for Eq. (7.16) with current λ
  end for
  if $Q(\boldsymbol{x}) > Q_{\max}$ **then**
    $\hat{\lambda} \leftarrow \lambda$ and $Q_{\max} \leftarrow Q(\boldsymbol{x})$
  end if
  $\lambda \leftarrow 10^{\log \lambda + \Delta\lambda}$
end while
while SCG convergence criterion not fulfilled **do**
  Update $\boldsymbol{x}$ by SCG iteration for Eq. (7.16) with $\lambda = \hat{\lambda}$
end while

where $Q : \mathbb{R}^N \to \mathbb{R}_0^+$ denotes a *no-reference* image quality measure that quantifies the level of noise and sharpness for a given image. Note that higher measures $Q(\boldsymbol{x})$ indicates a favorable image quality. Given the regularization weight λ, we denote by $\boldsymbol{x}(\lambda)$ the image reconstructed under this parameter according to the minimization:

$$\boldsymbol{x}(\lambda) = \operatorname*{argmin}_{\boldsymbol{x}} \left\{ L(\boldsymbol{x}) + \lambda R(\boldsymbol{x}) \right\}, \tag{7.16}$$

where:

$$L(\boldsymbol{x}) = \sum_{k=1}^{K} \phi_{\text{data}} \left(\boldsymbol{y}^{(k)} - \boldsymbol{\gamma}_m^{(k)} \odot \boldsymbol{W}^{(k)} \boldsymbol{x} - \gamma_a^{(k)} \boldsymbol{1} \right), \tag{7.17}$$

and $\phi_{\text{data}} : \mathbb{R}^M \to \mathbb{R}_0^+$ is a loss function related to the underlying noise model. It is worth noting that this optimization scheme is independent on the implementations of the observation and prior model. Hence, we omit their definitions in this general derivation.

For the joint estimation of the optimal regularization weight and the high-resolution image with consideration of their interdependence, the proposed algorithm nests the solution of Eq. (7.15) and Eq. (7.16), see Algorithm 7.1. In order to determine the regularization weight, Eq. (7.15) is approximated by a one-dimensional discrete search. The reconstruction of the desired high-resolution image in Eq. (7.16) is accomplished by T_{scg} iterations of SCG optimization. This seeks the stationary point:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}) + \lambda \nabla_{\boldsymbol{x}} R(\boldsymbol{x}) = \boldsymbol{0}, \tag{7.18}$$

where the gradient of the data fidelity term is given by:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}) = \sum_{k=1}^{K} \boldsymbol{\gamma}_m^{(k)} \odot \boldsymbol{W}^{(k)} \psi_{\text{data}} \left(\boldsymbol{y}^{(k)} - \boldsymbol{\gamma}_m^{(k)} \odot \boldsymbol{W}^{(k)} \boldsymbol{x} - \gamma_a^{(k)} \boldsymbol{1} \right), \tag{7.19}$$

and $\psi_{\text{data}}(\boldsymbol{z}) = \nabla_{\boldsymbol{x}} \phi_{\text{data}}(\boldsymbol{z})$ denotes the gradient of $\phi_{\text{data}}(\boldsymbol{z})$. For this optimization scheme, the initial guess for the high-resolution image is obtained by the temporal median of the geometrically and photometrically registered low-resolution

frames followed by bicubic interpolation according to the desired magnification factor. The regularization weight is initialized by the log-transformed search range $[\log \lambda_l, \log \lambda_u]$. Once the optimal regularization weight $\hat{\lambda}$ under the given quality measure is determined, the high-resolution image $\boldsymbol{x}(\hat{\lambda})$ is refined by SCG iterations for Eq. (7.16) until convergence.

7.3.3 No-Reference Quality Measure for Retinal Imaging

Quality self-assessment requires a reliable measure of image quality. In general, such measures can be divided into classification based and continuous scores. The classification based approach aims at predicting a discrete quality measure from discriminative image features. In the simplest case, this is restricted to two classes to discriminate low-quality images from good ones. In retinal imaging, this can be achieved by supervised learning using features of diagnostic significance [Niem 06, Paul 10]. However, these classification methods are not well suited for continuous optimization within the proposed quality self-assessment scheme.

In contrast to supervised learning, continuous quality measures are inferred from low-level features in an unsupervised way such that the resulting measure correlates with the human visual perception. Some well known features are the image entropy [Gaba 07], spatial and spectral properties [Vu 12], or the image gradient [Zhu 10]. Such features have previously been used for retinal image analysis [Marr 11b]. In this work, we focus on continuous measures that are applicable for quality self-assessment.

Derivation of the Quality Measure. The quality measure that is employed in this work is based on the *coherence* feature proposed by Zhu et al. [Zhu 10] for natural images that has been later adopted to retinal imaging by Köhler et al. [Kohl 13a]. To derive the measure for a given image $\boldsymbol{x}$, we decompose $\boldsymbol{x}$ in disjoint $N_p \times N_p$ patches $\boldsymbol{p} \in \mathbb{R}^{N_p^2}$ with domain $\Omega(\boldsymbol{p}) \subset \mathbb{R}^2$. We aim at quantifying noise and sharpness based on two features that are related to the image gradient and the curvature.

In terms of the gradient information as proposed in [Zhu 10], a local gradient matrix is constructed for the patch $\boldsymbol{p}$ according to:

$$\boldsymbol{G}(\boldsymbol{p}) = \begin{pmatrix} [\boldsymbol{Q}_u \boldsymbol{p}]_1 & [\boldsymbol{Q}_v \boldsymbol{p}]_1 \\ \vdots & \vdots \\ [\boldsymbol{Q}_u \boldsymbol{p}]_{N_p^2} & [\boldsymbol{Q}_v \boldsymbol{p}]_{N_p^2} \end{pmatrix}, \tag{7.20}$$

where $\boldsymbol{Q}_u$ and $\boldsymbol{Q}_v$ denote discrete derivative filters for the coordinate directions u and v, respectively. For this local gradient matrix, we calculate the singular value decomposition (SVD):

$$\boldsymbol{G}(\boldsymbol{p}) = \boldsymbol{U}(\boldsymbol{p}) \begin{pmatrix} s_1(\boldsymbol{p}) & 0 \\ 0 & s_2(\boldsymbol{p}) \end{pmatrix} \boldsymbol{V}(\boldsymbol{p})^\top, \tag{7.21}$$

where $\boldsymbol{U}(\boldsymbol{p})$ and $\boldsymbol{V}(\boldsymbol{p})$ are orthogonal matrices, and $s_1(\boldsymbol{p})$ and $s_2(\boldsymbol{p})$ denote the singular values of the gradient matrix associated with the patch $\boldsymbol{p}$. In the noise and sharpness measure presented below, the singular values are used as basic features.

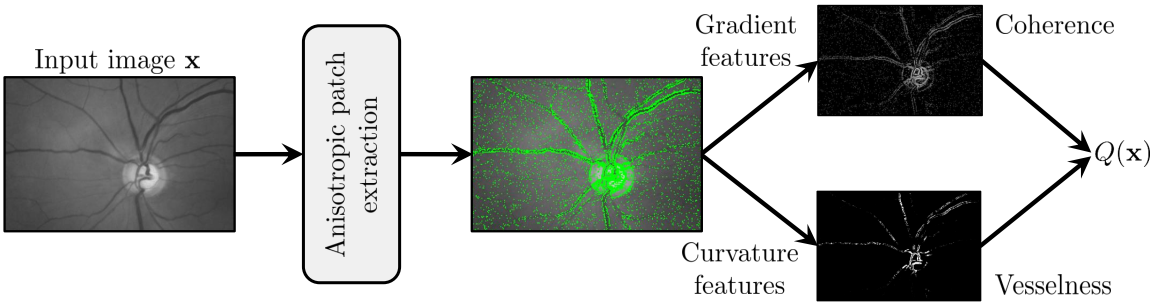


Figure 7.4: Computation of the no-reference quality measure $Q(\boldsymbol{x})$ for retinal fundus images. For the given input image $\boldsymbol{x}$, anisotropic patches are detected. Afterwards, $Q(\boldsymbol{x})$ is determined on a patch level from the coherence and the vesselness features that are obtained from the image gradient and the curvature, respectively.

In terms of the curvature information as proposed in [Kohl 13a], we compute the Hessian matrix in a pixel-wise manner according to:

$$\boldsymbol{H}_i(\sigma_j) = \begin{pmatrix} \left[\boldsymbol{Q}_{uu}(\sigma_j)\boldsymbol{x}\right]_i & \left[\boldsymbol{Q}_{uv}(\sigma_j)\boldsymbol{x}\right]_i \\ \left[\boldsymbol{Q}_{uv}(\sigma_j)\boldsymbol{x}\right]_i & \left[\boldsymbol{Q}_{vv}(\sigma_j)\boldsymbol{x}\right]_i \end{pmatrix}, \tag{7.22}$$

where $\boldsymbol{Q}_{uu}(\sigma_j)$, $\boldsymbol{Q}_{vv}(\sigma_j)$ and $\boldsymbol{Q}_{uv}(\sigma_j)$ denote discrete Laplacian of Gaussian filters with the kernel standard deviation σ_j for the coordinate directions u and v. The Hessian is employed to determine the *vesselness*, which represents a probability map that enables the detection of tubular structures. In retinal imaging, the vesselness provides a blood vessel detection and image quality assessment is steered by the detected vessel tree. The vesselness filter used in this work is based on the approach of Frangi et al. [Fran 98] to detect dark tubular structures[2] according to:

$$V_i(\sigma_j) = \begin{cases} \exp\left(-\frac{1}{2V_\beta^2}\frac{\lambda_{1,i}(\sigma_j)^2}{\lambda_{2,i}(\sigma_j)^2}\right)\left(1-\exp\left(-\frac{\lambda_{1,i}(\sigma_j)^2+\lambda_{2,i}(\sigma_j)^2}{2V_c^2}\right)\right) & \text{if } \lambda_{1,i}(\sigma_j) \geq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{7.23}$$

where $\lambda_{1,i}(\sigma_j)$ and $\lambda_{2,i}(\sigma_j)$ with $|\lambda_{1,i}(\sigma_j)| \leq |\lambda_{2,i}(\sigma_j)|$ are the eigenvalues of the Hessian at the i-th pixel associated with the kernel standard deviation σ_j. The parameters V_β and V_c are thresholds to control the vesselness filter response. The filter responses over a set of N_σ kernel standard deviations are used to compute the local variance of the vesselness in the patch $\boldsymbol{p}$ according to:

$$V(\boldsymbol{p}) = \frac{1}{N_p^2}\sum_{i\in\Omega(\boldsymbol{p})}\left(V_i^* - \frac{1}{N_p^2}\sum_{i\in\Omega(\boldsymbol{p})} V_i^*\right)^2, \tag{7.24}$$

$$V_i^* = \max_{j=1,\ldots,N_\sigma} V_i(\sigma_j). \tag{7.25}$$

[2]Without loss of generality, we limit our consideration to dark tubular structures as these structures correspond to blood vessels in fundus images.

The proposed quality measure combines both feature types, i. e. the gradient information and the vesselness, to assess the level of noise and sharpness in a given image, see Fig. 7.4. In principle, noise and sharpness in a patch $\boldsymbol{p}_i$ are characterized by the singular values $s_1(\boldsymbol{p}_i)$ and $s_2(\boldsymbol{p}_i)$ of the local gradient matrix [Zhu 10]. The local variance $V(\boldsymbol{p}_i)$ is used to guide this measurement based on the hypothesis that patches $\boldsymbol{p}_i$ located on the boundaries of tubular structures, i. e. blood vessels, should have a higher contribution to the local quality measure. We compute the local quality $q(\boldsymbol{p}_i)$ associated with the patch $\boldsymbol{p}_i$ according to:

$$q(\boldsymbol{p}_i) = V(\boldsymbol{p}_i) s_1(\boldsymbol{p}_i) c(\boldsymbol{p}_i), \tag{7.26}$$

where $c(\boldsymbol{p}_i)$ denotes the coherence that is computed from the singular values:

$$c(\boldsymbol{p}_i) = \frac{s_1(\boldsymbol{p}_i) - s_2(\boldsymbol{p}_i)}{s_1(\boldsymbol{p}_i) + s_2(\boldsymbol{p}_i)}. \tag{7.27}$$

Then, the global quality measure $Q(\boldsymbol{x})$ for the entire image is given by:

$$Q(\boldsymbol{x}) = \sum_{\boldsymbol{p}_i \in \mathcal{A}(\boldsymbol{x})} q(\boldsymbol{p}_i), \tag{7.28}$$

where $\mathcal{A}(\boldsymbol{x})$ denotes a set of *anisotropic* patches. These anisotropic patches are characterized by a dominant orientation of the image gradient and are meaningful to characterize noise and sharpness. In accordance to [Zhu 10], these patches are detected automatically by statistical significance testing of the local coherence $c(\boldsymbol{p}_i)$. This leads to the thresholding procedure:

$$\mathcal{A}(\boldsymbol{x}) = \{\boldsymbol{p}_i \,:\, c(\boldsymbol{p}_i) \geq \tau_c\}, \tag{7.29}$$

$$\tau_c = \sqrt{\left(1 - \alpha_c^{\frac{1}{N_p^2 - 1}}\right)\left(1 + \alpha_c^{\frac{1}{N_p^2 - 1}}\right)^{-1}}, \tag{7.30}$$

with threshold τ_c that is determined from the significance level α_c.

Correlation to Full-Reference Quality Assessment. Let us now investigate the validity of the proposed measure for image quality self-assessment. For this purpose, the agreement of the no-reference quality measure $Q(\boldsymbol{x})$ to full-reference quality assessment is studied on simulated data. Figure 7.5 shows the progress of the no-reference measure over the search range of the unknown regularization parameter λ in Algorithm 7.1 averaged over 40 simulated fundus image sequences. In addition, the PSNR for the super-resolved images associated with the different parameter settings is depicted as an example full-reference measure. The relationship between both measures confirms a reasonable agreement between no-reference and full-reference assessment. Note that both measures result in comparable solutions in terms of the optimal regularization weight. Furthermore, a Spearman rank correlation of 0.70 ± 0.34 averaged over all simulated datasets indicates a reasonable correlation between both measures. This validates the proposed no-reference measure as a surrogate for full-reference quality assessment in the absence of ground truth data. For a comprehensive evaluation of the no-reference measure in retinal fundus imaging and comparisons to other state-of-the-art methods, we refer to [Kohl 13a].

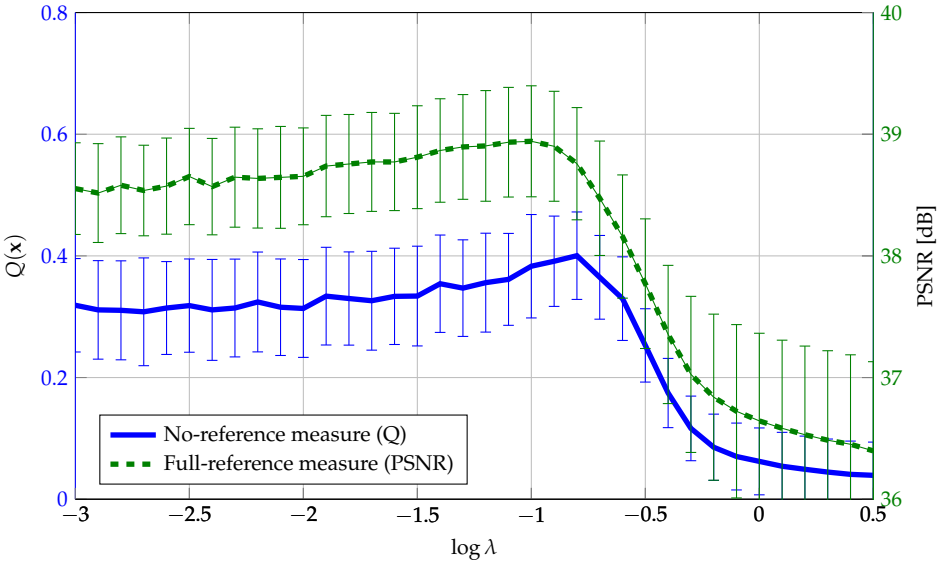


Figure 7.5: Correlation analysis between no-reference and full-reference quality assessment. Blue, solid line: mean ± standard deviation of the no-reference measure $Q(\boldsymbol{x})$ used for quality self-assessment versus the regularization weight λ on 40 image sequences. Green, dotted line: mean ± standard deviation of the PSNR relative to the ground truth. Both measures reach their optimal value within the range $-1.0 \leq \log\lambda \leq -0.8$. The Spearman rank correlation between both measures is 0.70 ± 0.34.

7.4 Experiments and Results

The experimental evaluation for the proposed framework is divided into three parts. In the first part, super-resolution is quantitatively evaluated on simulated fundus images to investigate the potential of the proposed framework in retinal imaging. The second part addresses real data experiments with the target to gain high-resolution fundus images from low-resolution video sequences acquired with a low-cost camera. The third part presents *super-resolved mosaicing* [Kohl 16a] as a novel application of super-resolution in ophthalmic imaging workflows.

Throughout all experiments, super-resolution was applied with the L_1 norm error model and a BTV prior with $N_{\mathrm{BTV}} = 1$ and $\alpha_{\mathrm{BTV}} = 0.4$. The regularization weight selection was performed in the search range given by $\log\lambda_l = -3.0$ and $\log\lambda_u = 0$ with $\Delta\log\lambda = 0.15$ and $T_{\mathrm{scg}} = 50$ SCG iterations. Quality assessment was performed with patch size $N_p = 8$ and significance level $\alpha_c = 0.001$ to detect anisotropic patches. The vesselness filter parameters were set to $V_\beta = 0.5$ and $V_c = 15$ for $N_\sigma = 4$ different filter standard deviations $\sigma_i \in \{1, 3, 5, 8\}$.

7.4.1 Experiments on Simulated Fundus Images

For the sake of a quantitative evaluation, simulated images generated from the DRIVE database [Staa 05] were used. For this task, excerpts of 40 reference images of size 360×360 px served as ground truth data. The green color channels were used to generate sequences of $K = 15$ monochromatic frames of size 120×120 px from the reference color images. Eye movements were simulated by uniformly

Table 7.1: Performance of super-resolution on the DRIVE database [Staa 05]. The PSNR and SSIM statistics were determined for 40 simulated images relative to ground truth data. The sensitivity and specificity statistics were determined for automatic vessel segmentation [Buda 13] relative to a gold standard. All measures were evaluated for low-resolution data, the initial guess for super-resolution, and the final super-resolved image.

	Original	SR (initial)	SR (final)	Ground truth
PSNR [dB]	35.19 ± 1.07	36.65 ± 1.55	$\mathbf{38.64 \pm 1.00}$	-
SSIM	0.84 ± 0.01	0.89 ± 0.02	$\mathbf{0.91 \pm 0.01}$	-
Sensitivity [%]	59.00 ± 6.08	66.66 ± 4.95	69.41 ± 5.49	$\mathbf{74.96 \pm 5.87}$
Specificity [%]	93.13 ± 1.26	$\mathbf{95.04 \pm 1.02}$	94.44 ± 1.27	94.48 ± 1.16

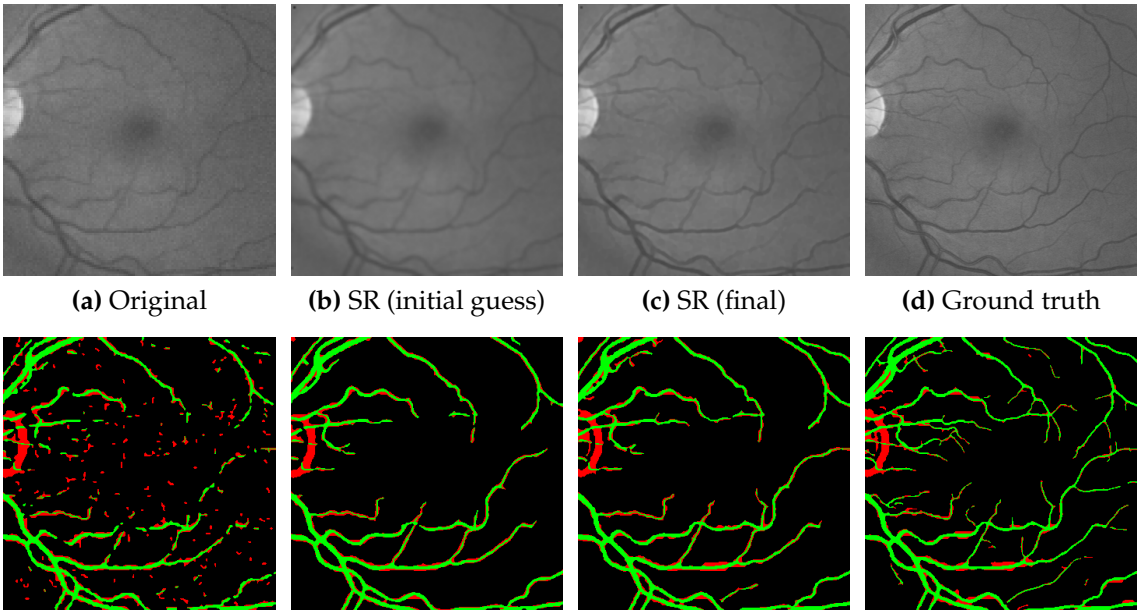
distributed inter-frame translations (-4 to $+4$ px) and rotation angles ($-1°$ to $+1°$) relative to a the first frame. The formation of each frame considered a Gaussian PSF ($\sigma_{\text{PSF}} = 0.5$) and additive, zero-mean Gaussian noise ($\sigma_{\text{noise}} = 0.01$).

The impact of super-resolution was quantitatively assessed by four evaluation measures. On the one hand, PSNR and SSIM were used to assess the fidelity of a reconstruction relative to the ground truth. On the other hand, super-resolution was studied in combination with automatic blood vessel segmentation. This was done by applying the proposed framework as preprocessing for the state-of-the-art segmentation method introduced by Budai et al. [Buda 13]. Super-resolution was assessed by analyzing the sensitivity and specificity of the automatic segmentation relative to a gold standard segmentation provided by a human expert. The statistics of these measures are summarized in Tab. 7.1 for simulated low-resolution data, the initial guess for the iterative super-resolution algorithm obtained by temporal median filtering as well as the final super-resolved image. On average, super-resolution improved the PSNR by 3.5 dB and the SSIM by 0.07 compared to the original video data. The sensitivity of vessel segmentation was enhanced by 10 % at a comparable specificity in comparison to a direct segmentation on the low-resolution data. This reveals the potential performance boost achieved by super-resolution in this application.

Figure 7.6 compares low-resolution data and super-resolution on one example dataset along with the corresponding vessel segmentations. Notice that the gain of super-resolution is revealed by a recovery of fine structures on the retina, e. g. blood vessels, that are barely visible in low-resolution data. Consequently, vessel segmentation achieved a higher sensitivity based on preprocessing by means of super-resolution compared to a segmentation on low-resolution data.

7.4.2 Experiments on Real Fundus Videos

In order to conduct experiments on real images, we used fundus video data captured with the low-cost and mobile camera developed by Tornow et al. [Torn 15]. This camera system is based on a monochromatic charge-coupled device (CCD) sensor and provides a spatial resolution of 640×480 px, a FOV of $20°$ in horizontal direction and a temporal resolution of up to 50 Hz. In this study, the left eyes of different human subjects including healthy subjects and glaucoma pa-



(a) Original **(b)** SR (initial guess) **(c)** SR (final) **(d)** Ground truth

(e) Se: 0.64, Sp: 0.93 **(f)** Se: 0.71, Sp: 0.95 **(g)** Se: 0.75, Sp: 0.95 **(h)** Se: 0.82, Sp: 0.95

Figure 7.6: Super-resolution on simulated fundus images generated from the DRIVE database [Staa 05] ($K = 15$ frames, magnification $s = 3$). (a) - (d) Low-resolution data, the initial guess (temporal median), the super-resolved image, and the ground truth image. (e) - (h) Blood vessel segmentation [Buda 13] along with the sensitivity (Se) and specificity (Sp). True-positive and false-positive pixels are color-coded in green and red, respectively.

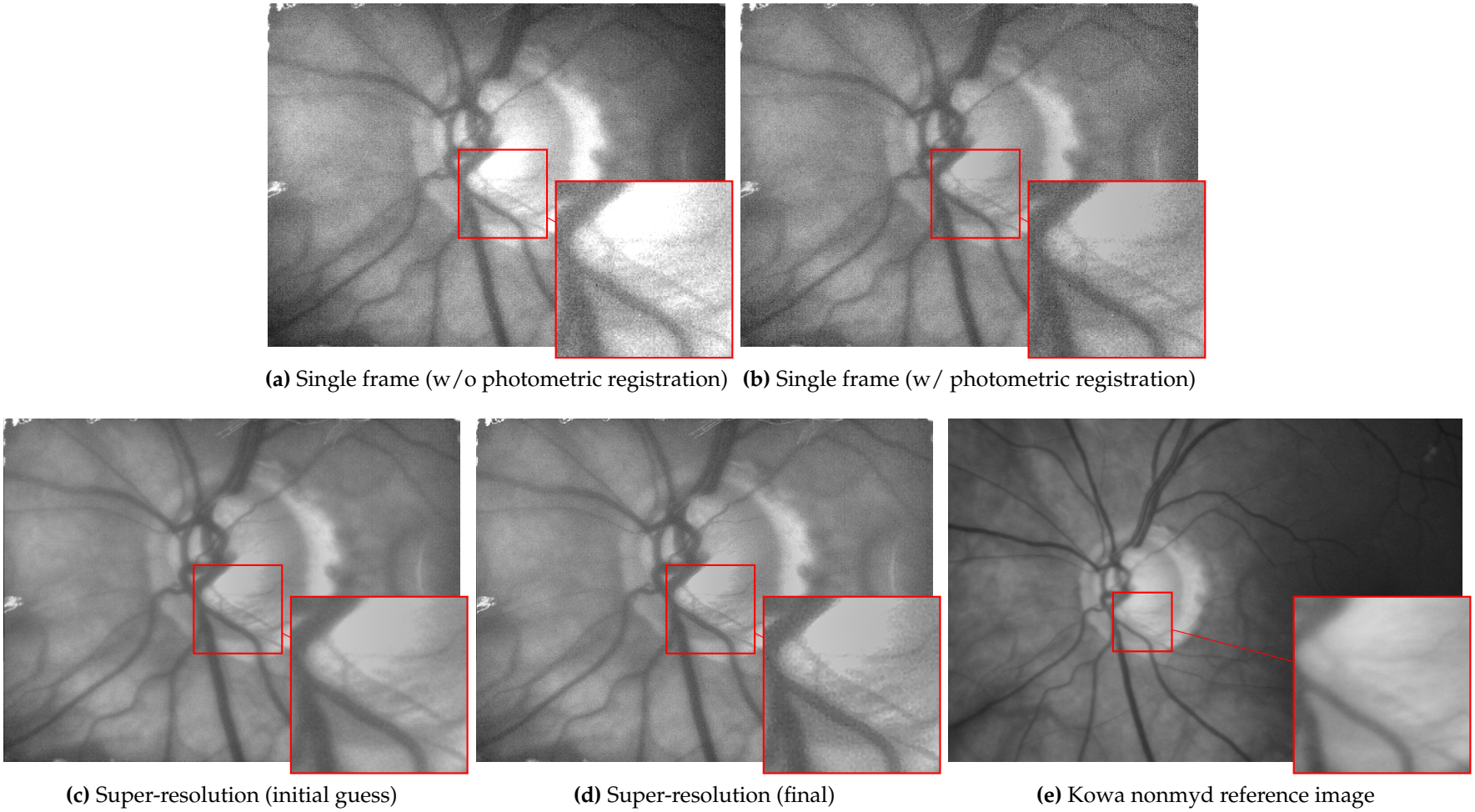
tients were examined. All examinations were done without dilating the pupil, i. e. non-mydriatically. The acquired video sequences have durations between 5 and 15 seconds[3]. Super-resolution was applied on subsequences extracted from these videos by processing $K = 8$ successive frames in a sliding window approach with magnification $s = 2$. Throughout all experiments, the unknown PSF was approximated by an isotropic Gaussian kernel ($\sigma_{\text{PSF}} = 0.8$).

Comparison to High-Resolution Reference Images. For the sake of a qualitative comparison to super-resolved data, a commercially available Kowa nonmyd camera[4] was employed to capture color fundus images. This single-shot camera features a spatial resolution of 1600×1216 px with a FOV of 25° and was used to gain high-resolution reference photographs of the same subjects that were examined with the low-cost camera. For fair comparisons to monochromatic video data, the green channels of the color photographs were used in this study.

The reconstruction of high-resolution fundus images from low-resolution video was investigated for anatomical regions that are relevant for diagnostic purposes and contain fine structures to outline the impact of super-resolution. Therefore, the optic nerve head region that captures the optic disk and the cup as two relevant structures for glaucoma detection [Bock 10, Josh 11] was examined.

[3] The data acquisition for this study was done in collaboration with Dr.-Ing. Ralf-Peter Tornow at the Department of Ophthalmology, Eye Clinics Erlangen, Germany

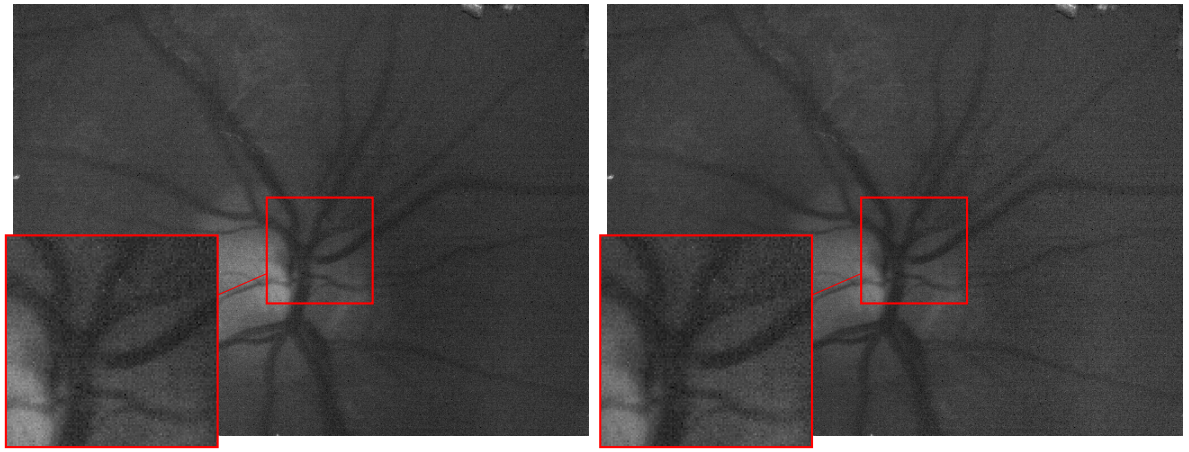
[4] http://www.kowamedical.com/



(a) Single frame (w/o photometric registration) **(b)** Single frame (w/ photometric registration)

(c) Super-resolution (initial guess) **(d)** Super-resolution (final) **(e)** Kowa nonmyd reference image

Figure 7.7: Super-resolution on low-cost fundus video frames for a glaucoma patient. (a) - (b) Single low-resolution frames without (w/o) and with (w/) photometric registration used in the proposed framework. (c) - (d) Initial guess determined by the temporal median of the registered low-resolution frames as well as the final super-resolved image. (e) Reference photograph captured with a Kowa nonymd camera.



(a) Single frame (w/o photometric registration) **(b)** Single frame (w/ photometric registration)

(c) Super-resolution (initial guess) **(d)** Super-resolution (final) **(e)** Kowa nonmyd reference image

Figure 7.8: Super-resolution on low-cost fundus video frames for a healthy subject. (a) - (b) Single low-resolution frames without (w/o) and with (w/) photometric registration used in the proposed framework. (c) - (d) Initial guess determined by the temporal median of the registered low-resolution frames as well as the final super-resolved image. (e) Reference photograph captured with a Kowa nonymd camera.

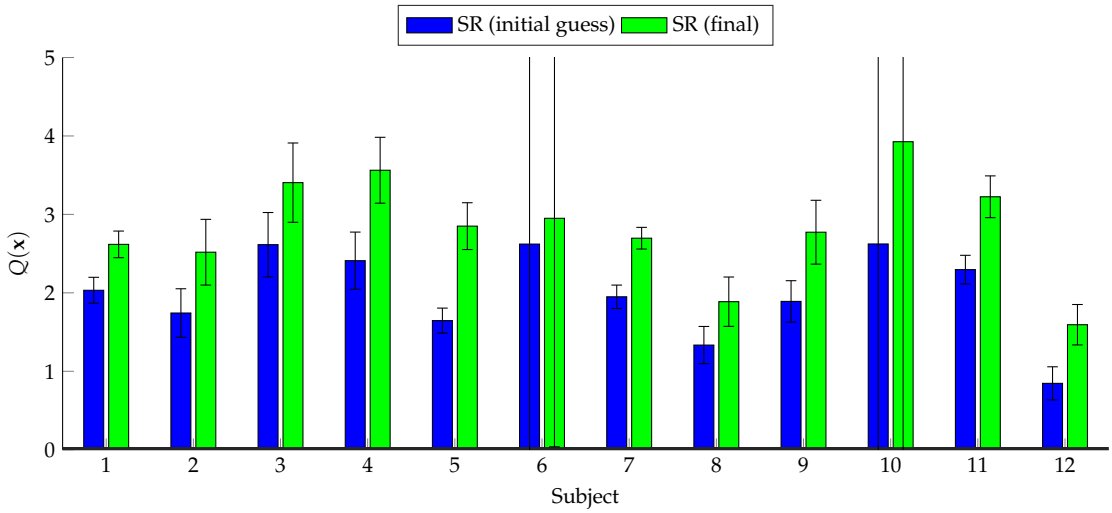
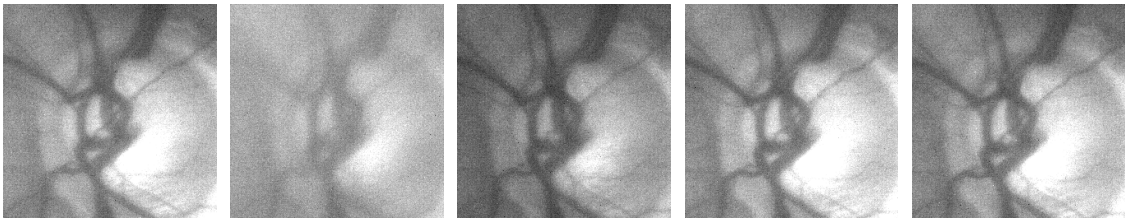


Figure 7.9: Quality measure $Q(\boldsymbol{x})$ for six healthy subjects (1 - 6) and six glaucoma patients (7 - 12). For each dataset, ten consecutive sequences in a sliding window scheme were processed. The quality assessments were compared for temporal median filtering used as initial guess and the final super-resolved image. Notice that $Q(\boldsymbol{x})$ is normalized by the quality measurement of the corresponding low-resolution reference image.

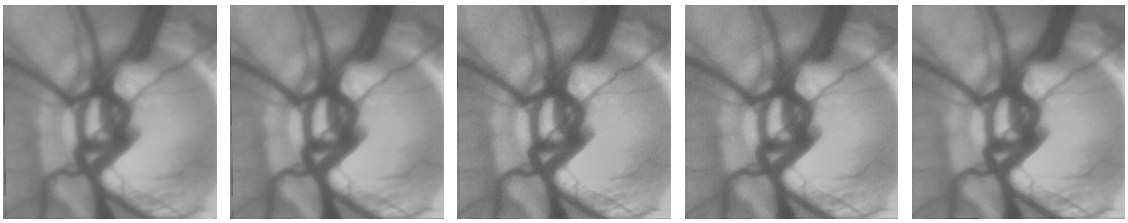
Figure 7.7 compares original video data and different intermediate results of the proposed framework applied in this region of a glaucoma patient to a reference photograph captured with the Kowa camera. The comparison among a single low-resolution frame in Fig. 7.7a and the frame in Fig. 7.7b depicts the impact of photometric registration as an initial stage of the super-resolution framework. Here, the photometric registration compensated for spatially and temporally varying illumination. Figure 7.7c depicts eye movement compensation implemented by the geometric registration and shows the temporal median of K registered frames that is used as an initial guess of iterative super-resolution. The final super-resolved image is shown in Fig. 7.7d. Note that super-resolution substantially enhanced the appearance of anatomical structures, e. g. thin blood vessels, which are barley visible in noisy low-resolution frames. This resulted in a visual appearance that is comparable to the Kowa reference image in Fig. 7.7e. Figure 7.8 depicts the same comparison on an example dataset captured from a healthy subject.

In order to validate this quality enhancement quantitatively, the gain in terms of the proposed no-reference quality measure was analyzed. The distribution of $Q(\boldsymbol{x})$ normalized by the quality of the reference low-resolution frames is summarized in Fig. 7.9 for the optic nerve head regions of six healthy subjects and six glaucoma patients. For each subject, ten consecutive image sequences extracted in a sliding window scheme were analyzed. This comparison among the temporal median and the final super-resolved image confirms that super-resolution improved noise and sharpness characteristics compared to raw video data.

Super-Resolution Under Photometric Variations. Let us now examine the reliability of super-resolution under challenging conditions in retinal imaging. One common issue is a severe photometric variation during an examination that is



(a) Low-resolution video frames with spatial and temporal illumination variations



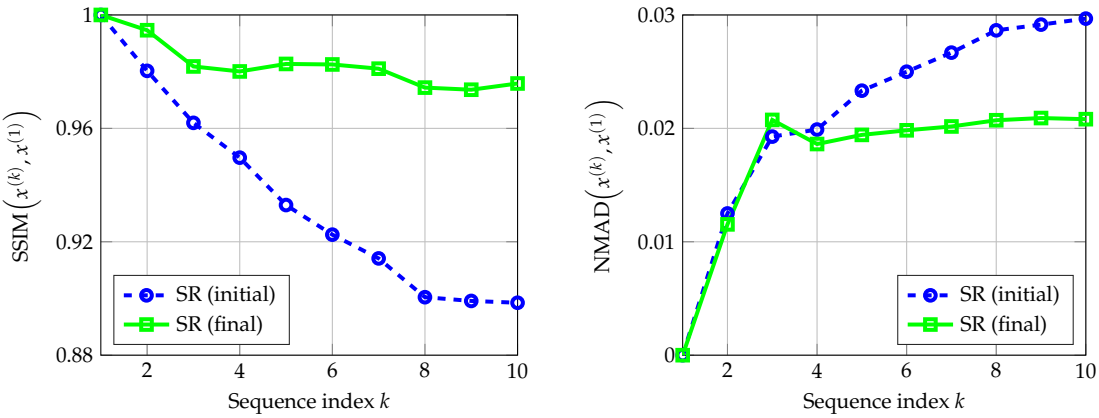
(b) Super-resolution under photometric variations

Figure 7.10: Super-resolution under spatial and temporal photometric variations across video frames. (a) Different low-resolution frames taken from a sequence with photometric variations. (b) Super-resolution for five different subsequences taken from the input video with $K = 8$ frames using the same reference frame.

caused by the light source of the camera and the patient anatomy. Notice that photogeometric registration cannot entirely compensate such variations, e. g. in case of oversaturations of the intensities.

Figure 7.10 (top row) shows this issue for a subset of video frames captured from a glaucoma patient with brightness and contrast variations over time. Super-resolution was applied to $K = 8$ frames in a sliding window scheme but using the same reference frame for each window. This led to a set of L super-resolved images $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(L)}$ obtained from L different windows but reconstructed in the same reference coordinate grid. The sensitivity of super-resolution regarding brightness and contrast variations was examined by assessing the reproducibility [Krau 17] of the super-resolved images $\boldsymbol{x}^{(l)}$, $l > 1$ relative to the first image $\boldsymbol{x}^{(1)}$.

In Fig. 7.11, this reproducibility is depicted by the SSIM and the normalized mean absolute deviation (NMAD) over ten super-resolved images corresponding to ten frame windows. Thus, a SSIM equal to one and a NMAD equal to zero indicate a perfect reproducibility on two disjoint input sequences. In terms of both measures, it is noticeable that inconsistencies among super-resolved images increases with shorter temporal overlap and hence a higher variability of the illumination. However, super-resolution achieved a reasonable reproducibility with a SSIM of above 0.85 and NMAD below 0.03. Compared to the initial guess determined by the temporal median, the proposed iterative algorithm resulted a better reproducibility. This behavior is also noticeable by visual comparisons among super-resolved images reconstructed from different frame windows as shown in Fig. 7.10 (bottom row). Notice that severe photometric variations in the input video were successfully compensated in super-resolved data, which confirms the robustness of the proposed framework.



(a) SSIM reproducibility **(b)** NMAD reproducibility

Figure 7.11: Sensitivity of super-resolution against photometric variations. The sensitivity was assessed by the reproducibility of super-resolved images for ten subsequences with photometric variations relative to the super-resolved image reconstructed from the first (variation-free) sequence. Reproducibility was measured by the SSIM and the NMAD.

Super-Resolution Under Eye Accommodation. Another condition of practical relevance is eye accommodation, which impairs the reliability of super-resolution.

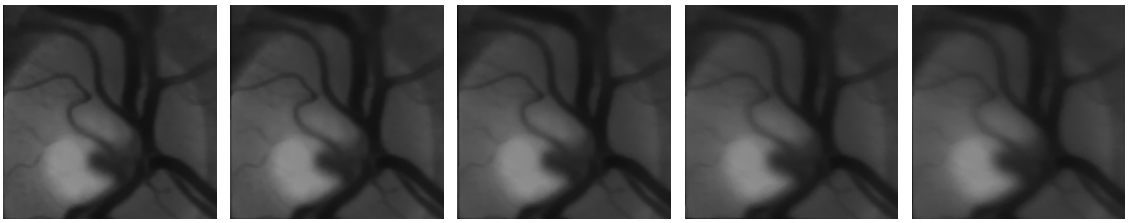
In Fig. 7.12 (top row), eye accommodation is shown for five video frames captured from a healthy subject. This resulted in out-of-focus blur that is increasing over time. Similar to the previous experiment, super-resolution was performed in a sliding window scheme based on $K = 8$ frames but with a fixed reference for each window to study its sensitivity regarding this effect.

Figure 7.13 depicts the reproducibility measures for L super-resolved images obtained in this experiment. The reproducibility characterized by these measures was dropped for larger amounts of out-of-focus blur related to accommodation. However, super-resolution provided a better reproducibility compared to its initial guess, which indicates a lower sensitivity regarding accommodation. More specifically, moderate levels of accommodation were successfully compensated as depicted for the first three cases in Fig. 7.12 (bottom row). In this experiment, severe levels of eye accommodation that are related to a substantial amount of time variant blur as shown for the last two cases could not be compensated.

7.4.3 Application to Super-Resolved Mosaicing

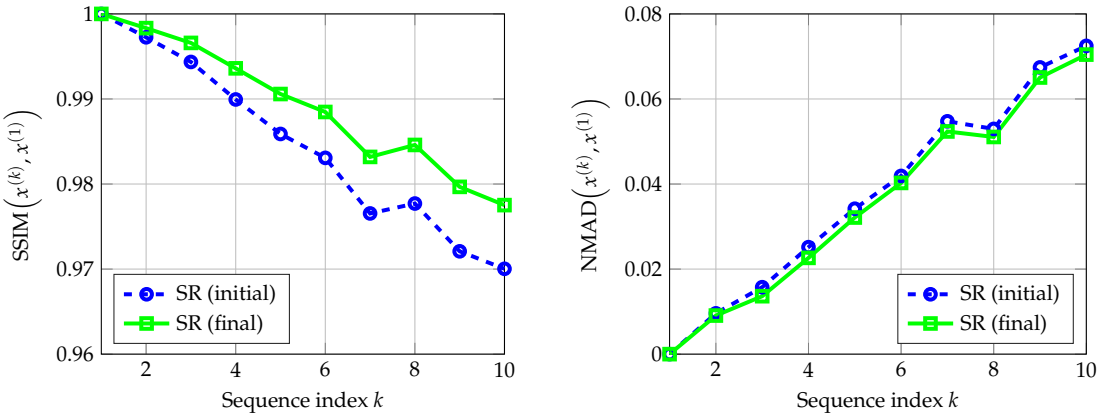
Besides the spatial resolution, another quality criterion of ophthalmic imaging systems is their FOV. In order to get a comprehensive view of the human retina for diagnostic or interventional purposes, there is a strong need to capture retinal images with a wide FOV. This applies to technologies like the slit lamp, scanning laser ophthalmoscopy or digital fundus cameras. In practice, however, this is challenging due the finite size of the human pupil and the fact that the pupil needs to be dilated to increase the FOV. For this reason, software-based image registration and mosaicing [Can 02, Catt 06, Adal 14, Zhen 14] have been proposed, which aims at stitching multiple views of the retina.

(a) Low-resolution video frames with eye accommodation over time



(b) Super-resolution under increasing eye accommodation

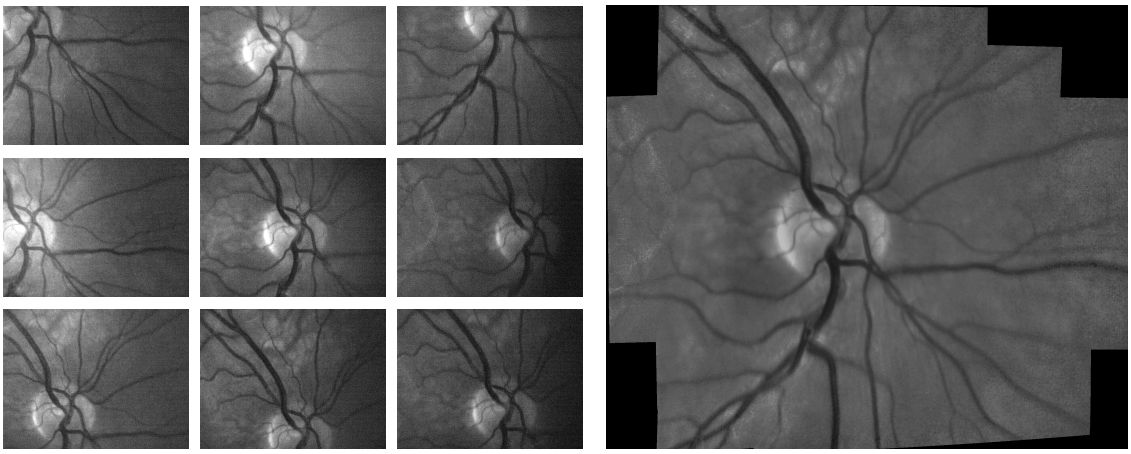
Figure 7.12: Super-resolution under out-of-focus blur due to eye accommodation over time. (a) Low-resolution frames taken from a sequence with accommodation and increasing out-of-focus blur from the first to the last frame. (b) Super-resolution for five different subsequences taken from the input video with $K = 8$ frames using the same reference frame and increasing out-of-focus blur from the first to the last frame.



(a) SSIM reproducibility **(b)** NMAD reproducibility

Figure 7.13: Sensitivity of super-resolution against out-of-focus blur due to eye accommodation. The sensitivity was assessed by measuring reproducibility of super-resolved images associated with ten subsequences that show increasing out-of-focus blur relative to the first subsequence. Reproducibility was measured by the SSIM and the NMAD.

This section demonstrates a novel combination of multi-frame super-resolution with mosaicing techniques to enable super-resolved mosaicing. Unlike related methods, this joint approach recovers a single mosaic view from low-resolution video while simultaneously enhancing the spatial resolution. For the study of super-resolved mosaicing, we employ the method proposed in [Kohl 16a] that reconstructs a single retinal mosaic from multiple super-resolved images. This approach exploits a set of low-resolution frames $\mathcal{Y}$ that consists of n disjoint subsets



(a) Low-resolution views **(b)** Super-resolved mosaic

Figure 7.14: Application of multi-frame super-resolution for image mosaicing. (a) Single low-resolution images for nine different regions of the human retina extracted from video data of a healthy subject. The different regions were scanned by asking the subject to fixate nine different positions on a fixation target. (b) Super-resolved mosaic reconstructed from the original video data of the nine regions. Figure reused from [Kohl 16a] with the publisher's permission ©2016 IEEE.

$\mathcal{Y}_i$, $i = 1, \ldots, n$. Each subset $\mathcal{Y}_i$ is represented by K_i consecutive frames taken from $\mathcal{Y}$ and is referred to as a *view*. These views capture complementary regions on the human retina due to eye motion during the examination.

In summary, super-resolved mosaicing is described by the following three-stage procedure:

1. Eye tracking is used for a fully automatic selection of n views. This is done in real-time using the optic disk as a robust feature for tracking [Kurt 14].

2. For each view $\mathcal{Y}_i$ that is selected according to the tracking procedure, K_i frames are utilized to reconstruct the corresponding super-resolved view $\boldsymbol{x}_i$.

3. The super-resolved views $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are first geometrically and photometrically registered and then stitched to a mosaic by adaptive averaging.

Figure 7.14 demonstrates super-resolved mosaicing on video data acquired from one healthy subject. In this experiment, we examined the left eye without dilating the pupil and asked the subject to fixate nine different positions on a fixation target. This resulted in eye movements across the frames in the acquired video sequence, and hence a scan of different regions of the retina as depicted in Fig. 7.14. We employed super-resolution with magnification $s = 2$ for these views with $K_i = 8$ frames as embedded in the proposed mosaicing framework. The final mosaic was assembled from nine different views. On the one hand, the super-resolution stage enhanced the spatial resolution of the original video data. On the other hand, stitching of super-resolved views enlarged the FOV from $\approx 15^\circ$ in the video data to $\approx 30^\circ$ in the mosaic image.

7.5 Conclusion

This chapter introduced a new super-resolution framework for retinal fundus video imaging as a novel diagnostic technique in ophthalmology. In order to reconstruct high-resolution fundus images from a low-resolution video, natural human eye movements during an examination are exploited. An image formation model was introduced that models eye movements by affine image-to-image transformations and considers spatial and temporal photometric variations across multiple frames. Based on this model, an iterative super-resolution algorithm was introduced that is steered by image quality self-assessment for the automatic selection of regularization hyperparameters. Quality self-assessment is based on a continuous quality measure that characterizes the level of sharpness and noise. The proposed measure has a mean Spearman rank correlation of 0.70 w. r. t. the PSNR, which shows that it can act as surrogate for full-reference quality assessment in the absence of ground truth data.

In a quantitative study, super-resolution enhanced the PSNR by 3.5 dB and the SSIM by 0.07 compared to low-resolution data. Moreover, the sensitivity of automatic blood vessel segmentation was improved by 10 %. Super-resolution on video data acquired with a mobile low-cost fundus camera provided images of comparable quality to those of commercially available, but expensive and stationary cameras. This encourages the use of the proposed method within clinical workflows, where cost-efficiency and mobility are essential, e. g. computer-aided screening. In addition, super-resolved mosaicing was presented to reconstruct high-resolution fundus images with enlarged FOV.

Future work needs to study the impact of the proposed framework to the diagnostic usability of fundus video imaging. One promising direction for future research is the adoption of super-resolution within machine learning techniques for computer-aided diagnosis regarding prevalent eye diseases [Bock 10, Abra 15].

CHAPTER 8

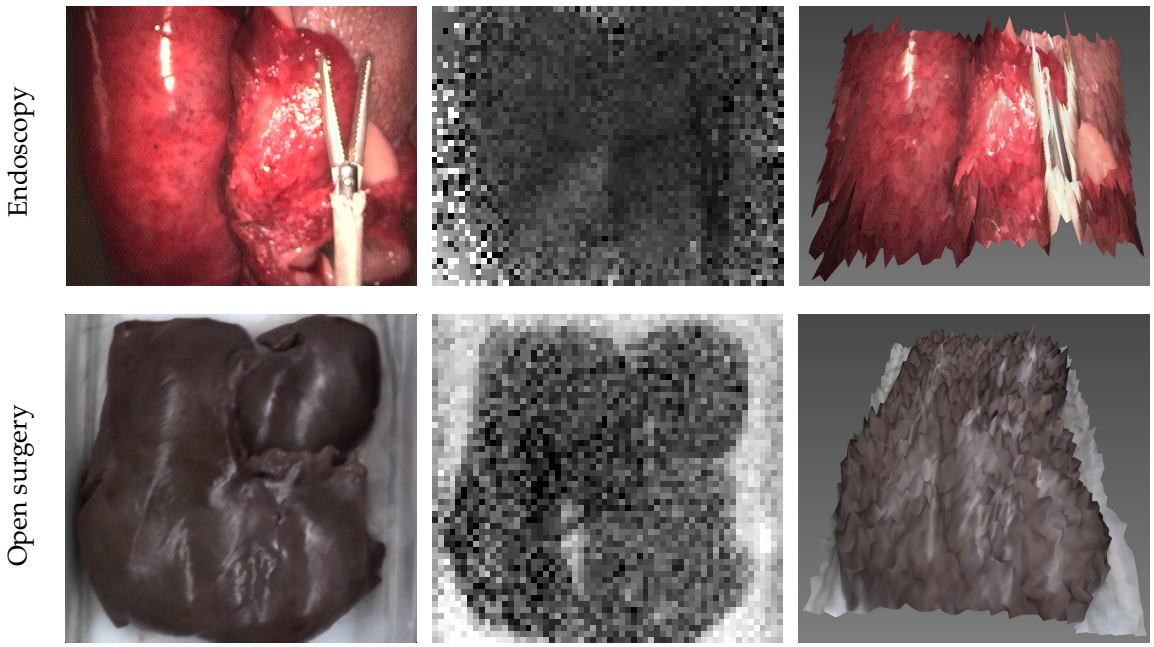
Applications in Image-Guided Surgery

This chapter investigates new applications of super-resolution to facilitate interventional medical imaging. In this context, one emerging field of research is the development of image guidance systems by means of hybrid range imaging to assist surgeons during minimally invasive or open surgical procedures. These systems can be implemented based on active range sensor technologies that enable the joint acquisition of surface data besides photometric information to provide a comprehensive view of the underlying scene. However, one common issue of these technologies is the low spatial resolution of todays range sensors, which limits their applicability in medical workflows. In order to enhance the reliability of image guidance, this chapter adopts the multi-sensor super-resolution framework presented in Chapter 5. The following complementary imaging setups are examined: 1) 3-D endoscopy to enhance minimally invasive surgical procedures as well as 2) 3-D image guidance for open surgery. This chapter presents system calibration approaches for both setups as a prerequisite for multi-sensor super-resolution. In addition, a comprehensive evaluation for super-resolution based on synthetic and ex-vivo datasets in both applications is reported.

An early study of super-resolution in 3-D endoscopy has been published by Köhler et al. [Kohl 13b] and Haase [Haas 16]. These concepts have been later extended in [Kohl 14b] and [Kohl 15b] including their application in open surgery.

8.1 Introduction and Medical Background

In the area of interventional medical imaging, one recent trend is the usage of range imaging technologies to gain 3-D surface information of patient anatomy in addition to 2-D photometric data [Baue 13]. If both approaches are aggregated, the combined setup enables intra-operative hybrid imaging of the anatomy. Compared to pure 2-D imaging, the existence of additional range information features various advantages for medical interventions. One of the most obvious benefits is



(a) Color image **(b)** Range image **(c)** Textured 3-D mesh

Figure 8.1: Color and ToF range measurements of porcine organs along with textured mesh representations to visualize sensor data fusion. Top row: ex-vivo data captured with a hybrid 3-D endoscope in minimally invasive surgery (see Section 8.3.2). Bottom row: ex-vivo data acquired with an imaging setup applicable for open surgery (see Section 8.3.3).

the expectation that range data holds the potential to offer the surgeon a more comprehensive view of patient anatomy in order to enhance the safety and efficiency of surgical procedures. In addition, it initiated the development of novel techniques for computer-assisted interventions to aid the surgeon. Some prominent examples for such applications in the area of minimally invasive surgery include automatic localization and collision avoidance for surgical instruments [Haas 13d, Wang 14]. More recently, 3-D abdomen reconstruction using range satellite cameras has been proposed to improve orientation and navigation during minimally invasive procedures [Haas 13a]. Another use case widely investigated for open surgery is the multi-modal registration of pre-operative, tomographic planning data with intra-operative range information [Mers 11]. This has widespread applications for augmented reality to aid surgeries or forensic medicine [Kilg 15].

In terms of the technical implementation of hybrid range imaging, there exist various approaches with individual pros and cons in image-guided surgery [Maie 13, Maie 14]. One of the historically first approaches to gain range data is stereo vision. Stereoscopy features a *passive* approach that utilizes geometric correspondences across two views of the same scene, e. g. pairs of corresponding points, to triangulate range data. The advantage of stereoscopy is that it can capture highly accurate measurements under ideal situations and has been engineered in stereo-based endoscopes [Fiel 09]. However, under realistic conditions in image-guided surgery it is error prone due to repetitive image structures or texture-less surfaces. In image-guided surgery, *active* senor technologies such as

ToF [Penn 09] or structured light [Schm 12] provide a promising alternative and hold the potential to capture dense range images in real-time. Unfortunately, one of their major shortcomings are their spatial resolutions that are rather low compared to modern color cameras. This means a major barrier to employ such sensors in clinical workflows. Figure 8.1 depicts ex-vivo ToF measurements alongside with high-resolution color images to visualize this issue for endoscopy and open surgery. Here, the overlay of range and color data demonstrates the complementary natures of both technologies. While color sensors capture photometric information of high spatial resolution, range sensors provide 3-D information that is acquired at a lower resolution. Range sensors might also suffer from a low SNR due to random or systematic errors, which is common in case of current ToF sensors [Kolb 10, Furs 16]. In order to overcome the low spatial resolution of range sensors in these systems, this chapter adopts the multi-sensor super-resolution framework introduced in Chapter 5. Accordingly, we employ high-resolution color images as guidance to super-resolve low-resolution range data.

The remainder of this chapter is structured as follows. Section 8.2 introduces system calibration techniques to facilitate sensor data fusion for hybrid range imaging in image-guided surgery as prerequisite for multi-sensor super-resolution. Section 8.3 presents a quantitative simulation study along with ex-vivo experiments for super-resolution in minimally invasive and open surgery workflows. Section 8.4 draws a conclusion for these studies.

8.2 System Calibration and Sensor Data Fusion

The super-resolution method proposed in Chapter 5 has the goal to reconstruct high-resolution range data from a set of low-resolution range images $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}$ to facilitate accurate 3-D measurements for image-guided surgery. This framework is driven by high-resolution color images $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(K)}$ that encode photometric information of the same scene. According to Chapter 5, color images are exploited for motion estimation, spatially adaptive regularization as well as outlier detection in the underlying reconstruction algorithm, see Fig. 8.2. This requires a pixel-wise mapping across both modalities, which is unknown a priori. Accordingly, multi-sensor super-resolution necessitates a system calibration to establish the mapping.

In this section, two calibration schemes are presented that are applicable to hybrid range imaging systems in image-guided surgery. This includes a homography approach that is applicable to *beam splitter* setups as well as a *stereo vision* approach that involves intrinsic and extrinsic camera calibrations. For more technical details on these approaches and their comparative evaluation, we refer to [Haas 16].

8.2.1 Sensor Data Fusion using a Homography

The first approach assumes that a 3-D surface is measured by a single optical system that acquires range and photometric information simultaneously. This can be implemented by means of a beam splitter that decomposes incoming light into two parts according to the wavelength, see Fig. 8.3. Range and photometric data is captured by two separate sensors that have the same view to the underlying scene.

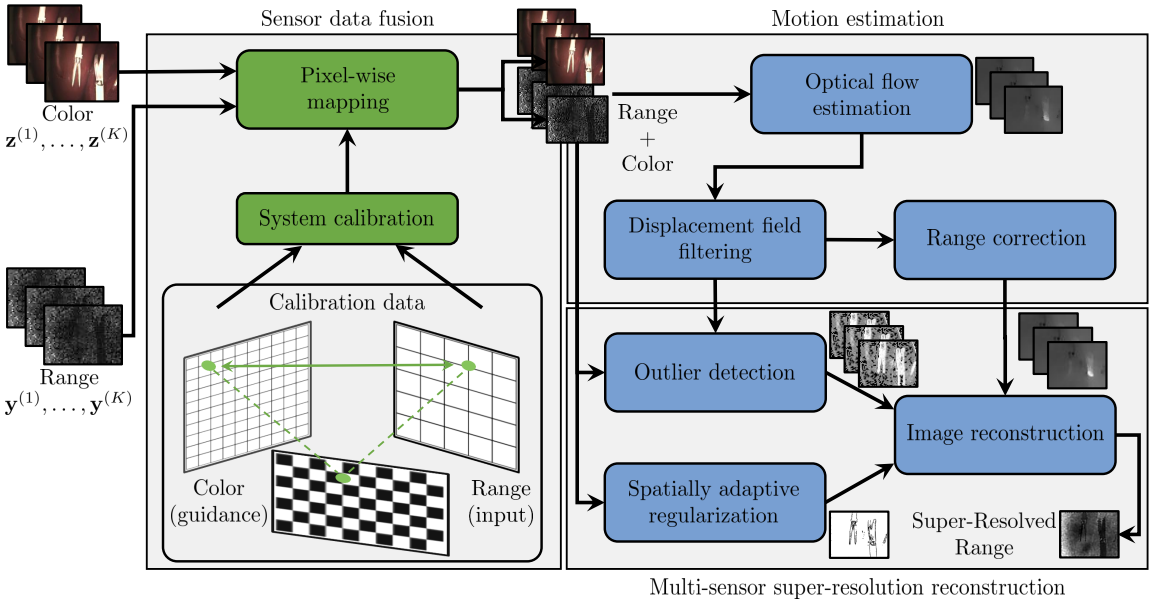


Figure 8.2: Flowchart of multi-sensor super-resolution for hybrid range imaging in image-guided surgery. The sensor fusion between range and color images is gained by system calibration using geometric correspondences among both modalities. Then, the fused images are used for motion estimation and the reconstruction of super-resolved range data.

In order to perform sensor data fusion, we employ the calibration approach introduced by Haase et al. [Haas 13b] that has been later adopted for multi-sensor super-resolution in [Kohl 13b]. For the task of system calibration, let $\boldsymbol{u}_y$ and $\boldsymbol{u}_z$ be a pair of corresponding points on a checkerboard calibration pattern in a range and a color image, respectively[1]. The relationship between these points is modeled according to:

$$\tilde{\boldsymbol{u}}_z \cong \boldsymbol{H}_{yz} \tilde{\boldsymbol{u}}_y, \tag{8.1}$$

where $\tilde{\boldsymbol{u}}_z \in \mathbb{R}^3$ and $\tilde{\boldsymbol{u}}_y \in \mathbb{R}^3$ denote the points $\boldsymbol{u}_z$ and $\boldsymbol{u}_y$ in homogeneous coordinates [Hart 04]. The homography $\boldsymbol{H}_{yz} \in \mathbb{R}^{3\times3}$ describes this mapping up to scale denoted by $\cong$. For system calibration, a set of point correspondences is identified by a self-encoded marker [Form 11] and the homography $\boldsymbol{H}_{yz}$ is found by least-squares estimation [Brad 00]. Then, the homography is used to fuse range and color images in a common coordinate system. In the proposed framework, each color image is warped towards the corresponding range image, up to a scale factor to preserve the spatial resolution.

The use of a homography offers a couple of useful properties. One essential property is the possible inversion of the mapping between both modalities. In addition, the homography enables sensor data fusion solely with corresponding point pairs without involving intrinsic or extrinsic camera calibration. In practice, the estimated homography leads to re-projection errors of subpixel accuracy [Haas 13b], which enables a highly accurate sensor data fusion. In this chapter, this

[1] In order to detect feature points in range data, the amplitude images provided by a ToF sensor after contrast enhancement and binarization as shown in [Haas 13b] can be used.

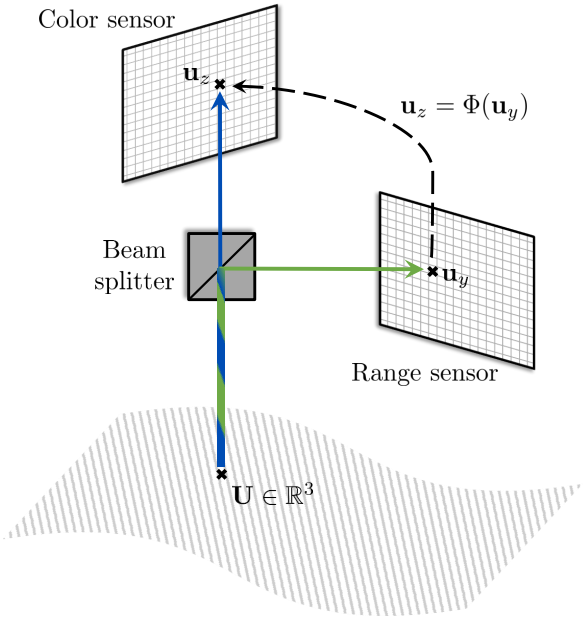


Figure 8.3: Geometry of the system setup for the simultaneous acquisition of range and photometric data with one common optical system. The beam splitter decomposes incoming light into a range and a photometric signal. In this approach, the mapping of photometric data towards a range image (and vice versa) is modeled by a homography.

approach is utilized for hybrid 3-D endoscopy. Here, it provides the acquisition of color and range data through one single endoscope equipped with a beam splitter.

8.2.2 Sensor Data Fusion using Stereo Vision

The second approach considers the case that a 3-D surface is captured with two separate cameras with respective optical systems. These cameras acquire color and range information of the same scene from different viewpoints. Using a temporal synchronization, we can combine both modalities. The advantage of this setup is that it is not necessary to combine range and color sensors by a common optical system, see Fig. 8.4. This is beneficial as it increases the flexibility in terms of the camera hardware. In the applications presented below, this setup is examined for image-guided open surgery.

Unlike in the beam splitter setup, the system calibration cannot be described by a homography. As a consequence, one needs to perform stereo camera calibration [Hart 04] to fuse color and range images. In this work, sensor fusion is accomplished according to the calibration method introduced in [Haas 12] that has been later used for multi-sensor super-resolution in [Kohl 15b]. First, a point $\boldsymbol{u} = (u_y, v_y)$ on the range camera image plane with the measured range value y is re-projected to the 3-D space according to:

$$\tilde{\boldsymbol{U}} \cong \boldsymbol{P}_y^{-1} \begin{pmatrix} u_y & v_y & y & 1 \end{pmatrix}^\top , \tag{8.2}$$

where $\boldsymbol{P}_y \in \mathbb{R}^{4\times4}$ is the range camera projection matrix as shown in the calibration approach of Park et al. [Park 11]. Subsequently, the re-projected point given in

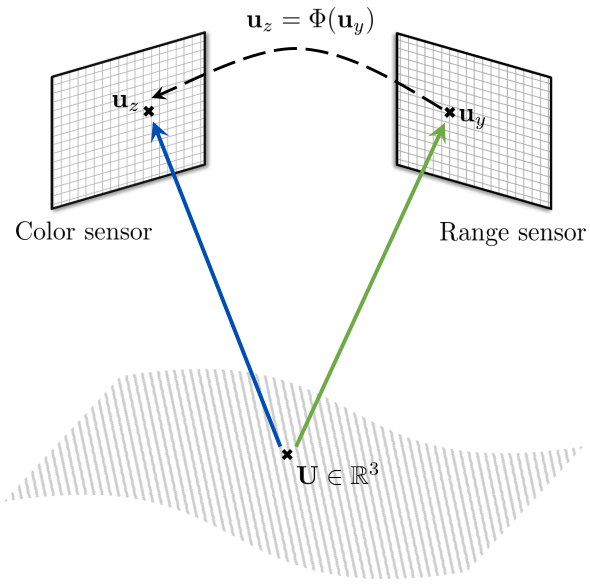


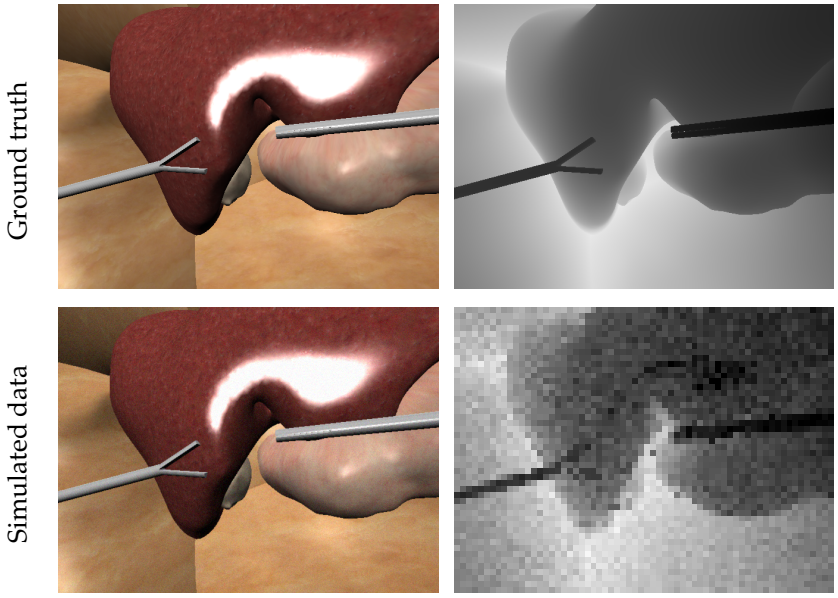
Figure 8.4: Geometry of the system setup for the simultaneous acquisition of range and photometric data with two separate sensors and optics. The mapping of photometric data towards a range image is determined by stereo calibration for both cameras.

homogeneous coordinates $\tilde{\boldsymbol{U}} \in \mathbb{R}^4$ can be projected onto the image plane of the color camera. Using the re-projected point in Eq. (8.2), the corresponding pixel position in homogenous coordinates is given by:

$$\begin{aligned} \tilde{\boldsymbol{u}}_z &\cong \boldsymbol{P}_z \tilde{\boldsymbol{U}} \\ &= \boldsymbol{P}_z \boldsymbol{P}_y^{-1} \begin{pmatrix} u_y & v_y & y & 1 \end{pmatrix}^\top , \end{aligned} \tag{8.3}$$

where $\boldsymbol{P}_z \in \mathbb{R}^{3\times4}$ denotes the projection matrix of the color camera. The projection matrices $\boldsymbol{P}_z$ and $\boldsymbol{P}_y$ in Eq. (8.3) are determined by intrinsic and extrinsic camera calibration. This is done using checkerboard calibration patterns to establish point correspondences for the calibration and least-squares optimization. The calibration procedure yields the intrinsic calibration matrices $\boldsymbol{K}_y \in \mathbb{R}^{3\times3}$ for the range camera and $\boldsymbol{K}_z \in \mathbb{R}^{3\times3}$ for the color camera as well as the extrinsic parameters given by the rotation matrix $\boldsymbol{R} \in \mathbb{R}^{4\times4}$ and the translation vector $\boldsymbol{t} \in \mathbb{R}^4$. Similar to the homography approach, this method is used to warp color images to the domain of the range data up to a scale factor. However, notice that the calibrated mapping is not invertible and is affected by occlusions.

Compared to the homography approach, a stereo calibration suffers from shortcomings in terms of its accuracy under practical conditions. In particular, the calibration accuracy is highly dependent on the reliability of the measured range data as these measurements are used for the re-projection in Eq. (8.2). In order to deal with random measurement noise, the calibration is performed on preprocessed range data using the filter pipeline proposed in [Wasz 11c]. Moreover, for the compensation of systematic errors in the range data [Kolb 10], the extrinsic parameters are further refined after stereo calibration. For this purpose, the translation vector $\boldsymbol{t}$ is refined by optimizing the normalized mutual information [Plui 03] between range and color data to alleviate a potential bias in the calibration.



(a) Color data **(b)** Range data

Figure 8.5: Color and range image obtained from an artificial laparoscopic scene. Simulated data (bottom row) is generated from the ground truth (top row) using subsampling, blurring as well as conditions of ToF endoscopy like specular highlights.

8.3 Experiments and Results

This section present an experimental evaluation of multi-sensor super-resolution for hybrid range imaging systems in image-guided surgery. For a quantitative evaluation, a comprehensive simulation study is presented to validate accuracy and robustness of super-resolution in the desired applications. Subsequently, the applicability of the proposed framework is demonstrated by ex-vivo experiments for hybrid 3-D endoscopy as well as image guidance in open surgery.

8.3.1 Simulated Data Experiments

In the following study, we used artificial hybrid range data from the publicly available Multi-Sensor Super-Resolution Datasets[2] for a quantitative evaluation. These range and color images were obtained from an artificial laparoscopic scene under the conditions of minimally invasive surgery using the RITK [Wasz 11a]. Ground truth range and color images were gained from the artificial 3-D model and both modalities were perfectly aligned to exclude the influence of calibration errors in this baseline experiment. Color images were encoded with a pixel resolution of 640×480 px and disturbed by a Gaussian PSF ($\sigma_{\text{PSF}} = 0.5$) as well as additive Gaussian noise ($\sigma_{\text{noise}} = 0.001$). The corresponding range images were simulated with a pixel resolution of 64×48 px. To analyze the influence of systematic errors, the simulation considered the following effects of ToF imaging and surgical interventions. First, as opposed to space invariant noise, Gaussian noise in range data

[2] https://www5.cs.fau.de/research/data/multi-sensor-super-resolution-datasets/

Table 8.1: Overview of range super-resolution algorithms along with their parameter settings. Multi-sensor super-resolution is employed with different algorithm profiles (MSR, AMSR, AMSR-OD). Single-sensor super-resolution (SSR) is considered as the baseline.

Reconstruction algorithm	**Algorithm properties**		
	Motion estimation	**Adaptive regularization**	**Outlier detection**
Single-sensor super-resolution (SSR)	direct	✗	✗
Multi-sensor super-resolution (MSR)	filter-based	✗	✗
Adaptive multi-sensor super-resolution (AMSR)	filter-based	✓ $\tau_0 = 0.025$, $N_{xz} = 7$	✗
Adaptive multi-sensor super-resolution with outlier detection (AMSR-OD)	filter-based	✓ $\tau_0 = 0.025$, $N_{xz} = 7$	✓ $\rho_0 = 0.5$

was simulated to be distance-dependent with a maximum standard deviation of $\sigma_{\text{noise}} = 10\,\text{mm}$ in the scene space. Second, specular highlights in color images were simulated as a common issue in endoscopy [Haas 14]. These specular highlights resulted in range measurements disturbed by Perlin noise in the affected image regions. Finally, range data was corrupted by *flying pixels* [Kolb 10] that were generated by randomly flipping 20 % of all pixels located on depth discontinuities.

The artificial model was used to generate image sequences of the scene from different perspectives along with surgical instruments, see Fig. 8.5. Movements of an hand-held endoscope were simulated by a randomly generated rigid motion of the virtual camera in the scene space. In addition, movements of surgical instruments and soft tissue were simulated to consider realistic conditions in minimally invasive surgery. We use the superposition of these 3-D movements that appear as 2-D subpixel motion in range and color images as a cue for super-resolution.

We examine the reconstruction methods that were previously introduced in Section 5.5 using a Huber prior with $\delta_{\text{Huber}} = 5 \cdot 10^{-4}$ and $\lambda = 0.8$. See Tab. 8.1 for an overview of the configurations of the different multi-sensor techniques. Throughout the following experiments, the single-sensor reconstruction algorithm (SSR) that works solely on the range data is considered as the baseline and compared to the different multi-sensor methods (MSR, AMSR, and AMSR-OD).

Accuracy Analysis. In order to conduct a baseline experiment, four artificial datasets (S1 - S4) were generated from the given laparoscopic scene. Throughout all experiments, super-resolution was performed with magnification $s = 4$ and $K = 31$ frames, where the central one was used as reference for variational optical flow computation [Liu 09]. The reconstructions were conducted in a sliding window scheme using K successive frames to obtain single super-resolved images.

The statistics of the PSNR and SSIM of super-resolved range data reconstructed by the different algorithms on ten randomly generated image sequences per dataset

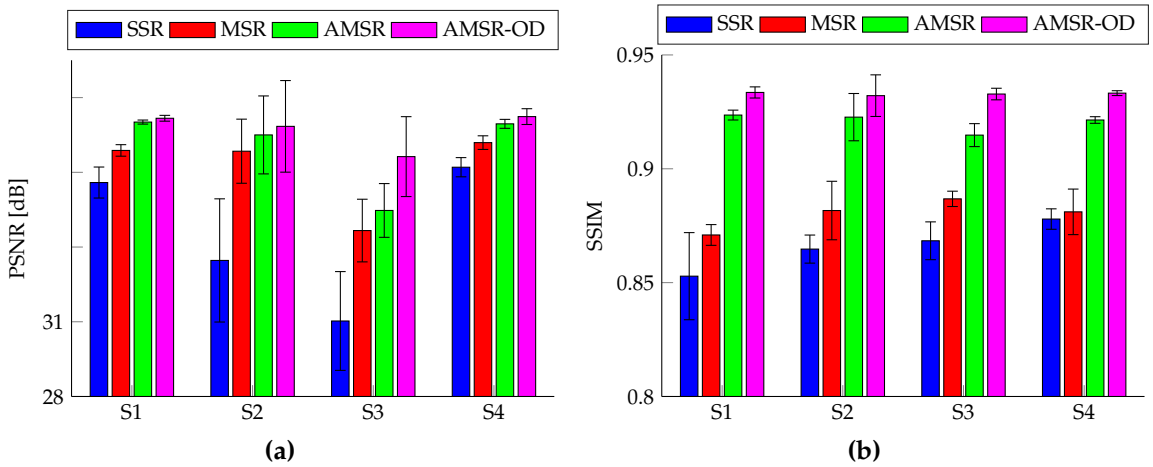
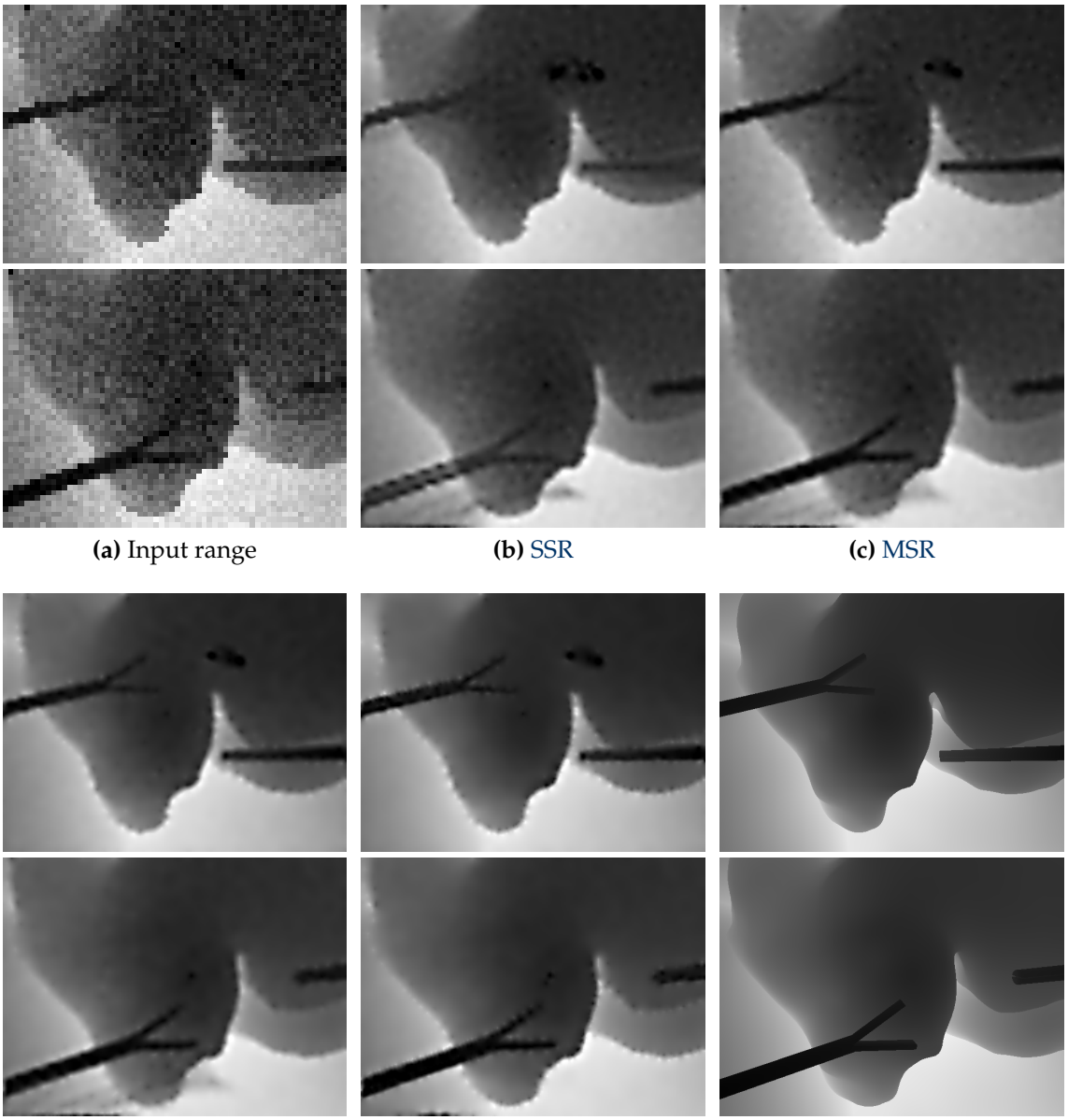


Figure 8.6: Single-sensor super-resolution (SSR) versus the different multi-sensor algorithms (MSR, AMSR, and AMSR-OD) on an artificial laparoscopic scene. The statistics (mean $\pm$ standard deviation) of the PSNR and SSIM measures were determined on four datasets (S1 - S4) with known ground truth range data. For each dataset, super-resolution was applied for ten consecutive image sequences using sliding window processing.

are reported in Fig. 8.6. Notice that multi-sensor super-resolution consistently outperformed the single-sensor approach on all datasets. The most substantial improvements were obtained under challenging situations for optical flow estimation due to motion of soft tissue and instruments along with movements of the virtual endoscope. The multi-sensor formulation considerably increased the accuracy of the motion estimate using the underlying filter-based technique, which resulted in accurate range super-resolution. In addition, spatially adaptive regularization (AMSR) leveraged the reconstruction of depth discontinuities compared to non-adaptive regularization (MSR). This affects the measurement of anatomical structures or surgical instruments. Moreover, outlier detection (AMSR-OD) enhanced the robustness against individual misregistered frames as well as outliers in the low-resolution range data, e. g. space variant random noise or systematic errors. This is notably for situations where the filter-based motion estimation cannot establish a reliable displacement field since optical flow computation on color images entirely failed. See Fig. 8.7 for a visual comparison on two example datasets (S2 and S3) in these situations. Here, the full combination of the proposed multi-sensor techniques (AMSR-OD) provides reliable range data including accurate reconstructions of soft tissue and surgical instruments.

Robustness Analysis. In terms of the robustness, multi-sensor super-resolution is affected by the conditions in image-guided surgery. Let us study two important issues that are related to motion estimation and system calibration.

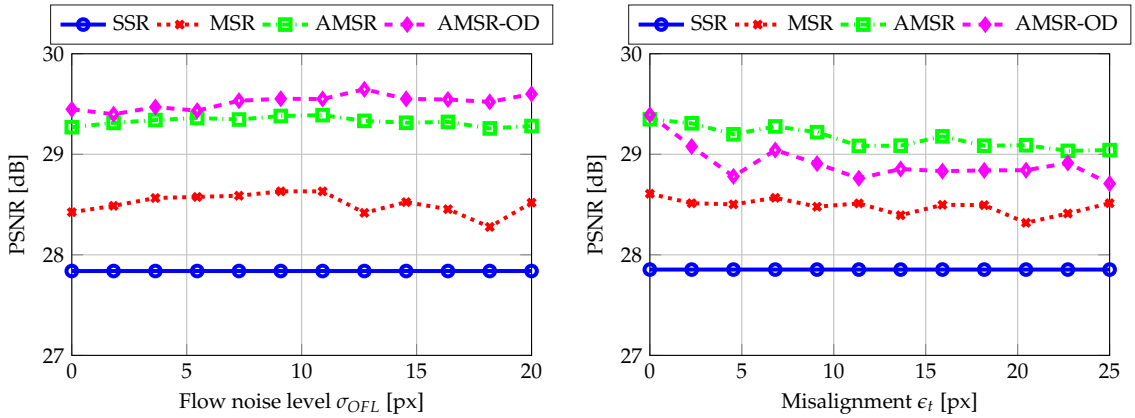
One important problem case is the uncertainty of optical flow estimation under realistic conditions. This issue was investigated by intentionally disturbing the optical flow determined on color images by zero-mean, normal distributed noise with standard deviation σ_{OFL} in each displacement component. In practice, this situation might appear in texture-less regions on organ surfaces resulting in un-



(a) Input range **(b)** SSR **(c)** MSR

(d) AMSR **(e)** AMSR-OD **(f)** Ground truth

Figure 8.7: Super-resolution reconstruction ($K = 31$ frames, $4\times$ magnification) on two datasets of an artificial laparoscopic scene. First and third row: comparison of low-resolution range data to single-sensor super-resolution (SSR) and the different multi-sensor approaches (MSR, AMSR, and AMSR-OD) for the dataset S2. Notice that direct motion estimation on range data as implemented by the SSR approach failed, which resulted in unreliable super-resolved range information. Second and fourth row: comparison for the dataset S3. Note that outlier detection as implemented by AMSR-OD compensated for outliers in optical flow that are related to difficult motion types, e. g. endoscope movements superimposed with independently moving surgical instruments.



(a) PSNR vs. flow noise level σ_{OFL} **(b)** PSNR vs. calibration error ϵ_t

Figure 8.8: Robustness analysis of different multi-sensor approaches (MSR, AMSR, and AMSR-OD) using the single-sensor approach (SSR) as baseline. (a) Sensitivity regarding optical flow corrupted by Gaussian noise with standard deviation σ_{OFL} measured in pixels of the color images. (b) Sensitivity regarding calibration errors simulated by translational misalignments of length ϵ_t measured in pixels of the color images.

reliable displacement fields. Figure 8.8a shows the impact of noisy optical flow for the dataset S1 at different noise levels σ_{OFL} measured in terms of pixels of the color images. Notice that even for large noise levels, the different multi-sensor approaches were nearly insensitive to noisy optical flow and considerably outperformed the single-sensor reconstruction. This behavior is related to the filter-based motion estimation as an integral part of multi-sensor super-resolution, which gets rid of the noise present in the optical flow of the color images.

So far, the fusion among color and range data was assumed to be exact. In practice, however, the actual accuracy is highly affected by unavoidable calibration errors. This influences the reliability of the multi-sensor reconstruction as it relies on accurate sensor data fusion. For a robustness analysis, small misalignments between both modalities were intentionally induced in the simulation process to consider calibration errors. These misalignments were obtained by randomly generated translations of the color images relative to the range data. The behavior of super-resolution at different misalignments measured by the translation length ϵ_t in terms of pixels of color data is shown in Fig. 8.8b for the dataset S1. Notice that in contrast to single-sensor super-resolution, the accuracy of the different multi-sensor approaches dropped under increasing misalignments. In particular, spatially adaptive regularization as well as outlier detection were sensitive regarding this effect, while filter-based motion estimation was less affected. However, the different multi-sensor approaches still outperformed the competing single-sensor approach confirming a reasonable robustness against calibration errors.

8.3.2 Application to Hybrid 3-D Endoscopy

This section demonstrates the application of multi-sensor super-resolution in the area of minimally invasive surgery. For this study, ex-vivo experiments were con-

ducted by measuring porcine organs with a hybrid 3-D endoscope[3]. Range data was captured with a ToF sensor that features a pixel resolution of 64×48 px at a frame rate of 30 Hz. The corresponding color sensor provides a resolution of 640×480 px. Both sensors are combined in a single optical system that is equipped with a beam splitter to synchronize the acquisition of range and color images [Haas 13b]. Therefore, the homography approach presented in Section 8.2.1 was used for calibration and sensor data fusion.

To induce motion, the endoscope was shifted over time relative to the organ surface. In addition, surgical instruments were moved to consider the conditions of minimally invasive procedures. Super-resolution was performed with magnification $s = 4$ and $K = 31$ consecutive frames, where the central frame was chosen as reference for optical flow [Liu 09].

Reconstruction Results. Figure 8.9 depicts a comparison among the different reconstruction algorithms to low-resolution range data on one example dataset. In this application, we are particularly interested in reliable reconstructions of soft tissue surfaces and artificial objects, e. g. the instrument tips. It is worth noting that these structures are difficult to detect in the measured range data.

In comparison to the single-sensor reconstruction (SSR), the proposed filter-based motion estimation substantially improved the reliability of the computed displacement fields. This resulted in a superior accuracy of the multi-sensor algorithm (MSR) in terms of the reconstruction of anatomical structures and surgical instruments. In contrast to the filter-based approach, direct optical flow estimation on range data as implemented for the single-sensor reconstruction was error prone and did not capture endoscope or instrument movements appropriately. Notice that spatially adaptive regularization (AMSR) enhanced the multi-sensor reconstruction even further. This translated into a superior recovery of depth discontinuities. Moreover, outlier detection (AMSR-OD) got rid of non-Gaussian noise related to systematic distance- and intensity-dependent errors in ToF imaging. This considerably improved the reconstruction of soft tissue surfaces.

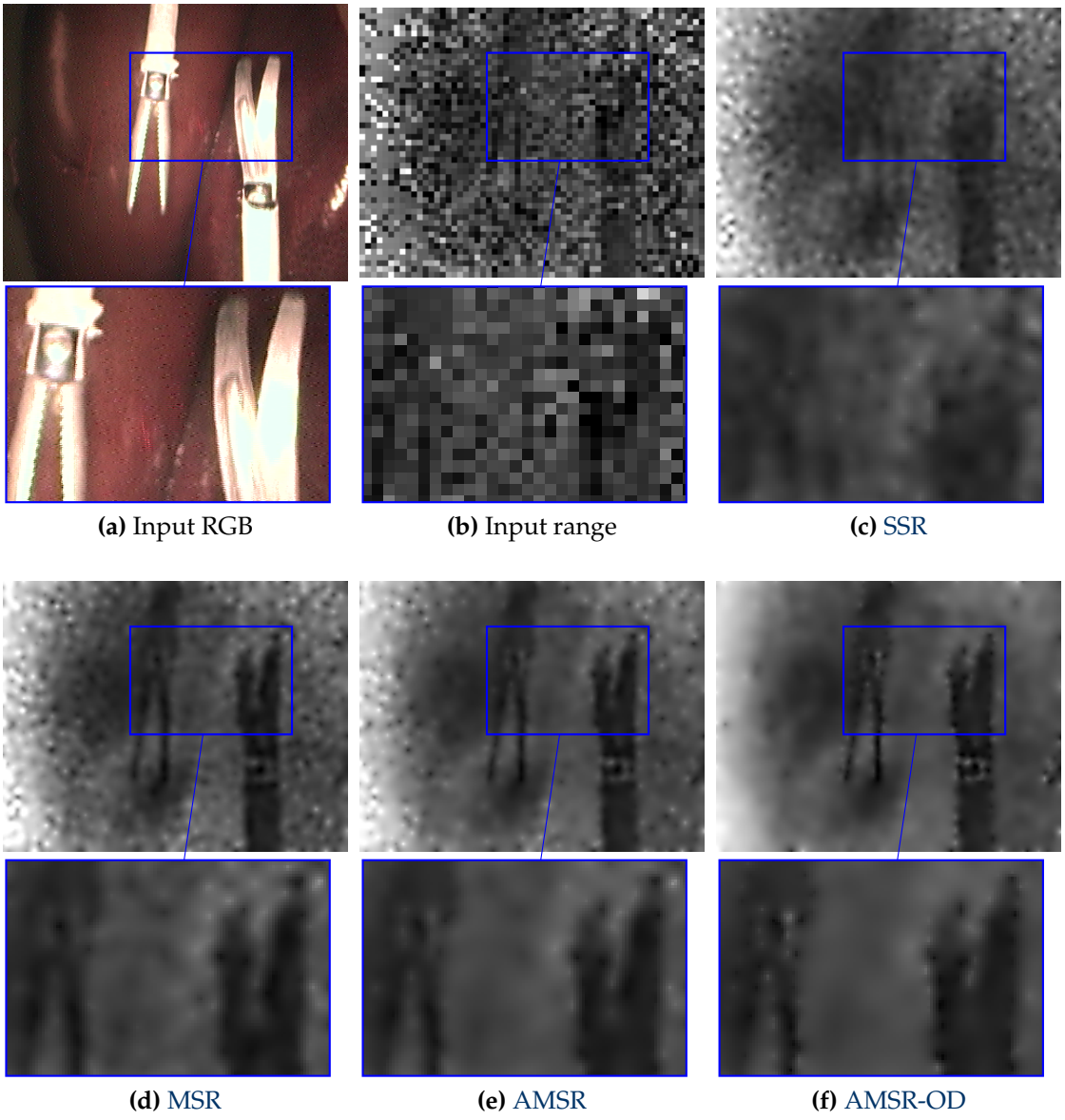
Range Data Quality Assessment. Two no-reference quality measures are used to quantitatively assess the reliability of range data. One criterion is noise reduction on reconstructed surfaces. To this end, a blind SNR estimation is performed. The measure that is used in this work is computed for flat surfaces according to:

$$Q_{\text{snr}} = 10 \log_{10} \left(\frac{\mu_{\text{flat}}}{\sigma_{\text{flat}}} \right), \tag{8.4}$$

where μ_{flat} and σ_{flat} denote the mean and standard deviation of the range measurements in a rectangular region of interest, respectively. This measure is defined in dB and the higher Q_{snr}, the more accurate the reconstruction of the surface.

Besides reliable surface reconstruction, another goal is the accurate reconstruction of object transitions. For this purpose, regions of interests that contain an

[3] All experiments for this study were conducted with the hybrid 3-D endoscope prototype manufactured by the Richard Wolf GmbH, Knittlingen, Germany.



(a) Input RGB **(b)** Input range **(c)** SSR

(d) MSR **(e)** AMSR **(f)** AMSR-OD

Figure 8.9: Super-resolution reconstruction ($K = 31$ frames, $4\times$ magnification) for hybrid ToF/RGB endoscopy on a porcine liver. First and third row: high-resolution color image and low-resolution range data in comparison to super-resolved range images obtained by single-sensor super-resolution (SSR) as well as the different multi-sensor algorithms (MSR, AMSR, and AMSR-OD). Second and fourth row: zoom-in for an image region that contains surgical instruments. Notice that single-sensor super-resolution failed to provide reliable range information, while the different multi-sensor algorithms considerably enhanced the accuracy of range information for soft tissue and the surgical instruments.

Table 8.2: Mean ± standard deviation of the no-reference quality measures Q_{snr} and Q_{edge} in the ex-vivo study for hybrid 3-D endoscopy. Both measures were determined using manually selected regions of interest in low-resolution and super-resolved range data gained by the different reconstruction algorithms. In total, nine datasets with six manually selected regions per dataset were used.

Measure	Low-res. data	Super-resolved data			
		SSR	MSR	AMSR	AMSR-OD
Q_{snr}	7.2 ± 2.8	10.6 ± 2.1	10.6 ± 2.1	11.4 ± 2.3	11.8 ± 2.4
Q_{edge}	1.9 ± 0.4	2.4 ± 0.7	2.7 ± 1.1	3.1 ± 1.7	3.2 ± 2.0

edge between two structures in the range data are analyzed. The range values are described by a Gaussian mixture model (GMM) consisting of two components that represent foreground and background range values, respectively. Then, the quality measure to assess the reconstruction of depth discontinuities is defined as:

$$Q_{\text{edge}} = \frac{w_b(\mu_b - \mu)^2 + w_f(\mu_f - \mu)^2}{w_b\sigma_b^2 - w_f\sigma_f^2}, \tag{8.5}$$

where μ denotes the mean range value in the selected region, and μ_b and μ_f are the mean values of the background and the foreground, respectively. Similarly, σ_b and σ_f are the standard deviations, and w_b and w_f are the weights associated with the GMM components. This model is fitted to the range values using k-means clustering ($k = 2$). Notice that lower estimates of σ_b and σ_f along with larger differences between μ_b and μ_f indicate a better discrimination between foreground and background. Accordingly, the higher Q_{edge} the better the underlying reconstruction.

Six image regions per dataset containing flat surfaces and depth discontinuities were manually selected. The respective statistics of Q_{snr} and Q_{edge} for nine datasets are summarized in Tab. 8.2. In comparison to raw range data, the combination of all multi-sensor techniques (AMSR-OD) leads to an increase of Q_{snr} and Q_{edge} by 64 % and 68 %, respectively. The different multi-sensor algorithms also improved the reconstruction of flat surfaces and depth discontinuities in comparison to a single-sensor approach (SSR). These properties express a higher reliability of range data to facilitate segmentation or object detection [Haas 13c].

8.3.3 Application to Image Guidance in Open Surgery

Let us now demonstrate the application of multi-sensor super-resolution for image guidance in open surgery. In contrast to hybrid 3-D endoscopy, a stereo camera setup was developed for ex-vivo measurements on a porcine liver. In order to measure the liver surface, range data was captured with a PMD CamCube 3 ToF camera that provides a pixel resolution of 200×200 px at a frame rate of 30 Hz. A Grasshopper2 camera with a resolution of 1200×1200 px was used to acquire color images and was temporally synchronized to the range sensor. Both cameras were coupled on a tripod with a baseline that was chosen as small as possible to minimize occlusions. This stereo setup was calibrated according to Section 8.2.2 and

Table 8.3: Mean $\pm$ standard deviation of the no-reference quality measures Q_{snr} and Q_{edge} in the ex-vivo study of image-guided open surgery. Both measures were determined using manually selected regions of interest in low-resolution and super-resolved range data gained by the different reconstruction algorithms. In total, four datasets with six manually selected regions per dataset were used.

Measure	Low-res. data	Super-resolved data			
		SSR	MSR	AMSR	AMSR-OD
Q_{snr}	17.8 ± 0.9	21.7 ± 1.5	22.0 ± 1.4	22.1 ± 1.4	22.1 ± 1.4
Q_{edge}	3.9 ± 0.9	5.0 ± 1.0	6.6 ± 1.3	6.7 ± 1.4	6.7 ± 1.4

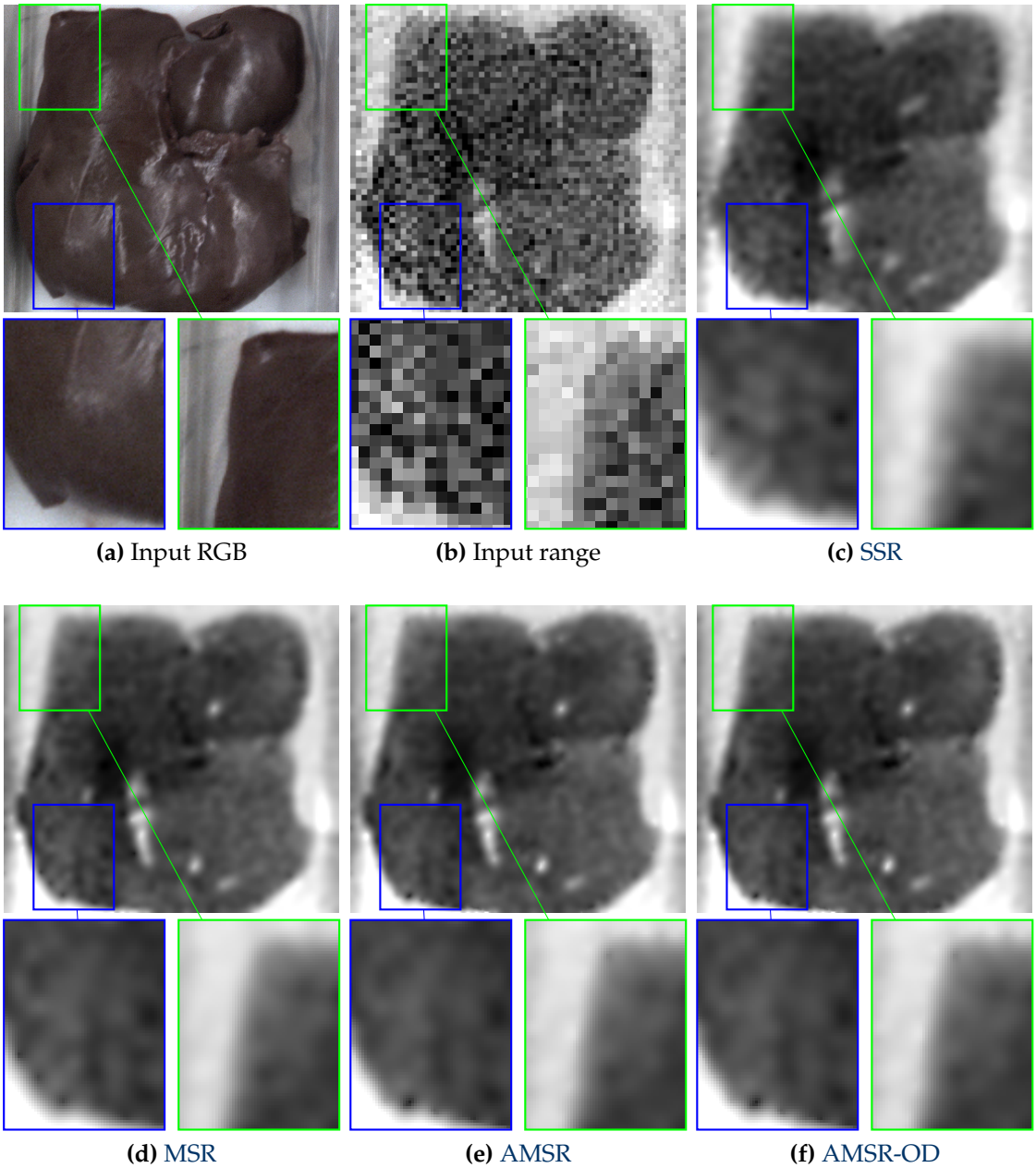
color images were fused with the range data based on the intrinsic and extrinsic camera parameters.

The motion required for super-resolution was induced by small vibrations of the tripod. Super-resolution was performed with magnification factor $s = 4$ using $K = 31$ consecutive range images, where the central one was chosen as reference for optical flow [Liu 09].

Reconstruction Results. Figure 8.10 shows a qualitative comparison among low-resolution range data acquired under this setup and super-resolution using the different reconstruction algorithms. In this use case, super-resolution aims at reconstructing reliable surface information of the measured porcine liver.

Similar to the ex-vivo experiments for minimally invasive surgery, one can observe that direct optical flow estimation on low-resolution range data was error prone. This resulted in a poor surface reconstruction provided by the single-sensor approach (SSR), which is particularly visible by blurred boundaries of the porcine liver. The different multi-sensor approaches (MSR, AMSR, and AMSR-OD) were less sensitive to this issue due to the higher reliability of the filter-based motion estimation driven by the color images. Note that in this application, the quality gain achieved by spatially adaptive regularization was limited since surfaces were typically more smooth and thus edges could not be exploited by the underlying regularization technique. Moreover, outlier detection did not achieved substantial quality gains due to the higher reliability of the motion estimate and the measured range data.

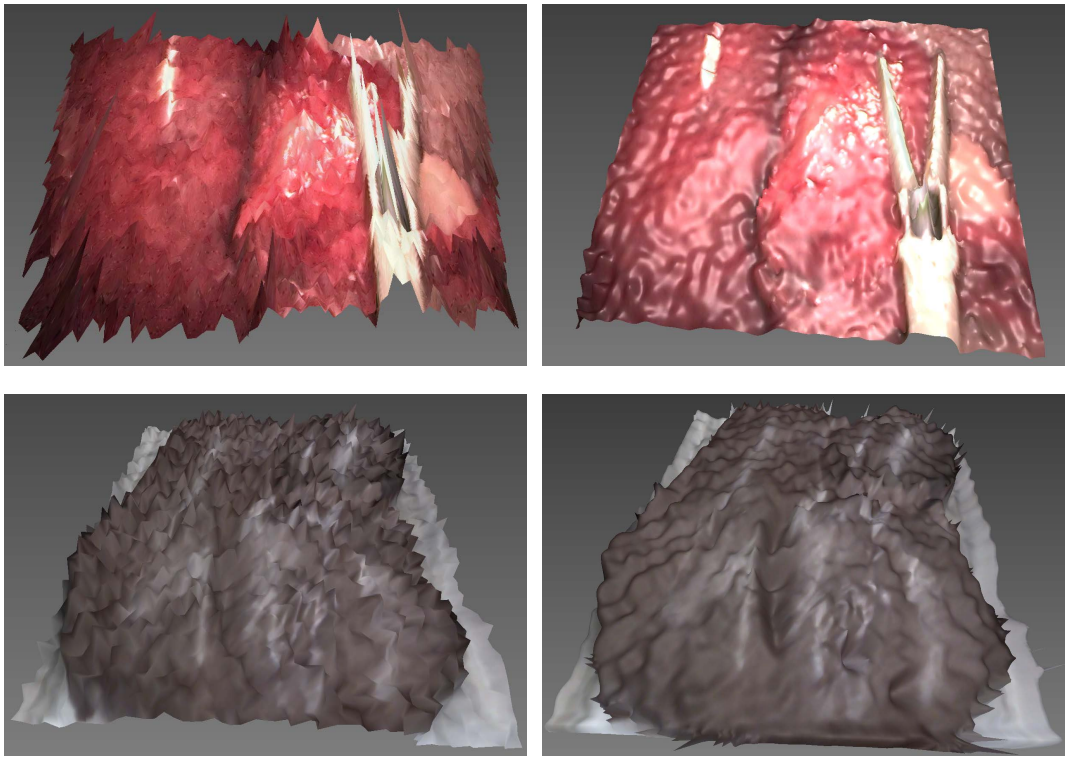
Range Data Quality Assessment. Super-resolved and low-resolution range data was quantitatively assessed in six image regions per dataset. The statistics of Q_{snr} and Q_{edge} evaluated on four datasets are summarized in Tab. 8.3. In comparison to raw range data, combining the multi-sensor techniques (AMSR-OD) leads to an increase of Q_{snr} and Q_{edge} by 24 % and 72 %, respectively. The multi-sensor reconstruction algorithms also outperformed the single-sensor approach (SSR) regarding the reconstruction of depth discontinuities. It is worth noting that this is essential for the further usage of range information for image guidance, e. g. in augmented reality [Mers 11, Kilg 15].



(a) Input RGB **(b)** Input range **(c)** SSR

(d) MSR **(e)** AMSR **(f)** AMSR-OD

Figure 8.10: Super-resolution reconstruction ($K = 31$ frames, $4\times$ magnification) for ex-vivo experiments in image-guided open surgery on a porcine liver. First and third row: high-resolution color image and low-resolution range data in comparison to super-resolved range images obtained by single-sensor super-resolution (SSR) as well as the different multi-sensor algorithms (MSR, AMSR, and AMSR-OD). Second and fourth row: zoom-in for different areas on the boundary of the porcine liver. Notice that multi-sensor super-resolution considerably enhanced the accuracy of the reconstructed liver surface.



(a) Original **(b)** Super-resolution (4× magnification)

Figure 8.11: Fusion of color and range information depicted as a textured 3-D mesh visualization in hybrid 3-D endoscopy (top row) and image guidance for open surgery (bottom row). (a) 3-D mesh based on low-resolution range data acquired with ToF sensors. (b) 3-D mesh based on super-resolved range data obtained by multi-sensor super-resolution (4× magnification). Figure reused from [Kohl 15b] with the publisher's permission.

8.4 Conclusion

This chapter investigated super-resolution in image-guided surgery based on hybrid range imaging. One of the fundamental limitations of this technology towards clinical applications is the low spatial resolution of current range sensors. In order to alleviate this issue, multi-sensor super-resolution was adopted to gain high-resolution range images from low-resolution ones using color images as guidance. To this end, domain-specific system calibration schemes to enable sensor data fusion among range and color images were introduced.

Two application areas were investigated in a simulation study as well as ex-vivo experiments on porcine organs: 1) hybrid 3-D endoscopy for minimally invasive surgery, and 2) image guidance in open surgery. In both areas, super-resolution enhanced the reliability of surface information and enabled the reconstruction of anatomical structures or artificial objects like surgical instruments that were barley detectable in low-resolution measurements. In ex-vivo experiments using ToF sensors, the proposed techniques improved the reliability of surface and depth discontinuity measurements compared to raw range data by more than 24 % and 68 %, respectively. Super-resolved range data can be augmented with

high-resolution color images for a comprehensive representation of the measured scene, see Fig. 8.11. This can be considered as a key requirement regarding applications for computer-assisted interventions, e. g. tracking [Haas 13d] and segmentation [Haas 13c] of surgical instruments as well as augmented reality [Mers 11] to name a few.

Different to related filter-based preprocessing techniques proposed for image-guided surgery [Wasz 11b], super-resolution appropriately models physical effects of image formation like motion among successive frames or the underlying camera PSF. The general-purpose model used in this work can be extended even further by domain-specific effects like specular highlights or intensity related uncertainty of range measurements [Reyn 11]. However, super-resolution comes to a higher computational effort, which means a practical limitation for specific workflows with real-time constraints. Therefore, future work needs to consider an efficient implementation to enable real-time processing. This might be achieved by the use of efficient motion estimation methods [Plye 14] or by a massively parallel implementation of the reconstruction algorithm [Wetz 13].

Part IV

Summary and Outlook

CHAPTER 9

Summary

Multi-frame super-resolution is a software-based approach to overcome physical limitations regarding the spatial resolution of digital sensor technologies. This pursues the objective of enabling high-resolution imagery based on cost-effective systems retrospectively. To this end, we investigated novel computational methods and their applications with emphasis on medical imaging workflows. This thesis covers both, classical super-resolution applied to image data of a single modality as well as multi-sensor super-resolution in hybrid imaging.

Chapter 2 of this work concerned a theoretical study of multi-frame super-resolution using the Fourier transform. More specifically, we described super-resolution from a signal processing point of view as a linear inverse problem in multi-channel sampling. This concept was utilized to show the relationship to the Nyquist-Shannon sampling theorem and to discuss the meaning of the magnification factor. Eventually, we proved and discussed conditions to achieve uniqueness of super-resolution reconstruction.

In the sequel, the main research findings of this thesis were divided into three parts.

Numerical Methods for Multi-Frame Super-Resolution. Part I concerned the development of super-resolution methods for a single imaging modality. We focused on multi-frame algorithms that aim at reconstructing single high-resolution images from sequences of low-resolution frames by exploiting subpixel motion across the input images.

Chapter 3 introduced the computational framework utilized throughout this work along with a review on current super-resolution paradigms. We introduced a mathematical model to describe the physics of digital imaging. This image formation model was discretized to make it accessible for the development of numerical algorithms from a Bayesian estimation perspective. More specifically, as two of the most fundamental approaches, we presented maximum likelihood (ML) and maximum a-posteriori (MAP) estimation to gain point estimates of a latent high-resolution image from noisy, low-resolution observations. Numerous algorithms, building upon the Bayesian paradigm, that have been developed over the past years are prone to fail under practical conditions. Most of these methods are particularly sensitive regarding model parameter uncertainties like inaccurate subpixel motion estimation.

Chapter 4 proposed a novel algorithm to meet the requirements regarding robustness in real-world applications. To this end, we introduced a weighted Bayesian observation model to consider outliers in the reconstruction algorithm.

Furthermore, we introduced a weighted prior distribution that encourages sparsity to model the statistical appearance of natural images. Super-resolution was implemented as iteratively re-weighted energy minimization to simultaneously estimate high-resolution images and latent model confidence weights. We showed the relationship of this iteration scheme to majorization-minimization (MM) algorithms and rigorously proved its convergence. In comparative experimental evaluations with focus on challenging real-world conditions like space variant noise, inaccurate motion estimation, or photometric variations, the proposed method outperformed the state-of-the-art. For instance, in a benchmark with inaccurate motion estimation, iteratively re-weighted minimization improved the peak-signal-to-noise ratio (PSNR) by 0.7 decibel (dB) and the structural similarity (SSIM) by 0.04 compared to related robust algorithms. The optimization algorithm also relies on a minimal amount of manual parameter tuning making it attractive for real applications. It was also further customized in the super-resolution algorithms developed in the remainder of this thesis.

Multi-Sensor Super-Resolution for Hybrid Imaging. Part II concerned super-resolution for multiple modalities. In this area, referred to as hybrid imaging, we studied two complementary problem statements.

Chapter 5 introduced multi-sensor super-resolution for a single modality under the guidance of a second one. We studied the case that both modalities are co-registered but complementary regarding their spatial resolutions. Accordingly, we proposed a guidance image driven framework comprising three key components: First, filter-based motion estimation is used to obtain displacement fields from optical flow on high-resolution guidance data to avoid error-prone motion estimation on low-resolution frames. Second, feature-based adaptive regularization is used to exploit correlations in terms of discontinuities between low-resolution and guidance data. Third, outlier detection using iteratively re-weighted minimization driven by image similarity assessment on the guidance data is employed. These techniques were validated for hybrid 3-D range imaging, where high-quality color images steer super-resolution of range data. Overall, the multi-sensor methodology led to gains of 0.9 dB in terms of peak-signal-to-noise ratio (PSNR) and 0.02 in terms of structural similarity (SSIM) over a straightforward application of super-resolution solely on the range data.

Chapter 6 generalized multi-sensor super-resolution to jointly super-resolve a set of modalities. To this end, we dropped the usage of guidance data to facilitate a wider range of hybrid imaging setups. This methodology builds on multi-channel images as the underlying mathematical concept. Its key notion is the consideration of mutual dependencies between image channels in a Bayesian model. Different to feature-based regularization, mutual dependencies are captured by a novel locally linear regression (LLR) prior. This model was used to develop an alternating minimization scheme building upon the robust algorithm presented in Chapter 4. It is applicable in color-, multispectral-, and range imaging as well as further applications beyond classical multi-frame super-resolution like joint segmentation and resolution enhancement. As the primary insight, multi-channel reconstructions outperformed sequential channel-wise reconstructions that essen-

tially ignore inter-channel dependencies. In color imaging as a classical use case, the proposed method led to a gain of 1.5 dB in terms of PSNR and 0.04 in terms of SSIM compared to channel-wise super-resolution of color images.

Super-Resolution in Medical Imaging. Part III addressed applications in medical imaging with focus on diagnostic and interventional use cases. The methods investigated in this part pursue the common goal of overcoming the resolution limitations of recently developed imaging technologies as important step towards their clinical use.

Chapter 7 presented a new framework to approach super-resolution in the area of retinal imaging. This framework targets at the reconstruction of a high-resolution retinal image from a low-resolution video acquired from the human eye background. For this purpose, we introduced an image formation model tailored to the conditions of retinal video imaging and exploited natural eye movements. Moreover, we presented a quality self-assessment scheme to estimate a high-resolution image driven by an objective no-reference measure of image noise and sharpness. This method was evaluated for low-cost retinal imaging on real video data of healthy subjects and glaucoma patients. In this study, it led to an image quality comparable to those of commercially available, but expensive and stationary cameras. Furthermore, super-resolution was able to enhance common image analysis tasks like automatic blood vessel segmentation, where it increased the sensitivity by 10 % compared to a direct segmentation on low-resolution images. The proposed method can serve as a valuable tool for high-resolution imagery in clinical workflows with high demands on cost-efficiency and mobility, e. g. screening applications.

Chapter 8 examined super-resolution to aid hybrid range imaging for image-guided surgery. This concerns an adoption of the multi-sensor framework introduced in Chapter 5, where high-resolution color images steer super-resolution on low-resolution range data. To make this method usable for this particular domain, we introduced two system calibration schemes for sensor data fusion among both modalities: a beam splitter setup to measure geometric and photometric information through a single optical system as well as a stereo vision setup that combines distinct cameras. We conducted comprehensive experiments in two fields of today's surgery using these setups, namely 3-D endoscopy for minimally invasive procedures and image guidance for open surgery. In ex-vivo experiments using Time-of-Flight (ToF) sensors, multi-sensor super-resolution improved the reliability of surface and depth discontinuity measurements compared to raw range data by more than 24 % and 68 %, respectively. This is an essential step towards reliable geometric measurements of anatomical structures or artificial objects like surgical instruments. In combination with high-resolution photometric information, this can provide surgeons a comprehensive view of the underlying scene.

CHAPTER 10

Outlook

Apart from the theory and applications investigated in this thesis, there is a great number of opportunities for future work. Below, we summarize several promising directions for further research that are related to this work.

Extension of the Image Formation Model. Throughout this work, we limited ourselves to the algorithm design for linear image formation models that build upon several idealizing assumptions. One of the pitfalls is the assumption of isotropic and space invariant blur related to the camera point spread function (PSF) that is known a priori. This assumption is reasonable in case of optical blur as mainly considered in the investigated applications but might be violated under different conditions. Some typical examples include atmospheric or motion blur, where the modeling by simple isotropic kernels is inappropriate. Consequently, the underlying image formation model needs to be revised to tackle these effects. However, recent attempts to handle motion blur [Ma 15] or more general space variant models [Sore 10] might provide a basis towards tackling these challenging situations.

Another crucial limitation is the assumption that raw data untouched by camera internal preprocessing is accessible by super-resolution algorithms via the camera interface. This is convincing for scientific or medical applications but might be violated by low-cost consumer cameras that employ compression codecs, which limits the efficiency of super-resolution. Thus, modeling data compression is important to break into new application areas. Related work considered this aspect by new image priors tailored to compressed video reconstruction [Bele 09].

Despite the simplicity of the underlying model and the aforementioned limitations, the modular design of the proposed algorithms developed from a Bayesian perceptive makes them flexible regarding revisions or extensions. This enables the tailoring of these algorithms to new domains either by adapting the image formation model or by considering new effects in the design of image priors.

Extension to Video Super-Resolution. This thesis considered the use case of reconstructing single images of enhanced spatial resolution from a set of low-resolution frames. Consequently, super-resolution buys an improved spatial resolution at the price of a decreased temporal one that is lost in the reconstruction. One interesting extension comprises *video super-resolution* that targets at the simultaneous estimation of an entire high-resolution video from a low-resolution one. Although this could be approached by a successive use of image super-resolution in a temporal sliding window mode, special algorithms have already been intro-

duced to solve this highly ill-posed problem by exploiting temporal consistencies in natural videos [Zibe 07, Dirk 16]. This can also be achieved by fast incremental algorithms [Su 11] to accelerate video super-resolution.

As most of the presented algorithms are extendable by these concepts, one promising direction for future work is their transfer from classical image to video super-resolution. In particular, the robust estimation techniques proposed in Chapter 4 provide a sensible basis to approach video super-resolution that involves additional issues regarding robustness. Such techniques might also contribute to new applications that would benefit from additional temporal information, e. g. diagnostic medical imaging investigated in Chapter 7.

Extension to Learning-Based Methods. In contrast to this work that approaches super-resolution in an unsupervised way, *learning-based* methods gained enormous interest over the past years. This class of algorithms aims at learning the mapping among low-resolution and high-resolution images from training data. This can be done via sparse signal representation and dictionary learning [Yang 10]. More recently, current deep learning architectures made their entrance into single-image super-resolution including deep convolutional neural networks [Dong 14, Kim 16a, Kim 16b] or generative adversarial learning [Ledi 16].

Such architectures are also extendable to the multi-frame case [Liao 15, Kapp 16]. In spite of their success as demonstrated in recent works, these methods heavily rely on the existence of large training datasets to learn the mappings among the low-resolution and high-resolution domains. Nevertheless, given a reliable training, they enable efficient resolution enhancement in contrary to algorithms that involve time-consuming numerical optimizations based on generative modeling. Hence, future research needs to consider *hybrid* super-resolution schemes by combining the individual strengths of these complementary paradigms.

Theoretical Considerations. An essential question in the research of image super-resolution is the question whether there exist fundamental limits of the investigated algorithms. More specifically, it is worthwhile to derive upper bounds regarding the spatial resolution reachable by super-resolution and thus an effective magnification factor. A basic study of this question under ideal conditions in the Fourier domain comprising noise-free sampling was presented in Chapter 2. This demonstrated that the effective magnification is bounded by the band limitation of the signal that needs to be reconstructed.

Several attempts have been made to derive fundamental limits in more general situations but based on simplifying assumptions. In [Bake 02], Baker and Kanade reported that classical reconstruction-based algorithms as studied in this thesis tend to be less profitable under an increasing magnification factor. Later, Lin and Shum [Lin 04] presented quantitative statements for this fact using the perturbation theory of linear systems under translational motion and a box shaped PSF. Tanaka and Okutomi [Tana 05] extended these studies for an arbitrary space invariant PSF by formulating the condition number theorem. However, in addition to a pure translation, the underlying assumption of an infinite number of low-resolution observations means an oversimplification. Thus, one unsolved aspect

in the research community is the derivation of tighter bounds regarding the performance of super-resolution. This needs to consider real-world conditions including more general motion models and different sources of error like image noise.

Practical Considerations. The practical aspects for future work concern the usability of the presented algorithms. This mainly includes two considerations.

One aspect is the amount of parameter calibration that comes along with the usage of super-resolution. This is due to the fact that most of the presented techniques involve several tuning parameters in the underlying image formation model or the employed numerical optimization algorithms. Some examples are the PSF kernel, scale parameters, or regularization weights. In Chapter 4, we introduced an optimization scheme that provides a scale and regularization parameter estimation to minimize the amount of manual parameter tuning. With a similar motivation, we proposed quality self-assessment in Chapter 7 to objectify the choice of a regularization weight in a particular application domain. Future work needs to extend these concepts to rigorously reduce the number of user-defined parameters following the same lines of thought as used in related field, e. g. image auto-denoising [Kong 13]. This also includes research in the area of objective quality assessment [Yega 12] as the main building block of these techniques. Such a fully automatic parameter tuning could lead to a further improved robustness, flexibility, and user-friendliness of super-resolution algorithms.

Another aspect is the consideration of the computational complexity. While the main scope of the proposed algorithms is accurate resolution enhancement in a retrospective way, their use is computationally demanding. Non-parallelized implementations of these methods yield run times in the range of seconds up to several minutes growing linearly with the most relevant parameters, i. e. the number of observed low-resolution pixels and the desired magnification. This might reduce the acceptability in time-critical environments and does not meet real-time constraints. Example applications concerned by this issue include computer-assisted interventions as addressed in Chapter 8. For these reasons, future work needs to study efficient implementations of the proposed algorithms. Promising opportunities towards an interactive use of super-resolution could exploit parallelizations of the reconstruction algorithm [Wetz 13] or the motion estimation [Plye 14] using modern graphics processing units. Other possibilities are hardware-based implementations, e. g. using field programmable gate arrays [Bowe 08].

CHAPTER A

Appendix

A.1 Multi-Frame Super-Resolution and the Sampling Theorem

A.1.1 Uniqueness for Ideal Sampling

In this appendix, we investigate the conditions to provide unique super-resolution for ideal and real sampling in the Fourier domain based on [Tsai 84, Kim 90, Teka 92]. Let us first consider the case of ideal sampling. The conditions to guarantee a unique solution are summarized by Theorem 2.2.

Theorem 2.2 (Uniqueness for ideal sampling). *Let $s = K$ be the super-resolution magnification factor and K be the number of channels in a multi-channel sampling process, where t_i with $i = 1, \ldots, K$ and $t_1 = 0$ are the corresponding channel offsets and T is the sampling pitch. Then, the solution of the linear inverse problem in Eq. (2.18) is unique if and only if: $t_j \neq c_1 t_i + c_2 T$ for all $1 \leq i < j \leq K$ and $c_1, c_2 \in \mathbb{Z}$.*

Proof. For a unique solution of the linear problem in Eq. (2.18), the matrix $\boldsymbol{W}$ needs to be non-singular. Due to the block structure of $\boldsymbol{W}$ according to Eq. (2.16), we consider the reconstruction of frequencies $X_{n,-L}, \ldots, X_{n,L-1}$ for $n = 1, \ldots, N$. This can be written as the solution of the linear equation system:

$$\begin{pmatrix} \mathcal{Y}^{(1)}[n] \\ \mathcal{Y}^{(2)}[n] \\ \vdots \\ \mathcal{Y}^{(K)}[n] \end{pmatrix} = \underbrace{\begin{pmatrix} W^{(1)}_{n,-L} & W^{(1)}_{n,-L+1} & \cdots & W^{(1)}_{n,L-1} \\ W^{(2)}_{n,-L} & W^{(2)}_{n,-L+1} & \cdots & W^{(2)}_{n,L-1} \\ \vdots & \vdots & & \vdots \\ W^{(K)}_{n,-L} & W^{(K)}_{n,-L+1} & \cdots & W^{(K)}_{n,L-1} \end{pmatrix}}_{\boldsymbol{W}_n} \begin{pmatrix} X_{n,-L} \\ X_{n,-L+1} \\ \vdots \\ X_{n,L-1} \end{pmatrix}, \tag{A.1}$$

where the elements $W^{(k)}_{n,m}$ are calculated according to Eq. (2.17) for $k = 1, \ldots, K$ and $m = -L, \ldots, L-1$. Notice that the overall system matrix $\boldsymbol{W}$ in Eq. (2.18) is non-singular iff the matrices $\boldsymbol{W}_n$ are non-singular for all $n = 1, \ldots, N$.

Based on Eq. (2.17), we can decompose $\boldsymbol{W}_n$ for an arbitrary sample index n with $n = 1, \ldots, N$ according to [Kim 90]:

$$\boldsymbol{W}_n = \boldsymbol{U}_n \boldsymbol{V}_n, \tag{A.2}$$

where $\boldsymbol{U}_n$ is a complex-valued diagonal matrix that contains the non-zero elements $U_{n,kk} = \exp\left(-j2\pi f_s t_k \left(\frac{n}{N} + L\right)\right)$ for $k = 1\ldots, K$. $\boldsymbol{V}_n$ is given by the Vandermonde matrix:

$$\boldsymbol{V}_n = \begin{pmatrix} 1 & \exp\left(j2\pi f_s t_1\right) & \exp\left(j2\pi f_s t_1\right)^2 & \ldots & \exp\left(j2\pi f_s t_1\right)^{2L-1} \\ 1 & \exp\left(j2\pi f_s t_2\right) & \exp\left(j2\pi f_s t_2\right)^2 & \ldots & \exp\left(j2\pi f_s t_2\right)^{2L-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \exp\left(j2\pi f_s t_K\right) & \exp\left(j2\pi f_s t_K\right)^2 & \ldots & \exp\left(j2\pi f_s t_K\right)^{2L-1} \end{pmatrix}, \tag{A.3}$$

where the nodes $\exp\left(j2\pi f_s t_k\right)$ are located on the complex unit circle.

Since $\boldsymbol{U}_n$ is always non-singular, $\boldsymbol{W}_n$ is non-singular iff $\boldsymbol{V}_n$ is non-singular. That is, the Vandermonde determinant $\det(\boldsymbol{V}_n)$ of $\boldsymbol{V}_n$ needs to be non-zero, which is the case for distinct nodes [Horn 12]. Thus, we have:

$$\begin{aligned} \det(\boldsymbol{V}_n) = \prod_{i=1}^{K} \prod_{j=i+1}^{K} \left(\exp\left(j2\pi f_s t_i\right) - \exp\left(j2\pi f_s t_j\right) \right) \neq 0 \\ \Leftrightarrow \exp\left(j2\pi f_s t_i\right) - \exp\left(j2\pi f_s t_j\right) \neq 0 \\ \Leftrightarrow \exp\left(j2\pi f_s t_i\right) \neq \exp\left(j2\pi f_s t_j\right), \end{aligned} \tag{A.4}$$

for all $1 \leq i < j \leq K$. In order to guarantee a non-zero determinant $\det(\boldsymbol{V}_n)$ and thus a unique solution of Eq. (2.18), it follows:

$$\begin{aligned} \det(\boldsymbol{V}_n) \neq 0 \Leftrightarrow \quad & \exp\left(j2\pi f_s t_i\right) \neq \exp\left(j2\pi f_s t_j\right) \\ \Leftrightarrow \quad & t_j \neq c_1 t_i + c_2 \frac{1}{f_s} \\ \Leftrightarrow \quad & t_j \neq c_1 t_i + c_2 T, \end{aligned} \tag{A.5}$$

for all $1 \leq i < j \leq K$ and $c_1, c_2 \in \mathbb{Z}$. This shows the desired condition and completes the proof. □

A.1.2 Uniqueness for Real Sampling

Let us now prove the conditions regarding uniqueness in case of real sampling. These conditions are summarized by the following theorem.

Theorem 2.3 (Uniqueness for real sampling). *Let $s = K$ be the super-resolution magnification factor and K be the number of channels in multi-channel sampling with offsets $t_i = 0$ for all $i = 1, \ldots, K$ and sampling pitch T. Each channel $x^{(i)}(t)$ is affected by a blur kernel $H^{(i)}(f)$ denoted by $\boldsymbol{H}^{(i)}$ in matrix notation. Then, the solution of the linear inverse problem in Eq. (2.20) is unique if and only if:*

1. $\sum_{i=1}^{K} c_i \boldsymbol{H}^{(i)} \neq \boldsymbol{0}$ *for all* $c_i \neq 0$ *and* $i = 1, \ldots, K$ *(linear independent blur kernels)*
2. $\sum_{i=1}^{K} \left| H^{(i)}\left(\frac{n}{N} f_s + m f_s\right) \right| \neq 0$ *for all* $m = -L, \ldots, L-1$ *(kernel cut-off frequency)*

Proof. If we consider the reconstruction of frequencies $X_{n,-L},\dots,X_{n,L-1}$ for an arbitrary n with $n = 1,\dots,N$ and exploit the block structure of the matrix $\boldsymbol{H}$ in Eq. (2.20), the corresponding linear system is given by:

$$\begin{pmatrix} \mathcal{Y}^{(1)}[n] \\ \mathcal{Y}^{(2)}[n] \\ \vdots \\ \mathcal{Y}^{(K)}[n] \end{pmatrix} = \underbrace{\begin{pmatrix} H^{(1)}_{n,-L} & H^{(1)}_{n,-L+1} & \dots & H^{(1)}_{n,L-1} \\ H^{(2)}_{n,-L} & H^{(2)}_{n,-L+1} & \dots & H^{(2)}_{n,L-1} \\ \vdots & \vdots & & \vdots \\ H^{(K)}_{n,-L} & H^{(K)}_{n,-L+1} & \dots & H^{(K)}_{n,L-1} \end{pmatrix}}_{\boldsymbol{H}_n} \begin{pmatrix} X_{n,-L} \\ X_{n,-L+1} \\ \vdots \\ X_{n,L-1} \end{pmatrix}, \tag{A.6}$$

where the elements $H^{(k)}_{n,m}$ in the quadratic matrix $\boldsymbol{H}_n$ are calculated according to the corresponding blur kernels for $k = 1,\dots,K$ and $m = -L,\dots,L-1$. Let us assume that the channel offsets in Eq. (2.20) are given by $t_i = 0$ for all $i = 1,\dots,K$. Then, we can assemble the matrix $\boldsymbol{H}_n$ according to:

$$\boldsymbol{H}_n = \begin{pmatrix} H^{(1)}\left(\left(\frac{n}{N}+L\right)f_s\right) & H^{(1)}\left(\left(\frac{n}{N}+(L-1)\right)f_s\right) & \dots & H^{(1)}\left(\left(\frac{n}{N}-(L-1)\right)f_s\right) \\ H^{(2)}\left(\left(\frac{n}{N}+L\right)f_s\right) & H^{(2)}\left(\left(\frac{n}{N}+(L-1)\right)f_s\right) & \dots & H^{(2)}\left(\left(\frac{n}{N}-(L-1)\right)f_s\right) \\ \vdots & \vdots & & \vdots \\ H^{(K)}\left(\left(\frac{n}{N}+L\right)f_s\right) & H^{(K)}\left(\left(\frac{n}{N}+(L-1)\right)f_s\right) & \dots & H^{(K)}\left(\left(\frac{n}{N}-(L-1)\right)f_s\right) \end{pmatrix}. \tag{A.7}$$

Note that $\boldsymbol{H}$ in Eq. (2.20) is non-singular iff $\boldsymbol{H}_n$ has full rank for all $n = 1,\dots,N$. A full rank of the quadratic matrix $\boldsymbol{H}_n$ is equivalent to linearly independent rows. That is:

$$\sum_{i=1}^{K} c_{n,i} \begin{pmatrix} H^{(i)}\left(\left(\frac{n}{N}+L\right)f_s\right) & \dots & H^{(i)}\left(\left(\frac{n}{N}-(L-1)\right)f_s\right) \end{pmatrix}^{\top} \neq \boldsymbol{0}, \tag{A.8}$$

for all $c_{n,i} \neq 0$. This translates into:

$$\sum_{i=1}^{K} c_i \boldsymbol{H}^{(i)} \neq \boldsymbol{0}, \tag{A.9}$$

for all $c_i \neq 0$, which is equivalent to independent blur kernels and shows the first condition in Theorem 2.3. Moreover, we need to guarantee that all columns of $\boldsymbol{H}_n$ are non-zero. That is:

$$\sum_{i=1}^{K} \left| H^{(i)}\left(\frac{n}{N}f_s + m f_s\right)\right| \neq 0 \tag{A.10}$$

for all $m = -L,\dots,L-1$. This shows the second condition in Theorem 2.3 and completes the proof.

□

A.2 Robust Multi-Frame Super-Resolution with Sparse Regularization

A.2.1 Relationship to Majorization-Minimization Algorithms

In this appendix, we prove Theorem 4.1 to establish the connection between iteratively re-weighted minimization and majorization-minimization (MM) algorithms. First, let us present several important properties of the Huber loss and the mixed L_1/L_p norm. The following lemma states that the Huber loss function can be written as the solution of a weighted minimization problem.

Lemma A.1. *The Huber loss function $\phi_{\text{Huber}}(z)$ in Eq. (4.43) can be written as a weighted minimization problem:*

$$\phi_{\text{Huber}}(z) = \min_{\beta \in \mathbb{R}_0^+} \left\{ \beta z^2 + \sigma_{\text{noise}}^2 \rho(\beta) \right\} \tag{A.11}$$

$$\rho(\beta) = \begin{cases} \frac{1}{\beta} - 1 & \text{if } 0 \leq \beta < 1 \\ 0 & \text{if } \beta \geq 1. \end{cases} . \tag{A.12}$$

Proof. Obviously, $\phi_{\text{Huber}}(z)$ is a convex function and when $\beta \geq 1$, it is monotonically increasing. Thus, the optimal weight is $\beta^* = 1$ in case of $z^2 \leq \sigma_{\text{noise}}^2$ or $\beta^* = \sigma_{\text{noise}}/|z|$ in case of $z^2 > \sigma_{\text{noise}}^2$, where the later comes from the first order optimality condition. Comparing the objective values we get the optimal weight:

$$\beta^* = \begin{cases} 1 & \text{if } |z| \leq \sigma_{\text{noise}} \\ \frac{\sigma_{\text{noise}}}{|z|} & \text{otherwise} \end{cases} . \tag{A.13}$$

Therefore, the solution of the weighted minimization in Eq. (A.11) yields:

$$\min_{\beta \in \mathbb{R}_0^+} \left\{ \beta z^2 + \sigma_{\text{noise}}^2 \rho(\beta) \right\} = \begin{cases} z^2 & \text{if } |z| \leq \sigma_{\text{noise}} \\ 2\sigma_{\text{noise}}|z| - \sigma_{\text{noise}}^2 & \text{otherwise} \end{cases} , \tag{A.14}$$

which coincides with the Huber loss $\phi_{\text{Huber}}(z)$ in Eq. (4.43). □

Next, let us employ the weighted minimization problem in (A.11) to define a majorizing function for the Huber loss. This function is provided by the following lemma.

Lemma A.2. *The Huber loss $\phi_{\text{Huber}}(z)$ in Eq. (4.43) is majorized at z^{t-1} by:*

$$\tilde{\phi}_{\text{Huber}}\left(z, z^{t-1}\right) = \begin{cases} z^2 & \text{if } |z^{t-1}| \leq \sigma_{\text{noise}} \\ \frac{\sigma_{\text{noise}}}{|z^{t-1}|} z^2 + \sigma_{\text{noise}}^2 \left(\frac{|z^{t-1}|}{\sigma_{\text{noise}}} - 1 \right) & \text{otherwise} \end{cases} . \tag{A.15}$$

Proof. For this proof, let us first assume that $|z^{t-1}| \leq \sigma_{\text{noise}}$. Then, $\tilde{\phi}_{\text{Huber}}(z, z^{t-1})$ coincides with $\phi_{\text{Huber}}(z)$ according to the definitions of both functions.

If $|z^{t-1}| > \sigma_{\text{noise}}$ and $\beta = \sigma_{\text{noise}}/|z^{t-1}| \in [0,1]$, the function $\tilde{\phi}_{\text{Huber}}(z, z^{t-1})$ can be reformulated according to Lemma A.1:

$$\begin{aligned}\tilde{\phi}_{\text{Huber}}\left(z, z^{t-1}\right) &= \beta z^2 + \sigma_{\text{noise}}^2 \rho(\beta) \\ &= \frac{\sigma_{\text{noise}}}{|z^{t-1}|} z^2 + \sigma_{\text{noise}} \left|z^{t-1}\right| - \sigma_{\text{noise}}^2 \\ &\geq \phi_{\text{Huber}}(z)\,,\end{aligned} \tag{A.16}$$

where the equality holds true for $z = z^{t-1}$. Thus, $\tilde{\phi}_{\text{Huber}}(z, z^{t-1})$ is a majorizing function for $\phi_{\text{Huber}}(z)$ at z^{t-1}. □

Next, we define a majorizing function for the mixed L_1/L_p norm. This function is established by the following lemma.

Lemma A.3. *The mixed L_1/L_p norm $\phi_p(z)$ in Eq. (4.44) for $p \in [0,1]$ is majorized at z^{t-1} by:*

$$\tilde{\phi}_p\left(z, z^{t-1}\right) = \begin{cases} |z| & \text{if } |z^{t-1}| \leq \sigma_{\text{prior}} \\ p\left(\frac{\sigma_{\text{prior}}}{|z^{t-1}|}\right)^{1-p} |z| + (1-p)\sigma_{\text{prior}}^{1-p} |z^{t-1}|^p & \text{otherwise} \end{cases}. \tag{A.17}$$

Proof. Let us first consider the case $|z^{t-1}| \leq \sigma_{\text{prior}}$. Then, $\tilde{\phi}_p(z^{t-1}, z^{t-1})$ coincides with $\phi_p(z^{t-1})$ according to the definitions of both functions.

Let us now consider the case $|z^{t-1}| > \sigma_{\text{prior}}$. Since $\phi_p(z)$ is monotone and concave downwards, the Taylor series expansion for $|z| > \sigma_{\text{prior}}$ yields the inequality:

$$\begin{aligned}\phi_p(z) &\leq \phi_p\left(z^{t-1}\right) + \left(z - z^{t-1}\right) \cdot \frac{d}{dz}\phi_p(z)\Big|_{z=z^{t-1}} \\ &= \sigma_{\text{prior}}^{1-p}\left|z^{t-1}\right|^p + \left(z - z^{t-1}\right) \cdot \text{sign}\left(z^{t-1}\right) \sigma_{\text{prior}}^{1-p} p \left|z^{t-1}\right|^{p-1} \\ &= p\left(\frac{\sigma_{\text{prior}}}{|z^{t-1}|}\right)^{1-p} |z| + (1-p)\sigma_{\text{prior}}^{1-p}\left|z^{t-1}\right|^p \\ &\leq \tilde{\phi}_p\left(z, z^{t-1}\right),\end{aligned} \tag{A.18}$$

and $\phi_p(z^{t-1}) = \tilde{\phi}_p(z^{t-1}, z^{t-1})$ if $z = z^{t-1}$. Hence, $\tilde{\phi}_p(z, z^{t-1})$ is a majorizing function for $\phi_p(z)$ at z^{t-1}. □

Notice that the majorizing functions $\tilde{\phi}_{\text{Huber}}(z, z^{t-1})$ and $\tilde{\phi}_p(z, z^{t-1})$ are non-negative and provide upper bounds for $\phi_{\text{Huber}}(z)$ and $\tilde{\phi}_p(z)$, respectively. Hence, minimization can be performed by an MM algorithm [Hunt 04]. Based on these properties, we can establish Theorem 4.1.

Theorem 4.1. *The convex energy function $\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1})$ in Eq. (4.45) is a majorizing function for the non-convex energy function $F(\boldsymbol{x})$ in Eq. (4.42) at $\boldsymbol{x} = \boldsymbol{x}^{t-1}$.*

Proof. According to Lemma A.2 and since the majorization relation is closed under the formation of a sum of non-negative terms [Hunt 04], it follows:

$$\begin{aligned}\tilde{L}\left(\boldsymbol{x}, \boldsymbol{x}^{t-1}\right) &= (\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x})^{\top}\boldsymbol{B}^{t}(\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x}) + \sum_{i=1}^{KM}\rho\left(\left[\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x}^{t-1}\right]_i\right) \\ &\geq L(\boldsymbol{x}),\end{aligned} \tag{A.19}$$

with equality for $\boldsymbol{x} = \boldsymbol{x}^{t-1}$, where $\rho(\cdot)$ is given by (4.46) and:

$$L(\boldsymbol{x}) = \sum_{i=1}^{KM}\phi_{\text{Huber}}([\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x}]_i). \tag{A.20}$$

Hence, the confidence-aware data fidelity term $\tilde{L}\left(\boldsymbol{x}, \boldsymbol{x}^{t-1}\right)$ is a majorizing function for $L(\boldsymbol{x})$. Similarly, Lemma A.3 yields:

$$\begin{aligned}\tilde{R}\left(\boldsymbol{x}, \boldsymbol{x}^{t-1}\right) &= \left|\left|\boldsymbol{A}^{t}\boldsymbol{S}\boldsymbol{x}\right|\right|_1 + \sum_{i=0}^{N_S}\tau\left(\left[\boldsymbol{S}\boldsymbol{x}^{t-1}\right]_i\right) \\ &\geq R(\boldsymbol{x}),\end{aligned} \tag{A.21}$$

with equality for $\boldsymbol{x} = \boldsymbol{x}^{t-1}$, where $\tau(\cdot)$ is given by (4.47) and:

$$R(\boldsymbol{x}) = \sum_{i=1}^{N_S}\phi_p([\boldsymbol{S}\boldsymbol{x}]_i). \tag{A.22}$$

Thus, the regularization term $\tilde{R}\left(\boldsymbol{x}, \boldsymbol{x}^{t-1}\right)$ is a majorizing function for $R(\boldsymbol{x})$. Then, $\tilde{F}(\boldsymbol{x}, \boldsymbol{x}^{t-1}) = \tilde{L}\left(\boldsymbol{x}, \boldsymbol{x}^{t-1}\right) + \lambda\tilde{R}\left(\boldsymbol{x}, \boldsymbol{x}^{t-1}\right)$ majorizes $F(\boldsymbol{x})$ in Eq. (4.42) at $\boldsymbol{x} = \boldsymbol{x}^{t-1}$ for $\lambda \geq 0$, which completes the proof. □

A.2.2 Convergence Analysis

This appendix provides the proof of Theorem 4.2 according to [Kohl 16b] to analyze the convergence of iteratively re-weighted minimization. Let us first present one important property of the L_1/L_p norm regularization term. For the sake of notational brevity, this regularization is reformulated according to:

$$R(\boldsymbol{z}) = \sum_{i\notin\mathcal{I}(\boldsymbol{z})}|z_i| + \sum_{i\in\mathcal{I}(\boldsymbol{z})}\sigma_{\text{prior}}^{1-p}|z_i|^p, \tag{A.23}$$

where the index set $\mathcal{I}(\boldsymbol{z})$ is defined as $\mathcal{I}(\boldsymbol{z}) = \{i : z_i > \sigma_{\text{prior}}\}$ based on the scale parameter σ_{prior}. Without loss of generality, we consider in the following analysis the case $\sigma_{\text{prior}} = 1$. Note that if $\sigma_{\text{prior}} \neq 1$, one can use a normalization of $\boldsymbol{z}$ to satisfy this condition. In this situation, for the L_1/L_p norm regularization term with the index set $\mathcal{I}(\boldsymbol{z})$, the following inequality is fulfilled:

Lemma A.4. *For all index sets $\mathcal{I}(\boldsymbol{z}) = \{i : z_i > 1\}$ and $\mathcal{I}'$ with the sparsity parameter p, where $p \in [0, 1]$, there is:*

$$\sum_{i\notin\mathcal{I}(\boldsymbol{z})}|z_i| + \sum_{i\in\mathcal{I}(\boldsymbol{z})}|z_i|^p \leq \sum_{i\notin\mathcal{I}'}|z_i| + \sum_{i\in\mathcal{I}'}|z_i|^p. \tag{A.24}$$

Proof. Subtracting the left hand side in Eq. (A.24) by the right hand side, we have:

$$\begin{aligned}
&\sum_{i\notin\mathcal{I}(\boldsymbol{z})}|z_i| - \sum_{i\notin\mathcal{I}'}|z_i| + \sum_{i\in\mathcal{I}(\boldsymbol{z})}|z_i|^p - \sum_{i\in\mathcal{I}'}|z_i|^p \\
&\quad= \sum_{i\in\mathcal{I}'\setminus\mathcal{I}(\boldsymbol{z})}|z_i| - \sum_{i\in\mathcal{I}(\boldsymbol{z})\setminus\mathcal{I}'}|z_i| + \sum_{i\in\mathcal{I}(\boldsymbol{z})\setminus\mathcal{I}'}|z_i|^p - \sum_{i\in\mathcal{I}'\setminus\mathcal{I}(\boldsymbol{z})}|z_i|^p \\
&\quad= \sum_{i\in\mathcal{I}'\setminus\mathcal{I}(\boldsymbol{z})}\left(|z_i| - |z_i|^p\right) - \sum_{i\in\mathcal{I}(\boldsymbol{z})\setminus\mathcal{I}'}\left(|z_i| - |z_i|^p\right).
\end{aligned} \tag{A.25}$$

Notice that for $p \in [0,1]$ it follows that $|z_i| \geq |z_i|^p$ if and only if $z_i \geq 1$, i. e., $i \in \mathcal{I}(\boldsymbol{z})$. Thus, $|z_i| - |z_i|^p < 0, \forall i \in \mathcal{I}'\setminus\mathcal{I}(\boldsymbol{z})$ and $|z_i| - |z_i|^p \geq 0, \forall i \in \mathcal{I}(\boldsymbol{z})\setminus\mathcal{I}'$. From these inequalities, it follows that:

$$\sum_{i\in\mathcal{I}'\setminus\mathcal{I}(\boldsymbol{z})}\left(|z_i| - |z_i|^p\right) - \sum_{i\in\mathcal{I}(\boldsymbol{z})\setminus\mathcal{I}'}\left(|z_i| - |z_i|^p\right) \leq 0.$$

Hence, the inequality in Eq. (A.24) is true for all index sets $\mathcal{I}(\boldsymbol{z})$ and $\mathcal{I}'$. □

Now, we can establish Theorem 4.2, which shows the convergence of iteratively re-weighted minimization in terms of the energy function value $F(\boldsymbol{x})$.

Theorem 4.2. *Let $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^T$ be an iteration sequence obtained by iteratively re-weighted minimization. Then, for all $t = 2, \ldots, T$ there exists a strict positive $\underline{\beta}$ such that:*

$$F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^t) \geq \underline{\beta}\left|\left|\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t\right|\right|_2^2. \tag{4.48}$$

Proof. For the sake of notational brevity, we present this proof by assuming the identity for the sparsifying transform, i. e. $\boldsymbol{S} = \boldsymbol{I}$. Note that in the general case where $\boldsymbol{S} \neq \boldsymbol{I}$, we can include the transform $\boldsymbol{S}$ to the optimization problem by re-formulation as a constrained problem.

According to Lemma A.1, we reformulate the energy function $F(\boldsymbol{x}^t)$ by writing the Huber loss as a weighted minimization problem:

$$\begin{aligned}
F(\boldsymbol{x}^t) &= \lambda R(\boldsymbol{x}^t) + \sum_{i=1}^{KM}\min_{\beta\in\mathbb{R}_0^+}\left\{\beta\left[\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}\right]_i^2 + \sigma_{\text{noise}}^2\rho(\beta)\right\} \\
&\leq \lambda R(\boldsymbol{x}^t) + \sum_{i=1}^{KM}\left\{\beta_i^t\left[\boldsymbol{W}\boldsymbol{x}^t - \boldsymbol{y}\right]_i^2 + \sigma_{\text{noise}}^2\rho(\beta_i^t)\right\},
\end{aligned} \tag{A.26}$$

where the weights β_i^t for $i = 1, \ldots, KM$ are computed from $\boldsymbol{x}^{t-1}$ according to Eq. (4.23) and σ_{noise} is the scale parameter of the observation model that is assumed to be constant over the iterations. Comparing the weights β_i^t given by Eq. (4.23) and Eq. (A.13), one can verify that:

$$F(\boldsymbol{x}^{t-1}) = \lambda R(\boldsymbol{x}^{t-1}) + \sum_{i=1}^{KM}\left\{\beta_i^t\left[\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{y}\right]_i^2 + \sigma_{\text{noise}}^2\rho(\beta_i^t)\right\}. \tag{A.27}$$

Hence, we can derive the inequality condition for the energy function value among successive iterations:

$$\begin{aligned} & F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^{t}) \\ & \quad \geq \left(\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{y}\right)^{\top} \boldsymbol{B}^{t} \left(\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{y}\right) - \left(\boldsymbol{W}\boldsymbol{x}^{t} - \boldsymbol{y}\right)^{\top} \boldsymbol{B}^{t} \left(\boldsymbol{W}\boldsymbol{x}^{t} - \boldsymbol{y}\right) \\ & \qquad + \lambda \left(\sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t-1}\right| + \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t-1}\right|^{p} - \sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t})} \left|x_i^{t}\right| - \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t})} \left|x_i^{t}\right|^{p} \right), \end{aligned} \tag{A.28}$$

where $\boldsymbol{B}^t$ is constructed as $\boldsymbol{B}^t = \operatorname{diag}\left(\beta_1^t, \ldots, \beta_{KM}^t\right)$. This inequality condition can be rearranged according to:

$$\begin{aligned} & F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^{t}) \\ & \quad \geq \left(\boldsymbol{x}^{t-1} - \boldsymbol{x}^{t}\right)^{\top} \boldsymbol{W}^{\top} \boldsymbol{B}^{t} \boldsymbol{W} \left(\boldsymbol{x}^{t-1} - \boldsymbol{x}^{t}\right) + 2\left(\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^{t}\right)^{\top} \boldsymbol{B}^{t} \left(\boldsymbol{W}\boldsymbol{x}^{t} - \boldsymbol{y}\right) \\ & \qquad + \lambda \left(\sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t-1}\right| + \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t-1}\right|^{p} - \sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t})} \left|x_i^{t}\right| - \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t})} \left|x_i^{t}\right|^{p} \right) \\ & \quad \geq \underline{\beta}^{t} \left\| \boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^{t} \right\|_2^2 + 2\left(\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^{t}\right)^{\top} \boldsymbol{B}^{t} \left(\boldsymbol{W}\boldsymbol{x}^{t} - \boldsymbol{y}\right) \\ & \qquad + \lambda \left(\sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t-1}\right| + \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t-1}\right|^{p} - \sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t}\right| - \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t}\right|^{p} \right), \end{aligned} \tag{A.29}$$

where the last inequality is based on Lemma A.4 and $\underline{\beta}^t = \min_i \underline{\beta_i}^t$. Then, the weight $\underline{\beta} = \min_t \underline{\beta}^t$ is strictly positive and:

$$\begin{aligned} & F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^{t}) \\ & \quad \geq \underline{\beta} \left\| \boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^{t} \right\|_2^2 + 2\left(\boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^{t}\right)^{\top} \boldsymbol{B}^{t} \left(\boldsymbol{W}\boldsymbol{x}^{t} - \boldsymbol{y}\right) \\ & \qquad + \lambda \left(\sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t-1}\right| + \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t-1}\right|^{p} - \sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t}\right| - \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \left|x_i^{t}\right|^{p} \right). \end{aligned} \tag{A.30}$$

Since $\boldsymbol{x}^t$ is the solution of iteratively re-weighted minimization at iteration t, it follows:

$$\boldsymbol{0} \in \frac{\partial}{\partial \boldsymbol{x}} \left\{ \left(\boldsymbol{W}\boldsymbol{x} - \boldsymbol{y}\right)^{\top} \boldsymbol{B}^{t} \left(\boldsymbol{W}\boldsymbol{x} - \boldsymbol{y}\right) + \lambda \sum_{i=1}^{N_S} \alpha_i^t |x_i| \right\} \Bigg|_{\boldsymbol{x}=\boldsymbol{x}^t}, \tag{A.31}$$

where α_i^t is computed from the weighting function in Eq. (4.25). Thus, it follows for the subgradient:

$$\begin{aligned} & 2\boldsymbol{W}^{\top} \boldsymbol{B}^{t} \left(\boldsymbol{W}\boldsymbol{x}^{t} - \boldsymbol{y}\right) + \lambda c_i^t \alpha_i^t = 0, \forall i \\ & \qquad c_i^t \in \begin{cases} \{1\}, & \text{if } x_i^t > 0 \\ [-1, 1] & \text{if } x_i^t = 0 \\ \{-1\}, & \text{if } x_i^t < 0 \end{cases}. \end{aligned} \tag{A.32}$$

Substituting the condition in Eq. (A.32) into Eq. (A.30) and using the fact that $c_i^t x_i^t = |x_i^t|$ and $|c_i^t| \leq 1$, leads to:

$$\begin{aligned}
F(\boldsymbol{x}^{t-1}) - F(\boldsymbol{x}^t) \geq\ & \underline{\beta} \left\| \boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t \right\|_2^2 \\
& + \lambda \sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \left(|x_i^{t-1}| - |x_i^t| + c_i^t (x_i^t - x_i^{t-1}) \right) \\
& + \lambda \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \left(|x_i^{t-1}|^p - |x_i^t|^p + p|x_i^{t-1}|^{p-1} c_i^t (x_i^t - x_i^{t-1}) \right) \\
\geq\ & \underline{\beta} \left\| \boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t \right\|_2^2 \\
& + \lambda \sum_{i \notin \mathcal{I}(\boldsymbol{x}^{t-1})} \left(|x_i^{t-1}| - |x_i^t| + (|x_i^t| - |x_i^{t-1}|) \right) \\
& + \lambda \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \left(|x_i^{t-1}|^p - |x_i^t|^p + p|x_i^{t-1}|^{p-1} (|x_i^t| - |x_i^{t-1}|) \right) \\
\geq\ & \underline{\beta} \left\| \boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t \right\|_2^2 \\
& + \lambda \sum_{i \in \mathcal{I}(\boldsymbol{x}^{t-1})} \left(|x_i^{t-1}|^{p-1} \left((1-p)|x_i^{t-1}| + p|x_i^t| - |x_i^{t-1}|^{1-p} |x_i^t|^p \right) \right) \\
\geq\ & \underline{\beta} \left\| \boldsymbol{W}\boldsymbol{x}^{t-1} - \boldsymbol{W}\boldsymbol{x}^t \right\|_2^2,
\end{aligned} \tag{A.33}$$

where the last inequality is according to Lemma 1 in [Chen 14] as corollary of Young's inequality, which completes the proof. □

Since $F(\boldsymbol{x})$ is a lower-bounded function, $F(\boldsymbol{x}^t)$ converges to an extreme value. If $\boldsymbol{x}^t$ also converges to an extreme value denoted by $\boldsymbol{x}^*$, this estimate satisfies:

$$\boldsymbol{0} \in \frac{\partial}{\partial \boldsymbol{x}} \left\{ (\boldsymbol{W}\boldsymbol{x} - \boldsymbol{y})^\top \boldsymbol{B}^t (\boldsymbol{W}\boldsymbol{x} - \boldsymbol{y}) + \lambda \sum_{i=1}^{N_S} \alpha_i^t |x_i| \right\} \Bigg|_{\boldsymbol{x}=\boldsymbol{x}^*}. \tag{A.34}$$

As a consequence, $\boldsymbol{x}^*$ is a stationary point of the non-convex problem in Eq. (4.42).

Notice that there can be the situation $F(\boldsymbol{x}^t) = F(\boldsymbol{x}^{t-1})$ but $\boldsymbol{x}$ does not converge to a stationary point, i. e. $\boldsymbol{x}^t \neq \boldsymbol{x}^{t-1}$. This is the case if the following conditions are fulfilled:

1. $\boldsymbol{x}^t - \boldsymbol{x}^{t-1}$ is in the null space of the system matrix, i. e. $\boldsymbol{W}(\boldsymbol{x}^{t-1} - \boldsymbol{x}^t) = \boldsymbol{0}$.
2. The index sets among successive iterations are identical, i. e. $\mathcal{I}(\boldsymbol{x}^t) = \mathcal{I}(\boldsymbol{x}^{t-1})$.
3. The estimates in the index set $\mathcal{I}(\boldsymbol{x}^t)$ among successive iterations are identical, i. e. $|x_i^t| = |x_i^{t-1}|$ for all $i \in \mathcal{I}(\boldsymbol{x}^t)$.
4. The objective value of the regularization term among successive iterations is identical, i. e. $R(\boldsymbol{x}^t) = R(\boldsymbol{x}^{t-1})$

However, in practice this situation can be avoided. For instance, if the system matrix $\boldsymbol{W}$ has full rank, the null space is $\{\boldsymbol{0}\}$, i. e. $\boldsymbol{x}^t = \boldsymbol{x}^{t-1}$ if $F(\boldsymbol{x}^t) = F(\boldsymbol{x}^{t-1})$.

A.3 Multi-Sensor Super-Resolution using Locally Linear Regression

A.3.1 Majorization-Minimization for Tukey's Biweight Loss

Let us consider the 1-D minimization problem:

$$\hat{z} = \underset{z}{\operatorname{argmin}}\, \phi_{\boldsymbol{x}_{ij}}(z)\,. \tag{A.35}$$

We can reformulate the minimization of the non-convex loss function $\phi_{\boldsymbol{x}_{ij}}(z)$ as an MM algorithm according to [Ochs 15]. This leads to the iteratively re-weighted least squares (IRLS) scheme:

$$z^t = \underset{z}{\operatorname{argmin}}\, \kappa(z^{t-1})z^2, \tag{A.36}$$

$$\kappa(z) = \frac{\frac{d}{dz}\phi_{\boldsymbol{x}_{ij}}(z)}{z}, \tag{A.37}$$

where $\kappa(z)$ is the underlying weighting function. In case of Tukey's biweight loss defined in Eq. (6.14), we can compute the gradient according to:

$$\frac{d}{dz}\phi_{\boldsymbol{x}_{ij}}(z) = \begin{cases} z\left(1 - \frac{z^2}{\sigma^2_{\mathrm{LLR},ij}}\right)^2 & \text{if } |z| \leq \sigma^2_{\mathrm{LLR},ij} \\ 0 & \text{otherwise} \end{cases}. \tag{A.38}$$

This leads to corresponding weighting function for IRLS:

$$\kappa(z) = \begin{cases} \left(1 - \frac{z^2}{\sigma^2_{\mathrm{LLR},ij}}\right)^2 & \text{if } |z| \leq \sigma^2_{\mathrm{LLR},ij} \\ 0 & \text{otherwise} \end{cases}. \tag{A.39}$$

A.3.2 Estimation of the Regression Coefficients

If we omit the iteration index for the sake of notational clarity, the regression coefficients associated with the k-th pixel in the image channels $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are estimated according to:

$$(\tilde{C}_{ij,k}, \tilde{b}_{ij,k}) = \underset{C_{ij,k}, b_{ij,k}}{\operatorname{argmin}}\, F(C_{ij,k}, b_{ij,k}), \tag{A.40}$$

where:

$$F(C_{ij,k}, b_{ij,k}) = \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} \left(C_{ij,k}x_{i,l} + b_{ij,k} - x_{j,l}\right)^2 + \epsilon_{ij}C^2_{ij,k}. \tag{A.41}$$

Computing the zero-crossings of the derivative of this energy function w. r. t. the unknown regression coefficient $b_{ij,k}$ yields:

$$\begin{aligned} \frac{\partial}{\partial b_{ij,k}} F(C_{ij,k}, b_{ij,k}) &= 2 \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} \left(C_{ij,k}x_{i,l} + b_{ij,k} - x_{j,l}\right) \\ &= 0. \end{aligned} \tag{A.42}$$

If we rearrange this condition, we can compute the regression coefficient $b_{ij,k}$ in closed-form:

$$\begin{aligned} b_{ij,k} &= \frac{1}{Z_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{K}_{ij})} \left(\sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{j,l} - C_{ij,k} \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l} \right), \\ &= \mathrm{E}_{\omega_{\mathrm{LLR}}(k)} \left(\boldsymbol{x}_j, \boldsymbol{K}_{ij} \right) - C_{ij,k} \cdot \mathrm{E}_{\omega_{\mathrm{LLR}}(k)} \left(\boldsymbol{x}_i, \boldsymbol{K}_{ij} \right), \end{aligned} \tag{A.43}$$

where:

$$\mathrm{E}_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{z}, \boldsymbol{K}) = \frac{1}{Z_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{K})} \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_l z_l, \tag{A.44}$$

$$Z_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{K}) = \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_l. \tag{A.45}$$

Notice that the estimation of this regression coefficient involves box filtering operations denoted by $\mathrm{E}_{\omega_{\mathrm{LLR}}(k)}(\cdot, \cdot)$ for the channels $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ with a normalization of the filter kernel according to $Z_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{K})$. These box filters can be implemented efficiently using integral images, see e. g. [He 13, Hore 14].

Computing the zero-crossings of the derivative of the energy function w. r. t. the regression coefficient $C_{ij,k}$ yields:

$$\begin{aligned} \frac{\partial}{\partial C_{ij,k}} F(C_{ij,k}, b_{ij,k}) &= 2 \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} \left(C_{ij,k} x_{i,l} + b_{ij,k} - x_{j,l} \right) x_{i,l} + 2 \epsilon_{ij} C_{ij,k} \\ &= 0. \end{aligned} \tag{A.46}$$

If we substitute the expression for the regression coefficient $b_{ij,k}$ given by Eq. (A.43), this condition can be rearranged according to:

$$\begin{aligned} & C_{ij,k} \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l}^2 + b_{ij,k} \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l} - \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l} x_{j,l} + \epsilon_{ij} C_{ij,k} \\ &= C_{ij,k} \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l}^2 - \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l} x_{j,l} + \epsilon_{ij} C_{ij,k} \\ &\quad + \frac{1}{Z_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{K}_{ij})} \left(\sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{j,l} - C_{ij,k} \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l} \right) \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l} \\ &= C_{ij,k} \left(\sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l}^2 - \frac{1}{Z_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{K}_{ij})} \left(\sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l} \right)^2 + \epsilon_{ij} \right) \\ &\quad + \frac{1}{Z_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{K}_{ij})} \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l} \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{j,l} - \sum_{l \in \omega_{\mathrm{LLR}}(k)} \kappa_{ij,l} x_{i,l} x_{j,l} \\ &= 0. \end{aligned} \tag{A.47}$$

Thus, we can compute the regression coefficient $C_{ij,k}$ in closed-form:

$$C_{ij,k} = \frac{\mathrm{E}_{\omega_{\mathrm{LLR}}(k)} \left(\boldsymbol{x}_i \odot \boldsymbol{x}_j, \boldsymbol{K}_{ij} \right) - \mathrm{E}_{\omega_{\mathrm{LLR}}(k)} \left(\boldsymbol{x}_i, \boldsymbol{K}_{ij} \right) \cdot \mathrm{E}_{\omega_{\mathrm{LLR}}(k)} \left(\boldsymbol{x}_j, \boldsymbol{K}_{ij} \right)}{\mathrm{E}_{\omega_{\mathrm{LLR}}(k)} \left(\boldsymbol{x}_i \odot \boldsymbol{x}_i, \boldsymbol{K}_{ij} \right) + \frac{1}{Z_{\omega_{\mathrm{LLR}}(k)}(\boldsymbol{K}_{ij})} \epsilon_{ij}}, \tag{A.48}$$

where we can again use box filtering of the image channels $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

List of Symbols

Chapter 2

$x(t)$	1-D continuous signal	10
$y(t)$	Sampling of $x(t)$ as a continuous signal	10
$y[n]$	Discretization of the sampled signal $y(t)$	10
T	Sampling pitch	10
$\mathcal{D}_T\{\cdot\}$	Sampling operator with sampling pitch T	10
$\delta(t)$	Dirac delta impulse	11
j	Imaginary unit of complex number	11
$\star$	Convolution operator	11
$\mathcal{F}\{\cdot\}$	Continuous Fourier transform (CFT)	11
$X(f)$	CFT of a continuous signal $x(t)$	11
f_s	Sampling frequency (sampling rate)	11
f_0	Band limitation frequency	12
$\mathcal{F}^{-1}\{\cdot\}$	Inverse of the CFT	12
$h(t)$	Linear and shift invariant blur kernel	15
$x^{(k)}(t)$	1-D continuous signal of the k-th channel	16
$y^{(k)}(t)$	Sampling of $x^{(k)}(t)$ as a continuous signal	16
$y^{(k)}[n]$	Discretization the sampled signal $y^{(k)}(t)$	16
t_k	Offset of the k-th channel	16
$\mathcal{Y}[n]$	Discrete Fourier transform (DFT) of $y[n]$	17
$\boldsymbol{\mathcal{Y}}$	Samples of the DFT $\mathcal{Y}[n]$ in vector notation	18
$\boldsymbol{X}$	Samples of the CFT $X(f)$ in vector notation	18
$\boldsymbol{W}$	System matrix in the Fourier domain	18

Chapter 3

$\boldsymbol{u} = (u, v)^\top$	Position in 2-D space	32
$x(\boldsymbol{u})$	High-resolution image as continuous function	32
$y(\boldsymbol{u})$	Low-resolution image as continuous function	32
K	Number of frames in a sequence	32
$(\cdot)^{(k)}$	Frame index	32
$\mathcal{M}^{(k)}\{\cdot\}$	Motion model	32
$m^{(k)}(\boldsymbol{u})$	Displacement vector field	32
$h(\boldsymbol{u})$	Linear and shift invariant blur kernel	33
$\mathcal{D}\{\cdot\}$	Sampling model	33
$\epsilon(\boldsymbol{u})$	Stochastic noise signal	33
$\Omega_y \subset \mathbb{R}^2$	Domain of low-resolution data	34
$M_u \times M_v$	Pixel resolution of low-res. image (width $\times$ height)	34
$\boldsymbol{y}^{(k)} \in \mathbb{R}^M$	Low-resolution image in vector notation	34

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

List of Abbreviations

AMSR adaptive multi-sensor super-resolution 100

AMSR-OD adaptive multi-sensor super-resolution with outlier detection . . . 100

BTV bilateral total variation 44, 209

CCD charge-coupled device 2, 150

CFA color-filter array 39

CFT continuous Fourier transform 11

CG conjugate gradient 41

CMOS complementary metaloxide semiconductor 2, 14

CT computed tomography 84

dB decibel . . . . . iii, v, 67, 182

DCT discrete cosine transform 28

DFT discrete Fourier transform 17

ECC enhanced correlation coefficient 68

EM expectation maximization 49

FFT Fast Fourier Transform 20

FOV field of view 139

GCV generalized cross validation 60

GMM Gaussian mixture model 174

IDP inter-color dependency penalty 124

IRLS iteratively re-weighted least squares 198

LLR locally linear regression 106, 182, 210

LMI local mutual information 90

LSI linear shift invariant 15

MAD median absolute deviation 60

MAP maximum a-posteriori 31, 181, 209

ML maximum likelihood 30, 181, 209

MM majorization-minimization 56, 182, 192, 209

MRI magnetic resonance imaging 84

MSAC M-estimator sample consensus 96

MSR multi-sensor super-resolution 100

NCC normalized cross correlation 92

NMAD normalized mean absolute deviation 155

OCT optical coherence tomography 137

PCA principal component analysis 108

PDF probability density function 40

PET positron emission tomography 84

pixel picture element 2

POCS projection onto convex sets 31

PSF point spread function 2, 22, 185

PSNR peak-signal-to-noise ratio . . . . . iii, 66, 182, 209

RITK Range Imaging Toolkit 99

SCG scaled conjugate gradient 62

SNR signal-to-noise ratio 85

SSIM structural similarity 66, 182, 209

SSR single-sensor super-resolution 100

SVD singular value decomposition 146

ToF Time-of-Flight . . . . . iv, vi, 73, 183, 210

TV total variation 44

WBTV weighted bilateral total variation 54

List of Figures

List of Tables

Bibliography

[Abra 10] M. D. Abramoff, M. K. Garvin, and M. Sonka. "Retinal Imaging and Image Analysis". *IEEE Reviews in Biomedical Engineering*, Vol. 3, No. 1, pp. 169–208, Dec. 2010.

[Abra 15] M. D. Abramoff and M. Niemeijer. "Mass Screening of Diabetic Retinopathy Using Automated Methods". In: G. Michelson, Ed., *Teleophthalmology in Preventive Medicine*, pp. 41–50, Springer, Berlin, Heidelberg, 2015.

[Adal 14] K. M. Adal, R. M. Ensing, R. Couvert, P. van Etten, J. P. Martinez, K. A. Vermeer, and L. van Vliet. "A Hierarchical Coarse-to-Fine Approach for Fundus Image Registration". In: *International Workshop on Biomedical Image Registration (WBIR)*, pp. 93–102, LNCS Vol. 8545, Springer, London, UK, July 2014.

[Ague 06] M. L. Aguena and N. D. Mascarenhas. "Multispectral Image Data Fusion using POCS and Super-Resolution". *Computer Vision and Image Understanding*, Vol. 102, No. 2, pp. 178–187, May 2006.

[Akgu 05] T. Akgun, Y. Altunbasak, and R. Mersereau. "Super-Resolution Reconstruction of Hyperspectral Images". *IEEE Transactions on Image Processing*, Vol. 14, No. 11, pp. 1860–1875, Nov. 2005.

[Akht 15] N. Akhtar, F. Shafait, and A. Mian. "Bayesian Sparse Representation for Hyperspectral Image Super Resolution". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3631–3640, IEEE, Boston, MA, June 2015.

[Alam 00] M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda. "Infrared Image Registration and High-Resolution Reconstruction using Multiple Translationally Shifted Aliased Video Frames". *IEEE Transactions on Instrumentation and Measurement*, Vol. 49, No. 5, pp. 915–923, Oct. 2000.

[Anas 09] A. Anastassopoulos and P. Vassilis. "Regularized Super-Resolution Image Reconstruction Employing Robust Error Norms". *Optical Engineering*, Vol. 48, No. 11, pp. 117004, 1–14, Nov. 2009.

[Ba 14] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown. "Convergence and Stability of Iteratively Re-weighted Least Squares Algorithms". *IEEE Transactions on Signal Processing*, Vol. 62, No. 1, pp. 183–195, Jan. 2014.

[Baba 11] S. D. Babacan, R. Molina, and A. K. Katsaggelos. "Variational Bayesian Super Resolution". *IEEE Transactions on Image Processing*, Vol. 20, No. 4, pp. 984–999, Apr. 2011.

[Bake 02] S. Baker and T. Kanade. "Limits on Super-Resolution and How to Break Them". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 9, pp. 1167–1183, Sep. 2002.

[Batz 15] M. Bätz, A. Eichenseer, J. Seiler, M. Jonscher, and A. Kaup. "Hybrid Super-Resolution Combining Example-based Single-Image and Interpolation-based Multi-Image Reconstruction Approaches". In: *International Conference on Image Processing (ICIP)*, pp. 58–62, IEEE, Quebec, Canada, Sep. 2015.

[Batz 16] M. Bätz, A. Eichenseer, and A. Kaup. "Multi-Image Super-Resolution using a Dual Weighting Scheme based on Voronoi Tessellation". In: *International Conference on Image Processing (ICIP)*, pp. 2822–2826, IEEE, Phoenix, AZ, Sep. 2016.

[Baue 13] S. Bauer, A. Seitel, H. G. Hofmann, T. Blum, J. Wasza, M. Balda, H.-P. Meinzer, N. Navab, J. Hornegger, and L. Maier-Hein. "Real-Time Range Imaging in Health Care: A Survey". In: *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pp. 228–254, LNCS Vol. 8200, Part III, Springer, 2013.

[Bele 09] S. P. Belekos, N. P. Galatsanos, S. D. Babacan, and A. K. Katsaggelos. "Maximum a posteriori super-resolution of compressed video using a new multichannel image prior". In: *International Conference on Image Processing (ICIP)*, pp. 2797–2800, IEEE, Cairo, Egypt, Nov. 2009.

[Berc 16] C. Bercea, A. Maier, and T. Köhler. "Confidence-Aware Levenberg-Marquardt Optimization for Joint Motion Estimation and Super-Resolution". In: *International Conference on Image Processing (ICIP)*, pp. 1136–1140, IEEE, Phoenix, AZ, Sep. 2016.

[Bert 03] M. Bertero and P. Boccacci. "Super-resolution in Computational Imaging". *Micron*, Vol. 34, No. 6-7, pp. 265–273, Oct. 2003.

[Besa 91] J. Besag, J. York, and A. Mollié. "Bayesian Image Restoration, with Two Applications in Spatial Statistics". *Annals of the Institute of Statistical Mathematics*, Vol. 43, No. 1, pp. 1–20, March 1991.

[Beye 00] T. Beyer, D. W. Townsend, T. Brun, P. E. Kinahan, M. Charron, R. Roddy, J. Jerin, J. Young, L. Byars, and R. Nutt. "A Combined PET/CT Scanner for Clinical Oncology". *Journal of Nuclear Medicine*, Vol. 41, No. 8, pp. 1369–1379, Aug. 2000.

[Bhav 12] A. V. Bhavsar and A. N. Rajagopalan. "Range Map Superresolution-Inpainting, and Reconstruction from Sparse Data". *Computer Vision and Image Understanding*, Vol. 116, No. 4, pp. 572–591, Apr. 2012.

[Biou 06] J. Bioucas-Dias. "Bayesian Wavelet-Based Image Deconvolution: a GEM Algorithm Exploiting a Class of Heavy-Tailed Priors". *IEEE Transactions on Image Processing*, Vol. 15, No. 4, pp. 937–951, Apr. 2006.

[Bish 06] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer New York, 4th Ed., 2006.

[Bock 10] R. Bock, J. Meier, L. G. Nyúl, J. Hornegger, and G. Michelson. "Glaucoma Risk Index: Automated Glaucoma Detection from Color Fundus Images". *Medical Image Analysis*, Vol. 14, No. 3, pp. 471–81, June 2010.

[Bork 09] A. Borkowski, Z. Zalevsky, and B. Javidi. "Geometrical Superresolved Imaging using Nonperiodic Spatial Masking". *Journal of the Optical Society of America A*, Vol. 26, No. 3, pp. 589–601, March 2009.

[Bork 11] A. Borkowski, Z. Zalevsky, E. Marom, and B. Javidi. "Enhanced Geometrical Superresolved Imaging with Moving Binary Random Mask". *Journal of the Optical Society of America A*, Vol. 28, No. 4, pp. 566–575, Apr. 2011.

[Borm 04] S. Borman. *Topics in Multiframe Superresolution Restoration*. PhD thesis, University of Notre Dame, 2004.

[Bowe 08] O. Bowen and C. Bouganis. "Real-Time Image Super Resolution using an FPGA". In: *International Conference on Field Programmable Logic and Applications*, pp. 89–94, Heidelberg, Germany, Sep. 2008.

[Brad 00] G. Bradski. *The OpenCV Library*. Dr. Dobb's Journal of Software Tools, 2000.

[Buda 13] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson. "Robust Vessel Segmentation in Fundus Images". *International Journal of Biomedical Imaging*, Vol. 2013, No. 1, pp. 1–11, Article ID 154860, Sep. 2013.

[Can 02] A. Can, C. Stewart, B. Roysam, and H. Tanenbaum. "A Feature-Based, Robust, Hierarchical Algorithm for Registering Pairs of Images of the Curved Human Retina". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 3, pp. 347–364, March 2002.

[Cand 08] E. J. Candès, M. B. Wakin, and S. P. Boyd. "Enhancing Sparsity by Reweighted L1 Minimization". *Journal of Fourier Analysis and Applications*, Vol. 14, No. 5-6, pp. 877–905, Oct. 2008.

[Cape 00] D. Capel and A. Zisserman. "Super-Resolution Enhancement of Text Image Sequences". In: *International Conference on Pattern Recognition (ICPR)*, pp. 600–605, IEEE, Barcelona, Spain, Sep. 2000.

[Cape 03] D. P. Capel and A. Zisserman. "Computer Vision Applied to Super Resolution". *IEEE Signal Processing Magazine*, Vol. 20, No. 3, pp. 75–86, May 2003.

[Cape 04] D. P. Capel. *Image Mosaicing and Super-Resolution*. PhD thesis, University of Oxford, 2004.

[Catt 06] P. C. Cattin, H. Bay, L. Van Gool, and G. Székely. "Retina Mosaicing Using Local Features". In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 185–192, LNCS Vol. 4191, Part II, Springer, Copenhagen, Denmark, Oct. 2006.

[Chak 11] A. Chakrabarti and T. Zickler. "Statistics of Real-World Hyperspectral Images". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 193–200, IEEE, Colorado Springs, CO, June 2011.

[Chan 05] R. H. Chan, C. W. Ho, and M. Nikolova. "Salt-and-Pepper Noise Removal by Median-Type Noise Detectors and Detail-Preserving Regularization". *IEEE Transactions on Image Processing*, Vol. 14, No. 10, pp. 1479–1485, Oct. 2005.

[Chan 98] T. F. Chan and C. K. Wong. "Total Variation Blind Deconvolution". *IEEE Transactions on Image Processing*, Vol. 7, No. 3, pp. 370–375, Jan. 1998.

[Char 94] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. "Two Deterministic Half-Quadratic Regularization Algorithms for Computed Imaging". In: *International Conference on Image Processing (ICIP)*, pp. 168–172, IEEE, Austin, TX, Nov. 1994.

[Chen 00] T. Chen, P. B. Catrysse, A. El Gamal, and B. A. Wandell. "How Small Should Pixel Size Be?". In: *Proc. SPIE 3965, Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, pp. 451–459, San Jose, CA, Jan. 2000.

[Chen 07] J. Chen and C.-K. Tang. "Spatio-Temporal Markov Random Field for Video Denoising". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, IEEE, Minneapolis, MN, June 2007.

[Chen 14] X. Chen and W. Zhou. "Convergence of the Reweighted L1 Minimization Algorithm for L2-Lp Minimization". *Computational Optimization and Applications*, Vol. 59, No. 1-2, pp. 47–61, Apr. 2014.

[Cho 11] S. Cho, J. Wang, and S. Lee. "Handling Outliers in Non-Blind Image Deconvolution". In: *International Conference on Computer Vision (ICCV)*, pp. 495–502, IEEE, Barcelona, Spain, Nov. 2011.

[Chri 15] V. Christlein, D. Bernecker, A. Maier, and E. Angelopoulou. "Offline Writer Identification Using Convolutional Neural Network Activation Features". In: *German Conference on Pattern Recognition (GCPR)*, pp. 540–552, LNCS Vol. 9358, Springer, Aachen, Germany, Oct. 2015.

[Cool 65] J. W. Cooley and J. W. Tukey. "An Algorithm for the Machine Calculation of Complex Fourier Series". *Mathematics of Computation*, Vol. 19, No. 90, pp. 297–301, Apr. 1965.

[Cool 69] J. W. Cooley, P. A. W. Lewis, and P. D. Welch. "The Finite Fourier Transform". *IEEE Transactions on Audio and Electroacoustics*, Vol. 17, No. 2, pp. 77–85, June 1969.

[Crow 84] F. C. Crow. "Summed-Area Tables for Texture Mapping". In: *ACM SIGGRAPH Computer Graphics*, pp. 207–212, Minneapolis, MN, July 1984.

[Daub 10] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk. "Iteratively Reweighted Least Squares Minimization for Sparse Recovery". *Communications on Pure and Applied Mathematics*, Vol. 63, No. 1, pp. 1–38, Jan. 2010.

[Dekk 97] A. J. den Dekker and A. van den Bos. "Resolution: a Survey". *Journal of the Optical Society of America A*, Vol. 14, No. 3, pp. 547–557, March 1997.

[Demp 77] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 1–38, 1977.

[Dirk 16] H. Dirks, J. Geiping, D. Cremers, and M. Moeller. "Multiframe Motion Coupling via Infimal Convolution Regularization for Video Super Resolution". *arXiv preprint 1611.07767v1*, Nov. 2016.

[Dong 14] C. Dong, C. C. Loy, K. He, and X. Tang. "Learning a Deep Convolutional Network for Image Super-Resolution". In: *European Conference on Computer Vision (ECCV)*, pp. 184–199, LNCS Vol. 8692, Springer, Zurich, Switzerland, Sep. 2014.

[Dono 06] D. L. Donoho. "Compressed Sensing". *IEEE Transactions on Information Theory*, Vol. 52, No. 4, pp. 1289–1306, Apr. 2006.

[Drig 05] R. Driggers, K. Krapels, and S. Young. "The Meaning of Super-Resolution". *Proceedings of the SPIE, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVI*, Vol. 5784, No. 1, pp. 103–106, May 2005.

[Duch 07] A. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer, 2007.

[El G 05] A. El Gamal and H. Eltoukhy. "CMOS Image Sensors". *IEEE Circuits and Devices Magazine*, Vol. 21, No. 3, pp. 6–20, May 2005.

[El Y 08a] N. A. El-Yamany and P. E. Papamichalis. "Robust Color Image Super-resolution: An Adaptive M-Estimation Framework". *EURASIP Journal on Image and Video Processing*, Vol. 2008, No. 1, pp. 1–12, Dec. 2008.

[El Y 08b] N. A. El-Yamany and P. E. Papamichalis. "Using Bounded-Influence M-Estimators in Multi-Frame Super-Resolution Reconstruction: A Comparative Study". In: *International Conference on Image Processing (ICIP)*, pp. 337–340, IEEE, San Diego, CA, 2008.

[Elad 01] M. Elad and Y. Hel-Or. "A Fast Super-Resolution Reconstruction Algorithm for Pure Translational Motion and Common Space-Invariant Blur". *IEEE Transactions on Image Processing*, Vol. 10, No. 8, pp. 1187–1193, Aug. 2001.

[Elad 97] M. Elad and A. Feuer. "Restoration of a Single Superresolution Image from Several Blurred, Noisy, and Undersampled Measured Images". *IEEE Transactions on Image Processing*, Vol. 6, No. 12, pp. 1646–1658, Dec. 1997.

[Elda 06] Y. Eldar and M. Unser. "Nonideal Sampling and Interpolation from Noisy Observations in Shift-Invariant Spaces". *IEEE Transactions on Signal Processing*, Vol. 54, No. 7, pp. 2636–2651, July 2006.

[Erso 06] O. K. Ersoy. *Diffraction, Fourier Optics and Imaging*. John Wiley & Sons, 30 Ed., 2006.

[Espa 11] S. Espana-Boquera, M. J. Castro-Bleda, J. Gorbe-Moya, and F. Zamora-Martinez. "Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 4, pp. 767–779, Apr. 2011.

[Evan 08] G. D. Evangelidis and E. Z. Psarakis. "Parametric Image Alignment using Enhanced Correlation Coefficient Maximization". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 10, pp. 1858–1865, Oct. 2008.

[Fang 06] B. Fang and Y. Tang. "Elastic Registration for Retinal Images Based on Reconstructed Vascular Trees". *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 6, pp. 1183–1187, June 2006.

[Fara 13] E. Faramarzi, D. Rajan, and M. P. Christensen. "Unified Blind Method for Multi-Image Super-Resolution and Single/Multi-Image Blur Deconvolution". *IEEE Transactions on Image Processing*, Vol. 22, No. 6, pp. 2101–2114, June 2013.

[Fars 03] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. "Robust Shift and Add Approach to Superresolution". *Proceedings of the SPIE, Applications of Digital Image Processing XXVI*, Vol. 5203, No. 1, pp. 121–130, Nov. 2003.

[Fars 04a] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. "Advances and Challenges in Super-Resolution". *International Journal of Imaging Systems and Technology*, Vol. 14, No. 2, pp. 47–57, Aug. 2004.

[Fars 04b] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. "Fast and Robust Multiframe Super Resolution". *IEEE Transactions on Image Processing*, Vol. 13, No. 10, pp. 1327–1344, Oct. 2004.

[Fars 06] S. Farsiu, M. Elad, and P. Milanfar. "Multiframe Demosaicing and Super-Resolution of Color Images". *IEEE Transactions on Image Processing*, Vol. 15, No. 1, pp. 141–159, Jan. 2006.

[Fars 16] S. Farsiu, D. Robinson, and P. Milanfar. "Multi-Dimensional Signal Processing Research Group (MDSP) – Super-Resolution And Demosaicing Datasets". http://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html, 2016.

[Fatt 14] R. Fattal. "Dehazing using Color-Lines". *ACM Transactions on Graphics*, Vol. 34, No. 1, pp. 1–14, Nov. 2014.

[Fers 13] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof. "Image Guided Depth Upsampling Using Anisotropic Total Generalized Variation". In: *International Conference on Computer Vision (ICCV)*, pp. 993–1000, IEEE, Sydney, Australia, Dec. 2013.

[Fiel 09] M. Field, D. Clarke, S. Strup, and W. B. Seales. "Stereo Endoscopy as a 3-D Measurement Tool". In: *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5748–5751, IEEE, Minneapolis, MN, Sep. 2009.

[Form 11] C. Forman, M. Aksoy, J. Hornegger, and R. Bammer. "Self-Encoded Marker for Optical Prospective Head Motion Correction in MRI". *Medical Image Analysis*, Vol. 15, No. 5, pp. 708–719, Oct. 2011.

[Fran 07] R. Fransens, C. Strecha, and L. Van Gool. "Optical Flow Based Super-Resolution: A Probabilistic Approach". *Computer Vision and Image Understanding*, Vol. 106, No. 1, pp. 106–115, Apr. 2007.

[Fran 98] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. "Multiscale Vessel Enhancement Filtering". In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 130–137, LNCS Vol. 1496, Springer, Cambridge, MA, Oct. 1998.

[Free 00] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. "Learning Low-Level Vision". *International Journal of Computer Vision*, Vol. 40, No. 1, pp. 25–47, Oct. 2000.

[Furs 16] P. Fürsattel, S. Placht, M. Balda, C. Schaller, H. Hofmann, A. Maier, and C. Riess. "A Comparative Error Analysis of Current Time-of-Flight Sensors". *IEEE Transactions on Computational Imaging*, Vol. 2, No. 1, pp. 27–41, March 2016.

[Gaba 07] S. Gabarda and G. Cristóbal. "Blind Image Quality Assessment Through Anisotropy". *Journal of the Optical Society of America A*, Vol. 24, No. 12, pp. B42–B51, Dec. 2007.

[Gala 91] N. Galatsanos and R. Chin. "Restoration of Color Images by Multichannel Kalman Filtering". *IEEE Transactions on Signal Processing*, Vol. 39, No. 10, pp. 2237–2252, Oct. 1991.

[Garc 06] J. García, Z. Zalevsky, and C. Ferreira. "Superresolved Imaging of Remote Moving Targets". *Optics Letters*, Vol. 31, No. 5, pp. 586–588, March 2006.

[Ghes 14] F. C. Ghesu, T. Köhler, S. Haase, and J. Hornegger. "Guided Image Super-Resolution: A New Technique for Photogeometric Super-Resolution in Hybrid 3-D Range Imaging". In: *German Conference on Pattern Recognition (GCPR)*, pp. 227–238, LNCS Vol. 8753, Springer, Münster, Germany, Sep. 2014.

[Glas 09] D. Glasner, S. Bagon, and M. Irani. "Super-Resolution from a Single Image". In: *International Conference on Computer Vision (ICCV)*, pp. 349–356, IEEE, Kyoto, Japan, Sep. 2009.

[Goto 04] T. Gotoh and M. Okutomi. "Direct super-resolution and registration using raw CFA images". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 600–607, IEEE, Washington D.C., June 2004.

[Gree 08] H. Greenspan. "Super-Resolution in Medical Imaging". *The Computer Journal*, Vol. 52, No. 1, pp. 43–63, Feb. 2008.

[Guev 10] A. Guevara and R. Mester. "Signal Reconstruction from Noisy, Aliased, and Nonideal Samples: What Linear MMSE Approaches Can Achieve". In: *European Signal Processing Conference (EUSIPCO)*, pp. 1291–1295, IEEE, Aalborg, Denmark, Aug. 2010.

[Guo 08] W. Guo and F. Huang. "A Local Mutual Information Guided Denoising Technique and Its Application to Self-calibrated Partially Parallel Imaging". In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 939–947, LNCS Vol. 5242, Part II, Springer, New York, NY, Sep. 2008.

[Haas 12] S. Haase, C. Forman, T. Kilgus, R. Bammer, L. Maier-Hein, and J. Hornegger. "ToF/RGB Sensor Fusion for Augmented 3D Endoscopy using a Fully Automatic Calibration Scheme". In: *Bildverarbeitung für die Medizin (BVM)*, pp. 111–116, Springer, Berlin, Germany, March 2012.

[Haas 13a] S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, A. Schneider, M. Kranzfelder, H. Feussner, and J. Hornegger. "3-D Operation Situs Reconstruction with Time-of-Flight Satellite Cameras Using Photogeometric Data Fusion". In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 356–363, LNCS Vol. 8149, Part I, Springer, Nagoya, Japan, Sep. 2013.

[Haas 13b] S. Haase, C. Forman, T. Kilgus, R. Bammer, L. Maier-Hein, and J. Hornegger. "ToF/RGB sensor fusion for 3-D endoscopy". *Current Medical Imaging Reviews*, Vol. 9, No. 2, pp. 113–119, May 2013.

[Haas 13c] S. Haase, T. Köhler, T. Kilgus, L. Maier-Hein, J. Hornegger, and H. Feussner. "Instrument Segmentation in Hybrid 3-D Endoscopy using Multi-Sensor Super-Resolution". In: *Computer- und Roboter Assistierte Chirurgie (CURAC)*, pp. 194–197, Innsbruck, Austria, Nov. 2013.

[Haas 13d] S. Haase, J. Wasza, T. Kilgus, and J. Hornegger. "Laparoscopic Instrument Localization using a 3-D Time-of-Flight/RGB Endoscope". In: *Workshop on Applications of Computer Vision (WACV)*, pp. 449–454, IEEE, Tampa, FL, Jan. 2013.

[Haas 14] S. Haase, J. Wasza, M. Safak, T. Kilgus, L. Maier-Hein, H. Feussner, and J. Hornegger. "Patch based Specular Reflection Removal for Range Images in Hybrid 3-D Endoscopy". In: *International Symposium on Biomedical Imaging (ISBI)*, pp. 509–512, IEEE, Beijing, China, Apr. 2014.

[Haas 16] S. Haase. *Hybrid RGB/Time-of-Flight Sensors in Minimally Invasive Surgery*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2016.

[Ham 15] B. Ham, M. Cho, and J. Ponce. "Robust Image Filtering using Joint Static and Dynamic Guidance". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4823–4831, IEEE, Boston, MA, June 2015.

[Hama 13] K. Hamada, R. Nakashima, K. Takahashi, and T. Naemura. "Super-Resolution with Adaptive Pixel Weighting Scheme and Its Application to Super-Resolved Free-Viewpoint Image Synthesis". In: *International Conference on Signal-Image Technology & Internet-Based Systems*, pp. 757–764, IEEE, Kyoto, Japan, Jan. 2013.

[Han 13] J. Han, L. Shao, D. Xu, and J. Shotton. "Enhanced Computer Vision with Microsoft Kinect Sensor: A Review". *IEEE Transactions on Cybernetics*, Vol. 43, No. 5, pp. 1318–1334, Oct. 2013.

[Hard 07] R. C. Hardie. "A Fast Image Super-Resolution Algorithm Using an Adaptive Wiener Filter". *IEEE Transactions on Image Processing*, Vol. 16, No. 12, pp. 2953–2964, Dec. 2007.

[Hard 12] R. C. Hardie and K. J. Barnard. "Fast Super-Resolution using an Adaptive Wiener Filter with Robustness to Local Motion". *Optics Express*, Vol. 20, No. 19, pp. 21053–21073, Sep. 2012.

[Hard 97] R. C. Hardie, K. J. Barnard, and E. E. Armstrong. "Joint MAP Registration and High-Resolution Image Estimation using a Sequence of Undersampled Images". *IEEE Transactions on Image Processing*, Vol. 6, No. 12, pp. 1621–1633, Dec. 1997.

[Harm 10] S. Harmeling, S. Sra, M. Hirsch, and B. Scholkopf. "Multiframe Blind Deconvolution, Super-Resolution, and Saturation Correction via Incremental EM". In: *International Conference on Image Processing (ICIP)*, pp. 3313–3316, IEEE, Hong Kong, Sep. 2010.

[Hart 04] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd Ed., 2004.

[Hast 09] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2nd Ed., 2009.

[He 06] H. He and L. P. Kondi. "An Image Super-Resolution Algorithm for Different Error Levels per Frame". *IEEE Transactions on Image Processing*, Vol. 15, No. 3, pp. 592–603, March 2006.

[He 07] Y. He, K.-H. Yap, L. Chen, and L.-P. Chau. "A Nonlinear Least Square Technique for Simultaneous Image Registration and Super-Resolution". *IEEE Transactions on Image Processing*, Vol. 16, No. 11, pp. 2830–2841, Nov. 2007.

[He 10] K. He, J. Sun, and X. Tang. "Guided Image Filtering". In: *European Conference on Computer Vision (ECCV)*, pp. 1–14, LNCS Vol. 6311, Springer, Hersonissos, Greece, Sep. 2010.

[He 13] K. He, J. Sun, and X. Tang. "Guided Image Filtering". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 6, pp. 1397–409, June 2013.

[Hirs 07] H. Hirschmüller and D. Scharstein. "Evaluation of Cost Functions for Stereo Matching". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, IEEE, Minneapolis, MN, June 2007.

[Hofm 05] M. Hofmann, C. Eggeling, S. Jakobs, and S. W. Hell. "Breaking the Diffraction Barrier in Fluorescence Microscopy at Low Light Intensities by using Reversibly Photoswitchable Proteins". *Proceedings of the National Academy of Sciences*, Vol. 102, No. 49, pp. 17565–17569, Dec. 2005.

[Hohe 15] B. Höher, G. Michelson, P. Voigtmann, and B. Schmauss. "Low-Cost Non-mydriatic Color Video Imaging of the Retina for Nonindustrialized Countries". In: G. Michelson, Ed., *Teleophthalmology in Preventive Medicine*, pp. 51–62, Springer, Berlin, Heidelberg, 2015.

[Hore 14] J. Horentrup and M. Schlosser. "Confidence-Aware Guided Image Filter". In: *International Conference on Image Processing (ICIP)*, pp. 3243–3247, IEEE, Paris, France, 2014.

[Horn 12] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

[Horn 81] B. K. Horn and B. G. Schunck. "Determining Optical Flow". *Artificial Intelligence*, Vol. 17, No. 1-3, pp. 185–203, Aug. 1981.

[Hou 06] Z. Hou. "A Review on MR Image Intensity Inhomogeneity Correction". *International Journal of Biomedical Imaging*, Vol. 2006, No. 1, pp. 1–11, Article ID 49515, Jan. 2006.

[Huan 91] D. Huang, E. Swanson, C. Lin, J. Schuman, W. Stinson, W. Chang, M. Hee, T. Flotte, K. Gregory, C. Puliafito, and J. G. Fujimoto. "Optical Coherence Tomography". *Science*, Vol. 254, No. 5035, pp. 1178–1181, Nov. 1991.

[Huan 99] J. Huang and D. Mumford. "Statistics of Natural Images and Models". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 541–547, IEEE, Fort Collins, CO, June 1999.

[Hunt 04] D. R. Hunter and K. Lange. "A Tutorial on MM Algorithms". *The American Statistician*, Vol. 58, No. 1, pp. 30–37, Dec. 2004.

[Hurl 09] N. Hurley and S. Rickard. "Comparing Measures of Sparsity". *IEEE Transactions on Information Theory*, Vol. 55, No. 10, pp. 4723–4741, Oct. 2009.

[Ilov 14] A. Ilovitsh and Z. Zalevsky. "Super Resolved Passive Imaging of Remote Moving Object on Top of Sparse Unknown Background". *Applied Optics*, Vol. 53, No. 28, pp. 6340–6343, Oct. 2014.

[Iran 91] M. Irani and S. Peleg. "Improving Resolution by Image Registration". *CVGIP Graphical Models and Image Processing*, Vol. 53, No. 3, pp. 231–239, May 1991.

[ISO 00] "ISO 12233:2000 – Photography - Electronic Still-Picture Cameras – Resolution Measurements". International Organization for Standardization (ISO), 2000.

[Jord 14] J. Jordan, E. Angelopoulou, and A. Robles-Kelly. "An Unsupervised Material Learning Method for Imaging Spectroscopy". In: *International Joint Conference on Neural Networks (IJCNN)*, pp. 2428–2435, IEEE, Beijing, China, July 2014.

[Josh 11] G. D. Joshi, J. Sivaswamy, and S. R. Krishnadas. "Optic Disk and Cup Segmentation from Monocular Color Retinal Images for Glaucoma Assessment". *IEEE Transactions on Medical Imaging*, Vol. 30, No. 6, pp. 1192–205, June 2011.

[Jude 08] M. S. Judenhofer, H. F. Wehrl, D. F. Newport, C. Catana, S. B. Siegel, M. Becker, A. Thielscher, M. Kneilling, M. P. Lichy, M. Eichner, K. Klingel, G. Reischl, S. Widmaier, M. Röcken, R. E. Nutt, H.-J. Machulla, K. Uludag, S. R. Cherry, C. D. Claussen, and B. J. Pichler. "Simultaneous PET-MRI: A new Approach for Functional and Morphological Imaging". *Nature Medicine*, Vol. 14, No. 4, pp. 459–65, Apr. 2008.

[Kapp 16] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos. "Video Super-Resolution With Convolutional Neural Networks". *IEEE Transactions on Computational Imaging*, Vol. 2, No. 2, pp. 109–122, June 2016.

[Kats 93] A. K. Katsaggelos, K. Lay, and N. Galatsanos. "A General Framework for Frequency Domain Multi-Channel Signal Processing". *IEEE Transactions on Image Processing*, Vol. 2, No. 3, pp. 417–420, July 1993.

[Kell 11] S. H. Keller, F. Lauze, and M. Nielsen. "Video Super-Resolution using Simultaneous Motion and Intensity Calculations". *IEEE Transactions on Image Processing*, Vol. 20, No. 7, pp. 1870–84, July 2011.

[Kenn 07] J. A. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, and H. Azhari. "Improved Image Fusion in PET/CT using Hybrid Image Reconstruction and Super-Resolution". *International Journal of Biomedical Imaging*, Vol. 2007, No. 1, pp. 1–10, Article ID 46846, Jan. 2007.

[Kere 98] D. Keren and A. Gotlib. "Denoising Color Images Using Regularization and Correlation Terms". *Journal of Visual Communication and Image Representation*, Vol. 9, No. 4, pp. 352–365, Dec. 1998.

[Kere 99] D. Keren and M. Osadchy. "Restoring Subsampled Color Images". *Machine Vision and Applications*, Vol. 11, No. 4, pp. 197–202, Dec. 1999.

[Kiec 13] M. Kiechle, S. Hawe, and M. Kleinsteuber. "A Joint Intensity and Depth Co-sparse Analysis Model for Depth Map Super-resolution". In: *International Conference on Computer Vision (ICCV)*, pp. 1545–1552, IEEE, Sydney, Australia, Dec. 2013.

[Kilg 15] T. Kilgus, E. Heim, S. Haase, S. Prüfer, M. Müller, A. Seitel, M. Fangerau, T. Wiebe, J. Iszatt, H. P. Schlemmer, J. Hornegger, K. Yen, and L. Maier-Hein. "Mobile Markerless Augmented Reality and its Application in Forensic Medicine". *International Journal of Computer Assisted Radiology and Surgery*, Vol. 10, No. 5, pp. 573–586, May 2015.

[Kim 16a] J. Kim, J. K. Lee, and K. M. Lee. "Accurate Image Super-Resolution Using Very Deep Convolutional Networks". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654, IEEE, Las Vegas, NV, June 2016.

[Kim 16b] J. Kim, J. K. Lee, and K. M. Lee. "Deeply-Recursive Convolutional Network for Image Super-Resolution". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1637–1645, Las Vegas, NV, June 2016.

[Kim 90] S. Kim, N. Bose, and H. Valenzuela. "Recursive Reconstruction of High Resolution Image from Noisy Undersampled Multiframes". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 38, No. 6, pp. 1013–1027, June 1990.

[Kimm 99] R. Kimmel. "Demosaicing: Image Reconstruction from Color CCD Samples". *IEEE Transactions on Image Processing*, Vol. 8, No. 9, pp. 1221–1228, Sep. 1999.

[Kohl 12] T. Köhler, J. Hornegger, M. Mayer, and G. Michelson. "Quality-Guided Denoising for Low-Cost Fundus Imaging". In: *Bildverarbeitung für die Medizin (BVM)*, pp. 292–297, Springer, Berlin, Germany, March 2012.

[Kohl 13a] T. Köhler, A. Budai, M. F. Kraus, J. Odstrčilík, G. Michelson, and J. Hornegger. "Automatic No-Reference Quality Assessment for Retinal Fundus Images using Vessel Segmentation". In: *International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 95–100, IEEE, Porto, Portugal, Oct. 2013.

[Kohl 13b] T. Köhler, S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, H. Feussner, and J. Hornegger. "ToF Meets RGB: Novel Multi-Sensor Super-Resolution for Hybrid 3-D Endoscopy". In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 139–146, LNCS Vol. 8149, Part I, Springer, Nagoya, Japan, Sep. 2013.

[Kohl 14a] T. Köhler, A. Brost, K. Mogalle, Q. Zhang, C. Köhler, G. Michelson, J. Hornegger, and R. P. Tornow. "Multi-Frame Super-Resolution with Quality Self-Assessment for Retinal Fundus Videos". In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 650–657, LNCS Vol. 8673, Part I, Springer, Cambridge, MA, Sep. 2014.

[Kohl 14b] T. Köhler, S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, H. Feussner, and J. Hornegger. "Outlier Detection for Multi-Sensor Super-Resolution in Hybrid 3D Endoscopy". In: *Bildverarbeitung für die Medizin (BVM)*, pp. 84–89, Springer, Aachen, Germany, March 2014.

[Kohl 15a] T. Köhler, R. Bock, J. Hornegger, and G. Michelson. "Computer-Aided Diagnostics and Pattern Recognition: Automated Glaucoma Detection". In: G. Michelson, Ed., *Teleophthalmology in Preventive Medicine*, pp. 93–104, Springer, Berlin, Heidelberg, 2015.

[Kohl 15b] T. Köhler, S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, C. Stock, J. Hornegger, and H. Feussner. "Multi-Sensor Super-Resolution for Hybrid Range Imaging with Application to 3-D Endoscopy and Open Surgery". *Medical Image Analysis*, Vol. 24, No. 1, pp. 220–234, July 2015.

[Kohl 15c] T. Köhler, J. Jordan, A. Maier, and J. Hornegger. "A Unified Bayesian Approach to Multi-Frame Super-Resolution and Single-Image Upsampling in Multi Sensor Imaging". In: *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 143.1–143.12, BMVA Press, Swansea, UK, Sep. 2015.

[Kohl 15d] T. Köhler, A. Maier, and V. Christlein. "Binarization Driven Blind Deconvolution for Document Image Restoration". In: *German Conference on Pattern Recognition (GCPR)*, pp. 91–102, LNCS Vol. 9358, Springer, Aachen, Germany, Oct. 2015.

[Kohl 16a] T. Köhler, A. Heinrich, A. Maier, J. Hornegger, and R. P. Tornow. "Super-Resolved Retinal Image Mosaicing". In: *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1063 – 1067, IEEE, Prague, Czech Republic, Apr. 2016.

[Kohl 16b] T. Köhler, X. Huang, F. Schebesch, A. Aichert, A. Maier, and J. Hornegger. "Robust Multiframe Super-Resolution Employing Iteratively Re-Weighted Minimization". *IEEE Transactions on Computational Imaging*, Vol. 2, No. 1, pp. 42 – 58, March 2016.

[Kohl 17] T. Köhler, M. Bätz, F. Naderi, A. Kaup, A. K. Maier, and C. Riess. "Benchmarking Super-Resolution Algorithms on Real Data". *arXiv preprint arXiv:1709.04881*, pp. 1–10, Sep. 2017.

[Kola 11] R. Kolar, J. Odstrčilík, J. Jan, and V. Harabis. "Illumination Correction and Contrast Equalization in Colour Fundus Images". In: *European Signal Processing Conference (EUSIPCO)*, pp. 298–302, IEEE, Barcelona, Spain, Aug. 2011.

[Kola 15] R. Kolar, R. P. Tornow, and J. Odstrcilik. "Retinal Image Registration for Eye Movement Estimation". In: *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5247–5250, IEEE, Milan, Italy, Aug. 2015.

[Kola 16] R. Kolar, R. P. Tornow, J. Odstrcilik, and I. Liberdova. "Registration of Retinal Sequences from New Video-Ophthalmoscopic Camera". *BioMedical Engineering OnLine*, Vol. 15, No. 1, p. 57, Dec. 2016.

[Kolb 10] A. Kolb, E. Barth, R. Koch, and R. Larsen. "Time-of-Flight Cameras in Computer Graphics". *Computer Graphics Forum*, Vol. 29, No. 1, pp. 141–159, March 2010.

[Kong 13] X. Kong, K. Li, Q. Yang, L. Wenyin, and M.-H. Yang. "A New Image Quality Metric for Image Auto-Denoising". In: *International Conference on Computer Vision (ICCV)*, pp. 2888–2895, IEEE, Sydney, Australia, Dec. 2013.

[Kopf 07] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. "Joint Bilateral Upsampling". *ACM Transactions on Graphics*, Vol. 26, No. 3, pp. 1–6, Article 96, July 2007.

[Kote 13] J. Kotera, F. Šroubek, and P. Milanfar. "Blind Deconvolution Using Alternating Maximum a Posteriori Estimation with Heavy-Tailed Priors". In: *International Conference on Computer Analysis of Images and Patterns (CAIP)*, pp. 59–66, IEEE, York, UK, Aug. 2013.

[Krau 17] M. Kraus. *Motion Correction and Signal Enhancement in Optical Coherence Tomography*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2017.

[Kris 09] D. Krishnan and R. Fergus. "Fast Image Deconvolution using Hyper-Laplacian Priors". In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 1033–1041, Curran Associates, Inc., Vancouver, Canada, Dec. 2009.

[Kurt 14] A. Kürten, T. Köhler, A. Budai, R. P. Tornow, G. Michelson, and J. Hornegger. "Geometry-Based Optic Disk Tracking in Retinal Fundus Videos". In: *Bildverarbeitung für die Medizin (BVM)*, pp. 120–125, Springer, Aachen, Germany, March 2014.

[Lana 15] C. Lanaras, E. Baltsavias, and K. Schindler. "Hyperspectral Super-Resolution by Coupled Spectral Unmixing". In: *International Conference on Computer Vision (ICCV)*, pp. 3586–3594, IEEE, Santiago, Chile, Dec. 2015.

[Ledi 16] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". *arXiv preprint arXiv:1609.04802v5*, 2016.

[Lee 03] E. S. Lee and M. G. Kang. "Regularized Adaptive High-Resolution Image Reconstruction Considering Inaccurate Subpixel Registration.". *IEEE Transactions on Image Processing*, Vol. 12, No. 7, pp. 826–37, Jan. 2003.

[Lela 15] B. Lelandais and F. Duconge. "Deconvolution Regularized using Fuzzy C-Means Algorithm for Biomedical Image Deblurring and Segmentation". In: *International Symposium on Biomedical Imaging (ISBI)*, pp. 1457–1461, IEEE, New York, NY, Apr. 2015.

[Levi 09] A. Levin, Y. Weiss, F. Durand, and W. Freeman. "Understanding and Evaluating Blind Deconvolution Algorithms". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1964–1971, IEEE, Miami, FL, June 2009.

[Li 10] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning. "A Multi-Frame Image Super-Resolution Method". *Signal Processing*, Vol. 90, No. 2, pp. 405–414, Feb. 2010.

[Li 16] Y. Li, J. B. Huang, N. Ahuja, and M. H. Yang. "Deep Joint Image Filtering". In: *European Conference on Computer Vision (ECCV)*, pp. 154–169, LNCS, Vol. 9908, Springer, Amsterdam, Netherlands, Oct. 2016.

[Li 17] D. Li and Z. Wang. "Video Super-Resolution via Motion Compensation and Deep Residual Learning". *IEEE Transactions on Computational Imaging*, Vol. 3, No. 4, pp. 749 – 762, Dec. 2017.

[Liao 15] R. Liao, X. Tao, R. Li, Z. Ma, and J. Jiaya. "Video Super-Resolution via Deep Draft-Ensemble Learning". In: *International Conference on Computer Vision (ICCV)*, IEEE, Santiago, Chile, Dec. 2015.

[Lin 04] Z. Lin and H.-Y. Shum. "Fundamental Limits of Reconstruction-Based Superresolution Algorithms under Local Translation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 1, pp. 83–97, Jan. 2004.

[Lind 12] J. Lindberg. "Mathematical Concepts of Optical Superresolution". *Journal of Optics*, Vol. 14, No. 8, pp. Article 083001, 1–24, July 2012.

[Liu 09] C. Liu. *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. PhD thesis, Massachusetts Institute of Technology, 2009.

[Liu 11] C. Liu and D. Sun. "A Bayesian Approach to Adaptive Video Super Resolution". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 209–216, IEEE, Colorado Springs, CO, June 2011.

[Liu 14] C. Liu and D. Sun. "On Bayesian Adaptive Video Super Resolution". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 2, pp. 346–360, Feb. 2014.

[Lord 79] F. R. S. Lord Rayleigh. "Investigations in Optics, with Special Reference to the Spectroscope". *Philosophical Magazine Series 5*, Vol. 8, No. 49, pp. 261–274, 1879.

[Ma 13] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu. "Constant Time Weighted Median Filtering for Stereo Matching and Beyond". In: *International Conference on Computer Vision (ICCV)*, pp. 49–56, IEEE, Sydney, Australia, Dec. 2013.

[Ma 15] Z. Ma, R. Liao, X. Tao, L. Xu, J. Jia, and E. Wu. "Handling Motion Blur in Multi-Frame Super-Resolution". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5224–5232, IEEE, Boston, MA, June 2015.

[Maie 13] L. Maier-Hein, P. Mountney, A. Bartoli, H. Elhawary, D. Elson, A. Groch, A. Kolb, M. Rodrigues, J. Sorger, S. Speidel, and D. Stoyanov. "Optical techniques for 3D Surface Reconstruction in Computer-Assisted Laparoscopic Surgery". *Medical Image Analysis*, Vol. 17, No. 8, pp. 974–96, Dec. 2013.

[Maie 14] L. Maier-Hein, A. Groch, A. Bartoli, S. Bodenstedt, G. Boissonnat, P.-L. Chang, N. T. Clancy, D. S. Elson, S. Haase, E. Heim, J. Hornegger, P. Jannin, H. Kenngott, T. Kilgus, B. Müller-Stich, D. Oladokun, S. Röhl, T. R. Dos Santos, H.-P. Schlemmer, A. Seitel, S. Speidel, M. Wagner, and D. Stoyanov. "Comparative Validation of Single-Shot Optical Techniques for Laparoscopic 3-D Surface Reconstruction". *IEEE Transactions on Medical Imaging*, Vol. 33, No. 10, pp. 1913–30, Oct. 2014.

[Mair 08] J. Mairal, M. Elad, and G. Sapiro. "Sparse Representation for Color Image Restoration". *IEEE Transactions on Image Processing*, Vol. 17, No. 1, pp. 53–69, Jan. 2008.

[Mall 99] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic press, 1999.

[Marq 63] D. W. Marquardt. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". *Journal of the Society for Industrial and Applied Mathematics*, Vol. 11, No. 2, pp. 431–441, June 1963.

[Marr 11a] A. G. Marrugo and M. S. Millán. "Retinal Image Analysis: Preprocessing and Feature Extraction". *Journal of Physics: Conference Series*, Vol. 274, No. 1, pp. Article 012039, 1–9, Jan. 2011.

[Marr 11b] A. G. Marrugo, M. S. Millán, G. Cristóbal, S. Gabarda, and H. C. Abril. "No-reference Quality Metrics for Eye Fundus Imaging". In: *International Conference on Computer Analysis of Images and Patterns (CAIP)*, pp. 486–493, Springer, Seville, Spain, Aug. 2011.

[Marr 11c] A. G. Marrugo, M. Sorel, F. Sroubek, and M. S. Millán. "Retinal Image Restoration by Means of Blind Deconvolution". *Journal of Biomedical Optics*, Vol. 16, No. 11, pp. 116016–1–116016–11, Oct. 2011.

[Meer 91] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim. "Robust Regression Methods for Computer Vision: A Review". *International Journal of Computer Vision*, Vol. 6, No. 1, pp. 59–70, Apr. 1991.

[Mers 11] S. Mersmann, M. Müller, A. Seitel, F. Arnegger, R. Tetzlaff, J. Dinkel, M. Baumhauer, B. Schmied, H. Meinzer, and L. Maier-Hein. "Time-of-Flight Camera Technique for Augmented Reality in Computer-Assisted Interventions". In: *Proc. SPIE 7964, Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling*, pp. 79642C–1–79642C–9, Orlando, FL, Feb. 2011.

[Mila 10] P. Milanfar. *Super-Resolution Imaging*. CRC Press, 2010.

[Mitz 09] D. Mitzel, T. Pock, T. Schoenemann, and D. Cremers. "Video Super Resolution Using Duality Based TV-L1 Optical Flow". In: *DAGM Symposium*, pp. 432–441, LNCS Vol. 5748, Springer, Jena, Germany, Sep. 2009.

[Moli 03] R. Molina, J. Mateos, A. K. Katsaggelos, and M. Vega. "Bayesian Multichannel Image Restoration using Compound Gauss-Markov Random Fields". *IEEE Transactions on Image Processing*, Vol. 12, No. 12, pp. 1642–1654, Dec. 2003.

[Muri 11] S. Murillo, S. Echegaray, G. Zamora, P. Soliz, and W. Bauman. "Quantitative and Qualitative Image Quality Analysis of Super Resolution Images from a Low Cost Scanning Laser Ophthalmoscope". In: *Proc. SPIE 7962, Medical Imaging 2011: Image Processing*, pp. 79624T–79624T–9, 2011.

[Nabn 02] I. T. Nabney. *NETLAB: Algorithms for Pattern Recognition*. Springer, 1st Ed., 2002.

[Nasr 14] K. Nasrollahi and T. B. Moeslund. "Super-resolution: A Comprehensive Survey". *Machine Vision and Applications*, Vol. 25, No. 6, pp. 1423–1468, June 2014.

[Ng 07] M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang. "A Total Variation Regularization Based Super-Resolution Reconstruction Algorithm for Digital Video". *EURASIP Journal on Advances in Signal Processing*, Vol. 2007, No. 1, pp. 1–17, Article ID 74585, June 2007.

[Nguy 01a] N. Nguyen, P. Milanfar, and G. Golub. "Efficient Generalized Cross-Validation with Applications to Parametric Image Restoration and Resolution Enhancement". *IEEE Transactions on Image Processing*, Vol. 10, No. 9, pp. 1299–1308, Sep. 2001.

[Nguy 01b] N. Nguyen, P. Milanfar, and G. Golub. "A Computationally Efficient Superresolution Image Reconstruction Algorithm". *IEEE Transactions on Image Processing*, Vol. 10, No. 4, pp. 573–583, Apr. 2001.

[Niem 06] M. Niemeijer, M. D. Abramoff, and B. van Ginneken. "Image Structure Clustering for Image Quality Verification of Color Retina Images in Diabetic Retinopathy Screening". *Medical Image Analysis*, Vol. 10, No. 6, pp. 888–898, Dec. 2006.

[Ochs 13] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock. "An Iterated L1 Algorithm for Non-smooth Non-convex Optimization in Computer Vision". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1759–1766, IEEE, Portland, OR, June 2013.

[Ochs 15] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock. "On Iteratively Reweighted Algorithms for Nonsmooth Nonconvex Optimization in Computer Vision". *SIAM Journal on Imaging Sciences*, Vol. 8, No. 1, pp. 331–372, Jan. 2015.

[Oliv 09] J. P. Oliveira, J. Bioucas-Dias, and M. A. Figueiredo. "Adaptive Total Variation Image Deblurring: A Majorization-Minimization Approach". *Signal Processing*, Vol. 89, No. 9, pp. 1683–1693, Sep. 2009.

[Omer 04] I. Omer and M. Werman. "Color Lines: Image Specific Color Representation". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 946–953, IEEE, Washington D.C., June 2004.

[Ono 16] S. Ono and I. Yamada. "Color-Line Regularization for Color Artifact Removal". *IEEE Transactions on Computational Imaging*, Vol. 2, No. 3, pp. 204–217, Sep. 2016.

[Oppe 99] A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-Time Signal Processing*. Prentice-hall Englewood Cliffs, 2nd Ed., 1999.

[Papo 77] A. Papoulis. "Generalized Sampling Expansion". *IEEE Transactions on Circuits and Systems*, Vol. 24, No. 11, pp. 652–654, Nov. 1977.

[Park 03] S. C. Park, M. K. Park, and M. G. Kang. "Super-Resolution Image Reconstruction: A Technical Overview". *IEEE Signal Processing Magazine*, Vol. 20, No. 3, pp. 21–36, May 2003.

[Park 11] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon. "High Quality Depth Map Upsampling for 3D-TOF Cameras". In: *International Conference on Computer Vision (ICCV)*, pp. 1623–1630, IEEE, Barcelona, Spain, Nov. 2011.

[Pata 07] V. Patanavijit and S. Jitapunkul. "A Lorentzian Stochastic Estimation for a Robust Iterative Multiframe Super-Resolution Reconstruction with Lorentzian-Tikhonov Regularization". *EURASIP Journal on Advances in Signal Processing*, Vol. 2007, No. 1, pp. 1–21, Article ID 34821, July 2007.

[Patr 16] C. Patrizio and K. Egiazarian. *Blind Image Deconvolution: Theory and Applications*. CRC Press, 2016.

[Patt 01] A. J. Patti and Y. Altunbasak. "Artifact Reduction for Set Theoretic Super Resolution Image Reconstruction with Edge Adaptive Constraints and Higher-Order Interpolants". *IEEE Transactions on Image Processing*, Vol. 10, No. 1, pp. 179–186, Jan. 2001.

[Patt 06] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan, and I. J. Constable. "Retinal Image Analysis: Concepts, Applications and Potential". *Progress in Retinal and Eye Research*, Vol. 25, No. 1, pp. 99–127, Jan. 2006.

[Paul 10] J. Paulus, J. Meier, R. Bock, J. Hornegger, and G. Michelson. "Automated Quality Assessment of Retinal Fundus Photos". *International Journal of Computer Assisted Radiology and Surgery*, Vol. 5, No. 6, pp. 557–564, Nov. 2010.

[Penn 09] J. Penne, K. Höller, M. Stürmer, T. Schrauder, A. Schneider, R. Engelbrecht, H. Feussner, B. Schmauss, and J. Hornegger. "Time-of-Flight 3-D Endoscopy". In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 467–474, LNCS Vol. 5761, Part I, Springer, London, UK, Sep. 2009.

[Pham 06] T. Q. Pham, L. J. van Vliet, and K. Schutte. "Robust Fusion of Irregularly Sampled Data Using Adaptive Normalized Convolution". *EURASIP Journal on Advances in Signal Processing*, Vol. 2006, No. 1, pp. 1–12, Article ID 83268, Dec. 2006.

[Pham 08] T. Q. Pham, L. J. v. Vliet, and K. Schutte. "Robust Super-Resolution without Regularization". *Journal of Physics: Conference Series*, Vol. 124, No. 1, pp. 1–20, Article ID 012036, July 2008.

[Pick 07a] L. C. Pickup. *Machine Learning in Multi-frame Image Super-resolution*. PhD thesis, University of Oxford, 2007.

[Pick 07b] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. "Overcoming Registration Uncertainty in Image Super-Resolution: Maximize or Marginalize?". *EURASIP Journal on Advances in Signal Processing*, Vol. 2007, No. 1, pp. 1–14, Article ID 23565, Aug. 2007.

[Plui 03] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. "Mutual Information based Registration of Medical Images: A Survey". *IEEE Transactions on Medical Imaging*, Vol. 22, No. 8, pp. 986–1004, Aug. 2003.

[Plye 14] A. Plyer, G. Le Besnerais, and F. Champagnat. "Massively Parallel Lucas Kanade Optical Flow for Real-Time Video Processing Applications". *Journal of Real-Time Image Processing*, Vol. 11, No. 4, pp. 713–730, Apr. 2014.

[Rahm 11] M. Rahman. *Applications of Fourier Transforms to Generalized Functions*. WIT Press, 2011.

[Raja 03] A. N. Rajagopalan and V. P. Kiran. "Motion-Free Superresolution and the Role of Relative Blur". *Journal of the Optical Society of America A*, Vol. 20, No. 11, pp. 2022–2032, Nov. 2003.

[Rama 08] S. Ramani, D. Van De Ville, T. Blu, and M. Unser. "Nonideal Sampling and Regularization Theory". *IEEE Transactions on Signal Processing*, Vol. 56, No. 3, pp. 1055–1070, March 2008.

[Rayn 98] K. Rayner. "Eye Movements in Reading and Information Processing: 20 Years of Research". *Psychological Bulletin*, Vol. 124, No. 3, pp. 372–422, Nov. 1998.

[Reyn 11] M. Reynolds, J. Dobos, L. Peel, T. Weyrich, and G. J. Brostow. "Capturing Time-of-Flight Data with Confidence". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 945–952, IEEE, Colorado Springs, CO, June 2011.

[Rhee 99] S. Rhee and M. G. Kang. "Discrete Cosine Transform Based Regularized High-Resolution Image Reconstruction Algorithm". *Optical Engineering*, Vol. 38, No. 8, pp. 1348–1356, Aug. 1999.

[Robi 10] M. D. Robinson, S. J. Chiu, C. A. Toth, J. A. Izatt, J. Y. Lo, and S. Farsiu. "New Applications of Super-resolution in Medical Imaging". In: P. Milanfar, Ed., *Super-Resolution Imaging*, pp. 383–412, CRC Press, 2010.

[Rudi 92] L. I. Rudin, S. Osher, and E. Fatemi. "Nonlinear Total Variation Based Noise Removal Algorithms". *Physica D: Nonlinear Phenomena*, Vol. 60, No. 1-4, pp. 259–268, Nov. 1992.

[Scal 88] J. A. Scales and A. Gersztenkorn. "Robust Methods in Inverse Theory". *Inverse Problems*, Vol. 4, No. 4, pp. 1071–1091, Oct. 1988.

[Scha 03] D. Scharstein and R. Szeliski. "High-Accuracy Stereo Depth Maps using Structured Light". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 195–202, IEEE, Madison, WI, 2003.

[Scha 07] D. Scharstein and C. Pal. "Learning Conditional Random Fields for Stereo". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, IEEE, Minneapolis, MN, June 2007.

[Schi 17] F. Schirrmacher, T. Köhler, L. Husvogt, J. G. Fujimoto, J. Hornegger, and A. K. Maier. "QuaSI: Quantile Sparse Image Prior for Spatio-Temporal Denoising of Retinal OCT Data". In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 83–91, LNCS Vol. 10434, Part II, Quebec City, Canada, Sep. 2017.

[Schm 12] C. Schmalz, F. Forster, A. Schick, and E. Angelopoulou. "An Endoscopic 3D Scanner based on Structured Light". *Medical Image Analysis*, Vol. 16, No. 5, pp. 1063–72, July 2012.

[Schu 08] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. "High-Quality Scanning using Time-of-Flight Depth Superresolution". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop*, pp. 1–7, IEEE, Anchorage, AK, June 2008.

[Schu 09] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. "LidarBoost: Depth Superresolution for ToF 3D Shape Scanning". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 343–350, IEEE, Miami, FL, June 2009.

[Schu 94] R. R. Schultz and R. L. Stevenson. "A Bayesian Approach to Image Expansion for Improved Definition". *IEEE Transactions on Image Processing*, Vol. 3, No. 3, pp. 233–242, May 1994.

[Schu 95] R. R. Schultz and R. L. Stevenson. "Stochastic Modeling and Estimation of Multispectral Image Data". *IEEE Transactions on Image Processing*, Vol. 4, No. 8, pp. 1109–1119, Aug. 1995.

[Schu 96] R. R. Schultz and R. L. Stevenson. "Extraction of High-Resolution Frames from Video Sequences". *IEEE Transactions on Image Processing*, Vol. 5, No. 6, pp. 996–1011, June 1996.

[Shan 48] C. E. Shannon. "A Mathematical Theory of Communication". *Bell System Technical Journal*, Vol. 27, No. 3, pp. 379–423, July 1948.

[Shei 16] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. "LIVE Image Quality Assessment Database Release 2". http://live.ece.utexas.edu/research/quality, 2016.

[Shen 07] H. Shen, L. Zhang, B. Huang, and P. Li. "A MAP Approach for Joint Motion Estimation, Segmentation, and Super Resolution". *IEEE Transactions on Image Processing*, Vol. 16, No. 2, pp. 479–490, Feb. 2007.

[Shen 15] X. Shen, C. Zhou, L. Xu, and J. Jia. "Mutual-Structure for Joint Filtering". In: *International Conference on Computer Vision (ICCV)*, pp. 3406–3414, IEEE, Santiago, Chile, Dec. 2015.

[Song 10] H. Song, L. Zhang, P. Wang, K. Zhang, and X. Li. "An Adaptive L1-L2 Hybrid Error Model to Super-Resolution". In: *International Conference on Image Processing (ICIP)*, pp. 2821–2824, IEEE, Hong Kong, Sep. 2010.

[Sore 10] M. Sorel, F. Šroubek, and J. Flusser. "Towards Super-Resolution in the Presence of Spatially Varying Blur". In: P. Milanfar, Ed., *Super-Resolution Imaging*, pp. 1–38, CRC Press, 2010.

[Spar 16] C. M. Sparrow. "On Spectroscopic Resolving Power". *Astrophysical Journal*, Vol. 44, No. 1, pp. 76–86, 1916.

[Sriv 03] A. Srivastava, A. Lee, E. Simoncelli, and S.-C. Zhu. "On Advances in Statistical Modeling of Natural Images". *Journal of Mathematical Imaging and Vision*, Vol. 18, No. 1, pp. 17–33, Jan. 2003.

[Srou 07] F. Sroubek, G. Cristobal, and J. Flusser. "A Unified Approach to Superresolution and Multichannel Blind Deconvolution". *IEEE Transactions on Image Processing*, Vol. 16, No. 9, pp. 2322–2332, Sep. 2007.

[Staa 05] J. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken. "Ridge based Vessel Segmentation in Color Images of the Retina". *IEEE Transactions on Medical Imaging*, Vol. 23, No. 4, pp. 501–509, Apr. 2005.

[Star 89] H. Stark and P. Oskoui. "High-Resolution Image Recovery from Image-Plane Arrays, using Convex Projections". *Journal of the Optical Society of America A*, Vol. 6, No. 11, pp. 1715–1726, Nov. 1989.

[Su 11] H. Su, Y. Wu, and J. Zhou. "Adaptive Incremental Video Super-Resolution with Temporal Consistency". In: *International Conference on Image Processing (ICIP)*, pp. 1149–1152, IEEE, Brussels, Belgium, Sep. 2011.

[Take 07] H. Takeda, S. Farsiu, and P. Milanfar. "Kernel Regression for Image Processing and Reconstruction". *IEEE Transactions on Image Processing*, Vol. 16, No. 2, pp. 349–366, Feb. 2007.

[Tana 05] M. Tanaka and M. Okutomi. "Theoretical Analysis on Reconstruction-Based Super-Resolution for an Arbitrary PSF". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 947–954, IEEE, San Diego, CA, June 2005.

[Tana 10] M. Tanaka and M. Okutomi. "Toward Robust Reconstruction-Based Super-Resolution". In: P. Milanfar, Ed., *Super-Resolution Imaging*, pp. 219–246, CRC Press, 2010.

[Tao 17] X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia. "Detail-Revealing Deep Video Super-Resolution". In: *International Conference on Computer Vision (ICCV)*, p. to appear, Venice, Italy, Oct. 2017.

[Teka 92] A. Tekalp, M. Ozkan, and M. Sezan. "High-Resolution Image Reconstruction from Lower-Resolution Image Sequences and Space-Varying Image Restoration". In: *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 169–172, IEEE, San Francisco, CA, March 1992.

[Thap 14] D. Thapa, K. Raahemifar, W. R. Bobier, and V. Lakshminarayanan. "Comparison of Super-Resolution Algorithms Applied to Retinal Images". *Journal of Biomedical Optics*, Vol. 19, No. 5, pp. 056002–1–056002–16, May 2014.

[Tipp 03] M. E. Tipping and C. M. Bishop. "Bayesian Image Super-resolution". In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 1303–1310, MIT Press, Vancouver, Canada, Dec. 2003.

[Tom 94] B. Tom, A. Katsaggelos, and N. Galatsanos. "Reconstruction of a High Resolution Image From Registration and Restoration of Low Resolution Images". In: *International Conference on Image Processing (ICIP)*, pp. 553–557, IEEE, Austin, TX, Nov. 1994.

[Tom 95] B. Tom and A. Katsaggelos. "Reconstruction of a High-Resolution Image by Simultaneous Registration, Restoration, and Interpolation of Low-Resolution Images". In: *International Conference on Image Processing (ICIP)*, pp. 539–542, IEEE, Washington D.C., Oct. 1995.

[Toma 98] C. Tomasi and R. Manduchi. "Bilateral Filtering for Gray and Color Images". In: *International Conference on Computer Vision (ICCV)*, pp. 839–846, IEEE, Bombay, India, Jan. 1998.

[Torn 15] R. P. Tornow, R. Kolár, and J. Odstrčilík. "Non-Mydriatic Video Ophthalmoscope to Measure Fast Temporal Changes of the Human Retina". In: A. Amelink and I. A. Vitkin, Eds., *Proc. SPIE 9540, Novel Biophotonics Techniques and Applications III*, p. 954006, Munich Germany, June 2015.

[Torr 00] P. H. S. Torr and A. Zisserman. "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry". *Computer Vision and Image Understanding*, Vol. 78, No. 1, pp. 138–156, Apr. 2000.

[Tsai 84] R. Tsai and T. S. Huang. "Multiframe Image Restoration and Registration". *Advances in Computer Vision and Image Processing*, Vol. 1, No. 2, pp. 317–339, 1984.

[Unse 97] M. Unser and J. Zerubia. "Generalized Sampling: Stability and Performance Analysis". *IEEE Transactions on Signal Processing*, Vol. 45, No. 12, pp. 2941–2950, Dec. 1997.

[Ur 92] H. Ur and D. Gross. "Improved Resolution From Subpixel Shifted Pictures". *CVGIP: Graphical Models and Image Processing*, Vol. 54, No. 2, pp. 181–186, March 1992.

[Vand 06a] P. Vandewalle. *Super-Resolution from Unregistered Aliased Images*. PhD thesis, Lausanne, EPFL, 2006.

[Vand 06b] P. Vandewalle, S. Süsstrunk, and M. Vetterli. "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution". *EURASIP Journal on Advances in Signal Processing*, Vol. 2006, No. 1, pp. 1–14, Article ID 71459, Feb. 2006.

[Vand 07] P. Vandewalle, L. Sbaiz, J. Vandewalle, and M. Vetterli. "Super-Resolution From Unregistered and Totally Aliased Signals Using Subspace Methods". *IEEE Transactions on Signal Processing*, Vol. 55, No. 7, pp. 3687–3703, July 2007.

[Vand 10] P. Vandewalle, L. Sbaiz, and M. Vetterli. "Registration for Super-Resolution: Theory, Algorithms, and Applications in Image and Video Enhancement". In: P. Milanfar, Ed., *Super-Resolution Imaging*, pp. 155–185, CRC Press, 2010.

[Vega 06] M. Vega, R. Molina, and A. K. Katsaggelos. "A Bayesian Super-Resolution Approach to Demosaicing of Blurred Images". *EURASIP Journal on Advances in Signal Processing*, Vol. 2006, No. 1, pp. 1–13, Article ID 25072, Dec. 2006.

[Vrig 12] M. Vrigkas, C. Nikou, and L. P. Kondi. "A Fully Robust Framework for MAP Image Super-Resolution". In: *International Conference on Image Processing (ICIP)*, pp. 2225–2228, IEEE, Orlando, FL, Sep. 2012.

[Vrig 14] M. Vrigkas, C. Nikou, and L. P. Kondi. "Robust Maximum a Posteriori Image Super-Resolution". *Journal of Electronic Imaging*, Vol. 23, No. 4, pp. 043016–1–043016–12, July 2014.

[Vu 12] C. T. Vu, T. D. Phan, and D. M. Chandler. "S3: A Spectral and Spatial Measure of Local Perceived Sharpness in Natural Images". *IEEE Transactions on Image Processing*, Vol. 21, No. 3, pp. 934–45, March 2012.

[Wang 04a] Z. Wang and F. Qi. "On Ambiguities in Super-Resolution Modeling". *IEEE Signal Processing Letters*, Vol. 11, No. 8, pp. 678–681, Aug. 2004.

[Wang 04b] 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*, Vol. 13, No. 4, pp. 600–612, Apr. 2004.

[Wang 14] C. Wang, R. Palomar, and F. A. Cheikh. "Stereo Video Analysis for Instrument Tracking in Image-Guided Surgery". In: *European Workshop on Visual Information Processing (EUVIP)*, pp. 1–6, IEEE, Paris, France, Dec. 2014.

[Wasz 11a] J. Wasza, S. Bauer, S. Haase, M. Schmid, S. Reichert, and J. Hornegger. "RITK: The Range Imaging Toolkit - A Framework for 3-D Range Image Stream Processing". In: *Vision, Modeling, and Visualization (2011)*, pp. 57–64, The Eurographics Association, Berlin, Germany, Oct. 2011.

[Wasz 11b] J. Wasza, S. Bauer, and J. Hornegger. "High Performance GPU-based Preprocessing for Time-of-Flight Imaging in Medical Applications". In: *Bildverarbeitung für die Medizin (BVM)*, pp. 324–328, Springer, Lübeck, Germany, March 2011.

[Wasz 11c] J. Wasza, S. Bauer, and J. Hornegger. "Real-Time Preprocessing for Dense 3-D Range Imaging on the GPU: Defect Interpolation, Bilateral Temporal Averaging and Guided Filtering". In: *International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 1221–1227, IEEE, Barcelona, Spain, Nov. 2011.

[Wen 12] Y.-W. Wen and R. H. Chan. "Parameter Selection for Total-Variation-Based Image Restoration using Discrepancy Principle". *IEEE Transactions on Image Processing*, Vol. 21, No. 4, pp. 1770–1781, Apr. 2012.

[Wetz 13] J. Wetzl, O. Taubmann, S. Haase, T. Köhler, M. Kraus, and J. Hornegger. "GPU-Accelerated Time-of-Flight Super-Resolution for Image-Guided Surgery". In: *Bildverarbeitung für die Medizin (BVM)*, pp. 21–26, Springer, Heidelberg, Germany, March 2013.

[Xiao 11] Y. Xiao, T. Zeng, J. Yu, and M. K. Ng. "Restoration of Images Corrupted by Mixed Gaussian-Impulse Noise via L1-L0 Minimization". *Pattern Recognition*, Vol. 44, No. 8, pp. 1708–1720, Aug. 2011.

[Yan 13] Q. Yan, X. Shen, L. Xu, S. Zhuo, X. Zhang, L. Shen, and J. Jia. "Cross-Field Joint Image Restoration via Scale Map". In: *International Conference on Computer Vision (ICCV)*, pp. 1537–1544, IEEE, Sydney, Australia, Dec. 2013.

[Yang 10] J. Yang, J. Wright, T. S. Huang, and Y. Ma. "Image Super-Resolution via Sparse Representation". *IEEE Transactions on Image Processing*, Vol. 19, No. 11, pp. 2861–2873, Nov. 2010.

[Yap 09] K.-H. Yap, Y. He, Y. Tian, and L.-P. Chau. "A Nonlinear L1-Norm Approach for Joint Image Registration and Super-Resolution". *IEEE Signal Processing Letters*, Vol. 16, No. 11, pp. 981–984, Nov. 2009.

[Yega 12] H. Yeganeh, M. Rostami, and Z. Wang. "Objective Quality Assessment for Image Super-Resolution: A Natural Scene Statistics Approach". In: *International Conference on Image Processing (ICIP)*, pp. 1481–1484, IEEE, Orlando, FL, Sep. 2012.

[Yin 96] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo. "Weighted Median Filters: A Tutorial". *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 43, No. 3, pp. 157–192, March 1996.

[Yuan 12] Q. Yuan, L. Zhang, and H. Shen. "Multiframe Super-Resolution Employing a Spatially Weighted Total Variation Model". *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 3, pp. 379–392, March 2012.

[Yuan 13] Q. Yuan, L. Zhang, and H. Shen. "Regional Spatially Adaptive Total Variation Super-Resolution with Spatial Information Filtering and Clustering". *IEEE Transactions on Image Processing*, Vol. 22, No. 6, pp. 2327–2342, June 2013.

[Yue 14] L. Yue, H. Shen, Q. Yuan, and L. Zhang. "A Locally Adaptive L1-L2 Norm for Multi-Frame Super-Resolution of Images with Mixed Noise and Outliers". *Signal Processing*, Vol. 105, No. 1, pp. 156–174, Dec. 2014.

[Zale 10] Z. Zalevsky. "Exceeding the Diffraction and the Geometric Limits of Imaging systems: A Review". In: *International Workshop on Optical Supercomputing*, pp. 119–130, LNCS Vol. 6748, Bertinoro, Italy, Nov. 2010.

[Zale 13] Z. Zalevsky, S. Gaffling, J. Hutter, L. Chen, W. Iff, A. Tobisch, J. Garcia, and V. Mico. "Passive Time-Multiplexing Super-Resolved Technique for Axially Moving Targets". *Applied Optics*, Vol. 52, No. 7, pp. C11–C15, March 2013.

[Zeng 13] X. Zeng and L. Yang. "A Robust Multiframe Super-Resolution Algorithm based on Half-Quadratic Estimation with Modified BTV Regularization". *Digital Signal Processing*, Vol. 23, No. 1, pp. 98–109, Jan. 2013.

[Zhan 08] X. Zhang and Z. Liu. "Superlenses to Overcome the Diffraction Limit". *Nature Materials*, Vol. 7, No. 6, pp. 435–441, June 2008.

[Zhan 12a] H. Zhang, L. Zhang, and H. Shen. "A Super-Resolution Reconstruction Algorithm for Hyperspectral Images". *Signal Processing*, Vol. 92, No. 9, pp. 2082–2096, Sep. 2012.

[Zhan 12b] X. Zhang, J. Jiang, and S. Peng. "Commutability of Blur and Affine Warping in Super-Resolution With Application to Joint Estimation of Triple-Coupled Variables". *IEEE Transactions on Image Processing*, Vol. 21, No. 4, pp. 1796–808, Apr. 2012.

[Zhan 14a] H. Zhang and L. Carin. "Multi-Shot Imaging: Joint Alignment, Deblurring, and Resolution-Enhancement". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2925–2932, IEEE, Columbus, OH, June 2014.

[Zhan 14b] Q. Zhang, L. Xu, and J. Jia. "100+ Times Faster Weighted Median Filter (WMF)". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2830–2837, IEEE, Columbus, OH, June 2014.

[Zhao 02] W. Zhao and H. S. Sawhney. "Is Super-Resolution with Optical Flow Feasible?". In: *European Conference on Computer Vision (ECCV)*, pp. 599–613, LNCS Vol. 2350, Springer, Copenhagen, Denmark, May 2002.

[Zhen 12] Y. Zheng, B. Vanderbeek, R. Xiao, E. Daniel, D. Stambolian, M. Maguire, J. O'Brien, and J. Gee. "Retrospective Illumination Correction of Retinal Fundus Images from Gradient Distribution Sparsity". In: *International Symposium on Biomedical Imaging (ISBI)*, pp. 972–975, IEEE, Barcelona, Spain, May 2012.

[Zhen 14] Y. Zheng, E. Daniel, A. A. Hunter, R. Xiao, J. Gao, H. Li, M. G. Maguire, D. H. Brainard, and J. C. Gee. "Landmark Matching based Retinal Image Alignment by Enforcing Sparsity in Correspondence Matrix". *Medical Image Analysis*, Vol. 18, No. 6, pp. 903–913, Aug. 2014.

[Zhu 10] X. Z. X. Zhu and P. Milanfar. "Automatic Parameter Selection for Denoising Algorithms Using a No-Reference Measure of Image Content". *IEEE Transactions on Image Processing*, Vol. 19, No. 12, pp. 3116–3132, Dec. 2010.

[Zibe 07] M. Zibetti and J. Mayer. "A Robust and Computationally Efficient Simultaneous Super-Resolution Scheme for Image Sequences". *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 10, pp. 1288–1300, Oct. 2007.

[Zome 00] A. Zomet and S. Peleg. "Efficient Super-Resolution and Applications to Mosaics". In: *International Conference on Pattern Recognition (ICPR)*, pp. 579–583, IEEE, Barcelona, Spain, Sep. 2000.

[Zome 01] A. Zomet, A. Rav-Acha, and S. Peleg. "Robust Super-Resolution". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Kauai, HI, Dec. 2001.

[Zome 02] A. Zomet and S. Peleg. "Multi-Sensor Super-Resolution". In: *Workshop on Applications of Computer Vision (WACV)*, pp. 27–31, IEEE, Orlando, FL, Dec. 2002.

[Zoub 12] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma. "Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts". *IEEE Signal Processing Magazine*, Vol. 29, No. 4, pp. 61–80, July 2012.