# Blind Deconvolution
## in Scientific Imaging & Computational Photography

**Dissertation**

der Mathematisch - Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Dipl.-Phys. Michael Stefan Hirsch

aus Neumarkt i.d.Opf

Tübingen

2011

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 25.06.2012 |
| Dekan: | Prof. Dr. Wolfgang Rosenstiel |
| 1. Berichterstatter: | Prof. Dr. Claus Zimmermann |
| 2. Berichterstatter: | Prof. Dr. Hanns Ruder |

# Zusammenfassung

Die Restaurierung digitaler Bilder ist aufgrund ihrer vielfältigen Anwendung sowohl in der wissenschaftlichen Bildgebung, als auch im Bereich der alltäglichen Fotografie ein Schlüsselbereich der Signal- und Bildverarbeitung. Ein wichtiger Teilbereich ist der Bereich der Bilddekonvolution, welcher aufgrund seiner theoretischen und praktischen Implikationen zunehmend Interesse seitens der Wissenschaft als auch der Industrie erfährt. Das Ziel der *klassischen* oder *nicht blinden* Dekonvolution ist die Wiederherstellung eines scharfen Bildes, unter der Annahme, dass der Unschärfeprozess bekannt ist. Gegenstand dieser Arbeit ist das sehr viel schwierigere aber auch realistischere Problem der *blinden* Bilddekonvolution, bei welchem, im Gegensatz zur klassischen Dekonvolution, keine genaue Kenntnis über den Unschärfevorgang vorliegt. Das Orginalbild muss allein aus seinen unscharfen und möglicherweise verrauschten Aufnahmen rekonstruiert werden. Das Ziel der vorliegenden Arbeit ist es, den Stand der Wissenschaft und Technik im Bereich der blinden Dekonvolution voranzutreiben und dadurch das Spektrum möglicher Anwendungen im Bereich der Alltagsfotografie, aber auch auf dem Gebiet der wissenschaftlichen Bildgebung, bspw. in der Astronomie oder Mikroskopie, zu erweitern.

Zu diesem Zweck erarbeiten wir ein mathematisch solides und physikalisch wohl motiviertes Framework, welches die Beschreibung und effiziente Berechnung von ortsabhängiger Unschärfe ermöglicht. Wir leiten unser sogenanntes Efficient Filter Flow Framework als diskrete Approximation der inkohärenten Abbildungsgleichung her und entwickeln Formeln zur effizienten Implementierung mittels einer zweidimensionalen Verallgemeinerung der Kurzzeit-Fourier-Transformation. Unser Framework verallgemeinert das gemeinhin verwendete Model, welches annimmt, dass die Unschärfe über das ganze Bild hinweg gleich ist, und erweitert damit den Anwendungsbereich blinder Dekonvolutionsmethoden erheblich.

In einer Reihe anspruchsvoller Anwendungen mit realen Daten stellen wir die Gültigkeit und vielseitige Anwendbarkeit unseres Ansatzes unter Beweis. Inbesondere wird sich seine Nützlichkeit bei der Rekonstruktion eines scharfen Bildes aus einer Sequenz von Kurzzeitaufnahmen zeigen, deren Qualität durch atmosphärische Luftturbulenzen beeinträchtigt wurde. Um Nutzen aus der großen Menge an verfügbaren Bilddaten in der astronomischen Bildgebung ziehen zu können, entwickeln wir einen Algorithmus zur blinden Dekonvolution, welcher die hohen Rechenanforderungen bisheriger Methoden umgeht und insbesondere nicht in der Anzahl verwendbarer Bilder eingeschränkt ist. Darüberhinaus erweitern wir das bildgebende Modell dahingehend, dass es nicht nur ortsabhängige Unschärfe, sondern auch den Effekt der Saturierung und Superresolution beschreibt.

Ein weiteres schwieriges Problem, welches den Wert unseres Frameworks unterstreicht,

ist die Korrektur verwackelter Bilder. Wir erweitern unser Modell, um den spezifischen Eigenschaften von Kameraverwacklung Rechnung zu tragen, und entwickeln einen leistungsfähigen Algorithmus, welcher bisherige Methoden sowohl in der Qualität der rekonstruierten Bilder als auch Rechenzeit übertrifft.

Zum Schluss stellen wir einen neuartigen blinden Dekonvolutionsalgorithmus zur Verbesserung der Auflösung dreidimensionaler Elektronendichten vor, wir zeigen auf, wie die blinde Dekonvolution auch im interessanten Feld der Cryo-Elektronenmikroskopie eingesetzt werden kann. Ergebnisse für sowohl simulierte als auch experimentelle Daten zeigen, dass unser Ansatz als flexibles und generisches Hilfsmittel zur molekularen Strukturbestimmung beiträgt.

# Summary

Digital image restoration is a key area in signal and image processing due to its many applications in both scientific imaging as well as everyday photography. An important subdiscipline, that is receiving an ever increasing interest from the academic as well the industrial world, is the field of *image deconvolution*, which enjoys this interest due to both its theoretical and practical implications. While *classical* or *non-blind* image deconvolution aims at restoring a sharp latent image assuming the blur is known, this thesis addresses the much harder but also more realistic problem of *blind* image deconvolution, where the degradation is unknown. An estimate of the original image must be obtained using only its blurred and possibly noise corrupted observations. The aim of this thesis is to advance the state-of-the-art in the field of blind deconvolution and thereby to broaden the applicability of blind deconvolution techniques in everyday photography but also in scientific imaging such as astronomical and microscopic imaging.

To this end, we develop a mathematically sound and physically well-motivated framework, which allows to express and efficiently compute spatially-varying blur. We derive our "Efficient Filter Flow" framework as a discrete approximation of the incoherent imaging equation and devise expressions for its efficient implementation using the short-time Fourier transform. By extending the commonly employed invariant blur model, our framework substantially broadens the application range of blind deconvolution methods.

In a number of challenging real-world applications we demonstrate both the validity and versatility of our approach. In particular, we utilise our model for reconstructing a sharp latent image from a sequence of short-exposure images degraded by atmospheric turbulence. To capitalise on the abundance of data available in astronomical imaging, we develop a blind deconvolution algorithm, which bypasses the computational burden of current blind deconvolution methods that are restricted in the number of observations they can process. In addition, we further extend the imaging model to not only include the effect of spatially-varying blur but also to account for the effects of saturation and super-resolution.

Another challenging application which proves the usefulness of our framework, is the problem of removing camera shake from a single image. We extend our model to incorporate the particularities of camera shake and develop an efficient algorithm that outperforms state-of-the-art methods in both restoration quality and computation time.

Finally, by presenting a novel blind deconvolution algorithm for improving the resolution of three-dimensional electron density maps, we promote the use of blind deconvolution in the interesting field of cryo-electron microscopy. Results on both simulated and experimental density maps demonstrate that our approach serves as a flexible and generic tool that facilitates structure elucidation of macromolecular assemblies.

# Acknowledgements

First and foremost, I would like to thank my supervisor Bernhard Schölkopf. It has been a great privilege and honour to work and study under his guidance. I am most thankful for his unreserved support, trust and empathy. It is his ingenuity, encouragement, and naturalness, which sustains an exceptional and most enjoyable research environment distinguished by sincerity and generosity in sharing ideas amongst his group members.

My deep gratitude also goes to my mentor Hanns Ruder for his continued and reliable support throughout these years. His benevolence, warmheartedness, and constructive criticism have been constant sources of encouragement.

The work presented in this thesis would not have been possible without the innumerable enriching discussions with my mentors and friends Stefan Harmeling, Michael Habeck and Suvrit Sra. Thank you very much indeed for your constant and valuable advice and mentoring, for fruitful research co-operations and not least for proof-reading parts of this thesis. Many thanks also to my friend Vincent Gimbert for carefully reading a final draft and his valuable suggestions that helped to improve the presentation.

Furthermore, I thank all my colleagues and PhD fellows for open and stimulating discussions, honest and critical feedback and not least for sharing their enthusiasm for science and research. It has been a great pleasure and I am very grateful that I got the chance to work with so many inspiring people.

I also thank Sebastian Stark, Martin Lang and especially Sabrina Rehbaum for their support in all technical and organisational issues as well as the technical workshop at the MPI for the pleasant and uncomplicated cooperation.

Not at least, I would like to express my deep gratitude to my current supervisor Sarah Bridle for her most valuable advice, trust and encouragement, kindness and patience.

My special thanks go to the lovely town of Tübingen. Its beautiful meadows and forests have always been a source of inspiration and vitality, and kept me appreciating the beauty of life.

The last word of thanks and my deepest gratitude goes to my family and my beloved wife Maria for their unconditional love and support, their unshakable trust and faith, their warmhearted understanding and constant encouragement.

# Contents

# List of Figures

# List of Algorithms



# List of Tables



# List of Videos

# Chapter 1

# Introduction

In 2009, Williard S. Boyle and George E. Smith received the Nobel Prize for Physics for an invention that revolutionised our perception of our environment and moreover of the entire universe: "an imaging semiconductor circuit – the Charged Coupled Device (CCD) sensor". Through a photosensitive semiconductor a CCD sensor allows a localised record of photons resulting in digital images with increased spatial, spectral and dynamic resolution when compared to traditional photo imaging techniques. Both the possibility to record single photons as well to integrate light extend the capabilities of the human visual system, which proved to have tremendous implications on our view of the world. For instance, the CCD sensor allows deeper views in the universe and thus in our past and has become an essential implement in astronomical observation for answering some of the most profound questions of mankind. But not only for elucidating the origin of our mere existence but also for life as such, the CCD sensor has become indispensable as a key instrument in microscopic imaging and life sciences in general.

However, since its invention in 1969, the CCD sensor has not only become an integral part of scientific imaging, but also of our everyday life. The advent of digital photo sensors heralded the dawn of today's *information society* and lastingly affected all areas of social life and culture: communication, education, entertainment and art, *etc*. Digital images and videos are ubiquitous and have become indisputably the main carrier of information over the last few decades. Undoubtedly, this development is also due to the availability of ever faster computer technology which has given rise to the establishment of new fields of research such as digital image processing, computer vision, and computational photography.

The importance of image processing techniques to enhance the quality of digital images was stressed by George E. Smith in his speech at the Nobel Banquet in the Stockholm City Hall[1] by noticing that "no device is ever perfect". Sources of image degradation can be manifold, ranging from manufacturing defects of the CCD sensor, readout noise to blurring stemming from imperfect optics or detrimental imaging conditions, such as unwanted camera motion or atmospheric turbulence, at the time of recording.

*Digital image restoration* as a key area of signal and image processing aims at computationally enhancing the quality of images by undoing the adverse effects of image degradation

[1]Available from: `http://www.nobelprize.org/nobel_prizes/physics/laureates/2009/smith-speech.html`

such as noise and blur and plays an ever important role in both scientific imaging and everyday photography. An important sub-discipline is the field of *image deconvolution*, that has recently attracted much attention due to both its theoretical and practical implications. Classical image deconvolution aims at restoring a sharp latent image from its noisy and blurry observation assuming the blur is known.

Probably the most prominent example underpinning the importance of image deconvolution is the optical aberration problem of the Hubble Space Telescope (Hubble Space Telescope (HST)) at the beginning of its mission in 1990. Since the primary mirror was ground to the wrong shape, the resolution obtained was drastically lower than expected, as revealed by the significant spread in images of point-like stars. This issue triggered considerable work in the field and led to the development of novel and effective techniques for image deconvolution. Since the aberrations of the HST optical system could be well characterised, image quality could be restored by classical deconvolution techniques to a large extent. For a detailed overview of this work, we refer to White (1992) and Adorf et al. (1995).

Since then, image deconvolution has proven to be a powerful tool also in many other imaging applications ranging from medical imaging to remote sensing, microscopy and everyday photography, thus receiving an ever increasing interest from the academic as well the industrial world.

While *classical* or *non-blind* image deconvolution aim at restoring a sharp latent image assuming the blur is known, this thesis addresses the much harder but also more realistic problem of *blind* image deconvolution, where the degradation is unknown. Blind Deconvolution (BD) is a severely ill-posed problem because there exists an infinite number of solutions and small perturbations in the data lead to large distortions in the estimated latent image. As such, BD involves many challenging problems, including modelling the image formation process, formulating tractable priors that constrain the space of admissible solutions by the incorporation of additional information and prior knowledge, as well as devising efficient inference and optimisation methods. This renders it an intriguing but also intricate task, which has recently seen much attention as well as progress in both the image and signal processing and also the computer vision and graphics community. The overall aim of this thesis is to advance the state-of-the-art in the field of BD and thereby broaden the applicability of BD techniques in everyday photography and also in scientific imaging such as astronomical and microscopic imaging.

## 1.1 Contributions

Our contributions presented in this thesis are twofold. *First*, based on the fundamental laws of physical optics, we develop a novel and generic framework that allows us to express and efficiently compute spatially varying blur. We derive our EFF framework as a discrete approximation of the incoherent imaging equation that describes the image formation process of a linear optical system under incoherent illumination. We discuss its numerical properties and demonstrate in a number of important real-world applications both the validity as well as the versatility of our framework. *Second*, we develop novel solutions to challenging BD problems in both scientific imaging and everyday photography, that surpass current state-of-

the-art methods. To this end, we elaborate various imaging models to incorporate the effects of spatially varying blur, super-resolution, saturation, and camera shake and derive novel inference methods that are able to recover a sharp latent image given blurred and noise corrupted image data only. In particular, only generic information about the degradation process is available without explicit knowledge of the underlying blurring parameters.

## 1.2 Overview - How to Read This Text

Besides the introduction and conclusion, this thesis can be structured in two parts: *the first part* comprises Chapter 2, and lays the theoretical foundations for the remainder of this thesis by providing a mathematically sound and physically well-motivated derivation of the EFF framework. *The second part* consists of Chapters 3, 4, and 5. In a number of interesting real-world applications, we present novel approaches to challenging BD problems, that surpass current state-of-the-art methods.

Although each chapter of the second part makes use of the findings that are developed in the first part, efforts have been made to make each chapter self-contained. Hence, readers who are primarily interested in the algorithmic and application-oriented contributions of this thesis, may skip Chapter 2.

## 1.3 Synopsis - About This Text

*The first part of this thesis* provides a detailed derivation of a novel approximation of the incoherent imaging equation called EFF, which offers increased expressiveness and enables the efficient computation of blur that smoothly varies across the image plane. It extends the well-known and commonly employed invariant convolutional model without sacrificing its numerical benefits: by employing the short-time Fourier transform, the forward model of the EFF framework can be computed almost as efficient as an ordinary convolution. As an approximation of the incoherent imaging equation, it features linearity in both the PSF and latent image parameters. Furthermore, we derive expressions for transposing the linear model, which allow efficient computation of the gradients with respect to the PSF and latent image parameters, respectively. As a consequence, gradient based numerical optimisation methods (quasi-Newton methods) can be employed, which facilitate and speed up inference.

*In the second part of this thesis*, we develop various BD algorithms for applications ranging from astronomical imaging to computational photography and cryo electron microscopy:

- In Chapter 3, we present a novel approach to the multi-frame BD problem, which is able to restore a potentially higher resolved latent image given a sequence of blur and noise corrupted images. By assuming that the underlying scene is static and by performing deconvolution on a frame per frame basis, it gradually improves the reconstructed latent image. In other words, the more images have been processed the better the restoration result. We show how to incorporate the effects of saturation and

super-resolution, which allows the restoration of a higher resolved image provided that enough observations are available. We demonstrate the capabilities of our approach for both invariant and spatially-varying blur on simulated data as well as a number of real-world astronomical imaging examples comprising stellar and planetary imaging data.

- In Chapter 4, we address the problem of removing camera shake from a single blurry image. We extend the EFF framework to account for the particularities of camera shake and develop a fast single image blind deconvolution algorithm that performs image restoration in two steps: 1) it analyses the image by estimating the latent camera motion via an efficient multi-hierarchical inference scheme; 2) it restores the underlying sharp image by a non-blind deconvolution with the spatially-varying blur kernel that corresponds to the camera motion estimated in the first step. To compensate for the imperfections of the estimated PSF and to counter noise in the blurry photo, we employ a regularisation term that promotes natural image statistics and effectively guides the deconvolution process. In a comprehensive comparison on a number of real-world examples we demonstrate the extended capabilities and improved performance of our approach.

- In Chapter 5, we derive a novel algorithm for improving the resolution of intermediate- and low-resolution three-dimensional density maps of macromolecular assemblies and large bio-molecular complexes obtained by cryo-electron microscopy. Our approach models the low-resolution density map as a blurred and noise corrupted version of a latent high-resolution density map which we seek to recover by BD. By assuming non-negativity and i.i.d. Gaussian noise we derive multiplicative updates that alternatingly and iteratively estimate the unknown PSF as well as the latent high-resolution density map and allow the incorporation of additional prior terms while preserving the non-negativity of the recovered densities. A hierarchical Bayes model that allows the estimation of latent model parameters renders our algorithm fully parameter-free. Results on both simulated and experimental density maps demonstrate the improved performance and versatility of our approach when compared with state-of-the-art solutions.

*In a final conclusion* in Chapter 6, we summarise the contributions of this thesis, discuss a few open problems and interesting future directions, and outline some ongoing projects that are directly related to the work presented in this thesis.

## 1.4 Supplementary Material

Throughout this thesis we included links to supplementary and multimedia content in the form of hyperlinked Quick Response (QR) codes. The QR codes can be scanned with any mobile device such as smart phones or tablet PCs with an appropriate software application installed. For MacOS and Android based devices we recommend the free app "Scan"[2].

[2]Available from: `http://www.qrcodecity.com`

Given a working internet connection, the scanned QR codes will direct to supplementary video files hosted by youtube.com. Reading the thesis as a pdf on screen, the links can be readily accessed by clicking on the QR code images as they are hyperlinked to the corresponding files. All videos are encoded with a H.264 codec which is fairly standard on most modern systems and hence shouldn't necessitate the installation of any additional software. All links have been tested on a PC, Apple iphone4 and a Android based Samsung GalaxyTab 10.1. However, if a link seems broken, please don't hesitate to contact the author of this thesis at michael.hirsch@tuebingen.mpg.de.

# Chapter 2

# Incoherent Imaging Equation and Efficient Numerical Approximations

An integral part in tackling any inference problem in science or engineering is to develop a good understanding of the underlying physics, with the aim of finding a faithful mathematical model or representation of the data generation process.

To this end, this chapter discusses the theoretical foundations of the remainder of this thesis. Starting from first principles, we derive the incoherent imaging equation, which describes the image formation process of most common optical systems, in Section 2.1. This will highlight the implicit assumptions that underly the work presented in this thesis. At the same time, it serves as the starting point for a comprehensive discussion of various imaging models for both invariant and space-variant point spread functions (PSFs) which we present in Section 2.2. In this context, we propose a novel approximation of the incoherent imaging equation for modelling spatially varying blur, which generalises the well-known invariant convolution model and comprises it as a special case. By discussing the benefits and shortcomings of existing approaches, we put our work into perspective. Finally, in Section 2.3, we detail an efficient implementation of our proposed imaging model, which allows rapid computation and facilitates both image and PSF estimation, both of which is needed for Blind Deconvolution (BD). A section on related work and a short summary will conclude this chapter.

## 2.1 Introduction

In this first introductory section we revise the basics of classical optics with the aim of deriving the incoherent imaging equation at the end of this section. The incoherent imaging equation is of utter importance as it describes the image formation process of almost all optical systems that are considered in the field of image processing. It is widely used and does also serve as the starting point of the exposition to follow. By doing so, we will learn about its underlying physical assumptions, understand its range of validity, and not least set the basis for the material covered in this thesis.

### 2.1.1 Scalar Theory of Light

Light is electromagnetic radiation and as such governed by Maxwell's equations – a set of partial differential equations that form the foundation of classical electrodynamics including classical optics. Although electric and magnetic fields are vectorial in nature, in many situations[1] their behaviour can be well described by a single scalar wave equation

$$(\nabla^2 - \frac{n^2}{c^2}\frac{\partial^2}{\partial t^2})\,\Phi(r,t) = 0, \tag{2.1}$$

where $\Phi(r,t)$ is any of the scalar field components of the electric or magnetic field and $n$ denotes the refractive index of the medium, within which the light is propagating. In particular, the scalar theory disregards any polarisation effects, *i.e.* any *coupling* between the electric and magnetic fields. Restricting ourselves to monochromatic light, the solution to (2.1) are planar waves of the form

$$\Phi(r,t) = \Re\left\{A(r)\exp(-i(2\pi\nu t + kr))\right\}, \tag{2.2}$$

where $A(r)$ denotes the wave amplitude at position $r$. The wave number $k$, frequency $\nu$ and wavelength $\lambda$ are related via the dispersion relation

$$k = \frac{2\pi\nu n}{c} = \frac{2\pi}{\lambda}.$$

Since (2.1) is a *linear* partial differential equation, any linear combination of its solutions yields another solution to Equation (2.1). The single property of linearity has major implications for the mathematical treatment of physical systems as it allows us to analyse a system by studying its response to a single point stimulus. Its effect to a complex input signal can then be obtained by considering the input signal being composed of point stimuli and adding up their known responses accordingly.

### 2.1.2 Linear Systems Theory

A system, defined as a mapping $\mathcal{S}$ from an input function $\Phi$ to an output function $\Psi$, is said to be linear if for any input functions $\Phi_1(r,t)$ and $\Phi_2(r,t)$ and any complex constant $\alpha$, it obeys the following two properties (Hecht, 2003):

1. *Homogeneity:*
$$\mathcal{S}\{\alpha\,\Phi_1(r,t)\} = \alpha\,\mathcal{S}\{\Phi_1(r,t)\} \tag{2.3}$$

2. *Additivity:*
$$\mathcal{S}\{\Phi_1(r,t) + \Phi_2(r,t)\} = \mathcal{S}\{\Phi_1(r,t)\} + \mathcal{S}\{\Phi_2(r,t)\}. \tag{2.4}$$

[1] More precisely, the scalar theory of electromagnetism is valid in linear, isotropic, homogeneous and non-dispersive dielectric media such as free space or a lens with constant refractive index, where all components of the electric and magnetic field behave identically (Goodman, 2005, page 35 ff).

Often, both properties are combined into a single condition known as the *superposition principle*. The virtue of linear systems is invoked by the *sifting property* (Goodman, 2005, page 20ff) of the delta function which enables the decomposition of an arbitrary signal $\Phi(r,t)$ into

$$\Phi(r,t) = \iint_{-\infty}^{\infty} \Phi(\rho,\tau)\,\delta(\rho - r;\tau - t)\;\, d\rho\, d\tau. \tag{2.5}$$

This allows us to express the response of a system $\mathcal{S}$ in terms of the system's response to a point stimulus, since

$$\mathcal{S}\{\Phi(r,t)\} = \iint_{-\infty}^{\infty} \Phi(\rho,\tau)\;\underbrace{\mathcal{S}\{\delta(\rho - r;\tau - t)\}}_{\equiv h(\rho,\tau;r,t)}\; d\rho\, d\tau, \tag{2.6}$$

where we introduced the function $h(\rho,\tau;r,t)$, which is called the *impulse response* of the system. Hence, we can fully determine and describe a system by studying its response to point sources located throughout the input domain and over time. By assuming *stationarity* in time and/or space, *i.e.* that the system's response does not change over time and/or is independent of the position of the point source, Equation (2.6) can be further simplified. The study of linear systems and their properties is subject of *Linear Systems Theory*, which plays a key role in many technical application domains including image processing.

### 2.1.3 Image Formation Under Incoherent Illumination

We are now able to mathematically describe the image formation process underlying many imaging applications including microscopical and astronomical imaging as well as photography. For ease of exposition, we consider monochromatic light and stick to the scalar theory thereby neglecting any polarisation effects, a valid and accurate assumption provided that the diffracting structures are large compared with the wavelength of light.

As depicted in Figure 2.1, let $\Sigma_1$ denote the *object plane* with coordinates $(\xi,\eta)$ and $\Sigma_2$ the *image plane* described by coordinates $(u,v)$. We assume a *linear* optical system $\mathcal{S}$ (*e.g.* a lens as shown in Figure 2.1) that is fully characterised by its impulse response or PSF $h(u,v;\xi,\eta)\colon \Sigma_1 \times \Sigma_2 \to \mathbb{R}$. According to the *superposition integral* (2.6), an object with a light distribution $\Phi(\xi,\eta;t)\colon \Sigma_1 \times \mathbb{R} \to \mathbb{R}$ is mapped via the optical system $\mathcal{S}$ to a light distribution $\Psi(u,v;t)\colon \Sigma_2 \times \mathbb{R} \to \mathbb{R}$ in the image plane via

$$\Psi(u,v;t) = \iint_{\Sigma_1} h(u,v;\xi,\eta)\;\Phi(\xi,\eta;t)\;d\xi\, d\eta. \tag{2.7}$$

For simplicity, we assume *stationarity* of the optical system $\mathcal{S}$, *i.e.* $\mathcal{S} \neq \mathcal{S}(t)$ thereby neglecting any temporal changes. In this model, the image is formed as the combination of the system's impulse responses, where each point in the object being imaged, is itself considered as a point source of light.

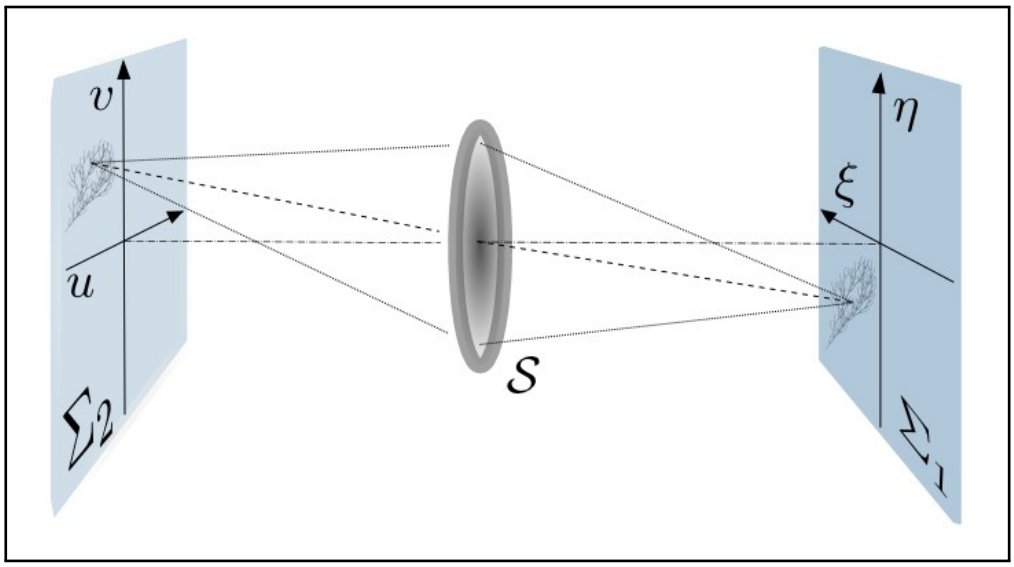



Figure 2.1: Image formation process of a linear optical system

For ease of exposition and without loss of generality, we will from here on restrict ourselves to one spatial dimension. The generalisation to two or more dimensions is straightforward and doesn't present any significantly new. Hence, Equation (2.7) becomes

$$\Psi(u,t) = \int_{\Sigma_1} h(u;\xi)\ \Phi(\xi,t)\ d\xi. \tag{2.8}$$

As discussed above, Equation (2.8) is valid for any scalar component of the electromagnetic field. However, optical detectors such as CCD sensors usually record intensities, *i.e.* the square of the field amplitude. Since the integration time is much longer than a single period of oscillation, we must average over time to obtain the measured quantities

$$\left\langle \Psi(u,t)\bar{\Psi}(u,t)\right\rangle = \iint_{\Sigma_1} h(u;\xi)\,\bar{h}(u;\xi')\ \left\langle \Phi(\xi,t)\,\bar{\Phi}(\xi',t)\right\rangle\ d\xi\,d\xi', \tag{2.9}$$

where $\langle . \rangle$ denotes temporal averaging. Here, we must take the coherence properties of the light into account and distinguish between *coherent* and *incoherent illumination*:

- In the case of *coherent illumination* , we cannot simplify Equation (2.9) any further without making any additional assumptions. The square of the complex field can lead to cancellations or other non-linear *interference* effects.

- In the case of *incoherent illumination*, the spatial correlation between any two light rays emitted from the scene is assumed to be negligible. Hence, the time average in (2.9) will only contribute to the integral for $\xi = \xi'$:

$$\left\langle \Phi(\xi,t)\,\bar{\Phi}(\xi',t)\right\rangle = |\Phi(\xi)|^2\ \delta(\xi-\xi') \equiv x(\xi)\ \delta(\xi-\xi') \tag{2.10}$$

$$\tag{2.11}$$

Plugging expression (2.11) into Equation (2.9) yields the *incohorent imaging equation*

$$y(u) = \int_{\Sigma_1} f(u;\xi)\ x(\xi)\ d\xi, \tag{2.12}$$

where we introduced $y(u)$ and $f(u;\xi)$ for $\langle|\Psi(u,t)|^2\rangle$ and $|h(u;\xi)|^2$, respectively. Both $x(\xi)$ and $y(u)$ correspond to intensities and $y(u)$ is the image of the observed object under the system $\mathcal{S}$. The impulse response $f(u;\xi)$ is called the *point spread function* of the imaging system $\mathcal{S}$ as it corresponds to the image of a point light source.

Although we had to make a number of assumptions to derive the incohorent imaging equation (2.12), it has been found to provide an accurate description for most typical imaging systems including astronomical, microscopical imaging and photography (Barnes, 1971).

## 2.2 Point Spread Function & Imaging Models

The notion of a *point spread function* (PSF) is invaluable in the description of optical systems and directly owes itself to the linearity of the underlying physical equations. As mentioned above, the PSF of an imaging system describes its response to a point source or point object. This can be readily seen by replacing $x(\xi)$ in (2.12) with $\delta(\xi - \xi')$, which describes an ideal point source located at $\xi'$ in the object plane:

$$y(u) = \int_{\Sigma_1} f(u;\xi)\ \delta(\xi - \xi')\ d\xi = f(u;\xi'). \tag{2.13}$$

The image $y(u)$ is fully determined by $f(u;\xi')$ alone. Only in an ideal, *i.e.* aberration-free imaging system, where

$$f(u;\xi') = \delta(u - \xi'). \tag{2.14}$$

each point of the object plane is mapped to a single point in the image plane, yielding a sharp image of the static scene being captured. However, in reality an ideal system does not exist and a point in the object plane will be spread or blurred according to the pattern determined by the *point spread function*.

Typically, this pattern will change with the position of the point source across the object plane, leading to a spatially varying PSF. Although a spatial variation of the PSF of any real optical system is the rule rather than an exception, it is only rarely discussed in the literature. Most often, it is argued that any imaging problem that exhibits a spatial variation of the PSF can be reduced to the invariant case by splitting it into multiple sub-problems, in each of which the assumption of an invariant PSF does hold. In contrast to this rather pragmatic viewpoint, we are aiming at a more formal treatment in the following section. This leads us to an improved mathematical description of spatially varying PSFs, which goes beyond the commonly employed approximation of a piecewise constant PSF.

### 2.2.1 Space-Invariant Systems

An important type of linear systems are spatially invariant systems, where the PSF is invariant under translations. In the invariant case, each point source of the object plane yields the same image, however shifted to the corresponding position of the point source, *i.e.*

$$f(u;\xi) = f(u-\xi). \tag{2.15}$$

By plugging expression (2.15) into the incoherent imaging equation (2.12) we obtain

$$y(u) = \int_{\Sigma_1} f(u-\xi)\; x(\xi)\; d\xi, \tag{2.16}$$

which can be readily identified as a convolution such that

$$y(u) = (f * x)(u). \tag{2.17}$$

where $*$ denotes the convolution operator. A convolution has a number of favourable properties, one of which is *e.g.* symmetry w.r.t. to its arguments, *i.e.* $f * x = x * f$. Most important in practice is the *convolution theorem*:

$$\mathscr{F}(f * x) = \mathscr{F}(f) \cdot \mathscr{F}(x), \tag{2.18}$$

which states that the convolution of two signals can be computed as the product of their Fourier transforms. This proves particularly useful for its efficient computation via the Fast Fourier Transform (FFT) (Press et al., 2007). Due to its mathematical and numerical amenities, invariant linear systems are widely studied and used. In imaging applications in particular, the convolution model is ubiquitous and modelling inaccuracies are often accepted for the sake of efficient computation.

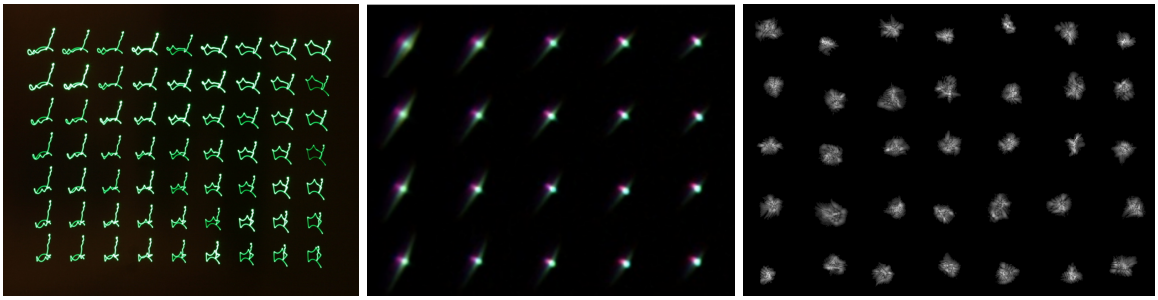
### 2.2.2 Space-Variant Systems

In many real-world imaging applications, the assumption of translational invariance does not hold. Figure 2.2 shows typical examples exhibiting spatially varying PSFs. The translation-invariant convolution model is too restrictive and insufficient for the mathematical description of such blur. In the following, we first discuss state-of-the-art approaches for modelling spatially varying PSFs before we develop a novel approximation in Subsection 2.2.2.3.

#### 2.2.2.1 Piecewise-constant PSF

A common strategy in image processing to deal with the spatial dependence of a PSF is to divide the image in small enough image regions where the assumption of an invariant PSF holds true. Mathematically, this approach corresponds to approximating the true PSF with a piecewise constant or simple function, *i.e.*

$$f(u;\xi) = \sum_{r=0}^{R-1} f^{(r)}(u-\xi)\, \chi^{(r)}(\xi) \tag{2.19}$$



Camera Shake | Optical Aberrations | Atmospheric Blur

Figure 2.2: Typical examples exhibiting spatially varying PSFs.

where $R > 0$, $f^{(r)}(u - \xi)$ are local invariant PSFs and $\chi^{(r)}(\xi)$ are indicator functions specifying the support or application domain of each local PSF, *i.e.*

$$\chi(\xi)^{(r)} = \begin{cases} 1 & \text{for } \xi \in \Omega^{(r)} \\ 0 & \text{for } \xi \notin \Omega^{(r)} \end{cases} \tag{2.20}$$

where $\Omega^{(r)} \subseteq \Sigma_1$ is the region of influence of the local PSF $f^{(r)}(u - \xi)$.

Inserting expression (2.19) into the incoherent imaging equation (2.12) yields

$$\begin{aligned} y(u) &= \int_{\Sigma_1} \underbrace{\sum_{r=0}^{R-1} f^{(r)}(u - \xi)\, \chi^{(r)}(\xi)}_{f(u;\xi)}\; x(\xi)\; d\xi \\ &= \sum_{r=0}^{R-1} \int_{\Omega^{(r)}} f^{(r)}(u - \xi)\; x(\xi)\; d\xi, \end{aligned} \tag{2.21}$$

where we made us of the fact that $\int_{\Sigma_1} \sum_{r=0}^{R-1} \chi^{(r)}(\xi)\, d\xi = \sum_{r=0}^{R-1} \int_{\Omega^{(r)}} d\xi$ (Königsberger, 2002, 359ff).

**Discussion.** Approximating a continuous function as a step function is often quite poor unless the number of support sites is chosen sufficiently high. Blocking artifacts at the patch boundaries are almost inevitable if the PSF variation involves a relative displacement of neighbouring PSF samples. An example is shown in Figure 2.3, where despite the smooth spatial variation of the PSF severe blocking artifacts are visible.

#### 2.2.2.2 Non-stationary Combination

To remedy this shortcoming, Nagy and O'Leary (1997) proposed to split the image into image patches that overlap each other. Each image patch is processed individually assuming a spatially invariant PSF. The final image is obtained by sewing the individual image patches.

Mathematically, this imaging model can be described as

$$y(u) = \sum_{r=0}^{R-1} w^{(r)}(u) \int_{\Sigma_1} f^{(r)}(u-\xi)\; x(\xi)\; \chi^{(r)}(\xi)\; d\xi, \tag{2.22}$$

where the finite set $\{w^{(r)} | r = 0, \ldots, R-1\}$ is a partition of the image plane, *i.e.*

$$\sum_{r=0}^{R-1} w^{(r)}(u) = \mathbf{1} \qquad \forall u \in \Sigma_2 \tag{2.23}$$

with each $w^{(r)}\colon \Sigma_2 \to \mathbb{R}$ being a continuous weighting function with finite support. Note, that this corresponds to approximating $f(u;\xi)$ as

$$f(u;\xi) = \sum_{r=0}^{R-1} w^{(r)}(u)\, f^{(r)}(u-\xi). \tag{2.24}$$

Although originally in (Nagy and O'Leary, 1997) the weighting functions have been chosen to correspond to piecewise-constant and linear interpolation only, the authors mention in a subsequent work (Nagy and O'Leary, 1998), that higher-order interpolation could also be used. Furthermore, the authors show how to efficiently compute (2.22) by making use of the *overlap-add* (OLA) and *overlap-save* (OLS) method of Stockham (1966).

**Discussion.** The imaging model (2.22) improves the constant-piecewise model and is better suited to describe spatially varying PSFs as can be seen in Figure 2.3. Furthermore, it has been successfully applied for the image restoration of simulated data of the Hubble Space Telescope (Nagy and O'Leary, 1998; Bardsley et al., 2005). Recently, Šorel and Šroubek (2009) employed this model for describing motion blur and show an example where they were able to remove camera shake from a blurry photo by using a noisy/blurry image pair.

It is worth mentioning, that to our knowledge Nagy and O'Leary (1997) were the first who proposed (2.22) as a global imaging model and stressed the importance of summing up the individual image patches in each model evaluation. Heretofore, the common practice has been to perform image restoration on each patch individually and sew the results only subsequently to obtain the final restored image (*e.g.* see Trussell and Fogel (1992); Adorf (1994)). According to Nagy and O'Leary (1998) this approach can lead to "visible discontinuities at the region boundaries".

Independently, Margrave (1998) proposed a similar approach to time varying filtering in the field of geophysics. He calls his approach *non-stationary combination*, a name which we adopted here. In addition, he discusses an alternative model for time-varying filtering which he calls *non-stationary convolution*. Closely related to the latter, is the imaging model we proposed independently in (Hirsch et al., 2010) and which is subject of the next paragraph.

#### 2.2.2.3 Non-Stationary Convolution

The incoherent imaging equation (2.12) describes each point in the image plane as a weighted average of neighbouring points in the object plane. The weighting of each average is determined by the PSF. It tells us how each point of the object plane is being imaged. For a spatially varying PSF this weighting or pattern changes across the object plane. Often, we can simply obtain or measure the PSF by placing a point light source at this position of interest in the object plane and capturing an image of it. Hence, assume we have measured the PSF $f$ at a number of discrete positions $\xi_0, \xi_1, \ldots, \xi_{R-1}$ to take the values $f(u, \xi_0), f(u, \xi_1), \ldots, f(u, \xi_{R-1})$. The upper left panel of Figure 2.4 illustrates this for a one-dimensional toy model. To evaluate the PSF at intermediate positions we can simply interpolate between nearby measurements

$$f(u;\xi) = \sum_{r=0}^{R-1} w^{(r)}(\xi)\, f^{(r)}(u-\xi), \tag{2.25}$$

where for convenience, we denoted $f(u, \xi_r)$ as $f^{(r)}(u - \xi)$. Inserting this expression into (2.12) yields

$$y(u) = \sum_{r=0}^{R-1} \int_{\Sigma_1} w^{(r)}(\xi)\ f^{(r)}(u-\xi)\ x(\xi)\ d\xi. \tag{2.26}$$

We can further simplify the integral by choosing the $w^{(r)}(\xi)$ to be a partition of unity, *i.e.* a finite set $\{w^{(r)} | r = 0, \ldots, R-1\}$ with the property

$$\sum_{r=0}^{R-1} w^{(r)}(\xi) = \mathbf{1} \qquad \forall \xi \in \Sigma_1 \tag{2.27}$$

with each $w^{(r)}\colon \Sigma_1 \to \mathbb{R}$ being a continuous weighting function with finite support $\Omega^{(r)} = \text{supp}(w^{(r)})$. Hence, with

$$f(u;\xi) = \sum_{r=0}^{R-1} w^{(r)}(\xi)\, \chi(\xi)^{(r)}\ f^{(r)}(u-\xi) \tag{2.28}$$

Integral (2.26) becomes

$$\begin{aligned} y(u) &= \sum_{r=0}^{R-1} \int_{\Sigma_1} w^{(r)}(\xi)\, \chi(\xi)^{(r)}\ f^{(r)}(u-\xi)\ x(\xi)\ d\xi \\ &= \sum_{r=0}^{R-1} \int_{\Omega^{(r)}} f^{(r)}(u-\xi)\ x(\xi)\ w^{(r)}(\xi)\ d\xi. \end{aligned} \tag{2.29}$$

where we again used the fact that we can break up the integral in smaller integration domains as discussed in (2.21).

**Choice of interpolation functions.** By now, our discussion has been fairly general. We had not to make any restrictive assumptions about the sampling density of the PSF nor about the functional form of $w^{(r)}$. In particular, the weights can be chosen to be arbitrarily non-linear functions or can be tailored to a specific problem by minimising some appropriate error metric as recently shown by Denis et al. (2011). In other cases, the weights might correspond to image masks that select certain objects in the scene which undergo a different blur than the remainder of the scene being captured.

As we will show in the following section, choosing the samples equidistantly on a uniform grid enables efficient computation via the short-time Fourier transform. For a uniform sampling of the PSF, different choices of possible interpolation schemes and their effect on the approximation are shown in Figure 2.4.

**Discussion** Non-stationary convolution naturally generalises the space-invariant convolution model to spatially varying PSFs. It models a non-stationary PSF at any location in the object plane by interpolating between nearby PSF samples. It is intuitively clear, that with an increased spatial variation, a higher number of support sites or PSF measurements are necessary to faithfully describe the true underlying PSF. This is verified by the experiment shown Figure 2.3, where an increased number of support sites indeed reduces the difference to the continous model. Furthermore, it reveals least artifacts when compared to the piecewise-constant model and non-stationary combination. We want to note, that in the case of a constant PSF, where $R = 1$ and $w^0(\xi) = 1$ for all points within the object plane $\Sigma_1$, it reduces to the invariant convolution model.

Despite the formal similarity of the superposition integrals of *non-stationary combination* (2.22) and *non-stationary convolution* (2.29), both approaches compute spatially varying blur rather differently: non-stationary combination first computes the effect of constant piece-wise PSFs and interpolates the resulting image patches to yield the final image. In contrast, non-stationary convolution interpolates between neighbouring PSF samples to yield a different PSF for each point of the object plane. Only then, it adds up the resulting convolved image patches.

While both models are able to express spatially varying blur and have been applied successfully to image restoration, Denis et al. (2011) argue in their recent work in favour of non-stationary convolution as it seems better suited for smoothly spatially varying PSF and exhibits a number of physically favourable properties.

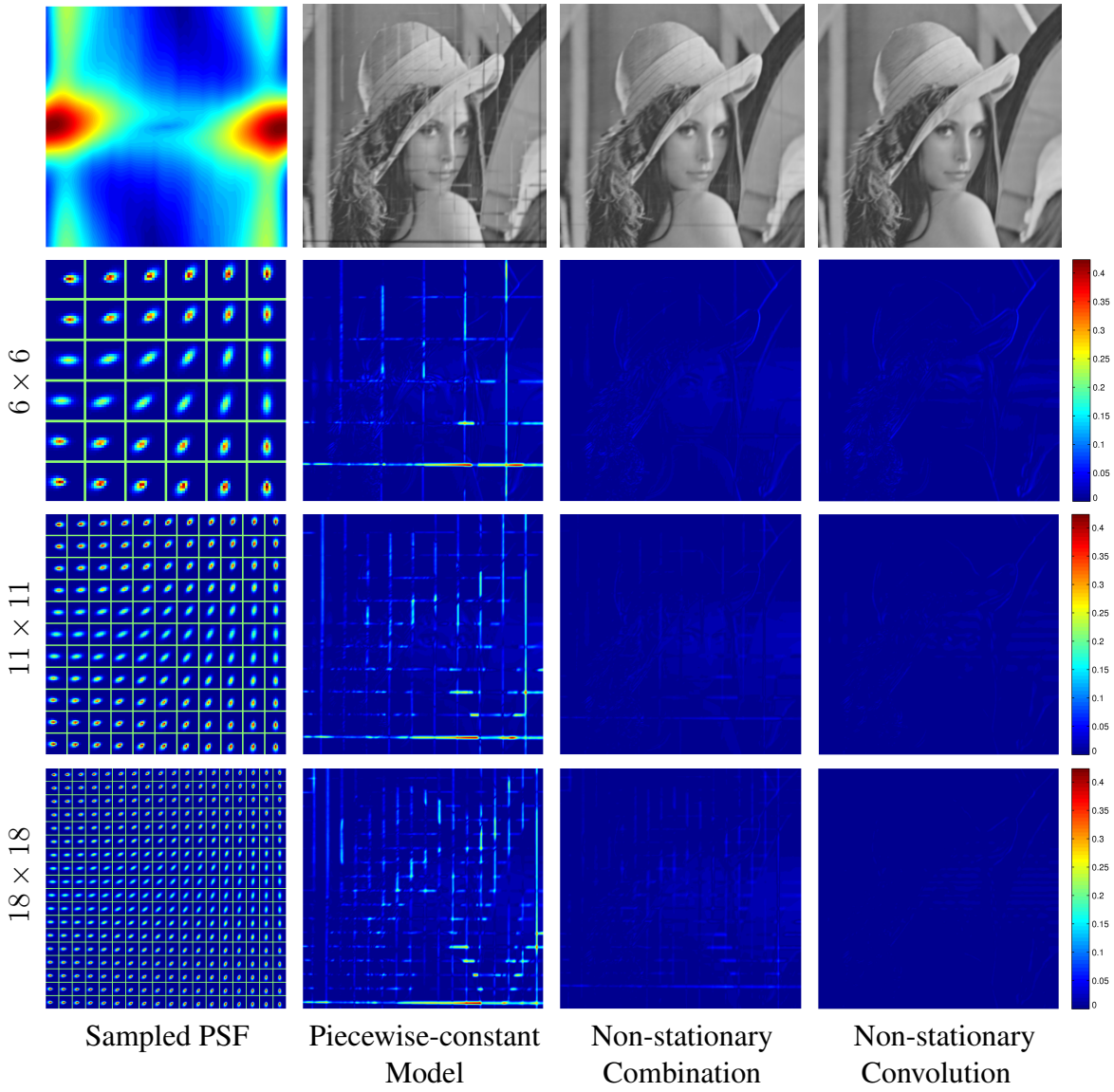



Figure 2.3: Comparison of various imaging models for spatially varying blur. The true PSF consists of as many samples as pixels in the sharp image, *i.e.* $512 \times 512$, and varies continously across the image plane. Each PSF sample is $15 \times 15$ pixels in size. The amount of spatial variation can be assessed by looking at the difference (in a L2 sense) between neighbouring PSF samples which is shown in the upleft corner. Besides, the top row shows the blurred image for the piecewise-constant model, non-stationary combination and non-stationary convolution. Row 2 to 4 show from left to right: the PSF samples used for computing the blurred images and the difference images between the result of the various approximations and the continuous model. The scale for the difference image has been kept constant for all cases to ease visual assessment. The scale refers to the relative error in each pixel. Please note, that this figure is best viewed on screen rather than in print.

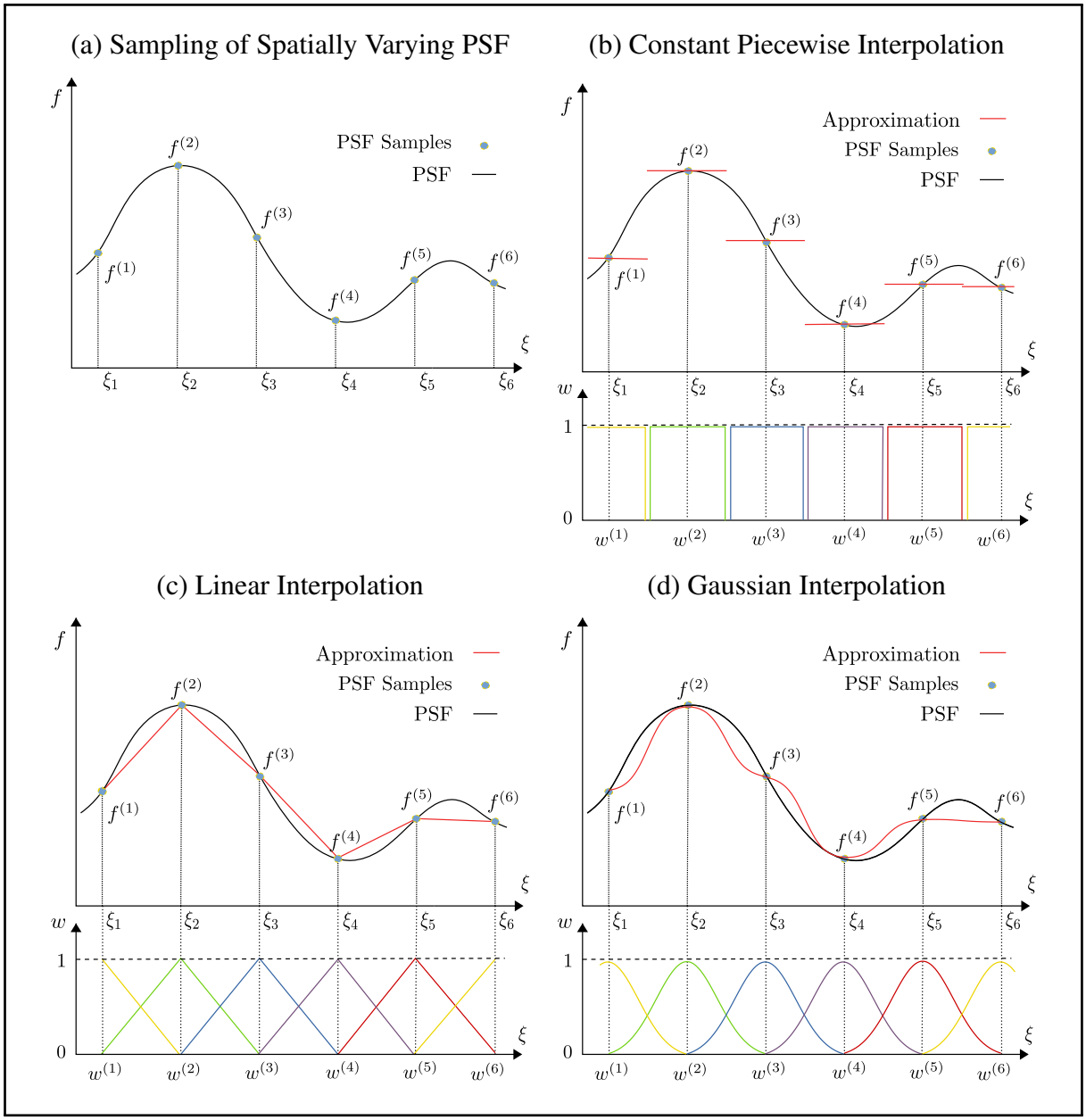



Figure 2.4: Toy example: Sampling of one-dimensional PSF and various interpolation schemes for its approximation.

## 2.3 Numerical Approximations

In this section, we show how to efficiently implement the non-stationary convolution model (2.29) by making use of the short-time Fourier transform. Ultimately motivated by facilitating space-variant BD, we derive expressions for all matrix vector multiplications (MVMs) that are needed for both image and PSF estimation. We call our approach *Efficient Filter Flow* (EFF) framework as it allows rapid MVMs while simultaneously being expressive enough to provide space-variant filtering.

For simplicity we introduce our framework for vector-valued images. The generalisation to matrix-valued images is straightforward. We quickly review space-invariant systems; this sets the notation and lays the foundation of our Efficient Filter Flow (EFF) framework.

### 2.3.1 Convolution Model

In practice, capturing a picture $y$ with a digital image sensor yields a finite set of intensity values. In particular, a digital image can be represented by a matrix whose dimensions correspond to the resolution of the image. As in the previous section and without loss of generality, we can consider $y$, $x$ and $f$ to be vectors of length $m$, $n$ and $k$ with entries $y_i$, $x_i$, and $f_{i,j}$, respectively. In the following, we consider only the valid part of the full convolution so that $m = n - k + 1$.

Replacing the integral by a summation, Equation (2.16) becomes

$$y_i = \sum_{j=0}^{k-1} f_{i-j}\ x_j \quad \text{for } 0 \leq i < m. \tag{2.30}$$

In digital signal processing $f$ is often called *filter*, a term which from now on we will use interchangeably with PSF. Since the transformation (2.30) is linear in $x$, it can be written as $y = Fx$ for $F_{i,i+j} = f_j$ for $0 \leq i < m$ and $0 \leq j < k$. In other words $F$ contains in each row a shifted copy of $f$. For such a structured $F$, MVMs can be performed in $O(n \log n)$ multiplications using FFTs with appropriate zero-padding (Press et al., 2007). If the signal is much longer than the PSF, *i.e.*, $n \gg k$, then the MVMs can be processed even faster by chunking the signal into patches and using the overlap-add (OLA) method of Stockham (1966). If $q$ is the size of the FFT for the patches, then OLA costs $O(n \log q)$. We explain OLA in greater detail below as it forms the basis of our framework for efficient space-variant linear filters.

### 2.3.2 Efficient Filter Flow (EFF)

Discretising Equation (2.29) is straightforward and yields

$$y_i = \sum_{r=0}^{R-1} \sum_{j=0}^{k-1} f_{i-j}^{(r)}\ w_j^{(r)}\, x_j \qquad \text{for } 0 \leq i < m, \tag{2.31}$$

$$\mathbf{y} = \mathbf{Z}_\mathbf{y}^\mathsf{T} \sum_{\mathbf{r}=\mathbf{0}}^{\mathbf{R}-\mathbf{1}} \mathbf{E}_\mathbf{r}^\mathsf{T} \mathcal{F}^\mathsf{H} \operatorname{diag}(\mathcal{F}\mathbf{Z}_\mathbf{f}\mathbf{f}^{(\mathbf{r})}) \mathcal{F} \operatorname{diag}(\mathbf{w}^{(\mathbf{r})}) \mathbf{C}_\mathbf{r}\, \mathbf{x}$$

Figure 2.5: Illustration of the matrix operations involved in the computation of the Efficient Filter Flow framework.

where we now have an additional sum over the set of localised uniform filters $\{f^{(r)}|r = 0 \dots R-1\}$. Their corresponding weighting functions $\{w^{(r)}|r = 0 \dots R-1\}$ must add up to one for each pixel in the object plane, *i.e.*

$$\sum_{r=0}^{R-1} w_i^{(r)} = 1 \quad \text{for } 0 \leq i < n. \tag{2.32}$$

Otherwise, artifacts might show up in the output image $y$ at the overlapping areas of the patches. Note that in practice, this property can always be enforced by normalising the window functions.

The key idea behind the efficient computation of (2.31) is to employ and modify the overlap-add (OLA) method of Stockham (1966), which was originally proposed for fast computation of convolution and correlations. The idea of the OLA method is to chop a image into overlapping patches, damp the borders of each patch with some windowing function, convolve each patch with the *same* filter, and then add the transformed patches to obtain the output image. However, if each patch is processed with its own corresponding filter we perform a non-stationary convolution.

In practice, it is most beneficial if the image patches are chosen to be uniform and a multiple power of two, as the FFT is fastest in such cases. For fixed patch size and fixed number of patches, the amount of overlap and the locations of the patches can be easily calculated. The parameters of the EFF are $R$ local uniform filters $f^{(0)}, \dots, f^{(R-1)}$, which equidistantly sample the PSF on a regular grid. These filters can either be measured (see *e.g.* Section 6.2, where we measured the space-variant PSF caused by lens aberrations) or estimated by minimising an appropriate cost function (see *e.g.* Chapter 4, where we blindly estimate a space-variant PSF caused by camera shake). For the estimation of the EFF parameters, it is advisable to resort to gradient-based optimisation techniques due to their improved rate of convergence. To this end, we need not only to know how to efficiently compute Equation (2.31), but also its transpose operation, which is the subject of the following paragraph.

### 2.3.3 Efficient MVMs for EFF

Now that we have defined the space-variant version of OLA, we come to the most important part: efficient implementation of the corresponding MVMs.

Since $x$ appears only linearly in Equation (2.31), we can write it as $y = Fx$. The filter matrix $F$ is given by

$$F = \sum_{r=0}^{R-1} F^{(r)} \operatorname{diag}\big(w^{(r)}\big), \tag{2.33}$$

where $F^{(r)}$ is the matrix corresponding to the convolution of $f^{(r)}$ ($0 \leq r \leq R-1$), and $\operatorname{diag}(v)$ is a diagonal matrix that has vector $v$ along its diagonal. However, this representation does not describe how MVMs with $F$, nor with $F^{\mathsf{T}}$, can be computed efficiently. For that we equivalently express $F$ as the following sum of a product of matrices

$$y = \underbrace{Z_y^{\mathsf{T}} \sum_{r=0}^{R-1} E_r^{\mathsf{T}} \mathcal{F}^{\mathsf{H}} \operatorname{diag}(\mathcal{F} Z_f f^{(r)}) \mathcal{F} \operatorname{diag}(w^{(r)}) C_r}_{F} x. \tag{2.34}$$

Equation (2.34) looks complicated, but is simple to understand:

(i) $C_r$ is a matrix that chops the $r$-th patch from a vector of length $n$

(ii) $Z_f$ is a zero-padding matrix that appends zeros to $f^{(r)}$ such that its size matches the patch size

(iii) $\mathcal{F}$ is the Discrete Fourier Transform (DFT) matrix (implemented by FFT)

(iv) $\mathcal{F}^{\mathsf{H}}$ is the Hermitian of the the DFT matrix (implemented by inverse FFT)

(v) $E_r^{\mathsf{T}}$ is a matrix that places the $r$-th patch back in a vector and

(vi) $Z_y$ is the zero-padding matrix that prepends zeros to a vector such that its size matches the size of the vector resulting from the summation.

Figure 2.5 illustrates the the most important steps of (2.34).

The proof that $F$ in Equation (2.33) is the same as in Equation (2.34) follows directly from the FFT implementation of convolution. Reading (2.34) from right to left this expression succinctly describes the steps needed to efficiently compute $Fx$. We can also read off $F^{\mathsf{T}}$ as

$$\begin{aligned} F^{\mathsf{T}} &= \sum_{r=0}^{R-1} C_r^{\mathsf{T}} \operatorname{diag}(w^{(r)}) \mathcal{F}^{\mathsf{T}} \operatorname{diag}(\mathcal{F} Z_f f^{(r)}) \overline{\mathcal{F}} E_r Z_y \\ &= \sum_{r=0}^{R-1} C_r^{\mathsf{T}} \operatorname{diag}(w^{(r)}) \mathcal{F}^{\mathsf{H}} \overline{\operatorname{diag}(\mathcal{F} Z_f f^{(r)})} \mathcal{F} E_r Z_y, \end{aligned} \tag{2.35}$$

where $\overline{F}$ is the component-wise complex conjugate of $F$. Equation (2.35) follows from (**??**) because $F$ is a real valued matrix, whereby $F = \overline{F}$. Reading (2.35) from right to left describes the steps needed to efficiently calculate MVM for $F^\mathsf{T}$. In words, we perform steps similar to EFF on $y$ but with windowing at the end instead of the beginning, and with complex conjugation of the FFT of the PSFs, resulting in calculating patch-wise correlations instead of convolutions.

For non-blind deconvolution with spatially varying PSFs, efficient MVMs with $F$ and $F^\mathsf{T}$ suffice. But for blind deconvolution we need more. Since Equation (2.31) is also linear in the PSF parameters, we next define a matrix $X$ such that $y = Fx = Xf$, where $f$ denotes the stacked sequence of PSF samples $f^{(0)}, \ldots, f^{(R-1)}$. Now we rewrite (2.34) using $\operatorname{diag}(v)w = \operatorname{diag}(w)v$ and some matrix $B_r$ that chops the $r$-th PSF from the vector $f$,

$$y = \underbrace{Z_y^\mathsf{T} \sum_{r=0}^{R-1} E_r^\mathsf{T} \mathcal{F}^\mathsf{H} \operatorname{diag}\big(\mathcal{F} \operatorname{diag}(w^{(r)}) C_r x\big) \mathcal{F} Z_f B_r}_{X} f. \tag{2.36}$$

This expression is not needed for implementing $Xf$ since we already know how to compute $Fx = Xf$ quickly. However, it allows us to derive an algorithm for efficient MVM with $X^\mathsf{T}$ simply by taking the transpose of the expression for $X$,

$$\begin{aligned} X^\mathsf{T} &= \sum_{r=0}^{R-1} B_r^\mathsf{T} Z_f^\mathsf{T} \mathcal{F}^\mathsf{T} \operatorname{diag}\big(\mathcal{F} \operatorname{diag}(w^{(r)}) C_r x\big) \overline{\mathcal{F}} E_r Z_y, \\ &= \sum_{r=0}^{R-1} B_r^\mathsf{T} Z_f^\mathsf{T} \mathcal{F}^\mathsf{H} \overline{\operatorname{diag}\big(\mathcal{F} \operatorname{diag}(w^{(r)}) C_r x\big)} \mathcal{F} C_r Z_y, \end{aligned} \tag{2.37}$$

where we again used $X = \overline{X}$, as $X$ is real. In words, the algorithm implied by (2.37) for $X^\mathsf{T} y$ consists of splitting $y$ into patches, correlating them with the patches from $x$, and finally summing up the results.

### 2.3.4 Computational Complexity

The computational complexity of EFF-based MVMs for $F$, $F^\mathsf{T}$, $X$, and $X^\mathsf{T}$ is the same as the OLA method (Stockham, 1966) for space-invariant filtering which is about $O(n \log q)$, where $q$ is the size of the FFT applied to the image patches. The overlap increases the computational cost by a constant factor and is thus omitted. Hence, the EFF framework implements space-variant convolutions which are as efficient to compute as space-invariant convolutions, while being much more expressive. Being parametrised by the local PSF samples $f^{(r)}$ that are much smaller than the full image $x$ of the observed scene, the memory requirement for storing space-variant PSFs is much less than $O(mn)$ needed by a general linear transformation.

### 2.3.5 Expressivity

As discussed in Section 2.2.2.3, the non-stationary convolution model yields the uniform imaging model in the limiting case of a single PSF sample. This is also obvious from

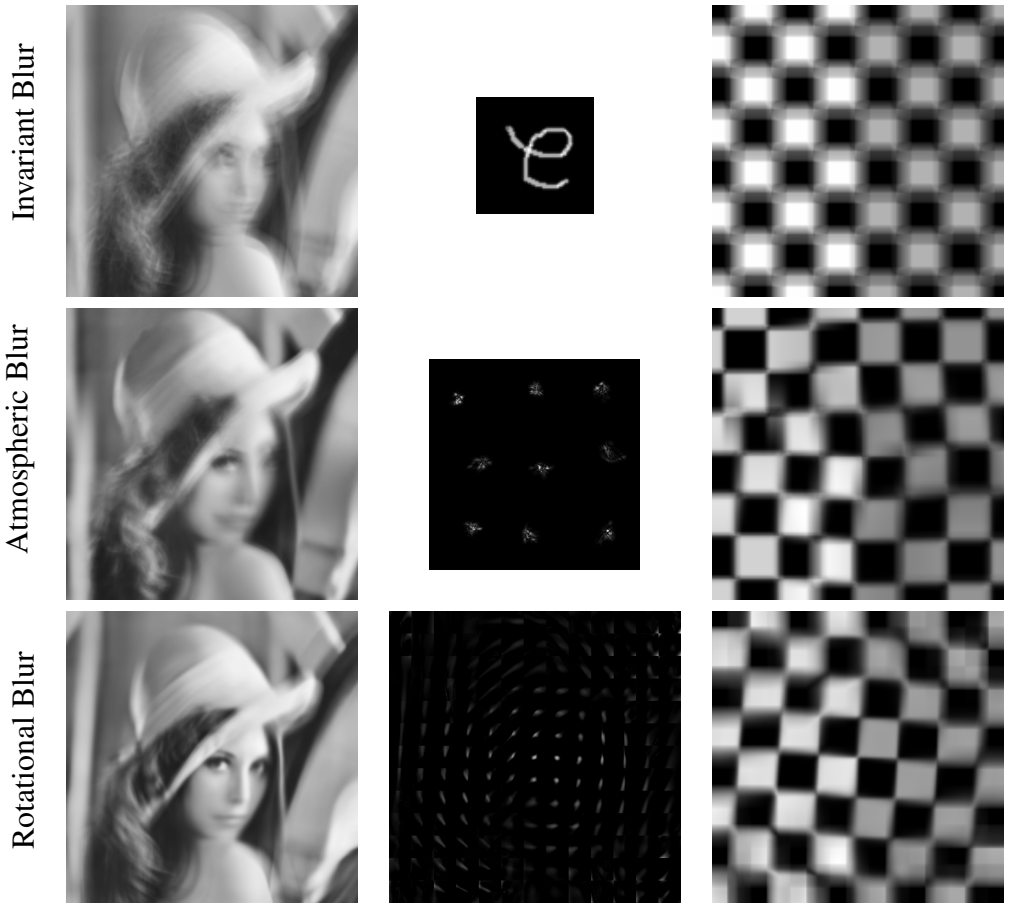



Figure 2.6: Expressivity of proposed EFF framework: (top to bottom) global motion blur, expressible by a single filter, atmospheric blur with $3 \times 3$ simulated speckle patterns, rotational transformation approximated by spatially-varying PSF with $13 \times 13$ filters each $41 \times 41$ pixels in size.

expression (2.31): when all PSFs $f^{(r)} = f$, using (2.32), we see that (2.31) reduces to (2.30).

At the other extreme, the EFF framework can implement *any* linear transformation $F$. To do so, we need $m$ image patches, one for each row of $F$. Then, we set all window functions to constant $1/m$, and the PSFs characterising the EFF to the rows of $F$. This case is degenerate as patches overlap completely and PSFs, as long as the signal, are only evaluated once. But it shows that EFF filtering actually covers the entire range from space-invariant filtering to arbitrary linear transformations, trading computational efficiency for being more expressive. Figure 2.6 shows further examples what kind of image transformations our framework is able to express and offers a link to an animated demonstration.

### 2.3.6 Related Work

An idea key to our framework was introduced by Stockham (1966), who presented the OLA method for fast convolution and correlation. For 1D signals, such as audio, Allen (1977) used OLA, aka short-time Fourier analysis and synthesis for time varying filtering. However, he considered $Fx$ only and did not show how to calculate $F^\mathsf{T}x$, which is required for deconvolution. Neither does he show how to calculate $X^\mathsf{T}y$, which is required for blind-deconvolution. For two dimensional signals, Hinman et al. (1984) generalised short-time Fourier analysis to short-space Fourier analysis, but did not consider synthesis needed for space-variant filtering. In his non-stationary combination approach to spatially varying filtering, Nagy and O'Leary (1997) does consider synthesis, but only for rectangular and triangular windows; he also considers $F$ and $F^\mathsf{T}$, thus the case of non-blind deconvolution only. Margrave (1998) discusses both non-stationary combination and convolution, however doesn't derive expressions for efficient numerical implementation.

## 2.4 Conclusion

In this chapter, we revised the basics of optical imaging under incoherent illumination, therewith providing the theoretical background for the material covered in this thesis. By approaching optical imaging via Linear Systems Theory (LST), we first revised the concept of a PSF as the impulse response of a linear optical system. Subsequently, we derived various approximations to the incoherent imaging equation and discussed their underlying assumptions:

1. Translational invariance of a optical system yields the well-known and widely used convolution model. Unfortunately, its computational amenities often come at the price of inaccurate modelling, since many real-world applications exhibit a significant spatial variation of the underlying PSFs.

2. Assuming smooth spatial variation of a PSF enabled us to derive our EFF framework, a novel approximation of the incoherent imaging equation, which allows increased expressiveness and efficient computation at the same time. It features linearity in both PSF parameters and image intensities and includes the convolution model as a special case. Our framework offers an efficient compromise between the extremes of space-invariance and full-dense linear transformations. We extended the overlap-add idea to the space-variant setting, thereby allowing us to compute matrix-vector-multiplications involving local PSF samples rapidly.

In the following two chapters we address the problems of removing spatially varying blur due to atmospheric turbulence (Chapter 3) and camera shake (Chapter 4), both of which will serve as a challenging testing ground for our model.

## 2.5 Acknowledgements

The work presented in this chapter was carried out in collaboration with Stefan Harmeling, Suvrit Sra and Bernhard Schölkopf of the Max Planck Institute for Intelligent Systems. Except otherwise explicitly stated, all mathematical derivations, algorithmic implementations and experimental evaluations were performed by the author of this thesis.

This chapter is based on the following publications:

- *Efficient Filter Flow for Space-Variant Multiframe Blind Deconvolution*, Michael Hirsch, Stefan Harmeling, Suvrit Sra and Bernhard Schölkopf. In 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010.

# Chapter 3

# Online Multi-Frame Blind Deconvolution Beyond The Convolution Model

Astronomical images taken by ground-based telescopes suffer degradation due to atmospheric turbulence. This degradation can be tackled by costly hardware-based approaches such as adaptive optics, or by sophisticated software-based methods such as lucky imaging, speckle imaging, or multi-frame deconvolution. Software-based methods process a sequence of images to reconstruct a deblurred high-quality image. However, existing approaches are limited in one or several aspects: (i) they process all images in batch mode, which for thousands of images is prohibitive; (ii) they do not reconstruct a super-resolved image, even though an image sequence often contains enough information; (iii) they are unable to deal with saturated pixels; (iv) they assume an invariant blur model and thus are restricted to the isoplanatic patch; and (v) they are usually non-blind, *i.e.*, they assume the blur kernels to be known. In this chapter, we present a new method for multi-frame deconvolution called Online Blind Deconvolution (OBD) that overcomes all these limitations simultaneously. Encouraging results on simulated and real astronomical images demonstrate that OBD yields deblurred images of comparable and often better quality than existing approaches.

The following chapter is organised as follows: in Section 3.1 we give a short introduction to the field of ground-based astronomical observation, followed by an overview of common approaches for overcoming the adverse effect of atmospheric turbulence. In Section 3.2 we discuss previous work that is directly related to the work presented in this chapter. After deriving our proposed OBD algorithm in Section 3.3, we extend our imaging model step by step to incorporate the effects of super-resolution, saturation and spatially varying blur. In Section 3.5 we carry out an extensive validation of our algorithm with simulated data where we have access to the ground truth and thus are able to quantitatively evaluate the performance of our approach. In Sections 3.7 and 3.8 we do a comprehensive comparison on both astronomical data but also non-astronomical imagery, followed by a final summary and conclusion in Section 3.9.

## 3.1 Introduction

Astronomical observation using ground-based telescopes is significantly degraded by diffraction-index fluctuations caused by atmospheric turbulence. This turbulence arises from local temperature and density inhomogeneities and results in a time- and space-variant point spread function (PSF). Often the PSF is assumed to be invariant within a short time-period and a small region of space, called an *isoplanatic patch*. The coherence time and the size of the isoplanatic patch depend on the strength of the turbulence that is usually quantified by Fried's parameter $r_o$ (Fried, 1978) ranging between 10–20 cm for visible wavelengths at astronomical telescope sites. The coherence time for atmospheric turbulence is effectively frozen for images with exposure times shorter than 5–15 ms. Longer exposures effectively perform a time average, and thereby irretrievably wipe out high frequency information, making them band-limited to angular frequencies smaller than $r_o/\lambda$, where $\lambda$ is the wavelength of the observed light. In contrast, short-exposures encapsulate information up to the diffraction-limited upper frequency bound (which is theoretically given by the ratio $D/\lambda$, where $D$ denotes the diameter of the telescope's primary mirror). Figure 3.1 depicts this issue for the simulated image of a single star and shows the radial averaged modular transfer function (MTF) for diffraction-limited, long- and short-exposure imaging.

The information carried by short exposures was first exploited by Labeyrie (1970), who proposed the averaging of the power spectra of a sequence of short exposures to retain diffraction-limited amplitude information. Shortly thereafter, Knox and Thompson (1974) extended Labeyrie's idea by suggesting a method for the recovery of the phase information, which is not preserved by Labeyrie's so-called stellar speckle interferometric method. These early works revolutionised ground-based astronomical observation with large telescopes and have since led to a number of improved signal-processing methods (Lohmann et al., 1983; Mikurda and Lühe, 2006; Stelzer and Ruder, 2007) widely referred to as speckle imaging techniques.

An alternative approach was proposed in the seminal work of Ayers and Dainty (1988), who presented a Blind Deconvolution (BD) algorithm for the problem of atmospherically degraded imaging. BD recovers object information from a blurry and noisy observation without any additional measurement of the distortion. The BD of a single observation is a severely ill-posed problem: there are an infinite number of solutions, and small perturbations of the data result in large deviations in the estimate of the object. The ill-posedness can be alleviated to some degree by confining the set of solutions to physically plausible ones by introducing additional constraints or prior knowledge. Another possibility is to use multiple images or to exploit the partial information about wavefront distortion obtained from wavefront-sensor data, as used in adaptive-optics based myopic deconvolution algorithms.

Since the work of Ayers and Dainty (1988), BD has grown to be a valuable tool in astronomical imaging and has been subject of numerous publications. Today a plethora of algorithms exist that primarily differ in: (i) the data used; (ii) the a-priori knowledge incorporated while deblurring; and (iii) the algorithmic approaches for estimating the object and its blur. For a good overview of BD in the domain of astronomical imaging we refer the reader to (Kundur and Hatzinakos, 1996; Molina et al., 2006; Pantin et al., 2007).

Recently, electron-multiplying CCD cameras have enabled capturing short-time expo-

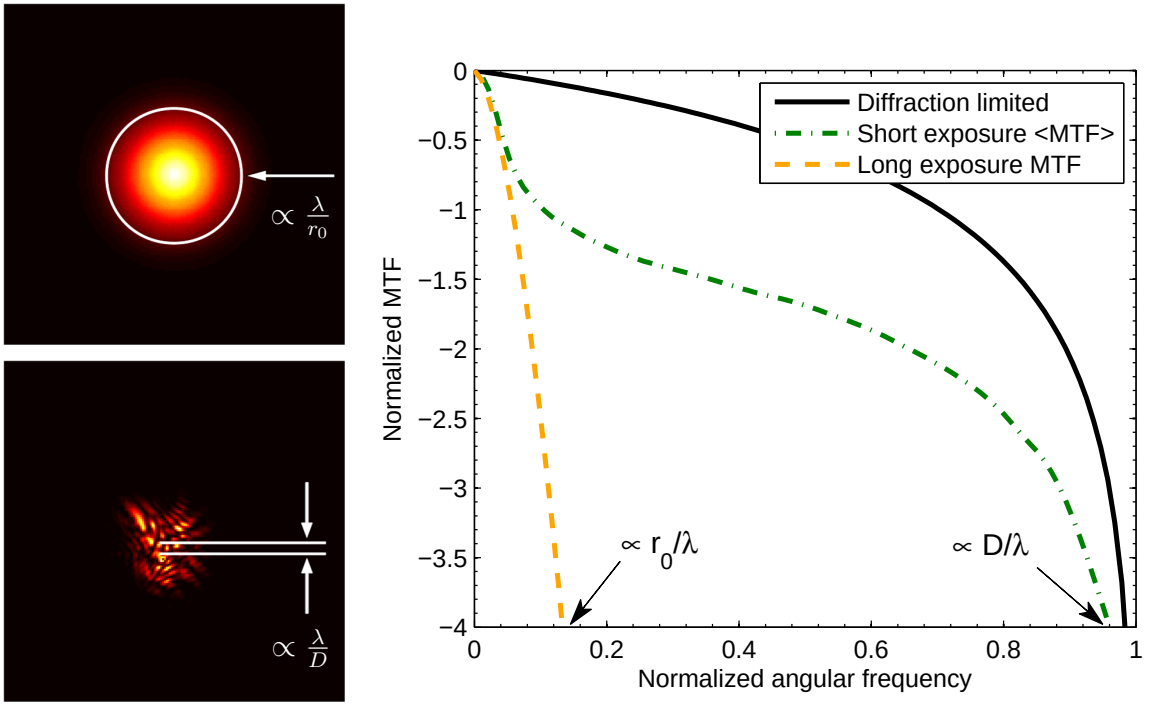



Figure 3.1: Imaging a single star through atmospheric turbulence: while the angular resolution of a long exposure image (top left) is limited by the strength of the atmospheric turbulence (commonly quantified by Fried's parameter $r_o$), a short exposure image (bottom left) encapsulates information up to the diffraction-limited upper frequency bound which is proportional to the diameter $D$ of the telescope's primary mirror. The right panel shows the radial averaged modular transfer function (MTF) for diffraction-limited, short and long exposure imaging. While the long exposure MTF falls to nearly zero at $r_0/\lambda$, the average short exposure MTF reaches intermediate levels up to the diffraction limited upper frequency bound. The short exposure MTF was averaged over 5000 trials. The simulation was performed for $D/r_o = 13.3$.

sures with negligible noise (Mackay et al., 2001). This in turn has led to a new method: *lucky imaging*, which can to some degree overcome atmospherically-induced resolution limitations of ground-based telescopes (Law et al., 2006; Oscoz et al., 2008; Hormuth et al., 2008). The lucky imaging idea is based on the work of Fried (1978) (who computed the probability of getting a *lucky frame*, *i.e.*, an image recorded at a time instant of exceptionally good seeing). This idea proposes to collect only the 'best' frames available in a recorded sequence. These 'best' frames are subsequently combined to obtain a final image of the object. Usually, out of a thousand images, only a few are selected for the final reconstruction and most of the observed frames are discarded.

This "wastage" can be avoided, and one can indeed use all the frames to obtain an improved reconstruction as we will see in Section 3.5.

Methods for Multi-frame Blind Deconvolution (MFBD) aim to recover the image of a fixed underlying object given a sequence of noisy, blurry observations. Each observation has a different and unknown blur, which makes the deconvolution task hard.

Previous approaches to MFBD process all observed frames simultaneously. Doing so limits the total number of frames that can be processed. We show how the computational burden can be greatly reduced by presenting OBD, our online algorithm that processes the input sequence one frame at a time. Each new frame helps to gradually improve the image reconstruction. This simplistic approach is not only natural, but also has several advantages over non-online methods, *e.g.*, lower resource requirements, highly competitive image restoration (Harmeling et al., 2009, 2010b; Hirsch et al., 2010; Hirsch et al., 2011), low to moderate dependence on regularisation or *a priori* information, and easy extension to super-resolution[1], saturation correction and spatially varying blur. In particular, the contributions presented in this chapter are as follows:

(a) we derive our MFBD algorithm in the framework of *stochastic gradient-descent*;

(b) we show how to incorporate *super-resolution* while simultaneously performing blind deconvolution;

(c) we tackle *saturation*, a nuisance familiar to anyone who works with astronomical images;

(d) we extend our imaging model to account for spatially varying blur due to atmospheric turbulence; and

(e) we present results on both simulated and real-world astronomical imagery taken in a simple astronomer setup, where one does not have access to sophisticated equipment (*e.g.*, adaptive optics), and computational resources might be limited.

Before describing further details, let us put our work into perspective by briefly surveying related work.

## 3.2 Related Work

### 3.2.1 Multi-Frame Blind Deconvolution

A multitude of multi-frame (or multiple-image) deblurring papers discuss the *non-blind* deconvolution setup, where, in addition to the image sequence the sequence of blur kernels must be known as well. We do not summarise such methods here because ours is a *blind* deconvolution method. Amongst multiple frame blind approaches, the method of Schulz (1993) is perhaps the earliest. Schulz used penalized likelihood maximisation based on a generalised expectation maximisation (GEM) framework. Closely related is Li et al. (2004), who also used a GEM framework, but focused on choosing a good objective function and regulariser for optimisation. In contrast to our work, both Schulz (1993) and Li et al. (2004) presented *batch* algorithms that are computationally prohibitive, which greatly limits the number of frames they can simultaneously process.

[1] Here, super-resolution refers to techniques that are able to enhance the resolution of an imaging system by exploiting the additional information introduced by sub-pixel shifts between multiple low resolution images of the same scene or object.

Sheppard et al. (1998) discussed the MFBD problem and presented a procedure that also processes all frames at the same time. They did, however, mention the possibility of incremental processing of frames, but gave an example only for the *non-blind* setup. Their blind-deconvolution algorithm was based on conjugate-gradients, for which they had to parametrise (*e.g.*, $x \to z^2$) the variables to enforce non-negativity. This reparametrisation has a long history in image deconvolution (Biraud, 1969), but numerically, the ensuing non-linearity can be damaging as it destroys the convexity of sub-problems.

More recently, Matson et al. (2008) also used the same nonlinear ($x \to z^2$) reparametrisation for solving MFBD with a parallel implementation of conjugate-gradients. Another approach is that of Zhang et al. (2009), who incorporated a low-pass filter into the MFBD process for suppressing noise, but again at the expense of convexity.

Further MFBD work includes: Anconelli et al. (2006) who considered methods for the reduction of boundary effects; Zhulina (2006) who discussed the Ayers-Dainty algorithm; and Löfdahl (2002) who permitted additional linear inequality constraints. We refer the reader to Matson et al. (2008) for even more references—including those to early works—and a nice summary of blind deconvolution for astronomy. Unlike our algorithm, all the above mentioned blind deconvolution methods are batch procedures; moreover *none* of them performs either super-resolution, saturation correction or spatially varying blur.

### 3.2.2 Super-Resolution

Numerous papers address the standard super-resolution problem. For good surveys we refer the reader to (Park et al., 2003; Farsiu et al., 2004). However, most of these works are based on the assumption that the blur is known, and only a few deal with the harder case of blind super-resolution.

The work most closely related to ours is Šroubek et al. (2007), who propose a unifying framework that simultaneously performs blind deconvolution and super-resolution. In Šroubek et al. (2007, 2008) the authors show how a high-resolution image can be obtained from multiple blurry and noise corrupted low-resolution frames. However, their model assumes *a priori* knowledge about both the image and the blur, and Šroubek et al. (2008) themselves note that their method suffers from numerical instabilities for super-resolution factors larger than 2.5. In contrast, our approach exploits the abundance of available data, which for moderate noise levels does not require imposing any image or blur prior (except non-negativity), leading to an overall simpler algorithm. Moreover, our method is computationally more efficient, since it is online.

### 3.2.3 Spatially Varying Blur

Early relevant work includes (Lohmann and Paris, 1965), where the authors clearly formalise space-variant imaging systems, and discuss basic special cases such as: piecewise space-invariant systems and geometric distortion. Another early work which presents an idea that is key to our framework was introduced by Stockham (1966), who presented the overlap-add (OLA) method for fast convolution and correlation. For one dimensonal signals, such as audio, Allen (1977) used OLA, aka short-time Fourier analysis and synthesis for

time-variant filtering. For two dimensional signals, Hinman et al. (1984) generalised short-time Fourier analysis to short-space Fourier analysis, but did not consider synthesis needed for space-variant filtering. Nagy and O'Leary (1998) does consider synthesis, but only the case needed for latent image estimation, thereby limited to the non-blind case. In (Bardsley et al., 2006) the approach of Nagy and O'Leary (1998) is extended to the blind setting, where the authors employ phase-diversity for PSF estimation to perform space-variant blind-deconvolution. Recently, Zhu and Milanfar (2011) presented a MFBD algorithm, that first corrects for the geometric distortions induced by atmospheric turbulence and subsequently performs a non-blind deconvolution with an invariant PSF for latent image reconstruction.

## 3.3 The OBD Algorithm

### 3.3.1 Problem Formulation

For simplicity of exposition, our description will focus on one-dimensional images and point spread functions (PSFs). In Appendix A we cover the generalisation to two-dimensions. Initially, for the derivation of the MFBD algorithm we will assume an invariant blur model. In Section 3.4.3 we will extend our approach to non-uniform blur.

Let each observed (blurry and noisy) frame be denoted by $y_t$, the 'true' unknown image by $x$, and each *unknown* PSF by $f_t$. Then, we use the observation model

$$y_t = f_t * x + n_t, \quad t = 1, 2, \ldots, T, \tag{3.1}$$

where $f_t * x$ represents convolution (circular or non-circular), and $n_t$ denotes measurement noise. Further, on physical grounds we assume both the image $x$ and the PSF $f_t$ to be non-negative.

### 3.3.2 Algorithm

First consider the case where given the next observation $y_t$ and the current image estimate $x_t$, we wish to compute the PSF $f_t$. Assuming the noise $n_t$ in equation (3.1) to be Gaussian distributed with zero mean and incorporating non-negativity, the PSF $f_t$ can be determined by solving a non-negative least-squares (NNLS) problem[2]. For a given observation frame $y_t$ and a current estimate $x_t$, we define the *loss*

$$\ell(y_t; x) = \min_{f_t \geq 0} \left\| y_t - f_t * x \right\|^2. \tag{3.2}$$

For a frame sequence $y_1, y_2, \ldots, y_T$, we aim to minimise the overall loss by computing the image $x$ that solves

$$\min_{x \geq 0} \; L_T(x) = \frac{1}{T} \sum_{t=1}^{T} \ell(y_t; x). \tag{3.3}$$

[2]This NNLS problem may be solved by various methods; we used the LBFGS-B algorithm (Byrd et al., 1995).

Problem (3.3) is not easy, because it is non-convex and its optimal solution requires computing both $x$ as well as the PSFs $f_1, \ldots, f_T$. Nevertheless, given our formulation, several methods could potentially be used for minimising $L_T(x)$. For example, an ordinary gradient-projection scheme would be

$$x_{t+1} = P_+\big(x_t - \alpha_t \nabla L_T(x_t)\big), \quad t = 0, 1, \ldots, \tag{3.4}$$

where $P_+$ denotes projection onto the non-negative orthant; $x_t$ denotes the current image estimate; and $\alpha_t$ is an appropriate step-size. However, when the number of frames $T$ is large, such an approach rapidly becomes computationally impractical. Hence we turn to a simpler method that processes the input one frame at a time.

### 3.3.3 Stochastic Gradient Descent

A simple and often effective method for minimising the overall loss in equation (3.3) is *stochastic gradient descent* (SGD). This method does not process all the frames simultaneously, but at step $t$ it picks (at random) some frame $y$ and updates the current image estimate $x_t$ as

$$x_{t+1} = P_+\big(x_t - \alpha_t \nabla \ell(y; x_t)\big), \tag{3.5}$$

where $P_+$ and $\alpha_t$ are as before; computing $\nabla \ell(y; x_t)$ requires solving equation (3.19). By processing only one frame at a time, stochastic gradient descent (SGD) leads to huge computational savings. However, there are two main difficulties: update rule (3.5) converges slowly; and more importantly, it is sensitive to the choice of the step-size $\alpha_t$; a popular choice is $\alpha_t = \beta/(t_0 + t)$, where the constants $t_0$ and $\beta$ must be tuned empirically.

We propose a practical modification to the step-size computation, wherein we instead use the scaled-gradient version

$$x_{t+1} = P_+\big(x_t - \alpha_t S_t \nabla \ell(y; x_t)\big), \tag{3.6}$$

where $S_t$ is a positive-definite matrix. Also update rule (3.6) can be shown to converge[3] under appropriate restrictions on $\alpha_t$ and $S_t$ (Kushner and Yin, 2003; Bottou, 1998). In general, the matrix $S_t$ is chosen to approximate the inverse of the Hessian of $L_T(x^*)$ for an optimal $x^*$, thereby yielding quasi-Newton versions of SGD. But a more straightforward choice is given by the diagonal matrix

$$S_t = \operatorname{diag}\big((x_t + \varepsilon)/(F_t^\mathsf{T} F_t x_t + \varepsilon)\big), \tag{3.7}$$

where the $\operatorname{diag}$ operator maps a vector $x$ to a diagonal matrix with elements of $x$ along its diagonal. Also note that the division in (3.7) is element-wise, $F_t$ is the matrix representation of the PSF $f_t$ (see Appendix A), and $\varepsilon > 0$ is a positive constant which ensures that $S_t$ remains positive definite and bounded (both requirements are crucial for convergence of the method). The choice (3.7) can be motivated with the help of auxiliary functions (*e.g.*, as in Harmeling et al., 2009),

[3] One can show *almost sure (a.s.) convergence* of the objective, and a.s. convergence of the gradient to the gradient at a stationary point.

**Remark:** We note in passing that if one were to use $\alpha_t = 1$, and set $\varepsilon = 0$, then although convergence is no longer guaranteed, iteration (3.6) takes a particularly simple form, namely,

$$x_{t+1} = x_t \odot (F_t^{\mathsf{T}} y)/(F_t^{\mathsf{T}} F_t x_t), \tag{3.8}$$

where $\odot$ denotes the Hadamard (element-wise) product of two vectors—this update may be viewed as an online version of the familiar ISRA (see Daube-Witherspoon and Muehllehner, 1986).

Note that for (3.7) the matrix $F$ corresponds to the PSF $f$ computed via the NNLS problem (3.19) with $y$ and $x = x_t$. We call the method based on iteration (3.6) online blind deconvolution (OBD) and provide pseudo-code as Algorithm 1. We further note that by assuming photon shot noise (Poisson-distributed) in equation (3.1) instead of additive noise, we can also design a Richardson-Lucy type iteration for solving equation (3.3).

**Algorithm 1**: Online Blind Deconvolution (OBD)

**Input**: Stream of images $y_t$ for $t \geq 1$
**Output**: Reconstructed image $x$
initialise $x_1$ with $y_1$;
**while** *another image* $y_{t+1}$ *available* **do**
  $t \leftarrow t + 1$;
  estimate $f_t$ by

$$f_t = \operatorname{argmin}_{f \geq 0} \left\| y_t - f * x_{t-1} \right\|^2 \tag{3.9}$$

  update $x_t$ by

$$x_t = P_+\big(x_{t-1} - \alpha_{t-1} S_{t-1} \nabla \ell(y; x_{t-1})\big) \tag{3.10}$$

**end**
**return** *last estimate* $x_t$

## 3.4 Extending OBD

### 3.4.1 Super-Resolution

In the OBD setup an entire sequence of frames is at our disposal. Can we exploit this sequence to improve the image reconstruction beyond mere blind deconvolution? The answer is 'yes'. With a small increase in computational costs we can augment the basic algorithm and perform super-resolution. For long-exposures that often lose higher-frequency structure (finer details) of the image due to averaging, such increased resolution is particularly desirable.

To incorporate super-resolution into our framework we introduce the *resizing* matrix

$$D_n^m = (I_n \otimes 1_m^T)(I_m \otimes 1_n)/n, \tag{3.11}$$

where $I_n$ is the $n \times n$ identity matrix, $1_n$ is an $n$ dimensional column vector of ones, and $\otimes$ denotes the Kronecker product. The matrix $D_n^m$ transforms a vector $v$ of length $m$ into

a vector of length $n$. The sum of $v$'s entries $1_m^T v = 1_n^T D_n^m v$ is preserved (formally verified by applying identity (A.5) twice). This is a favourable property for images, as the number of photons observed should not depend on the resolution. Note that even if the sizes $n$ and $m$ are not multiples of each other, $D_n^m$ will interpolate appropriately. Hence, the super-resolution factor, *i.e.*, the ratio $m/n$, is not restricted to be integral. Note that for $m \geq n$, *i.e.* downscaling, the matrix operation corresponds to integrating neighbouring pixels weighted by their overlap with the new pixel grid. Similarly for $m < n$, *i.e.* upscaling, the operation will take the nearest neighbour, if $n$ is divisible by $m$, or a weighted linear combination of close-by pixels.

To avoid double indexing let $n = l_y$ be the length of $y$. For super-resolution by a factor of $s$ we choose $x$ and $f$ large enough such that the vector $f * x$ has length $sn$. Then we replace the loss $\ell(y_t; x)$ by (cf. equation (3.19))

$$\ell(y_t; x) = \min_{f_t \geq 0} \left\| y_t - D_n^{sn}(f_t * x) \right\|^2. \tag{3.12}$$

For this loss, a derivation similar to that for (3.7) yields the diagonal matrix

$$S_t = \operatorname{diag}\big((x_t + \varepsilon)/((D_n^{sn} F_t)^\mathsf{T} D_n^{sn} F_t x_t + \varepsilon)\big), \tag{3.13}$$

where $F_t$ corresponds to $f_t$ obtained by solving (3.12).

### 3.4.2 Overexposed Pixels

For astronomical images a common problem is saturation of pixels due to overexposure, *i.e.*, some pixels receive so many photons that they exceed the peak intensity permitted by the hardware. This saturation can be particularly confounding if both bright and faint stars are present in the same image, especially when some stars are orders of magnitude brighter. Overexposed pixels impede not only deblurring but also super-resolution and applications such as estimation of star magnitudes.

However, since we have an entire sequence of observed frames, tackling overexposed pictures is feasible. Here the atmospheric blurring proves to be helpful, since it creates non-overexposed margins around a bright star whose centre pixels are overexposed. Our method is able to fit these margins and can approximate the true star magnitude. Our approach essentially consists of identifying saturated pixels and excluding them from the computation of the objective function. This approach might seem to be overly simple, but its success is deeply tied to the availability of multiple frames. Specifically, since each frame can have different pixels attaining saturation (different frames are aligned differently), we have to check at each iteration which pixels in the current image are saturated. To ignore these pixels we define a diagonal weighting matrix (per frame) with entries,

$$\Sigma_t = \begin{cases} 1 & \text{if } y_t < \rho_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{3.14}$$

along its diagonal. Hereby, we assume the value of a saturated pixel to be $\rho_{\max}$ (*e.g.* in the case of 16 bit images, $\rho_{\max} = 65535$). We can modify the updates to ignore saturated pixels

by replacing the Euclidean norm with a weighted norm $\|v\|_\Sigma^2 = v^\mathsf{T}\Sigma v$. We replace the loss $\ell(y_t; x)$ by

$$\ell(y_t; x) = \min_{f_t \geq 0} \left\| y_t - f_t * x \right\|_{\Sigma_t}^2. \tag{3.15}$$

For this loss, following a derivation similar to (3.7) yields the diagonal matrix

$$S_t = \operatorname{diag}\big((x_t + \varepsilon)/(F_t^\mathsf{T}\Sigma_t F_t x_t + \varepsilon)\big), \tag{3.16}$$

where, as before, $F_t$ corresponds to $f_t$ obtained by solving (3.12).

**Remark.** One might ask whether we can recover pixels in $x$ that are saturated in most of the frames. The answer is yes, and can be understood as follows. The photons corresponding to such a pixel in $x$ are spread by the PSF across a whole set of pixels in each observed frame. Thus, if not all these pixels are always saturated, the true value for the corresponding pixel in $x$ can be recovered.

### 3.4.3 Spatially Varying Blur

As mentioned above, the uniform or space-invariant imaging model is only valid within an isoplanatic patch whose size depends on the seeing conditions at the time of recording. In Section 2 we developed the so-called Efficient Filter Flow (EFF) framework, which allows the description and efficient computation of a PSF that smoothly varies across the image plane. The idea is (i) to cover the image with *overlapping* patches, (ii) to assign and apply to each patch a different PSF, and (iii) to add the patches to obtain a single large image. The $i$th pixel value $y_i$ in the blurred image $y$ can be written as a linear combination of the $R$ differently blurred patches,

$$y_i = \sum_{r=0}^{R-1} \sum_{j=0}^{k-1} f_j^{(r)} w_{i-j}^{(r)} \, x_{i-j} \text{ for } 0 \leq i < n, \tag{3.17}$$

where $w^{(r)} \geq 0$ is a fixed weighting vector which is non-zero only on the $r$th patch. Since the patches are usually chosen to overlap, these weights smoothly interpolate between neighbouring filters $f^{(r)}$. Note that the weighting vectors have to sum up to one, *i.e.*

$$\sum_{r=0}^{R-1} w_i^{(r)} = 1 \text{ for } 0 \leq i < n. \tag{3.18}$$

Note that this method does *not* simply apply a different PSF to different image regions, but instead yields a different PSF for *each* pixel. The reason is that the patches are usually chosen to overlap at least 50%, such that the PSF at each pixel is a certain linear combination of several filters. The set of weights $\{w^{(r)} | r = 0, \ldots R-1\}$ are chosen to smoothly interpolate between neighbouring filters $f^{(r)}$.

**Efficient implementation.** Evidently, equation (3.17), the EFF is linear in $x$ and in $f$, where $f$ is the vector obtained by stacking $f^{(0)}, \ldots, f^{(R-1)}$. This implies that there exist matrices

$F$ and $X$ such that $y = Fx = Xf$. Using Stockham's ideas (Stockham, 1966) to speed-up large convolutions, Hirsch et al. (2010) derive expressions for these matrices, namely

$$F = Z_y^\mathsf{T} \sum_{r=0}^{R-1} E_r^\mathsf{T} \mathcal{F}^\mathsf{H} \operatorname{diag}(\mathcal{F} Z_f f^{(r)}) \mathcal{F} C_r \operatorname{diag}(w^{(r)}), \tag{3.19}$$

$$X = Z_y^\mathsf{T} \sum_{r=0}^{R-1} E_r^\mathsf{T} \mathcal{F}^\mathsf{H} \operatorname{diag}\big(\mathcal{F} C_r \operatorname{diag}(w^{(r)}) x\big) \mathcal{F} B_r Z_f, \tag{3.20}$$

where $\operatorname{diag}(w^{(r)})$ is the diagonal matrix with vector $w^{(r)}$ along its diagonal, $C_r$, $E_r$ and $B_r$ are appropriate cropping matrices, $\mathcal{F}$ is the discrete Fourier transform matrix, $Z_f$ is a matrix that zero-pads $f^{(r)}$ to the size of the patch, $\mathcal{F}^\mathsf{H}$ performs the inverse Fourier transform, $Z_y^\mathsf{T}$ chops out the valid part of the space-variant convolution.

Reading Eqs. (3.19) and (3.20) forward and backward yields efficient implementations for $F$, $F^\mathsf{T}$, $X$, and $X^\mathsf{T}$ with running times $O(n \log q)$ where $q$ is the patch size, see (Hirsch et al., 2010) or Section 2 for details. The overlap increases the computational cost by a constant factor and is thus omitted. The EFF framework thus implements space-variant convolutions which are as efficient to compute as space-invariant convolutions, while being much more expressive.

Note that each of the matrix vector multiplications (MVMs) with $F$, $F^\mathsf{T}$, $X$, and $X^T$ is needed for blind deconvolution: $F$ and $F^\mathsf{T}$ for the estimation of $x$ given $f$, and $X$ and $X^T$ for the estimation of $f$.

## 3.5 Results on Simulated Data

To investigate how our OBD algorithm performs on atmospherically degraded short-exposure images, we first experiment in a controlled setting with simulated data, firstly assuming invariant blur only.

Following Harding et al. (1999), we generate a sequence of 200 PSFs with Kolmogorov random phase screens at a specified ratio $D/r_o$ of the telescope diameter to the atmospheric seeing parameter (Fried parameter) equal to 13.3. The strength of the turbulence is chosen to create images recorded by a 26-inch telescope through atmospheric turbulence of a coherence length of approximately $r_o = 5\,\text{cm}$.

Figure 3.2 shows the original object, one out of the 200 PSFs, and the noise-free short exposure image obtained by convolving the shown PSF with the object. The object is a rendered model of the satellite OCNR5 used by Sheppard et al. (1998) and was chosen because of its high dynamic range and its great level of detail.

Before corrupting the images with noise, we add a constant background $b$ to the blurred image $f_t * x$. To simulate photon noise we scale the pixel values (ranging between 0 and 255) of each short exposure to varying large numbers of photons, *i.e.* $\lambda(f_t * x + b)$ and sample a new image $z$ from the corresponding Poisson distribution, *i.e.*

$$z_t \sim \text{Poisson}(\lambda(f_t * x + b)). \tag{3.21}$$

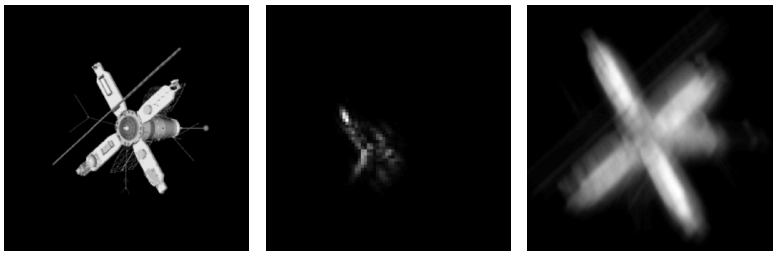


Figure 3.2: Simulation: From left to right: Original object image of OCNR5, typical PSF, blurred image.

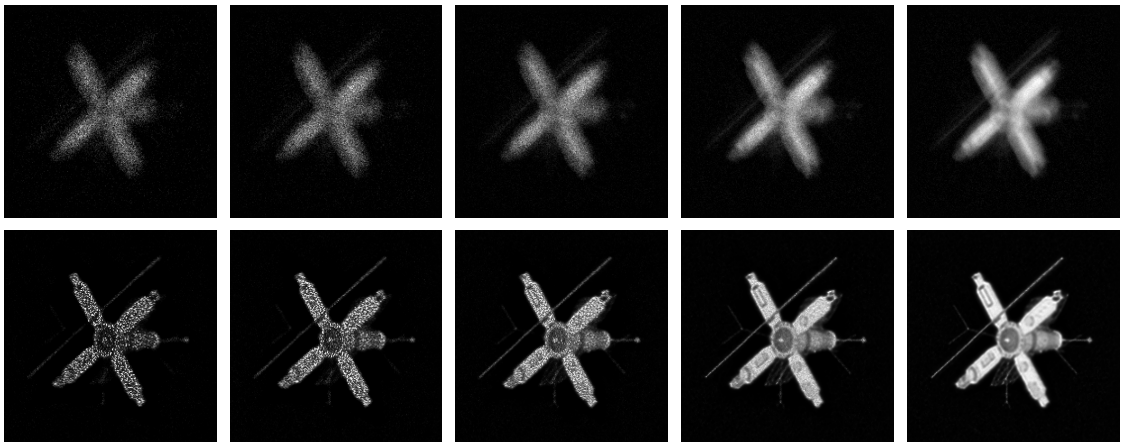



Figure 3.3: Simulation: Typical observed frames (top row) and reconstructed images (bottom row) after having processed a sequence of 200 blurred frames for different SNRs.

For differing $\lambda$ we can hereby simulate differing amounts of photon shot noise. After scaling down by $1/\lambda$, we add white Gaussian noise with zero mean and a variance $\sigma^2$ equal to two percent of the maximal image intensity of the whole sequence to model the readout noise common to CCD cameras,

$$n_t \sim \text{Gaussian}(0, \sigma^2) \tag{3.22}$$

$$y_t = \frac{z_t}{\lambda} + n_t. \tag{3.23}$$

To quantify the amount of image noise we define the following SNR,

$$\text{SNR}_t = 10 \log_{10} \frac{\text{Var}(x)}{\text{Var}(y_t - x * f_t)}, \tag{3.24}$$

where $x$ denotes the true satellite image, $y_t$ the noise-corrupted atmospherically degraded observation, and $f_t$ the PSF, respectively. $\text{Var}(x)$ denotes the variance of the pixel values of $x$. For an entire sequence $y_1$, $y_2$, ..., $y_{200}$ we average over the computed SNRs of all 200 frames, $\text{SNR} = \frac{1}{200} \sum_{t=1}^{200} \text{SNR}_t$. Table 3.1 shows the computed SNR for different parameter settings that we use in our experiments. Note that we use the SNR only to quantify the

Table 3.1: SNRs for different parameter settings of $\lambda$ and $\sigma^2$.

| $\lambda$ ($\times 10^3$) | 0.01 | 0.02 | 0.04 | 0.16 | 10.0 |
|---|---|---|---|---|---|
| $\sigma^2$ (%) | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| SNR in dB | 4.6 | 6.6 | 10.2 | 14.8 | 18.9 |

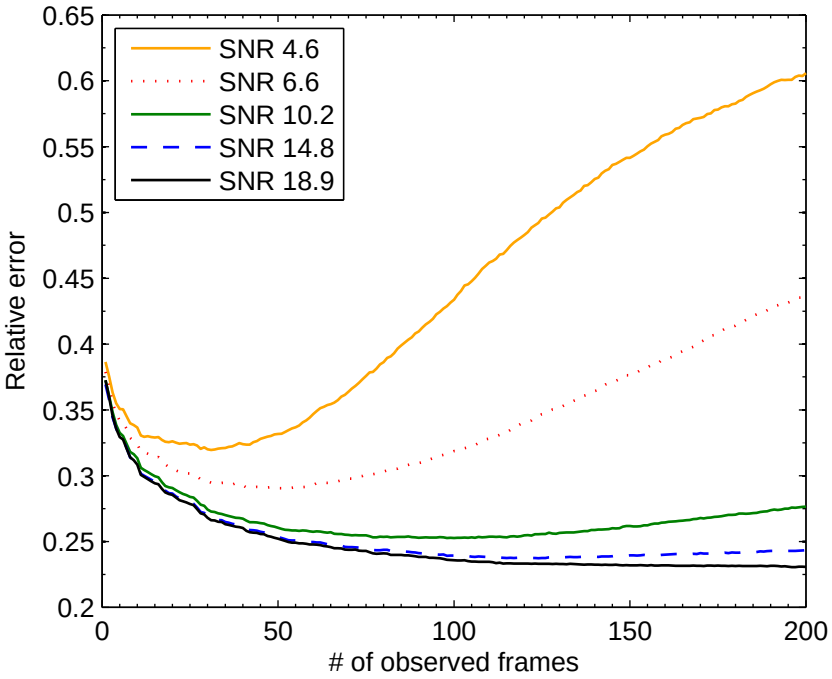



Figure 3.4: Simulation: Evaluation of relative reconstruction error for different SNRs.

amount of noise in the simulated data. To measure the quality of the reconstruction we use relative error (explained below).

Figure 3.3 shows typical frames for different SNRs, each $256 \times 256$ pixels in size, and the reconstructed object images of our basic algorithm after having processed all 200 frames within one sequence. The restored images shown are cropped to the size of the observations. As initial estimates for the PSFs we chose constant images of size $60 \times 60$ pixels, and as the initial estimate of the object, an average over the first twenty observed frames embedded in a $315 \times 315$ array of zeros.

As expected, the quality of the reconstruction suffers as the SNR decreases, which is also reflected quantitatively in Figure 3.4, where we plot the relative error $\|x - \hat{x}\|/\|x\|$ of the reconstructed image $\hat{x}$ as a function of observed frames and the corresponding SNR.

Evidently, for high SNRs the reconstruction error decreases the more observations that have been processed and saturates to a certain value dependent on the SNR. The error is higher the lower the SNR of the available observations. The error does not decrease strictly monotonically from frame to frame, but more in a (long-term) stochastic gradient manner. As expected, for lower SNRs, the unregularised reconstruction process can even diverge. In this noisy regime, additional prior knowledge about the object is necessary and

regularisation in the restoration process is inevitable.

Figure 3.5 illustrates that enforcing smoothness by employing Tikhonov regularisation on the gradients of the reconstructed image (*i.e.* a prior term $\eta\left\|\nabla x\right\|^2$ is added to the loss in (3.19)) is capable of suppressing noise amplification and stabilising the deconvolution process, even for low SNRs. As expected, when the regularisation parameter $\eta$ is too small, the reconstruction error still diverges (red dotted curve); similarly, when it is too large, the error is increased due to over-smoothing (blue dashed curve). A reasonable choice of the regularisation parameter may be obtained by setting it proportional to the noise variance. The colour framed image stamps show the reconstruction results for different values of the regularisation parameter.

To study the influence of the initialisation and the order of frames within one sequence, we reversed and randomly permuted the processing order of the input frames. Figure 3.6 shows restored object images and the corresponding error curves for a fixed SNR of 18.9 dB, respectively. As can be seen, the error evolution of the deconvolution process is almost independent of the particular ordering of the input frames. All curves converge to a similar value with small variance, and visually, only little (if at all) difference is discernible.

To numerically appraise the quality of our results, we did a quantitative comparison with various state-of-the-art reconstruction methods. Figure 3.7 shows the visually best observed frame, a reconstruction with AviStack (Theusner, 2009), a popular Lucky Imaging software. AviStack partitions the images into small image patches of variable sizes, evaluates the quality of all observed frames for all image patches and then aligns and stacks those image patches, that fulfil a certain quality threshold. For the final reconstruction only the best percent of observed frames was taken. Next to it, a Knox-Thompson reconstruction is shown, which was obtained using Speckle1, a reconstruction software by Stelzer (2009). For the reconstruction, 300 Knox-Thompson and 100 triple correlation phase pairs were used. Finally, the rightmost image shows the result of our basic algorithm without any additional regularisation. In all cases no further post-processing was performed.

For a single isoplanatic patch the reconstruction with AviStack is not substantially better than the visually best observed frame, which is also reflected in the relative error overlayed in white. In comparison, both the Knox-Thompson reconstruction and the result by the basic algorithm of our proposed method reveal much greater detail and higher spatial resolution. Subjectively, our result is comparable in quality and resolution to the Knox-Thompson reconstruction, which is quantitatively confirmed by the negligible difference in the reconstruction error. Regarding run-time, the C implementation of Stelzer (2009) takes about 15 minutes (when invoked carefully by an expert user) for the entire reconstruction on a single core of an Intel(R) Core(TM) i5 processor with 2.67 GHz. Our Matlab implementation that is however not optimised for speed and logs large quantities of intermediate results, takes about thrice as long. A Python implementation using PyCUDA (Klöckner et al., 2009) for GPU enabled computation of the discrete Fourier transform (see equation (A.2) and (A.3)) achieves a run-time of less than 10 minutes on a low-cost NVIDIA(R) GeForce(TM) GT 430.

Our final set of experiments with simulated data evaluates our algorithm's super-resolution abilities. We generated three sequences of atmospherically blurred, *differently* downsampled and noisy observations at a fixed SNR of 18.9 dB. Panel A of Figure 3.8 shows typical input

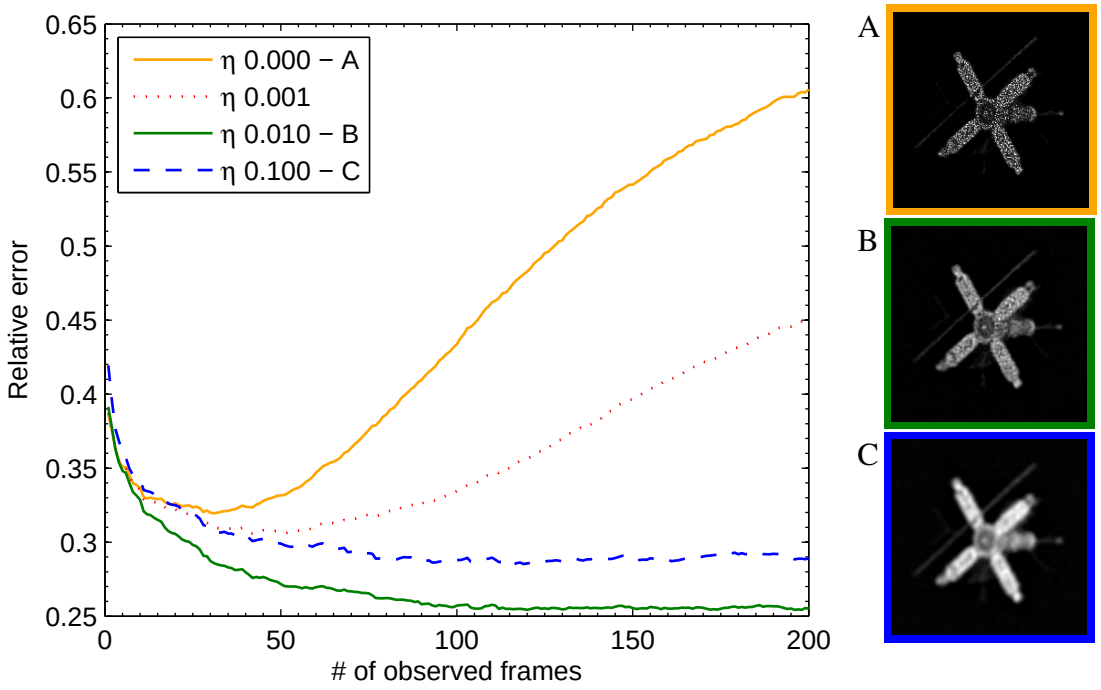



Figure 3.5: Simulation: Evaluation of the relative reconstruction error for different values of the regularisation constant $\eta$ at a fixed SNR of $4.6$ dB.

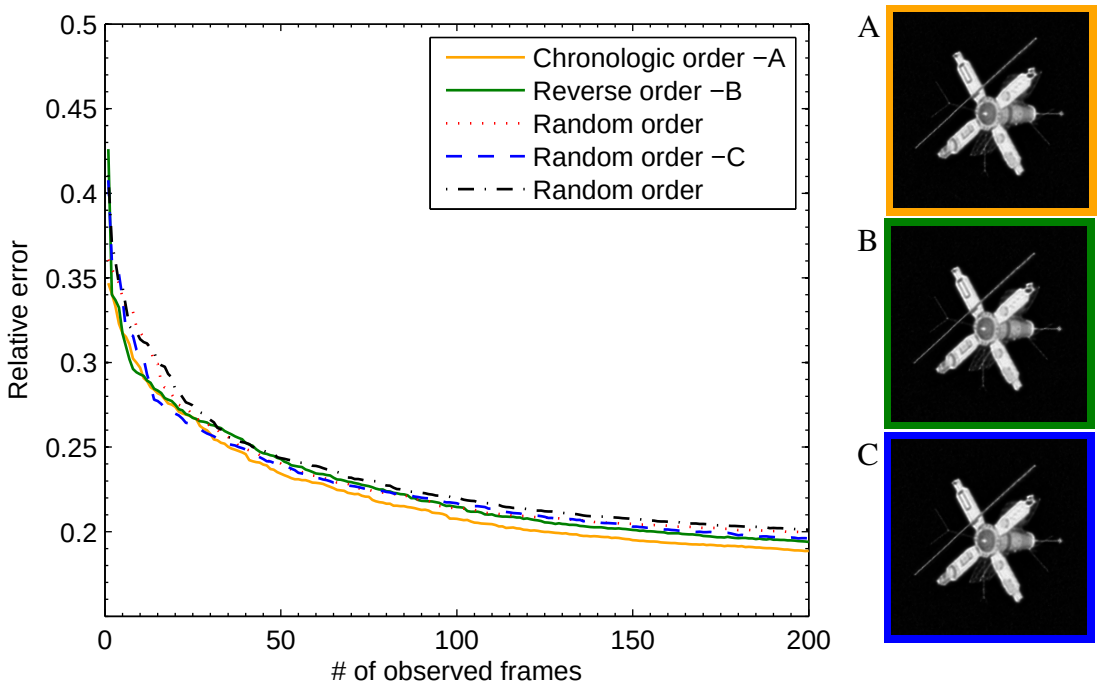



Figure 3.6: Simulation: Evaluation of the relative reconstruction error and final reconstructed images after having processed 200 frames in chronological, reverse and various random orders at a fixed SNR of 18.9 dB. The relative reconstruction error is overlayed in white.

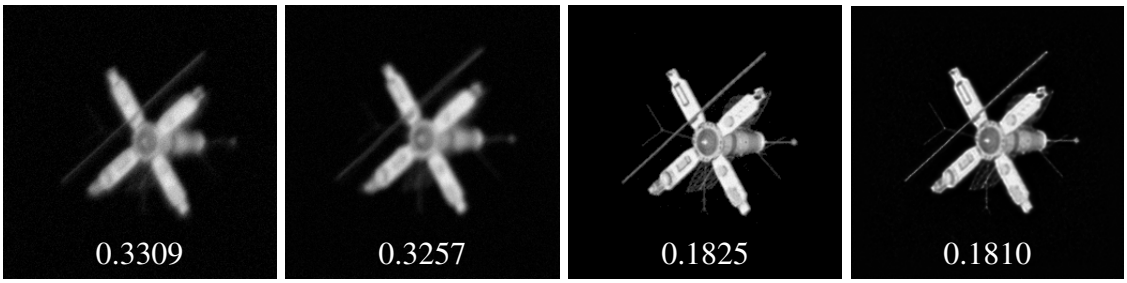



Figure 3.7: Simulation: Final reconstructed images after having processed 200 frames at a fixed SNR of 18.9 dB with Avistack, Knox-Thompson and our proposed method. The relative reconstruction error is overlayed in white.

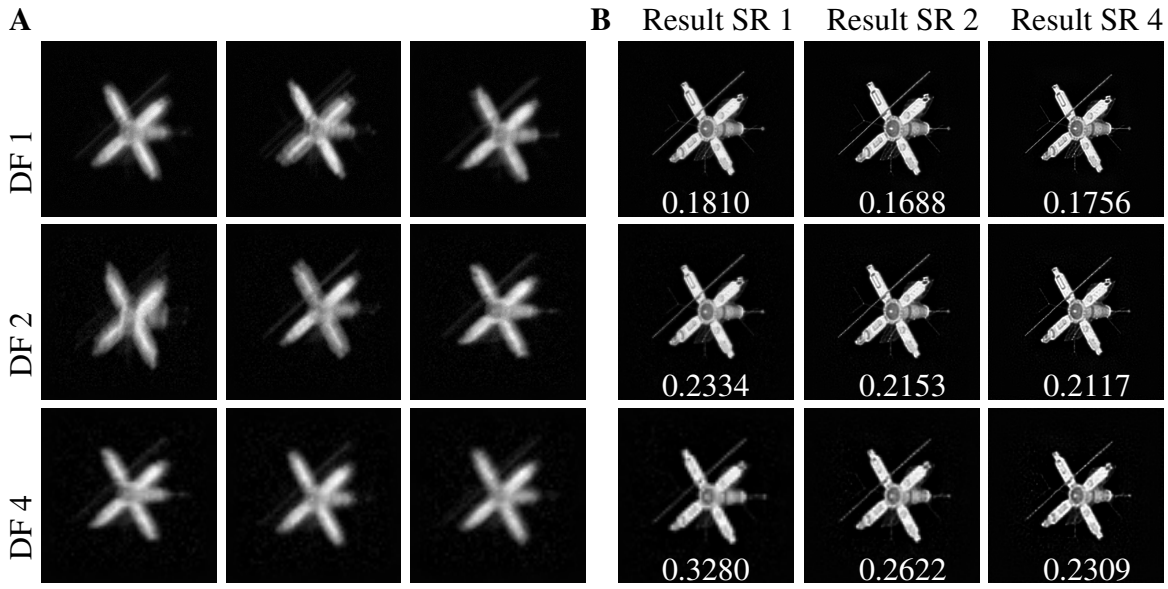
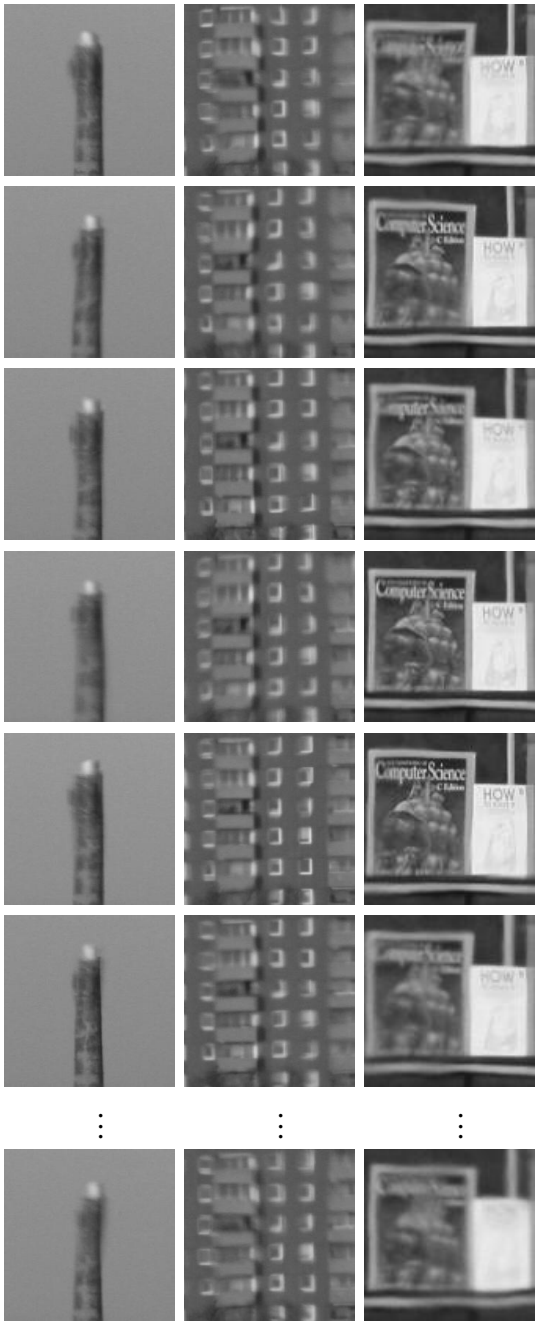



Figure 3.8: Simulation: Final reconstructed images (Panel B) after having processed 200 frames for differently downsampled input images (Panel A) and various super-resolution factors at a fixed SNR of 18.9 dB. The downsampling and super-resolution factor is abbreviated with DF and SR respectively. The displayed number corresponds to the relative reconstruction error. The corresponding video linked by the QR code shows the downsampled input sequence, the estimated PSFs and the result of our approach with 4x super-resolution.

images of these sequences together with their corresponding downsampling factors (DF in Figure 3.8). On each of these three simulations we ran our algorithm with various super-resolution factors. The results are shown in Panel B of Figure 3.8. The relative errors overlayed in white are computed by linearly interpolating the reconstructed images to the size of the ground truth image. The numbers suggest that incorporating super-resolution does improve the results beyond mere interpolation, which validates the merits of our approach.

## 3.6 Controlled Lab Experiments

To verify and evaluate our approach in the case of spatially varying blur, we recorded several image sequences under controlled conditions. For these experiments we mounted a Canon EOS 5D Mark II camera equipped with a 200 mm zoom lens on a tripod on a platform roof and captured a static scene through hot air exhausted by the building's vent, which could be closed to take sharp images of the same scene. The sequences consist of 100 frames degraded by spatially varying blur (each with an exposure time of 1/250s).



Chimney Building Books

Our approach Ground truth

Figure 3.9: Controlled Experiments: The left column shows typical recorded frames, the righ column compares the results from the OBD reconstruction to the ground truth images. The linked video shows the raw sequences and the result of our approach compared with the ground truth data.

Figure 3.9 shows typical frames, the reconstructed images as well as the ground truth images. The images of the *chimney* and the *building* sequences have size $237 \times 237$ pixels, the images of the *books* sequence have size $109 \times 109$ pixels. For all three sequences we choose a Bartlett-Hanning window with 50% overlap. We modelled the space-variant PSF with $3 \times 3$ local filters. The right panel of Figure 3.9 compares our reconstructed images (left column) with the sharp ground-truth images. For all three image sequences the reconstructed images reveal great faithfulness in detail and high-frequency structure confirming both our image model and the presented blind deconvolution algorithm.

## 3.7 Results on Astronomical Data

We now present results of our algorithm on a variety of actual astronomical data. Some of the images were taken with an off-the-shelf 12-inch f/10 MEADE LX200 ACF Schmidt-Cassegrain telescope, some with the 24-inch f/8 Hypergraph Cassegrain telescope "Ganymed" of the Capella Observatory located on Mount Skinikas in Crete, Greece. The data consists of short-exposure imagery of star constellations, the lunar Copernicus crater, as well as long-exposure deep-sky images. We compare our results against state-of-the-art methods used by both amateur and professional astronomers, and show that our method yields competitive if not superior results in all case studies.

### 3.7.1 Binary Star

The first dataset is an image sequence of the binary star system Epsilon Lyrae 2 of the constellation Lyra with an angular separation of 2.3" and a relative magnitude of 1.08. As we know precisely what to expect, our results on this dataset serve as an additional proof of concept. The sequence consists of 300 frames, each $132 \times 112$ pixels in size, taken with

Figure 3.10: Binary star system Epsilon Lyrae 2: typical observed image $y_{300}$ (left), reconstruction $x$ after 300 iterations (middle), estimated PSFs $f_{300}$ for each colour channel. Note the subtle differences in the PSFs due to wavelength dependent diffraction. Hence, the colour channels are not perfectly aligned.

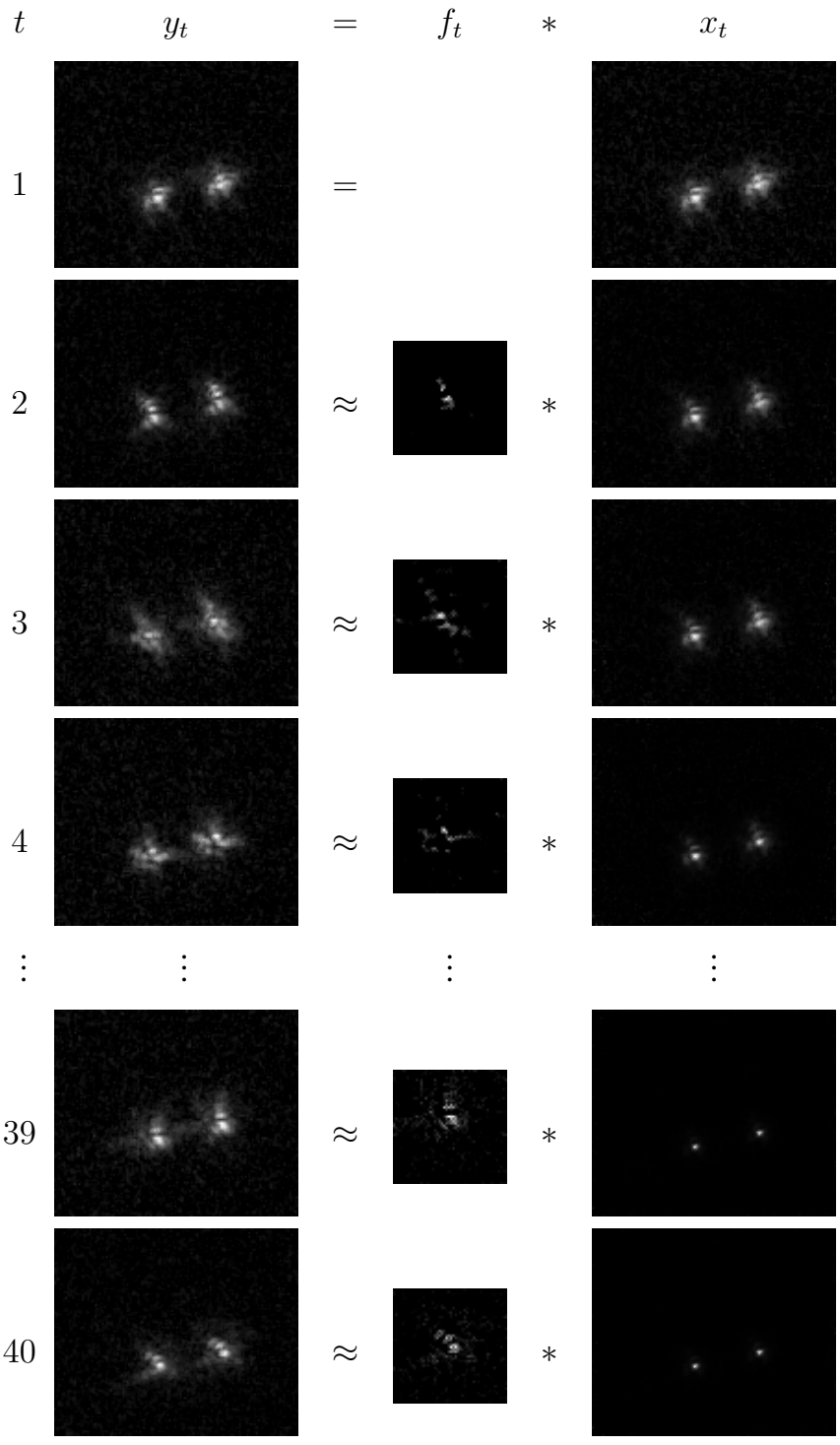



Figure 3.11: Binary star system Epsilon Lyrae 2: schematic illustration of the temporal evolution. From left to right: observed image $y_t$ (left), estimate of the corresponding PSF $f_t$ and reconstruction $x_t$ after $t$ time-steps. The linked video shows the original image sequence.

a 24-inch Cassegrain telescope at the Capella Observatory and an Imaging Source DFK 31BU03 CCD camera; the image scale was 0.06" per pixel. The seeing was estimated to FWHM $\approx$ 0.8", corresponding to a Fried parameter of $r_o \approx 20cm$ at a wavelength of $\lambda = 500nm$.

Figure 3.11 shows in three columns the first four and the last two of the first 40 frames of the processed sequence and illustrates schematically how our method works. Each row shows from left to right the observed image $y_t$, the corresponding PSF $f_t$ estimated by our algorithm and the current estimate $x_t$ of the true image we want to recover. The PSF is chosen to be of $30 \times 30$ pixels.

The image $x$ is initialised by the first observed frame $y_1$. Then $f_2$ is estimated from the second observed frame $y_2$ and the current estimate of $x$. After that we improve the estimate of $x$ by means of (3.10) and proceed with the next observed frame. Figure 3.11 shows nicely that already after 40 frames we obtain a good reconstruction.

Figure 3.10 shows an enlarged version of the result of our algorithm after 300 iterations along with the estimated PSFs for each colour channel. Note how blurry the observed image $y_t$ is (left), while our estimate of $x$ is almost free of any degradation (middle). Furthermore, we see that both stars have almost identical diffraction patterns which strongly resemble the estimated PSFs (shown on the right for each colour channel). This finding justifies our assumption about a constant PSF for the whole image. From the final reconstructed image we determined a separation of 2.28" and a magnitude ratio of 1.08, which is in excellent accordance with the literature.

### 3.7.2 Copernicus Crater

To evaluate our algorithm on an extended celestial object, we applied it to a sequence of short exposures of the Copernicus crater, a prominent lunar crater located in eastern Oceanus Procellarum. The original recording was taken with a 14-inch f/10 Celestron C14 and a DMK 31 AF03 CCD camera from Imaging Source at a frame rate of 30fps near Frankfurt, Germany (courtesy Mario Weigand). It consists of 2350 frames in total, where each frame is $1024 \times 768$ pixels in size. To begin with we focus on an isoplanatic patch. To meet our assumption of a constant PSF, we processed only a small image patch of $70 \times 70$ pixels, which corresponds to a angular size of 0.92". In this field of view the PSF is assumed to be constant, which is a valid assumption for the seeing conditions at the time of recording.

The top row of Figure 3.12 shows the selected region of the central peak in the Copernicus crater and typical observed frames. The image patches were aligned on a pixel scale before processing to reduce computational costs[4]. For reconstruction all 2350 observed frames were taken into account.

The bottom row of Figure 3.12 shows a comparison of different reconstruction methods. Panel (a) of Figure 3.12 shows the visually best observed frame, Panel (b) a reconstruction with AviStack (Theusner, 2009), for which the best ten frames were taken into account. In Panel (c) a Knox-Thompson reconstruction is shown, which was done with Stelzer (2009) using 300 Knox-Thompson and 100 triple correlation phase pairs. Finally, Panel (d) shows

[4]Note, that a PSF can account for translational motion but necessitates a PSF size as large as the translation amplitude, which might increase the computational cost for severe motion.

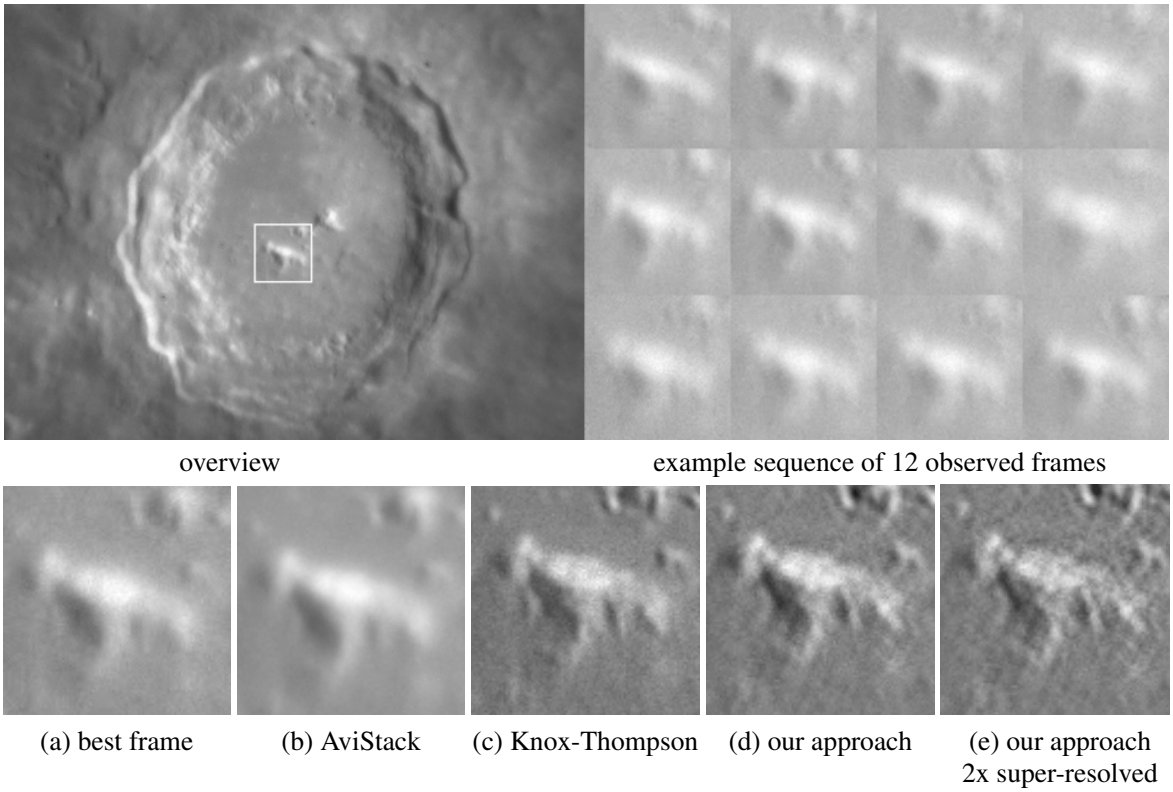



Figure 3.12: Copernicus Crater: Top panel: Full frame with extracted image patch marked by white square (left) and example sequence of 12 observed frames. Bottom panel: comparison of results of different reconstruction algorithms (from left to right): visually best frame, AviStack (best, Knox-Thompson, our approach and our approach with two times super-resolution. All image results are shown without any image enhancement. This figure is best viewed on screen, rather than in print.

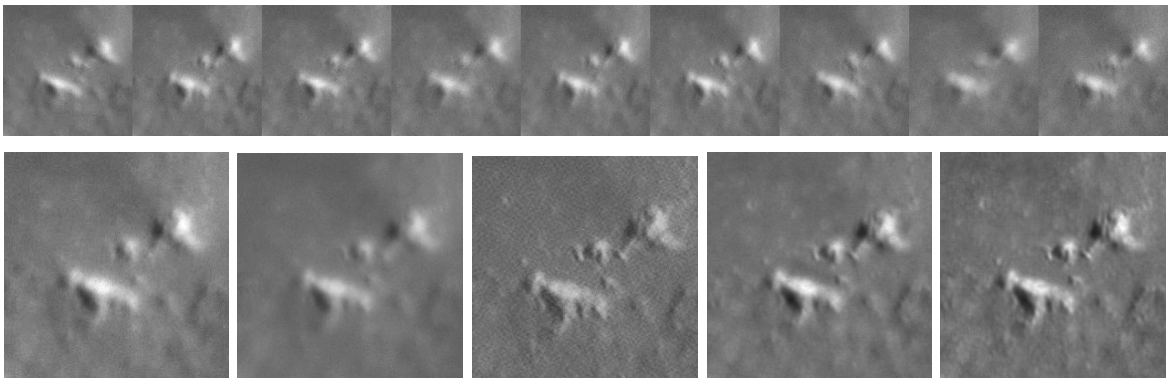
the result of our basic algorithm and Panel (e) the result two times super-resolved. In all cases no further post-processing was performed.

As before, within a single isoplanatic patch the result of AviStack does not to seem to be considerably better than the visually best observed frame. In contrast, the Knox-Thompson reconstruction reveals greater detail and higher spatial resolution. Subjectively, our result is comparable in quality and resolution to the Knox-Thompson reconstruction. The two times super-resolved reconstruction seems to reveal even more detail.

**Beyond the isoplanatic patch.** To investigate the capabilities of our approach assuming a non-uniform imaging model, we extended the field of view to $227 \times 227$ pixels in size, which corresponds to an angular size of approximately 16”. Despite the good seeing conditions at the time of recording, this field of view is beyond the isoplanatic patch and thus a spatially varying PSF is needed to describe the image deformation caused by the atmosphere. Again, the bottom row of Figure 3.13 shows from left to right the subjectively best observed frame, a reconstruction with AviStack (Theusner, 2009), a Knox-Thompson recon-

struction (Stelzer, 2009) (using 300 Knox-Thompson and 100 triple correlation phase pairs), the result of our approach assuming an invariant imaging model (with PSF size $31 \times 31$ pixels and Tikhonov regularisation for PSF estimation), and the result of our approach assuming a non-uniform imaging model. We modelled the recorded frames as an EFF with $3 \times 3$ local filters of size $31 \times 31$ pixels and a Bartlett-Hanning window of size $128 \times 128$ with 50% overlap to model the spatially varying blur. For none of the method further post-processing was performed.

Not surprisingly, the reconstruction with AviStack is only slightly better than the visually best observed frame. By comparison, the result of the Knox-Thompson method reveals greater detail and higher spatial resolution. Despite the violated assumption of isoplanacity, the reconstructed image modelled by a single PSF is comparable in quality to the Knox-Thompson reconstruction. Compared to the previous images, our estimated image under the assumption of a spatially varying PSF shows even more detail and reveals structure unresolved in the previous images.



Visually best frame | AviStack | Knox-Thompson | OBD with invariant PSF | OBD with EFF

Figure 3.13: Copernicus Crater: top row shows typical observed frames; the bottom row shows from left to right: the visually best frame, result of AviStack, a Knox-Thompson reconstruction, OBD with uniform PSF and non-uniform PSF. The linked video shows the input sequence as well as the result of our approach together with the estimated PSFs.

### 3.7.3 Orion Trapezium

In this experiment, we used a 12-inch f/10 Meade LX200 ACF Schmidt-Cassegrain telescope and a AVT PIKE F-032B uncooled CCD camera to record a short video (191 frames acquired at 120 fps) of the Trapezium in the constellation Orion. The exposure time of the individual frames was sufficiently short to "freeze" the atmospheric turbulence and thus retain the high-frequency information which is present in the atmospheric PSF–see Figure 3.14a for sample frames.

The Orion Trapezium is formed by four stars ranging in brightness from magnitude 5 to magnitude 8, with angular separations around 10" to 20". Here it should be mentioned that

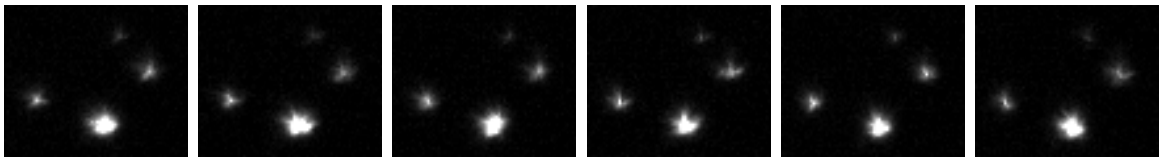
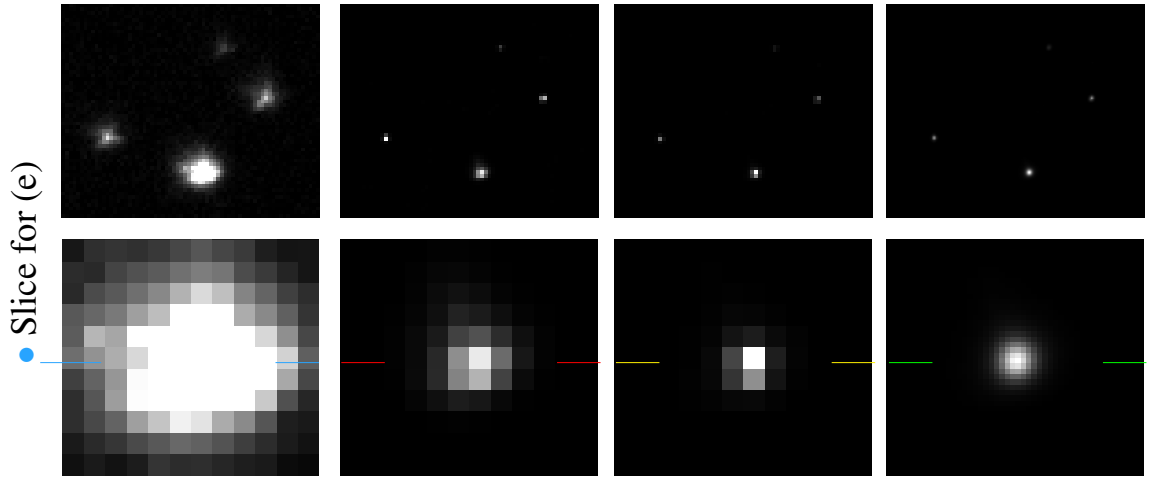


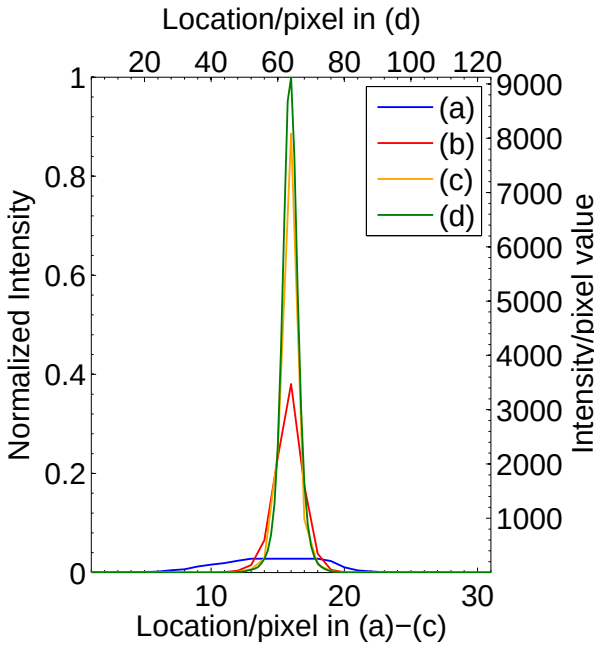
Figure 3.14: Orion Trapezium Cluster: example sequence of observed frames, $y_1, \ldots, y_6$.




(a) observed frame $y_1$ (b) plain OBD (c) OBD with SC (d) OBD with SC & SR

Figure 3.15: Orion Trapezium Cluster (from left to right): (a) the first observed frame, (b) $x_{191}$ for basic algorithm, (c) $x_{191}$ for saturation corrected, and (d) $x_{191}$ for saturation corrected and four times super-resolved. Top row shows the overall trapezium; bottom row shows the brightest star enlarged. Panel (e) shows the stellar profiles at the positions indicated by the coloured lines in plots (a)-(d). The linked video shows the original image sequence with a closeup of the overexposed image region.

(e)



our assumption of a constant PSF throughout the field of view is strongly violated. However, by resorting to early stopping in this case, we avoid over-fitting the PSF. The first row of Figure 3.15 shows from left to right (a) an enlarged unprocessed frame, (b) the deconvolution results obtained by our basic algorithm, (c) the result using the proposed method to handle saturation, and (d) the results if we additionally apply the proposed method for four times super-resolution. The bottom row shows a closeup of the brightest star within the Trapezium. Panel (e) of Figure 3.15 shows the star profiles obtained by slicing as indicated by the coloured lines in the image stamps (a)–(d).

An important application in astronomy is the measurement of the brightness of stars and other celestial objects (photometry). To this end, a linear sensor response is required (for our purposes, the used CCD sensor may be assumed linear). The intensity counts can then be translated into stellar magnitudes. Clearly, this is not directly possible for stars that saturate the CCD (*i.e.*, where so many photons are recorded that the capacities of the pixels are exceeded). However, we can use the proposed method for deconvolution with saturation correction and reconstruct the photon counts (image intensities) that would have been recorded had the pixels not been saturated; then we convert these into ratios between star intensities, *i.e.* differences between stellar magnitudes. For the difference between two star magnitudes, we use the formula $m_1 - m_2 = -2.5 \log_{10} p_1/p_2$ where $p_1$ and $p_2$ are the pixel values of two stars in the reconstructed image. We do this for all Trapezium stars relative to the brightest star $C$ and obtain encouraging results (see Table 3.2).

| Star | C (ref.) | A | B | D |
|---|---|---|---|---|
| True magnitude | 5.1 | 6.7 - 7.5 | 8.0 - 8.5 | 6.7 |
| | | A-C | B-C | D-C |
| True magnitude differences | | 1.6 - 2.4 | 2.9-3.4 | 1.6 |
| Est. diff., deconv. w/o sat. cor. | | 0.2936 | 1.4608 | -0.0964 |
| Est. diff., deconv. w. sat. cor. | | 1.1955 | 2.7718 | 0.8124 |

Table 3.2: True star magnitudes (note that stars A and B have variable magnitudes), true differences to star C, and estimated difference values estimated after deconvolution without and with saturation correction. Note that the results with saturation correction are closer to the true differences.

### 3.7.4 Globular Cluster M13

M13 is a globular cluster in the constellation Hercules, around 25,000 light years away, with an apparent size of around 20'. It contains several 100,000 stars, the brightest of which has an apparent magnitude of 12. Such faint stars cannot be imaged using our equipment for short exposures; however, long exposures with budget equipment typically incur tracking errors, caused by telescope mounts that do not perfectly compensate for the rotation of the earth. In our case, the tracking errors induced a significant motion blur in the images, which we attempted to remove using the same algorithm that we used above on short exposures. All

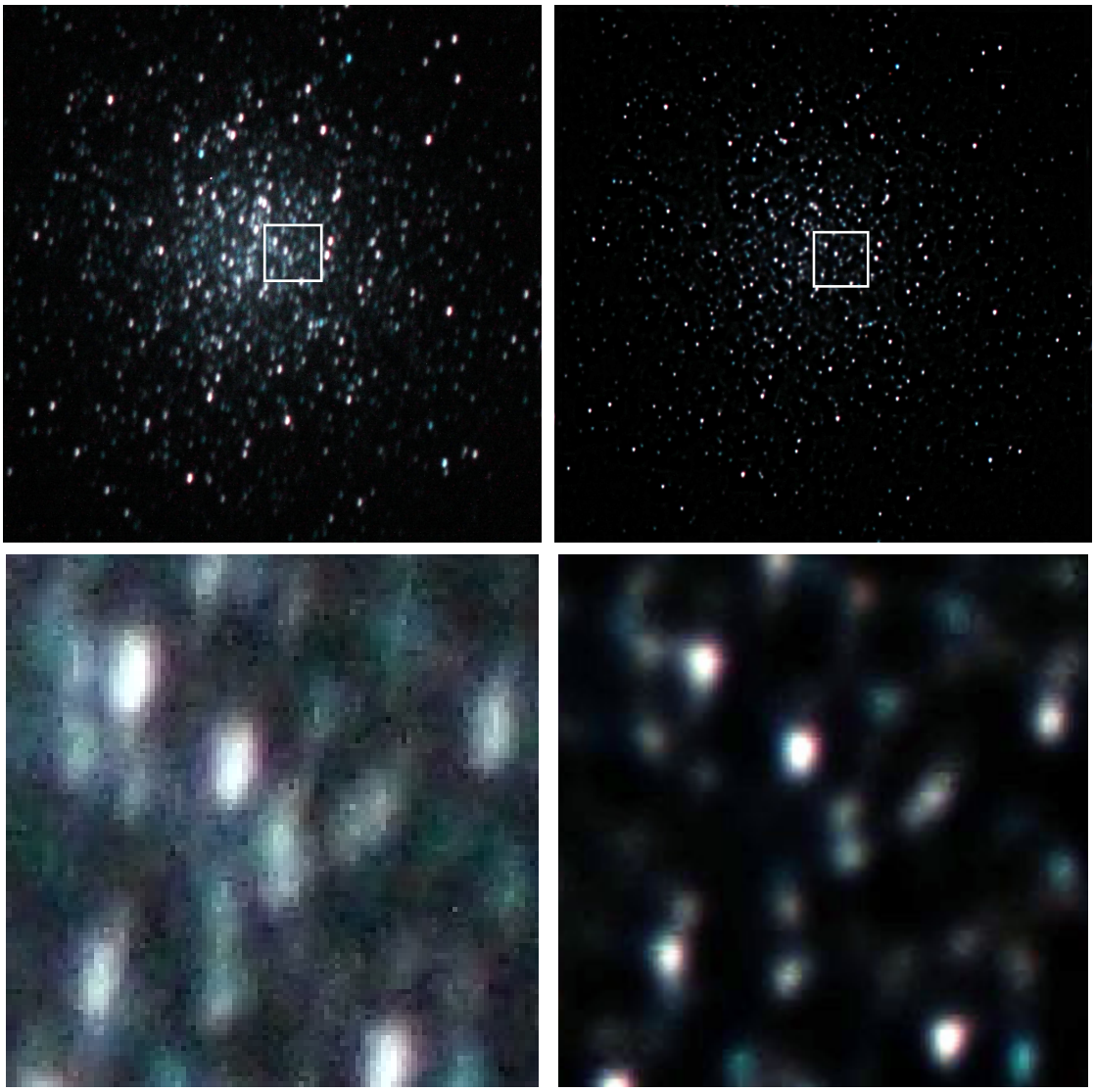


Figure 3.16: Globular cluster M13: (left) example observed frame, (right) result of saturation corrected, two times super-resolved multi-frame blind deconvolution; (top) overview, (bottom) closeup. For better display the images have been automatically gamma corrected.

raw images were recorded using a 12-inch f/10 MEADE LX200 ACF Schmidt-Cassegrain telescope and a Canon EOS 5D digital single lens reflex (DSLR) camera. The whole sequence consists of 26 images with an exposure time of 60s each. The top row of Figure 3.16 displays a long exposure with motion blur (left panel and the twice super-resolved result of our algorithm (right) applied to 26 motion degraded frames. In the bottom row we clearly see details in our reconstructed image (right) which where hidden in the recorded frames (left). However, note that in the bottom right panel there appear also some JPEG-like artifacts which might suggest that 26 frames were not enough for two times super-resolution.

## 3.8 Results on Non-Astronomical Data

In the following section, we show some results on non-astronomical imagery. First, we apply our algorithm to some benchmark data for MFBD with incorporated super-resolution and compare against state-of-the-art in the field. Secondly, we apply our proposed algorithm to medical imaging data, where we enhance a sequence of MRI images.

### 3.8.1 Benchmark Datasets

The method most closely related to ours is the state-of-the-art *blind* super-resolution approach of Šroubek et al. (2008). We compare their method against ours by showing results on some datasets of S. Farsiu and P. Milanfar[5]. We show Šroubek *et al.*'s results as reported in Šroubek et al. (2008). For brevity we consider only the *text* dataset (20 frames of size $57 \times 49$) and the *disk* dataset (20 frames of size $57 \times 49$).

The top rows of Figure 3.17 and Figure 3.18 show typical frames of these sequences. The bottom rows show the result of our method with increasing super resolution factors (one, two, four, and eight times). We used the results corresponding to lower super-resolution factors to initialise for the next higher factor. Already with a factor of two, our results compare favourably with Šroubek's, which is surprising because our method does not depend on any detailed image or blur priors like Šroubek's method. The image obtained using a factor eight super-resolution is clearly superior to the result of Šroubek et al. (2008), whose method could not super-resolve beyond a factor of two because of algorithmic instability as reported in (Šroubek et al., 2008).

### 3.8.2 MRI of Objects in Motion

The second application addresses the common problem of object motion in Magnetic Resonance Imaging (MRI). MRI is a medical imaging modality for visualising the internal structure and function of the human body and animals used in preclinical studies. Compared to Computed Tomography (CT), MRI provides much greater contrast between different soft tissues, that makes it especially useful in neurological (brain), musculoskeletal, cardiovascular, and oncological (cancer) imaging.

Figure 3.19 shows typical frames of two image sequences of a mouse's thorax in a preclinical study for contrast MRI. The sequences correspond to two transverse slices at different height. Both were taken with a 7 Tesla ClinScan of Bruker and consist of 200 frames, each $128 \times 128$ pixel in size. As can be seen from these frames, object motion produces large blurs and leads to a significant loss in image quality. Besides global object motion also the heart beat causes local distortions. Both global and local deformations can be described by our framework for space-variant filtering. Thus we applied our space-variant blind deconvolution algorithm with $4 \times 4$ PSFs of size $20 \times 20$ pixels (choosing a Bartlett-Hanning window of size $64 \times 64$ pixels with 50% overlap). For kernel estimation we imposed additional Tikhonov regularisation. Figure 3.19 shows the estimated images of our method.

[5]Available from: `http://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html`.

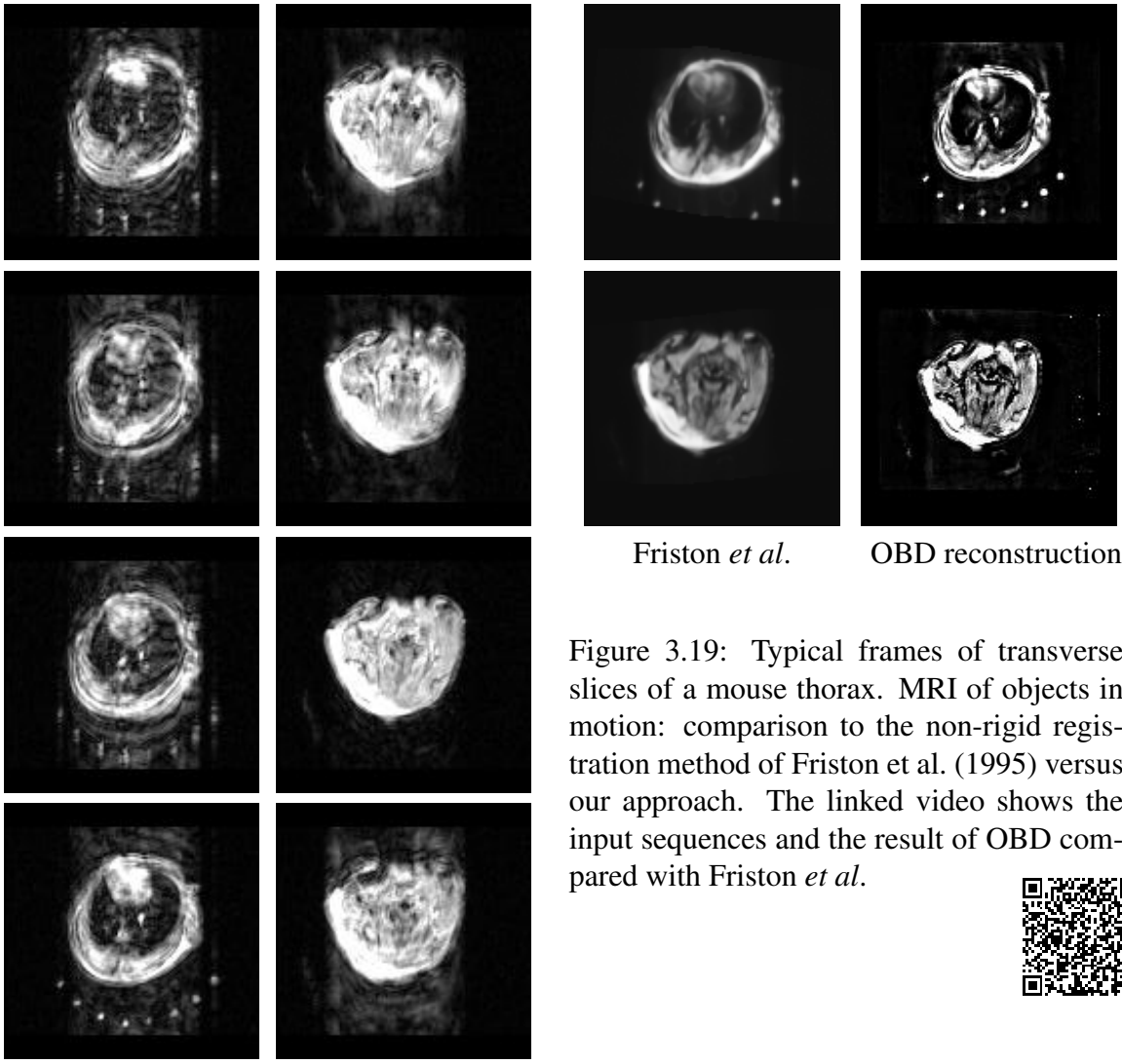

1x 2x 4x 8x Šroubek *et al.*

Figure 3.17: *Text* sequence: typical example frames (top row), results of our method for blind deconvolution with increasing super-resolution factor compared with Šroubek's results taken from Šroubek et al. (2008) (bottom row from left to right). Already at 2x our results appear to be better than Šroubek's. Note that Šroubek's pictures seem to be postprocessed as their background appears to be darker than that of the input sequence. The linked video shows the input sequence along with our reconstruction results and the estimated PSFs.


1x 2x 4x 8x Šroubek *et al.*

Figure 3.18: *Disk* data: typical example frames (top row), results of our method for blind deconvolution with increasing super-resolution factor compared with Šroubek's results taken from Šroubek et al. (2008) (bottom row from left to right). Already at 2x our results appear to be better than Šroubek's. The linked video shows the input sequence along with our reconstruction results and the estimated PSFs.

More interestingly, our method can be used for effective motion correction, as at each time step the estimated object image remains at the same position. Dropping the energy constraint on the kernel, the estimated PSFs give not only information about the object motion, but also about the intensity change, which is of mayor interest in contrast or functional MRI. The right panel of Figure 3.19 compares our results with a state-of-the-art method for non-rigid registration of Friston et al. (1995). It is clearly evident, that our method recovers more image details.



Friston *et al.* OBD reconstruction

Figure 3.19: Typical frames of transverse slices of a mouse thorax. MRI of objects in motion: comparison to the non-rigid registration method of Friston et al. (1995) versus our approach. The linked video shows the input sequences and the result of OBD compared with Friston *et al.*

## 3.9 Conclusions and Future Work

In this chapter, we proposed a simple, efficient, and effective multi-frame blind deconvolution algorithm. This algorithm restores an underlying static image from a stream of degraded and noisy observations by processing the observations in an online fashion. For moderate

signal-to-noise ratios our algorithm does not depend on any prior knowledge other than non-negativity of the PSFs and the images. Thus, in a sense our reconstruction is unbiased since no specific image model is enforced. Moreover, our formulation exploits the availability of multiple frames to incorporate super-resolution and saturation-correction.

We showed results on both simulated and real world astronomical data to verify and demonstrate the performance of our algorithm. We experimented with not only short-exposure images where the degradation is caused by atmospheric turbulence, but also with long exposures that suffer from saturation and additional blur arising from mechanical inaccuracies in the telescope mount. Our method yields results superior to or at worst comparable to existing frequently used reconstruction methods.

Future work includes further building on the simplicity of our method to improve it to work in real-time. This goal might be achievable by exploiting fast graphical processing unit (GPU) based computing. First attempts already yielded promising results (see Section 3.5. Beyond computing improvements, two other important aspects are: (i) to explore the spatio-temporal properties of the speckle pattern; and (ii) to incorporate and investigate additional regularisation within the reconstruction process.

## 3.10 Acknowledgements

The work presented in this chapter was carried out in collaboration with Stefan Harmeling, Suvrit Sra and Bernhard Schölkopf of the Max Planck Institute for Intelligent Systems. Stefan Harmeling helped with the derivation and an early implementation of the algorithm, the recording of the image sequence presented in Section 3.6 and the preparation of the manuscripts. Suvrit Sra gave valuable advice in all mathematical optimisation related issues as well scientific writing by greatly improving early drafts of the conference and journal papers. I am especially grateful to Bernhard Schölkopf who gave me an introduction to astronomical instrumentation and observation, joined observation nights whenever possible and fully supported the presented research by not only sharing ideas and giving valuable scientific advice but also by assisting the preparation of the manuscripts.

Except otherwise explicitly stated, all mathematical derivations, algorithmic implementations and experimental evaluations were performed by the author of this thesis.

This chapter is based on the following publications:

- *Multi-frame online blind deconvolution with super-resolution & saturation correction*, Michael Hirsch, Stefan Harmeling, Suvrit Sra and Bernhard Schölkopf. Astronomy and Astrophysics, 531, A9, 2011.
- *Efficient Filter Flow for Space-Variant Multi-frame Blind Deconvolution*, Michael Hirsch, Stefan Harmeling, Suvrit Sra, and Bernhard Schölkopf. In 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010.
- *Multi-frame Blind Deconvolution, Super-resolution, and Saturation Correction via Incremental EM*, Stefan Harmeling, Suvrit Sra, Michael Hirsch, and Bernhard Schölkopf. In 17th IEEE International Conference on Image Processing (ICIP) 2010.

- *Online blind deconvolution for Astronomy*, Stefan Harmeling, Michael Hirsch, Suvrit Sra and Bernhard Schölkopf. In 1st International Conference on Computational Photography (ICCP) 2009.

The author thanks Hanns Ruder, Josef Pöpsel and Stefan Binnewies for fruitful discussions, providing astronomical data and for their generosity regarding observation time at the Capella Observatory. Many thanks also to Karl-Ludwig Barth (IAS) and Mario Weigand for some of the presented astronomical data as well as Prof. Dr. Bernd Pichler from the UKT Tübingen for providing the MRI sequence shown in Section 3.8.2.

A Matlab implementation and source code is available from the accompanying webpage at `http://webdav.is.mpg.de/pixel/obd/`.

# Chapter 4

# Fast Removal of Non-Uniform Camera Shake

Motion blur due to camera shake is one of the predominant sources of degradation in photographic imagery. With a few exceptions such as panning photography, camera shake is unwanted, since it often severely limits the image quality by destroying details. Modelling camera shake as a space-invariant convolution simplifies the problem of removing camera shake, but it often insufficiently models actual motion blur as any camera rotation and motion outside the sensor plane will cause motion blur that spatially varies. Therefore, state-of-the-art methods for removing camera shake model the blur as a linear combination of homographically transformed versions of the true underlying image. While this is conceptually interesting, the resulting algorithms are computationally demanding. In this chapter, we develop a fast forward model based on the Efficient Filter Flow (EFF) framework developed in Section 2, that incorporates the particularities of camera shake. We derive an efficient algorithm for motion blur removal and show in a comprehensive comparison on a number of real-world examples that our approach is not only substantially faster, but also leads to better deblurring results.

This chapter, which is based on material contained in (Harmeling et al., 2010a) and (Hirsch et al., 2011) is organised as follows: we first give an introduction to the problem of single image blind deblurring and survey related work in Section 4.1. We start by summarising recent developments of single image Blind Deconvolution (BD) algorithms that assume uniform motion blur. By analysing real camera shake, we point out their limitations and stress the need for more sophisticated imaging models that are able to express non-uniform motion blur. Subsequently, we show results for a straightforward approach to the problem: we extend a state-of-the-art algorithm for removing uniform camera blur to being capable of expressing non-uniform blur through the incorporation of the EFF framework (Harmeling et al., 2010a). Despite being able to deblur real-world images, we demonstrate the limitations of this approach and point out its shortcomings. Other recent approaches, which remedy some of these shortcomings are based on the so-called Projective Motion Path Blur (PMPB) model, which was first proposed by Tai et al. (2009, 2010) and subsequently refined by Whyte et al. (2010) and Gupta et al. (2010). We give a quick review

of the PMPB model and discuss its benefits, but also work out its drawbacks. Before we show how to combine both of these recent developments in Section 4.2, we conclude the introduction by summarising other relevant work in the field. Afterwards by starting from first principles, we derive a fast image formation model, which unites the strengths of both the EFF framework as well as the PMPB model. Based on our fast forward model, we propose an efficient and robust deblurring algorithm in Section 4.3 and validate our method in a comprehensive comparison in Section 4.4. To objectively evaluate the performance of our algorithm, we present a novel experimental setup that allows both the capture of images with real camera shake and the simultaneous recording of the space-variant point spread function corresponding to that blur. Details on the implementation and the computational resources are given in Section 4.5. After a discussion of the presented results and its current limitations in Section 4.6, we conclude this chapter by giving an outline of future work in Section 4.7.

# 4.1 Introduction

Camera motion during exposure is a common problem in handheld photography, as it causes motion blur that destroys details in the captured photo. Especially images taken under low-light conditions without flash suffer from motion blur due to the necessity of longer exposure times.

Single image blind deconvolution or motion deblurring aims at restoring the sharp latent image from its blurred picture without knowing the camera motion that took place during the exposure. Blind deconvolution involves many challenging problems, including modelling the image formation process, formulating tractable priors incorporating generic image statistics, and devising efficient methods for optimisation. This renders it an intriguing but also intricate task, which has recently seen much attention as well as progress in the computer vision and graphics community.

## 4.1.1 Uniform Blur Models

Efforts focusing on the incorporation of prior terms that better model natural image statistics (Levin, 2006; Levin et al., 2009), as well as the refinement of inference methods (Fergus et al., 2006; Shan et al., 2008), paved the way for fast and robust deconvolution of motion blurred images yielding high-quality restoration results (Cho and Lee, 2009; Xu and Jia, 2010). These works assume a stationary or spatially invariant blur model, *i.e.* the image of any point source in the object plane is identical. In this case, the point spread function (PSF) is assumed to be translational invariant and the blurring process can be described mathematically by a convolution.

Making use of the *convolution theorem*, which states that a spatial convolution corresponds to a multiplication in Fourier space, a forward operation can be computed efficiently via the Fast Fourier Transform (FFT) in $O(n \log n)$ where $n$ denotes the length of the signal. The inverse operation, *i.e.* a deconvolution of the convolved image with a known PSF for the estimation of the sharp latent image corresponds to a division in Fourier space, be-

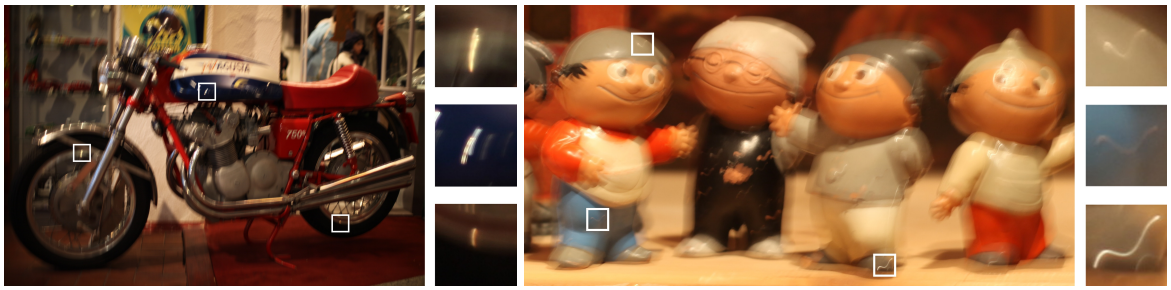


Figure 4.1: Photos taken by a handheld camera exhibiting typical blur due to camera shake during exposure. Point-like cues reveal a significant spatial variation of the unknown PSF.

ing of the same complexity as the forward model. However, noise in the observed image renders deconvolution even an ill-posed problem, for perfect knowledge of the PSF, leading to unwanted noise amplification and ringing artifacts in the reconstructed image (Levin, 2006). Nevertheless, recent works that presented fast algorithms for image and PSF estimation make heavy use of deconvolution via division in Fourier space and effectively suppress reconstruction artifacts by employing powerful nonlinear filtering techniques (Cho and Lee, 2009; Xu and Jia, 2010) or sophisticated prior terms (Krishnan and Fergus, 2009).

### 4.1.2 Real Camera Shake

Despite its computational amenities, the assumption of an invariant blur model is also a common drawback of these methods, as it has been demonstrated to be quite restrictive (Levin et al., 2009): the blur caused by camera shake usually involves a significant amount of rotational motion, rendering it non-stationary across the image plane as evident in typically blurred photos taken by a handheld camera in Figure 4.1.

To illustrate the variation of the PSF across the image plane, we built a device consisting of light fibres equidistantly affixed on a black hard plastic board (see left panel of Figure 4.2). A picture taken from front with a handheld camera under typical indoor conditions depicts the response of the camera to point-like objects which exactly corresponds to the PSF sampled at discrete sites. The middle and right panel of Figure 4.2 show typical images obtained with our experimental setup. Evidently, the common assumption of a stationary PSF is not true. However, what can also be seen is that the PSF varies smoothly across the image plane.

### 4.1.3 Non-Uniform Blur Models

The fact that camera shake causes non-uniform blur motivated other works focusing on more comprehensive image generation models that go beyond the spatially invariant convolution model and are able to describe space-variant motion blur.

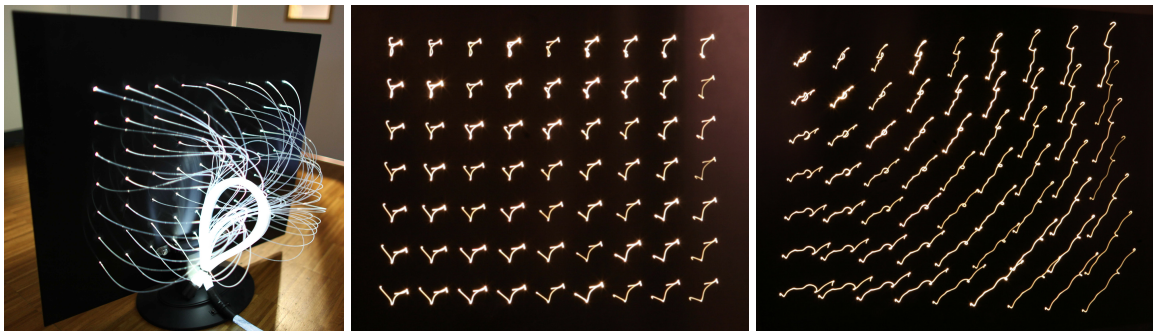


Figure 4.2: Left panel: Device for illustrating the PSF arising from motion blur caused by camera shake. Middle and right panel: Typical blur due to camera shake.

#### 4.1.3.1 Efficient Filter Flow Model

Our EFF framework presented in Chapter 2 has been proposed as an imaging model for expressing PSFs that vary smoothly across the image plane. Therefore, it seems well suited for the description of motion blur caused by camera shake. The idea is to describe a spatially varying PSF at discrete sites by a set of localised uniform filter kernels $\{f^{(r)}|r = 0 \dots R-1\}$, which are then – depending on the particular pixel position – linearly combined to yield a smoothly varying blur. The weighting coefficients for the interpolation of neighbouring filters can follow arbitrary nonlinear functions, being only restricted by the requirement that the sum of weights at each pixel has to be one.

For efficient computation the image is first partitioned into $p$ overlapping patches of size $q$, each of which is then modulated by its corresponding weighting function $w^{(r)}$. After having applied the assigned filter kernel $f^{(r)}$ at each image patch, the blurred image fragments are combined according to the overlap and add method of Stockham (1966), *i.e.*

$$y_i = \sum_{r=0}^{R-1} \sum_{j=0}^{k-1} f_j^{(r)} w_{i-j}^{(r)} \, x_{i-j} \text{ for } 0 \leq i < n \tag{4.1}$$

with $w^{(r)} \geq 0$ and $\sum_{r=0}^{R-1} w_i^{(r)} = 1$ for $0 \leq i < n$.

Although the interpolation weights $w^{(r)}$ can be arbitrary non-linear functions, the transformation itself is still *linear* in its parametrising filters $f^{(r)}$ and the image $x$, *i.e.* there exist matrices $F$ and $X$ such that $y = Fx = Xf$ in agreement with the linear system theory of incoherent imaging. Here, $f$ denotes the column vector obtained by stacking all filter kernels $\{f^{(r)}|r = 0 \dots R-1\}$. In (Hirsch et al., 2010) or Section 2.3.3 also expressions for efficient implementation of $F$ and $X$ and their transpose matrices are derived that allow computation in $O(n \log q)$ with $n$ and $q$ denoting the number of pixels in the image and each patch respectively. In summary, the EFF framework implements space-variant convolutions which are almost as efficient to compute as space-invariant convolutions, while at the same time being much more expressive.

**Discussion.** As demonstrated in Figure 4.3, the EFF framework is able to approximate real camera shake with as few as $3 \times 3$ filters. As we would expect, an increase in the number

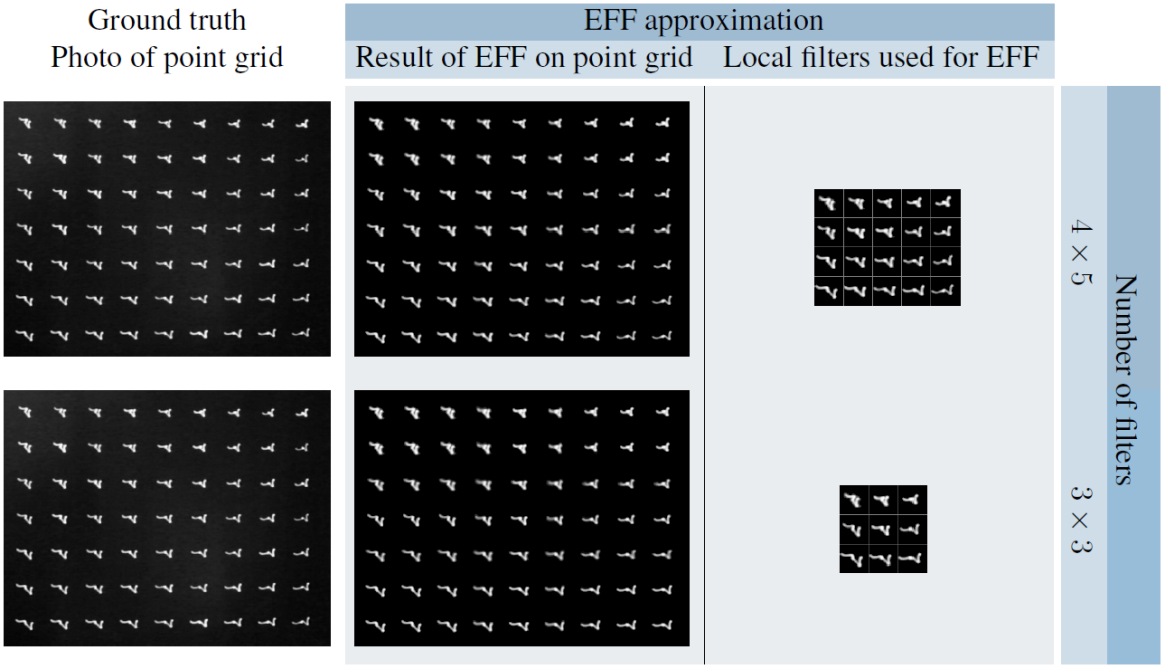



Figure 4.3: A small set of sampling sites or localised filters can parametrise smoothly varying blur: (left) grid photographed with real camera shake, (middle) regular point grid blurred by the EFF framework parametrised by $4 \times 5$ (top right) and $3 \times 3$ filters (bottom right).

of local support sites or filters used for parametrising the EFF improves the quality and accuracy of the approximation. Evidently, the higher the spatial variation of the motion blur, the more filters are needed to describe the corresponding PSF. Although only a single instance of a PSF due to camera shake is shown, the EFF model seems well suited for describing real motion blur.

In (Harmeling et al., 2010a), we extended the method of Cho and Lee (2009), who presented a very efficient and robust BD algorithm for uniform blur, by incorporating the EFF framework for being able to describe non-uniform motion blur. Although we succeeded in removing motion blur due to camera shake from several real-world images, our method performs poorly in cases where the spatial variation of the blur is too severe. This is due to the fact, that there is no global constraint on the estimated filters that parametrise the motion blur describing EFF. Despite the employment of a smoothness constraint on neighbouring filters and the introduction of a propagation step, that replaces bad estimates by their neighbouring filters, image regions with little structural information inevitably spoil PSF estimation. Another shortcoming of this approach is that the parameters of the EFF have to be adjusted manually. Admittedly, the method only works if a good trade-off between the patch size and overlap can be found, making it vulnerable and less robust.

Therefore, although being able to express camera shake, the EFF framework is too general in the sense that the space of admissible PSFs is too large. Enforcing similarity (in a L2 sense as in Harmeling et al. (2010a)) between neighbouring filters is not sufficient to rule out unphysical PSFs and to capture the smooth shape variation of real motion blur.

#### 4.1.3.2 Projective Motion Path Blur Model

The PMPB model tailored for capturing motion blur due to camera shake was first proposed in (Tai et al., 2009, 2011) and subsequently refined in (Whyte et al., 2010) and (Gupta et al., 2010). In this model, the observed blurred image is viewed as an integration of all images seen by the camera during exposure time. Under the pinhole model of a camera, all of these views can be obtained from the sharp scene by a 2D projective transformation (homography), each of which corresponds to a particular camera pose. Each camera pose is parametrised by a six-dimensional vector with three entries defining its position in space and three entries specifying its orientation. The space of possible camera poses has been named *Camera Pose Space* by Gupta et al. (2010), which we will denote by $\mathscr{P}$ in the following. While the shutter is open, the camera undergoes a sequence of different camera poses, which corresponds to a one-dimensional trajectory in Camera Pose Space (CPS). Discretising $\mathscr{P}$, the captured image $y$ is a summation of all the projectively transformed versions of the sharp image $x$, each weighted by $\mu_\theta$ corresponding to the proportion of time the camera spent in that pose, *i.e.*

$$y = \sum_{\theta \in \mathscr{P}} \mu_\theta \; x(\boldsymbol{H}_\theta \boldsymbol{u}). \tag{4.2}$$

Here, $\boldsymbol{H}_\theta$ denotes the projective transformation that warps the sharp latent image $x$ to the image seen by the camera in pose $\theta$. The coordinates in the image and object plane are related by the corresponding homography

$$\boldsymbol{H}_\theta = \boldsymbol{C} \left( \boldsymbol{R}_\theta + \boldsymbol{T}_\theta \frac{n}{d} \right) \boldsymbol{C}^{-1}, \tag{4.3}$$

where $\boldsymbol{C}$ denotes the parameter matrix of the camera intrinsics and $\boldsymbol{R}_\theta$ and $\boldsymbol{T}_\theta$ the rotation matrix and the translation vector for camera pose $\theta$, respectively; $n$ and $d$ define the normal vector of and the distance to the object plane. The set of all possible projective transformations $\{\boldsymbol{H}_\theta | \theta \in \mathscr{P}\}$ constitutes a basis which has been named the *Motion Response Basis* by Gupta et al. (2010) and which we abbreviate as $\mathscr{B}$. The Motion Response Basis (MRB) does not need to cover the entire six-dimensional CPS, but might consist of a lower dimensional subset. Whyte et al. (2010) restrict themselves to rotations only, while Gupta et al. (2010) consider in-plane translations (instead of *pitch* and *yaw*) and in-plane rotations (*roll*). Given a particular MRB, the weighting vector $\mu_\theta$ uniquely defines the motion blur by encoding the sequence of camera poses the camera passed through during exposure. Therefore, Whyte et al. (2010) calls it *blur kernel* analogue to the space-invariant case, while Gupta et al. (2010) refer to it as a *Motion Density Function*, a term which we will adopt in the following exposition.

**Discussion.** The PMPB model offers an intuitive and compact representation for space-varying motion blur due to camera shake. By modelling the camera motion explicitly, only physically meaningful blur is attained. PSF estimation is equivalent to finding the latent Motion Density Function (MDF) in camera pose space.

Despite its allure from a physical modelling point of view, this approach suffers from high computational cost. With each update of the latent image during kernel estimation as

many projective transformations as allowed camera poses (*i.e*. elements in the MRB) have to be computed, each of the order $O(n^2)$. Even for medium sized images, a computational time of several hours is reported in (Whyte et al., 2010) and (Gupta et al., 2010) for image deblurring, rendering these approaches practically unattractive especially when compared to the recently proposed fast motion deblurring algorithms for stationary motion blur (Cho and Lee, 2009; Xu and Jia, 2010), which operate in a few minutes even for Mega-pixel images.

### 4.1.4 Other Related Work

The problem of removing blur caused by space-invariant convolution, *i.e*. uniform blur has been studied extensively for a long time. Early works include those of *e.g*. Richardson (1972) and Lucy (1974), which date back to the early 70s. See (Kundur and Hatzinakos, 1996) for an overview of related methods.

For BD of single photographs, Fergus et al. (2006) combined the variational approach of Miskin and MacKay (2000) with natural image statistics (Field, 1994). Shan et al. (2008), Cho and Lee (2009) and Xu and Jia (2010) refined their approach using carefully chosen regularisation and fast optimisation techniques, see also (Levin et al., 2009) for a comprehensive overview of these and related approaches.

However, all these methods assume a uniform blur model based on space-invariant convolution, which is a severe limitation as discussed above. This motivated work on non-uniform blur models which we already discussed in detail in the previous section (Tai et al., 2009, 2011; Whyte et al., 2010; Gupta et al., 2010; Harmeling et al., 2010a).

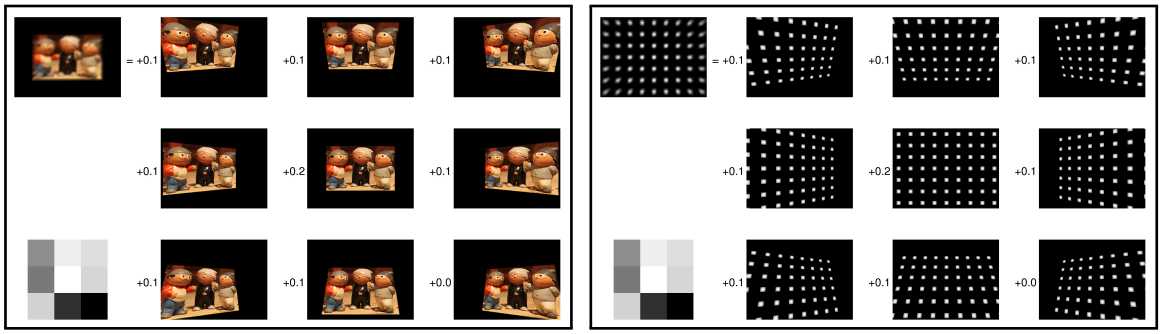
Other generalisations of the uniform blur model consider object motion instead of camera motion: Levin (2006) is able to deblur objects that move linearly, such as a bus that drives from left to right. Shan et al. (2007) focus on blurs in the image due to rotating objects, such as propellers.

Hardware approaches to obtain sharper images are based on manipulating the way images are taken, exemplary we mention: Yuan et al. (2008) reconstruct a single sharp image from a pair of blurred and noisy images. While Raskar et al. (2006) encodes the movement of objects by "fluttering" the shutter, Cho et al. (2010) is able to remove linear object motion by capturing two images of the scene with a parabolic motion in two orthogonal directions . Joshi et al. (2010) exploit inertial measurement sensor data to recover the true trajectory of the camera during exposure.

## 4.2 Fast Forward Model for Non-Uniform Camera Shake

Is there a way to combine the strengths of both approaches, *i.e*. the computational efficiency of the EFF framework and the benefit of the PMPB model arising from modelling the camera motion explicitly? In other words, is it possible to constrain the EFF such that the space of admissible PSFs corresponds to physically meaningful motion blur only?

Actually, the solution to this issue is directly related to the superposition principle of incoherent imaging: by knowing the impulse response of the optical system to point objects, the system is fully determined. Hence, by knowing how individual point objects transform



PMPB model applied to latent image PMPB model applied to regular point grid

Figure 4.4: The effect of the PMPB model (Equation 4.10) applied to a latent image (left panel) and a regular point grid (right panel) is visualised. Here we restricted the camera motion to pitch and yaw only. In this case, the Motion Density Function that specifies how much time the camera spent in a certain pose is two dimensional, depicted in the lower left corner of each panel. The result is obtained as a weighted sum of the different projections.

under motion blur, we can fully describe the imaging process. As we demonstrated in Figure 4.3, knowing the PSF at a few equidistant sites suffices for the EFF framework to model real motion blur. For this reason, all we have to study is how a point grid behaves under all possible projective transformations of the MRB, which the PMPB model readily tells us. Once we know this, we can compose the filters that parametrise the EFF as a weighted sum of the projectively transformed point grids, where the weighting corresponds to the MDF. Figure 4.4 visualises this idea and contrasts the effect of the PMPB model on both a latent image and a regular point grid. The weighted composition of the projectively transformed point grids forms the spatially varying PSF (upper left corner in right panel of Figure 4.4), whose effect on the latent image can then be computed by the EFF framework. Figure 4.5 provides a schematic overview of the proposed forward model. In other words, the set of projectively transformed point grids serves as a basis for the filters of the EFF, which allows globally consistent motion blurs only. We call this basis *Reduced Motion Response Basis*, as each of these projectively transformed point grids parametrises via the EFF an element $\boldsymbol{H}_\theta$ of the MRB. Another view point on this is that each projective transformation can be approximated by an EFF transformation. In the extreme, where we choose the patch size of the EFF to be a single pixel and take as many filters as pixels in the latent image to parametrise the EFF, we yield the MRB of Gupta et al. (2010). With having outlined the idea of our proposed fast forward model, we will provide a more formal derivation in the following section.

### 4.2.1 Derivation

The starting point of our derivation is the incoherent imaging equation that describes how an image $x(\xi, \eta) : \Sigma_1 \subset \mathbb{R}^2 \to \mathbb{R}$ is transformed by an optical system characterised by the impulse response or PSF $f(u, v; \xi, \eta)$: $\Sigma_1 \times \Sigma_2 \to \mathbb{R}$, yielding an image of the observed scene $y(u, v)$: $\Sigma_2 \subset R^2 \to \mathbb{R}$ in the image plane:

$$y(u, v) = \iint_{\Sigma_1} f(u, v; \xi, \eta)\ x(\xi, \eta)\ d\xi\, d\eta. \tag{4.4}$$

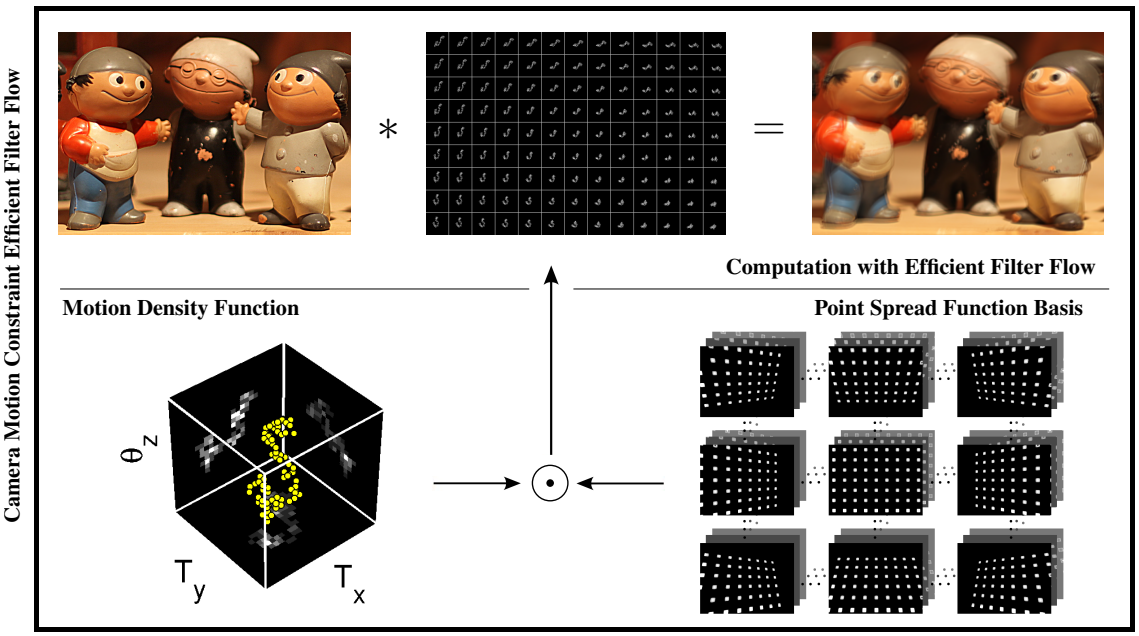



Figure 4.5: Illustration of our unified approach for modelling motion blur due to camera shake. The values of the Motion Density Function (bottom left, plotted with plot_nonuni_kernel.m from Whyte et al. (2010)) correspond to the time the camera spent in each pose. Linearly combined with the blur kernel basis (bottom right), it yields a non-uniform PSF (top middle) which parametrises the EFF transformation allowing fast computation. By construction, our forward model permits only physically plausible camera motions. The blur kernel basis has to be computed only once and allows a memory saving sparse representation. The dimensionality and size of the blur kernel basis depends on the motion considered. For translational motion only, the model reduces naturally to the uniform blur model. In this case the Motion Density Function equals the invariant PSF.

The pairs $(u, v)$ and $(\xi, \eta)$ denote the coordinates in the image and object plane, $\Sigma_1$ and $\Sigma_2$, respectively. For ease of exposition, we will abbreviate $(u, v)$ by $\boldsymbol{u}$ and $(\xi, \eta)$ by $\boldsymbol{\xi}$ in the following discussion. This equation is known as the *superposition integral* (Goodman, 2005), which states that the image of a static scene can be computed as the integral or weighted sum of point objects, *i.e.* as the weighted superposition of the object intensity $x$ and the PSF of the optical system.

**Model assumptions.** Since we want to focus on modelling the effect of camera shake on the PSF, we assume an ideal imaging system and disregard any effects of magnification and lens aberrations on the PSF in the following. In particular we assume that the PSF of a perfectly steady camera is translation invariant and a delta function, *i.e.*

$$f(\boldsymbol{u}; \boldsymbol{\xi}) = \delta(\boldsymbol{u} - \boldsymbol{\xi}) \equiv \delta(x - \xi, y - \eta). \tag{4.5}$$

In addition to the assumption of an ideal optical system, we assume that the underlying scene is static, *i.e.* $x \neq x(t)$ where $t$ denotes time. In other words, we consider no *object motion*

but restrict ourselves to *ego* or *camera motion* only. For such an *ideal*, *i.e*. aberration-free optical system, the object and the image are identical, because

$$y(\boldsymbol{u}) = \int_{-\infty}^{\infty} \delta(\boldsymbol{u} - \boldsymbol{\xi})\, x(\boldsymbol{\xi})\, d\boldsymbol{\xi} = x(\boldsymbol{u}). \tag{4.6}$$

Hence, each point in the object plane is mapped to a single point in the image plane, yielding a sharp image of the static scene being captured.

**Derivation of PMPB model.** When taking a photo, we expose the image sensor of the camera to the light emitting scene only for a certain period of time. Typically, if we take a photo under low light conditions or want to capture faint objects as *e.g*. in astronomy, we will choose a longer exposure time $T$[1]. While the shutter is open during exposure, each photosensitive element of the sensor or pixel accumulates photons arriving from the object plane through the optics on the camera sensor. However, if the camera and therewith the image sensor undergoes any motion during that time, a light ray emitted from a single point from the object plane will not converge to a single point in the image plane any more. Instead, its photons will hit and activate a trace of neighbouring pixels yielding a blurred image.

In general, camera shake can be viewed as moving the camera along some three-dimensional trajectory in space during exposure. Since at each position the camera can point in any direction, the full camera motion is described by a six-dimensional curve as already mentioned above. At each instance in time $t$ with $0 \leq t \leq T$, each point $\boldsymbol{u}$ in the object plane is mapped by some homography $\boldsymbol{H}(t)$ to a point $\boldsymbol{\xi}(t) = \boldsymbol{H}(t)\,\boldsymbol{u}$ in the image plane. By integrating along that trajectory we obtain the PSF that describes the camera shake, *i.e*.

$$f(\boldsymbol{u}; \boldsymbol{\xi}) = \int_0^T \delta(\boldsymbol{H}(t)\,\boldsymbol{u} - \boldsymbol{\xi})\; dt. \tag{4.7}$$

Note, that (4.7) is no longer spatially invariant in contrast to (4.5). Plugging expression (4.7) into Equation (4.4), integrating out the delta function w.r.t. $\boldsymbol{\xi}$, and discretising we obtain

$$y(\boldsymbol{u}) = \int_0^T x(\boldsymbol{H}(t)\,\boldsymbol{u})\; dt \;\approx \sum_{t=0}^{T} x(\boldsymbol{H}_t \boldsymbol{u}), \tag{4.8}$$

which is the forward model of Tai et al. (2009).

**Derivation of time-agnostic model.** The Riemannian integral in Equation (4.7) can be equivalently expressed as a Lebesgue integral,

$$f(\boldsymbol{u}; \boldsymbol{\xi}) = \int_{\mathscr{P}} \delta(\boldsymbol{H}(\theta)\,\boldsymbol{u} - \boldsymbol{\xi})\; d\mu(\theta), \tag{4.9}$$

[1]Of course, another option would be to increase the sensitivity of the sensor by choosing a higher ISO setting. However, this often leads to a undesirable noise in the recorded photo. There is a subtle trade-off between having to accept noise or blur when taking a photo under low light conditions. For a detailed discussion we refer to (Hasinoff, 2008).

where all (now pair-wise different) homographies $\boldsymbol{H}(\theta)$ are parametrised by $\theta$ (and not by a time index $t$) and the Lebesgue measure $d\mu(\theta)$ weights the contributions of the different homographies according to the amount of time the camera spent in the corresponding pose during exposure. The integration domain is the entire CPS. Plugging Equation (4.9) into Equation (4.4), integrating the delta function w.r.t. $\boldsymbol{\xi}$, and discretising we obtain

$$y(\boldsymbol{u}) = \int_{\mathscr{P}} x(\boldsymbol{H}(\theta)\,\boldsymbol{u})\; d\mu(\theta) \;\approx \sum_{\theta\in\mathscr{P}} \mu_\theta\; x(\boldsymbol{H}_\theta \boldsymbol{u}), \tag{4.10}$$

where we discretised CPS and $\mu_\theta$ is the weight of the corresponding homography. This is the forward model of Whyte et al. (2010) and Gupta et al. (2010). Whyte et al. (2010) calls the imaging model Equation (4.10) *time-agnostic* since all temporal information has been eliminated.

**PSF basis.** Despite its plausibility from a physical modelling point of view, the PMPB model (Tai et al., 2009; Whyte et al., 2010; Gupta et al., 2010) suffers from high computational cost since as many homographies as considered camera poses need to be computed for each forward modelling, as discussed above and sketched in the left panel of Figure (4.4). For realistic PSF sizes, the number of projective transformations is in the order of ten thousands, leading to a computation time in the order of minutes for realistic image sizes, even with compiled C code[2], which renders this approach practically unattractive. As outlined above, one possible solution to this issue is to decouple the image from the costly computation of its projectively transformed versions by studying the effect of the camera shake on a regular point grid. To this end, we first decompose the sharp image $x(\boldsymbol{\xi})$ into overlapping patches

$$x(\boldsymbol{\xi}) = \sum_{r=0}^{R-1} w^{(r)}(\boldsymbol{\xi})\, x(\boldsymbol{\xi}), \tag{4.11}$$

where the finite set $\{w^{(r)} | r = 0, \ldots, R-1\}$ is a partition of unity, *i.e.* $\sum_{r=0}^{R-1} w^{(r)}(\boldsymbol{\xi}) = \mathbf{1}$ and each $w^{(r)} \colon \Sigma_1 \to \mathbb{R}$ is a weighting function. In particular, we choose the partition to be a regular tiling over the integration domain such that the centre of each weighting mask $w^{(r)}$ lies on a regular grid and selects a small image region of $x$ only, thereby setting everything else to zero. Furthermore, we choose the $w^{(r)}$ to be smooth functions without any discontinuities or abrupt changes. To each image patch we can apply a convolution with a local delta function $\delta^{(r)}(\boldsymbol{\xi} - \boldsymbol{\xi}')$ centred at the support of each $w^{(r)}$ without actually changing anything. Hence, we obtain the identity:

$$x(\boldsymbol{\xi}) = \sum_{r=0}^{R-1} \int_{\Omega_r} \delta^{(r)}(\boldsymbol{\xi} - \boldsymbol{\xi}')\, w^{(r)}(\boldsymbol{\xi}')\, x(\boldsymbol{\xi}')\, d\boldsymbol{\xi}', \tag{4.12}$$

where $\Omega_r$ denotes the support of $w^{(r)}$, *i.e.* $\Omega_r = \operatorname{supp}\left(w^{(r)}\right)$[3]. Plugging this identity and expression (4.9) into Equation (4.4), then swapping integrals and summation, and integrating

[2]The blurring of an image of size $1200 \times 1600$ pixels with a PSF size of $21 \times 21$ pixels, takes about 180s with the compiled C code of Whyte et al. (2010).

[3]See (Königsberger, 2002, 359ff) for integration by using a partition of unity.

the delta peaks w.r.t. $\boldsymbol{\xi}$ we get

$$
\begin{aligned}
y(\boldsymbol{u}) &= \int_{\Sigma_1} \underbrace{\int_{\mathscr{P}} \delta(H(\theta)\,\boldsymbol{u} - \boldsymbol{\xi}) d\mu(\theta)}_{=f(\boldsymbol{u};\,\boldsymbol{\xi})} \underbrace{\sum_{r=0}^{R-1} \int_{\Omega_r} \delta^{(r)}(\boldsymbol{\xi} - \boldsymbol{\xi}')\, w^{(r)}(\boldsymbol{\xi}')\, x(\boldsymbol{\xi}')\, d\boldsymbol{\xi}'}_{=x(\boldsymbol{\xi})}\, d\boldsymbol{\xi} \\
&= \sum_{r=0}^{R-1} \int_{\Omega_r} \int_{\mathscr{P}} \delta^{(r)}(H(\theta)\,\boldsymbol{u} - \boldsymbol{\xi}')\, d\mu(\theta)\; w^{(r)}(\boldsymbol{\xi}')\, x(\boldsymbol{\xi}')\, d\boldsymbol{\xi}' \\
&= \sum_{r=0}^{R-1} \int_{\Omega_r} \int_{\mathscr{P}} \underbrace{\delta^{(r)}(H(\theta)\,\boldsymbol{u} - \boldsymbol{\xi}')}_{\equiv b_\theta^{(r)}(\boldsymbol{u};\boldsymbol{\xi}')}\, d\mu(\theta)\; w^{(r)}(\boldsymbol{\xi}')\, x(\boldsymbol{\xi}')\, d\boldsymbol{\xi}' \\
&= \sum_{r=0}^{R-1} \int_{\Omega_r} f^{(r)}(\boldsymbol{u};\boldsymbol{\xi}')\, w^{(r)}(\boldsymbol{\xi}')\, x(\boldsymbol{\xi}')\, d\boldsymbol{\xi}', \qquad (4.13)
\end{aligned}
$$

where the newly introduced function $b_\theta^{(r)}(\boldsymbol{u};\boldsymbol{\xi}')$ form the basis for PSFs due to camera shake. Linearly combined according to the measure $d\mu(\theta)$, they constitute the local filters $f^{(r)}(\boldsymbol{u};\boldsymbol{\xi}')$ that are patch-wise convolved with their corresponding image patches. Equation (4.13) allows efficient computation via the EFF framework, which will become more evident after discretising the involved quantities.

**Discretisation and EFF.** In practice, a picture $y$ taken with digital cameras is a matrix of sensor values, *i.e.* it is a discrete object. Without loss of generality we can consider $y$, $x$, $f$, and $w^{(r)}$ as vectors with entries $y_i$, $x_j$, $f_{i,j}$ and $w_j^{(r)}$, where we replaced $\boldsymbol{u}$ by $i$ and $\boldsymbol{\xi}$ by $j$. Similarly, discretising CPS will yield a vector $\mu$ with entries $\mu_\theta$. Assuming the PSF to have locally bounded support, we can also represent $b_\theta^{(r)}$ as finite-length vectors with entries $b_{\theta;i,j}^{(r)}$. This allows us to rewrite Equation (4.13) in its discrete form:

$$
y_i = \sum_{\theta,r,j} \mu_\theta\; b_{\theta;i,j}^{(r)}\; w_j^{(r)}\; x_j. \qquad (4.14)
$$

We note that this equation is linear in $x$ and also in $\mu$, *i.e.* we can define matrices $M$ and $F$, such that

$$
y \equiv \mu \diamond x = M\; x = F\; \mu, \qquad (4.15)
$$

where we introduced the symbol $\diamond$ for denoting our proposed fast forward model.

Equation (4.14) is an instance of an EFF. Therefore, analogously to Section 2.3.3, the forward model in Equation (4.14) can be written in matrix notation as

$$
y = \underbrace{Z_y^{\mathsf{T}} \sum_r E_r^{\mathsf{T}} \mathcal{F}^{\mathsf{H}} \operatorname{diag}\Big(\mathcal{F} Z_b B^{(r)} \mu\Big) \mathcal{F} C_r \operatorname{diag}(w^{(r)})}_{M}\, x, \qquad (4.16)
$$

where $B^{(r)}$ is the matrix with column vectors $b_\theta^{(r)}$ for varying $\theta$, *i.e.* $B^{(r)}\mu = \sum_\theta \mu_\theta b_\theta^{(r)}$. $C_r$ and $E_r$ are cropping matrices, $\mathcal{F}$ a Fourier transform matrix, and $Z_b$ and $Z_y$ are zero-padding

matrices. We see that as a linear map of the sharp image $x$, we can sum up the basis blur kernels $b_\theta^{(r)}$ for each $r$ and then apply the overlap-add trick for fast computation. Similarly we can write $y$ as a linear function of $\mu$,

$$y = \underbrace{Z_y^\mathsf{T} \sum_r E_r^\mathsf{T} \mathcal{F}^\mathsf{H} \operatorname{diag}\Big(\mathcal{F} C_r \operatorname{diag}(w^{(r)}) x\Big) \mathcal{F} Z_b B^{(r)}}_{F} \mu. \tag{4.17}$$

Similarly to Eqs. (4.16) and (4.17), we can obtain fast implementations of the MVMs with $M^\mathsf{T}$ and $F^\mathsf{T}$ which are needed for employing gradient-based optimisation techniques. All formulae for vector-valued images can be straightforwardly generalised to matrix-valued images.

### 4.2.2 Discussion

We succeeded in combining the structural constraints of the PMPB models and the efficiency of the EFF framework to obtain a fast forward model that is able to describe non-uniform blur caused by camera shake in a globally consistent way.

Its benefit becomes evident when inspecting Figure (4.6), where we compare the runtime of our forward model in dependence of both the image and blur size to Whyte et al. (2010). There, the computation of a forward model consists of making $d$ homographies on an image with $n$ pixels, which means a complexity of $O(n \cdot d)$. Since our model uses the EFF, the complexity is $O(n \cdot \log q)$ with the number $q$ of pixels in a patch (Harmeling et al., 2010a), which depends on the image and PSF sizes. The disadvantage in $\log q$ is easily outweighed even for a small number of homographies, leading to a speedup of up to several magnitudes.

Furthermore, Figure 4.7 shows that our fast forward model can approximate the non-stationary blur of Whyte and Gupta almost perfectly with as little kernels as $16 \times 12$ for an image of size $1600 \times 1200$ pixels.

In summary, the advantage of our approach is twofold:

- The Reduced MRB corresponds to a set of projectively transformed point grids that serves as a basis for the filters that parametrise the EFF instead of transformation matrices as required for the MRB in Gupta et al. (2010). The basis can be computed *in advance*, which Gupta et al. (2010) also does. However, instead of having to save $|\mathscr{B}|$ *transformation matrices* of the size $n^2$ ($n$ being the size of the image), we only have to save $|\mathscr{B}|$ *images* of the size $R \times k$ where $R$ is the number of filters to parametrise the EFF and $k$ the size of a single filter. Furthermore, our basis lends itself to a memory-saving sparse representation.

- Instead of computing the effect of *each* projective transformation and then adding the resulting images to form the blurred image, we first sum up the elements of the Reduced MRB to form the filters that parametrise the EFF and then compute its effect on the image by a *single* EFF transformation which can be done almost as efficiently as an ordinary convolution.

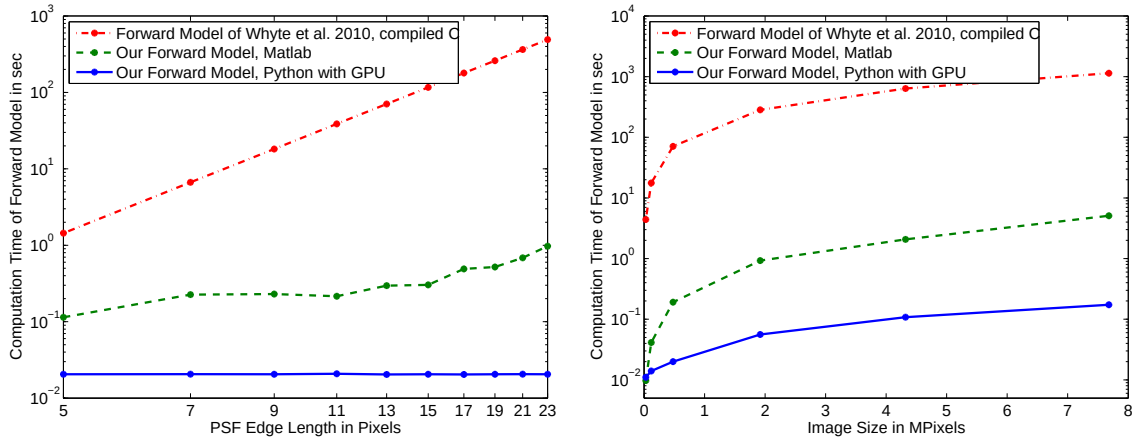



Figure 4.6: Run-time comparison of our forward model with the blurring model of (Whyte et al., 2010; Gupta et al., 2010) as a function of PSF (top) and image size (bottom). For an image of size $1600 \times 1200$ pixels our Matlab implementation is a factor 40 faster than the compiled C code of Whyte et al. (2010). Note that for fair comparison computation was performed on a single core machine as our Matlab implementation is able to take advantage of a multicore architecture by parallel computation while the implementation of Whyte et al. (2010) does not. A factor of 1000 can be gained by our Python implementation supporting GPU computation.

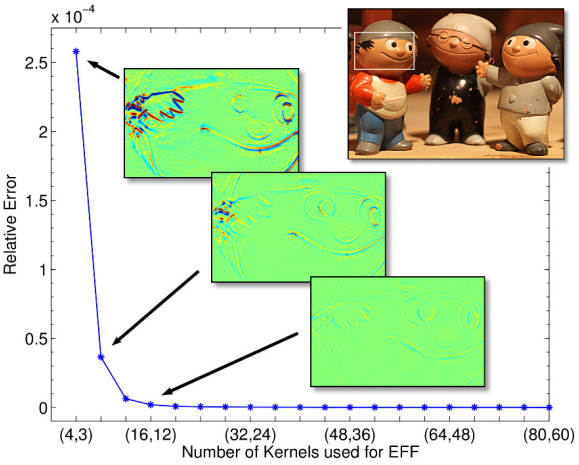



Figure 4.7: The curve shows the relative error of a homographically transformed image ($1600 \times 1200$ pixels) using the forward model of Whyte et al. (2010) and our fast forward model which approximates the homography by the camera motion constrained EFF framework. For some data points closeups of the difference images are shown. The relative error decreases the more kernels are used. With as little as $16 \times 12$ kernels the error is negligible.

## 4.3 Two-Phase Motion Deblurring

Given a blurred photograph taken with a handheld camera, we recover the unknown sharp image in two steps: (1) we analyse the motion blur and its corresponding non-uniform PSF, and (2) we recover the latent image by non-blind deconvolution that is tailored for spatially varying blur and promoting natural image statistics. First, we will give a brief overview of our algorithm, followed by a detailed description of both the blur estimation and the sharp image recovery phase.

### 4.3.1 Overview

**1. Blur estimation phase:**

Initialising $x$ with the blurry image $y$, the estimation of the camera shake blur parameters $\mu_\theta$, is performed by iterating over the following three steps from a coarse to fine image scale:

**(i) Prediction step:**

- remove noise in flat regions of $x$ by edge-preserving *bilateral filtering* and enhance edges by *shock filtering* following Cho and Lee (2009).
- compute gradient selection mask via *rmap approach* of Xu and Jia (2010) to use informative edges for motion blur estimation only. In particular, it neglects structures that are smaller in size than the local filters, which could be misleading for the blur parameter estimation.

**(ii) Blur parameter estimation step:**

- update the blur parameters given the blurry image $y$ and the current estimate of the predicted $\tilde{x}$ obtained by the prediction step (i).
- for a preconditioning effect use the gradient images of $x$ only (Cho and Lee, 2009).
- enforce smoothness of camera trajectory

$$\left\|\partial y - m_S \odot \partial(\mu \diamond \tilde{x})\right\|_2^2 + \lambda\left\|\mu\right\|_2^2 + \eta\left\|\partial\mu\right\|_2^2, \tag{4.18}$$

where $m_s$ is a mask (computed by *rmap approach*), that weights gradients according to their information content (see previous step). The regularisation constants $\lambda$ and $\eta$ balance the likelihood against the prior terms. The above optimisation problem is efficiently solved by gradient-based optimisation techniques (*e.g.* lbfgsb of Zhu et al. (1997) or Barzilai and Borwein (1988)).

**(iii) Latent image estimation step:**

- update the current deblurred image $x$ by solving a least-squares cost function using a smoothness prior on the gradient image via direct deconvolution

$$\|y - \mu \diamond x\|_2^2 + \alpha\|\partial x\|_2^2 \tag{4.19}$$

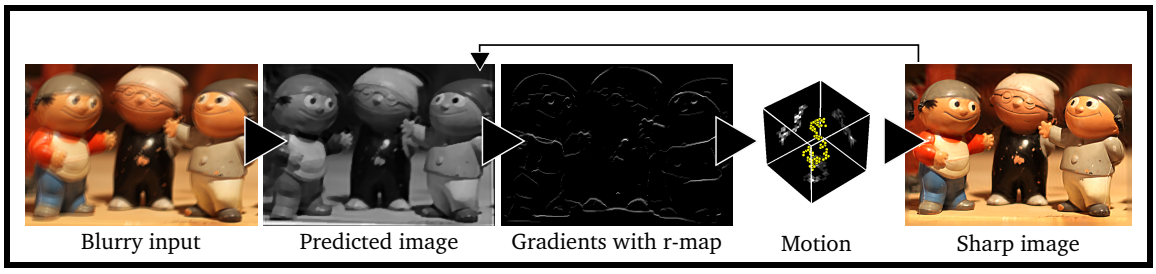



Figure 4.8: Overview of the blur estimation phase at a fixed image scale: the prediction step yields an enhanced version of the blurry input image by non-linear filtering. Via the rmap approach of Xu and Jia (2010) only informative edges are selected for the blur estimation step. After an estimate of the spatially varying PSF has been obtained, a latent image is computed via direct deconvolution. See text for details.

**2. Image Recovery Phase**

Non-blind deblurring (following Krishnan and Fergus, 2009): given the EFF parametrised by $\mu$, we yield the final image estimate by alternating between the following two steps:

- **Latent variable estimation:** estimate latent variables regularised with a sparsity prior that approximate the gradient of $x$. This can be efficiently solved with look-up tables as well as analytically, referred to as "w sub-problem" by Krishnan and Fergus (2009).
- **Image estimation step:** update the current deblurred image $x$ by directly solving a least-squares cost function while penalising the Euclidean norm of the gradient image to the latent variables of the previous step, referred to as "x sub-problem" by Krishnan and Fergus (2009).

### 4.3.2 Blur Estimation Phase

In the first phase of the algorithm, we try to recover the motion undertaken by the camera during exposure given the blurry photo only. To this end, we iterate the following three steps: (i) *prediction step* to reduce blur and enhance image quality by a combination of shock and bilateral filtering, (ii) *blur parameter estimation step* to find the MDF, which best explains the blurry picture from the predicted image of step (i), and (iii) *latent image estimation* via non-blind deconvolution.

To avoid local minima and make the blur estimation in the first phase more robust, we employ a coarse to fine approach (over several scales). In the beginning, the resolution of the recorded image $y$ is reduced and the blur estimation phase is performed. Then the lower resolved estimate of the spatially varying PSF initialises the next scale, and so on, up to the full resolution of the recorded image. At the coarsest scale we initialise the unknown sharp image $x$ by a down-sampled version of the blurry image $y$. For a fixed scale we iterate steps (i)-(iii), which are visualised in Figure 4.8 and will be detailed in the following, five times.

**(i) Prediction step:** The prediction step emphasises edges in $x$ by shock filtering (Osher and Rudin, 1990) and lowers the importance of flat noisy regions by edge-preserving bilateral filtering (Tomasi and Manduchi, 1998). This preprocessing step was first proposed by Money and Kang (2008) and subsequently refined by Cho and Lee (2009). It is a clever trick to replace expensive nonlinear optimisations which would otherwise be necessary if the same image features emphasised by the nonlinear filtering (*i.e.* shock and bilateral filtering) would have to be implemented by some image prior on $x$. Our blur estimation phase makes also use of this trick and we set all hyper-parameters exactly as Cho and Lee (2009).

Based on the blurred image $y$, we compute for each scale a weighting mask $m_S$ that selects only edges that are informative and facilitate kernel estimation. In particular, it neglects structures that are smaller in size than the local kernels, which could be misleading for kernel estimation as pointed out by Xu and Jia (2010). For computing $m_S$ we employ their r-map approach detailed in (Xu and Jia, 2010).

**(ii) Blur parameter update step:** The blur parameters $\mu$ are updated by minimising

$$\left\|\partial y - m_S \odot \partial(\mu \diamond \tilde{x})\right\|_2^2 + \lambda\left\|\mu\right\|_2^2 + \eta\left\|\partial\mu\right\|_2^2, \tag{4.20}$$

where we write the discrete derivative of $y$ symbolically as $\partial y$, *i.e.* $\partial y = [1, -1]^\mathsf{T} * y$. For matrix-valued images we consider the horizontal and vertical derivatives. Furthermore, $\tilde{x}$ denotes the outcome of the bilateral and shock filtering and $m_S$ is the *r-map* computed in the previous step (i).

The terms in Equation (4.20) can be motivated as follows: The first term is proportional to the log-likelihood, $\left\|\partial y - m_S \odot \partial(\mu \diamond \tilde{x})\right\|_2^2$ if we assume additive Gaussian noise. Considering the derivatives of $y$ and $\mu \diamond x$ brings several benefits: First, Shan et al. (2008) have shown that such terms with image derivatives help to reduce ringing artifacts by putting weight on edges. Secondly, it lowers the condition number of the optimisation problem Equation (4.20) and hence leads to faster convergence (Cho and Lee, 2009). The second summand $\left\|\mu\right\|_2^2$ penalises the $L_2$ norm of $\mu$ and helps to avoid the trivial solution by suppressing high intensity values in $\mu$. The third term $\left\|\partial\mu\right\|_2^2$ enforces smoothness of $\mu$, and thus favours connectedness in camera motion space, see also Gupta et al. (2010).

**(iii) Sharp image update step:** The sharp image estimate $x$ that is repeatedly updated during the blur estimation phase does not need to recover the true sharp image perfectly. However, it should guide the PSF estimation during the alternating updates, *i.e.* steps (i), (ii), and (iii). Since most computational time is spent in this first phase, the sharp image update step should be fast. This motivates to employ $L_2$ based regularisation terms for the sharp image, even though the resulting estimates might show some ringing and possibly other artifacts (which are dealt with in the prediction step). Thus we would like to minimise

$$\left\|y - \mu \diamond x\right\|_2^2 + \alpha\left\|\partial x\right\|_2^2 \tag{4.21}$$

with respect to $x$.

Cho and Lee (2009) gained large speed-ups for this step by replacing the iterative optimisation in $x$ by a pixel-wise division in Fourier space. They showed that such a non-iterative update step despite its known restoration artifacts is sufficient to guide the PSF estimation. We call such a pixel-wise division in Fourier space *Direct Deconvolution* (DD) and provide a similar update for our fast forward model for camera shake.

First, we adapt the matrix expression given in Hirsch et al. (2010) to obtain an explicit expression for $M$ introduced in Section 4.2,

$$y = \underbrace{Z_y^{\mathsf{T}} \sum_r E_r^{\mathsf{T}} \mathcal{F}^{\mathsf{H}} \operatorname{diag}\Big(\mathcal{F} Z_b B^{(r)} \mu\Big) \mathcal{F} C_r \operatorname{diag}(w^{(r)})}_{M} x, \tag{4.22}$$

where $B^{(r)}$ is the matrix with column vectors $b_\theta^{(r)}$ for varying $\theta$, *i.e.* $a^{(r)} = B^{(r)}\mu = \sum_\theta \mu_\theta b_\theta^{(r)}$, see also Equation (4.16). Matrices $C_r$ and $E_r$ are appropriately chosen cropping matrices, $\mathcal{F}$ is the discrete Fourier transform matrix, and $Z_b$ and $Z_y$ are appropriate zero-padding matrices. Furthermore, we denote by $\operatorname{diag}(v)$ the diagonal matrix with vector $v$ along its diagonal.

The basic idea for a direct update step of the image estimate is to combine the patch-wise pixel-wise divisions in Fourier space with reweighting and edge fading to minimise ringing artifacts. We use the following expression to approximately "invert" our forward model $y = Mx$:

$$\hat{x} \approx \operatorname{diag}(v) \sum_r \operatorname{diag}(w^{(r)})^{1/2} C_r^{\mathsf{T}} \mathcal{F}^{\mathsf{H}} \frac{\overline{\mathcal{F} Z_b B^{(r)} \mu} \odot (\mathcal{F} E_r \operatorname{diag}(w^{(r)})^{1/2} Z_y y)}{|\mathcal{F} Z_b B^{(r)} \mu|^2 + \frac{1}{2}|\mathcal{F} Z_l l|^2} \tag{4.23}$$

where $|z|$ for a vector $z$ with complex entries calculates the entry-wise absolute value, and $\overline{z}$ the entry-wise complex conjugate. The square root is taken pixel-wise. The term $\operatorname{diag}(v)$ is some additional weighting which we experimentally justify in the next paragraph. The fraction has to be implemented pixel-wise. The term $|\mathcal{F} Z_l l|^2$ in the denominator of the fraction originates from the regularisation in Equation (4.21) with $l = [-1, 2, -1]^{\mathsf{T}}$ corresponding to the discrete Laplace operator.

Note that the update formula in Equation (4.23) approximates the true sharp image $x$ given the blurry photograph $y$ and the blur parameters $\mu$ and can be implemented efficiently by reading it from right to left. The image rightmost in Figure 4.9 demonstrates how well Equation (4.23), *i.e.* direct deconvolution, *without* the additional weighting term (*i.e.* $v = 1$) approximates the true image, but also reveals artifacts stemming from the windowing. By applying the additional weighting term $v$, these artifacts can be suppressed effectively, as can be seen in the middle panel of Figure 4.9. The weighting $v$ is computed by applying Equation (4.23) *without* the additional weighting term to a constant image of the same size as the blurred image g. The deconvolved constant image reveals the same artifacts as present in the rightmost image of Figure 4.9. By taking its inverse pixel-wise, it serves as a corrective weighting term, which is able to remove most artifacts caused by the windowing and at the same time is fast to compute.

True image DD with corrective weighting DD w/o corrective weighting

Figure 4.9: Direct Deconvolution with and without corrective weighting for the blurred image shown in Figure 4.5. Note the artifacts stemming from improper treatment of overlapping parts which can be minimised by appropriate corrective weighting.

### 4.3.3 Sharp Image Recovery Phase

After having estimated and fixed the blur parameters $\mu$, we recover the final sharp image $x$ by replacing the $L_2$ image prior of the sharp image update step (4.21) by a natural image prior that is based on sparsity of the gradient images (*e.g.* Fergus et al. (2006)), *i.e.* we minimise

$$\left\|y - \mu \diamond x\right\|_2^2 + \nu \left\|\partial x\right\|_\alpha^\alpha, \tag{4.24}$$

where the $L_\alpha$ term represents a natural image prior for some $\alpha \leq 1$.

To minimise Equation (4.24), we adapt the approach of Krishnan and Fergus (2009) for stationary non-blind deconvolution in the non-stationary case: after introducing the auxiliary variable $v$ we alternatingly minimise

$$\min \left\|y - \mu \diamond x\right\|_2^2 + 2^t \left\|x - v\right\|_2^2 + \nu \left\|v\right\|_{2/3}^{2/3} \tag{4.25}$$

in $x$ and $v$. Note that the weight $2^t$ increases from 1 to 256 during nine alternating updates in $f$ and $v$ for $t = 0, 1, \ldots, 8$. Choosing $\alpha = 2/3$ allows an analytical formula for the update in $v$, see Krishnan and Fergus (2009) for details.

## 4.4 Empirical Evaluation

In this section, we will first present a method for simultaneous capture of a blurry image taken with a handheld camera as well as its corresponding motion blur. This will allow us to evaluate the *accuracy* of our proposed blur estimation method. Subsequently, we will evaluate the *performance* of our proposed algorithm in numerous real-world examples and compare against state-of-the-art in the field. Technical details about the experiments such as run-time, problem size and parameter settings will be given in Section 4.5.

### 4.4.1 Controlled Experiments

#### 4.4.1.1 Experimental Setup

To validate our BD algorithm on real-world data we followed the approach of Levin et al. (2009) who collected blurred data with ground truth for the case of uniform blur. In (Levin et al., 2009) it is already demonstrated that the spatially uniform blur model of most algorithms is an unrealistic assumption, a finding that we found confirmed in our own experiments. Since our approach drops this assumption, we are not restricted to in-plane translations. Hence, we were able to take pictures with a hand-held camera without the need of constraining the camera's motion. To obviate the effects of depth variation we captured images of a planar scene. For the reconstruction of the blur kernel, Levin et al. (2009) used a sharp version of the scene (captured by mounting the camera on a tripod) and solved a non-negatively constrained least-square problem to retain the invariant PSF of the underlying camera motion. Since in the case of non-uniform blur this optimisation problem is no longer well-defined, we pursue a different strategy to take record of the image blur.

Instead, we take advantage of the information capacity of colour images. We overlay a high-resolution gray-scale image that we store in the red colour channel of a RGB image by an equi-spaced grid of delta peaks embedded in the blue colour channel of the same image. The green colour channel is filled with zeros. Since a printout would destroy information of the gray-scale image at the locations of the delta peaks due to subtractive colour mixing, we display the image on a 20" computer monitor. To prevent the presence of Moiré pattern in the captured photo, the image scale has to be chosen such that the discrete structure of the computer screen cannot be resolved by the (discrete) image sensor of the camera. We also verified that the spectral characteristics of the screen and the camera's Bayer array filters are such that there is no cross-talk, *i.e.* the blue PSFs are not visible in the red colour image. Figure 4.10 shows the whole process.

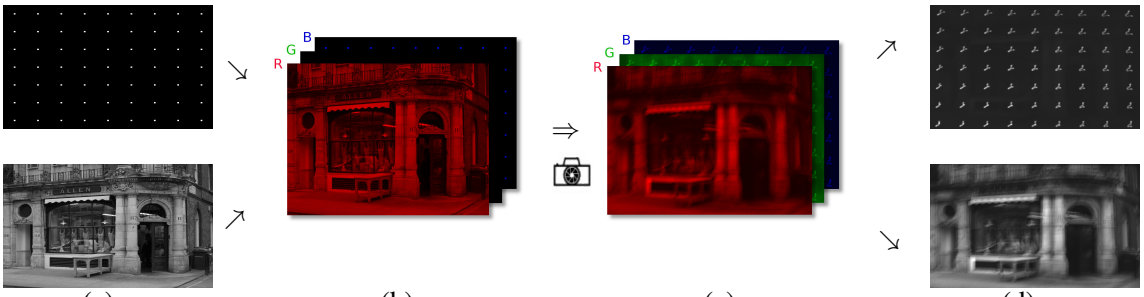



Figure 4.10: How to simultaneously capture an image blurred with real camera shake and its space-varying PSF; (a) the true image and a grid of dots is combined to (b) an RBG image, that is (c) photographed with camera shake, and (d) split into blue and red channel to separate the PSF depicting the blur and the blurred image.

#### 4.4.1.2 Experiments

We applied our method to three examples shown in Figures 4.11, 4.12 and 4.13, for which we captured the blurred image in the red channel, and true spatially varying PSFs in the blue channel, as explained in the previous subsection. For all experiments, photos were taken with a *hand-held* Canon EOS 1000D digital single lens reflex camera with a zoom lens (Canon zoom lens EF 24-70 mm 1:2.8 L USM). The exposure time was 1/4 seconds and the distance to the screen was about 2 meters. The input to the deblurring algorithm was the red channel of the RAW file only which we treat as it were a captured gray-scale image.

To validate and assess the accuracy of estimating the camera constrained EFF (*i.e.* of step (i) in Section 4.3) we compare our estimated PSFs evaluated on a regular grid of dots to the true blur which has been recorded in the blue colour channel of the RAW image during the camera shake.

Our experiments show that our approach is able to estimate and remove non-uniform motion blur from a single photo. The close similarity of the estimated PSFs to the ground truth is striking and provides strong validation for our approach. Furthermore, the recovered images reveal much greater detail and the high quality of the reconstructions is reflected by the fact that only little restoration artifacts are present in the final deconvolved images.

To appreciate the benefit of our non-uniform blur model, we compare with the method of Cho and Lee (2009) which we consider a state-of-art method for single image BD for space-invariant blurs. It is evident that the assumption of uniform motion blur is too restrictive for most real world camera shake as discussed in Section 4.1, experimentally verified with our hardware setup, and also discussed by Levin et al. (2009).

The advantage of constraining our EFF based forward model to physically plausible camera motions only, will be discussed in the following section when comparing the results of the proposed method with Harmeling et al. (2010a).

### 4.4.2 Comparative Evaluation

In the following, we show results on several challenging example images taken from the literature and do a comprehensive comparison against state-of-the-art algorithms for single image blind deblurring. We compare against both algorithms assuming uniform as well as non-uniform blur.

#### 4.4.2.1 Comparison with Harmeling et al. (2010a)

The benefit and gain in both restoration quality and modelling accuracy that is obtained by constraining the EFF to physically plausible camera motions only, is evident from Figs. 4.11, 4.12 and 4.13. While the local estimation of the EFF filters as suggested in Harmeling et al. (2010a) faces difficulties in regions with little edge information (as *e.g.* sky), the camera motion constrained EFF framework is able to use the textural information within the entire image to infer a globally consistent PSF. The benefit is not only apparent in the *Elephant* example, where structureless image regions (sky) impede motion blur estimation, but also in the subsequent example of the *Vintage Car* (paving in the bottom left image part). The

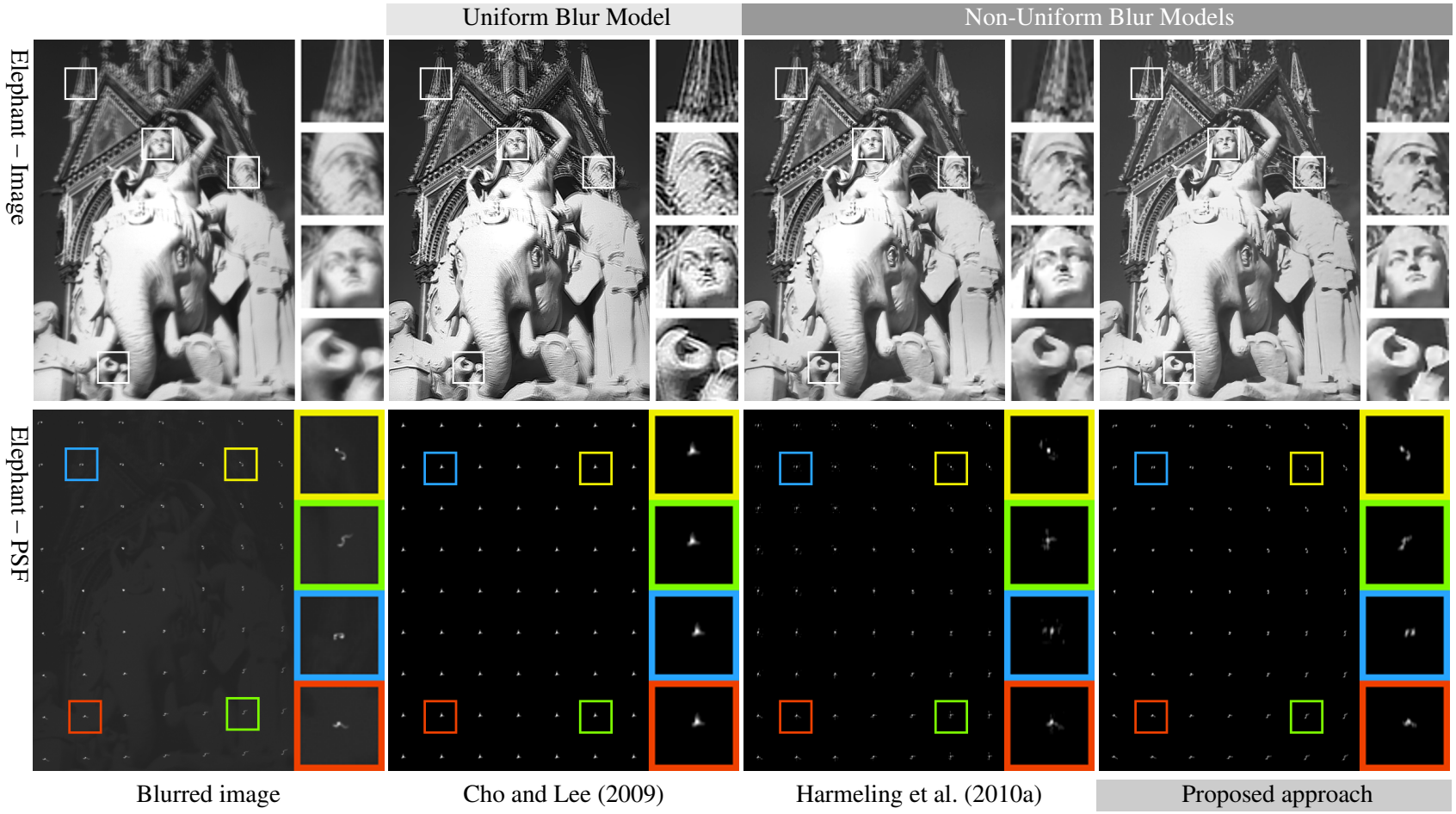



Figure 4.11: Controlled experiment: The left column shows the red and blue colour channels of a photo taken of a LCD screen with a handheld camera. While the red colour channel of the image that was displayed on the screen encoded a normal gray-scale image (top), the blue colour channel contained a point grid (bottom), which captured the motion blur during exposure and serves as a ground truth. Cho and Lee (2009) assume a uniform blur model which fails to describe the spatial variation of the PSF. In contrast, the EFF based method presented in Harmeling et al. (2010a) is able to do so, however fails in regions with little edge information (sky). The approach proposed in this thesis surpasses this limitation by constraining the EFF to allow for physically plausible motion blur only.

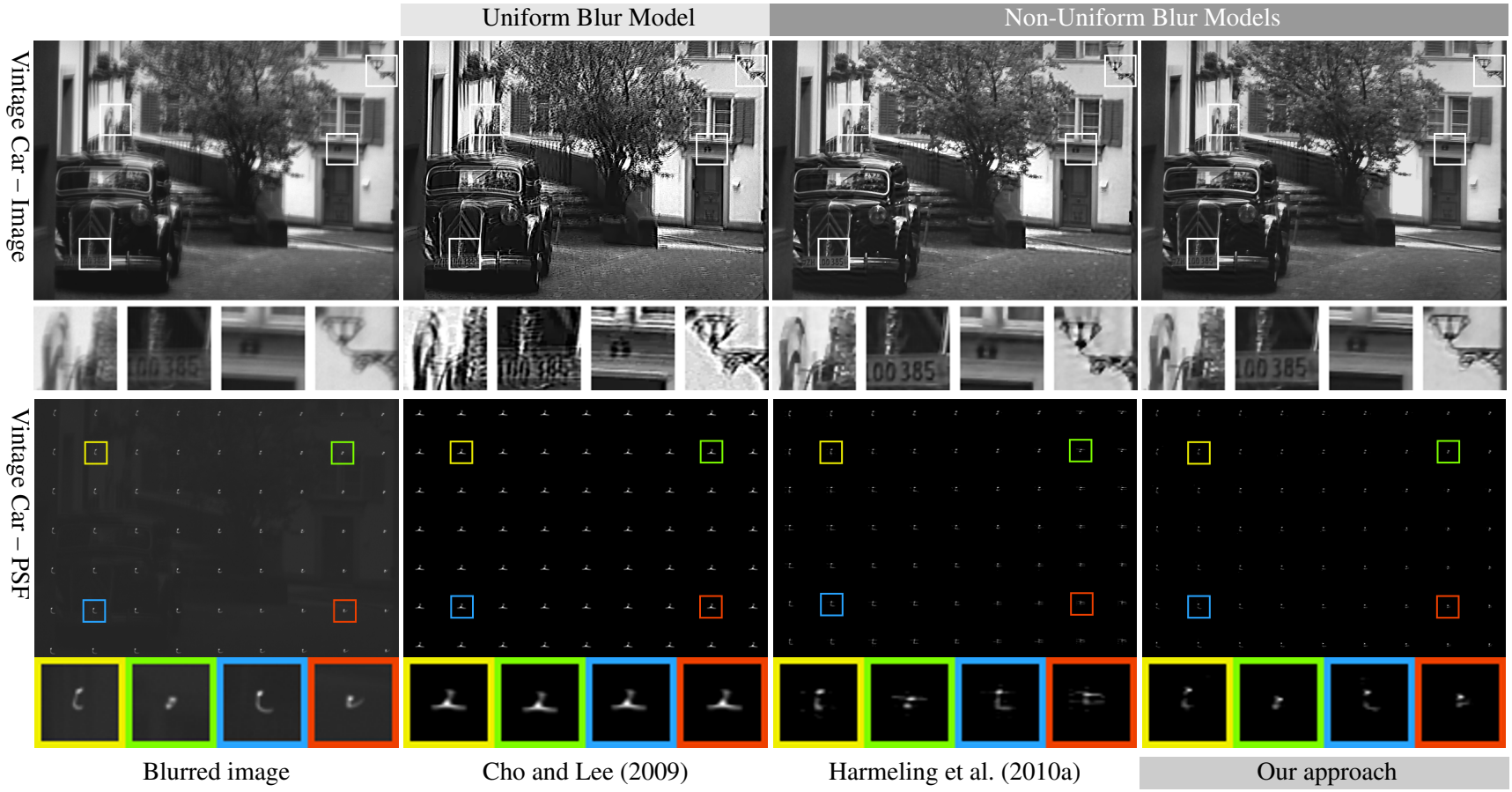



Figure 4.12: Controlled experiment: The left column shows the red and blue colour channels of a photo taken of a LCD screen with a handheld camera. While the red colour channel of the image that was displayed on the screen encoded a normal gray-scale image (top), the blue colour channel contained a point grid (bottom), which captured the motion blur during exposure and serves as a ground truth. Cho and Lee (2009) assume a uniform blur model which fails to describe the spatial variation of the PSF. In contrast, the EFF based method presented in Harmeling et al. (2010a) is able to do so, however fails in regions with little edge information (paving). The approach proposed in this thesis surpasses this limitation by constraining the EFF to allow for physically plausible motion blur only.

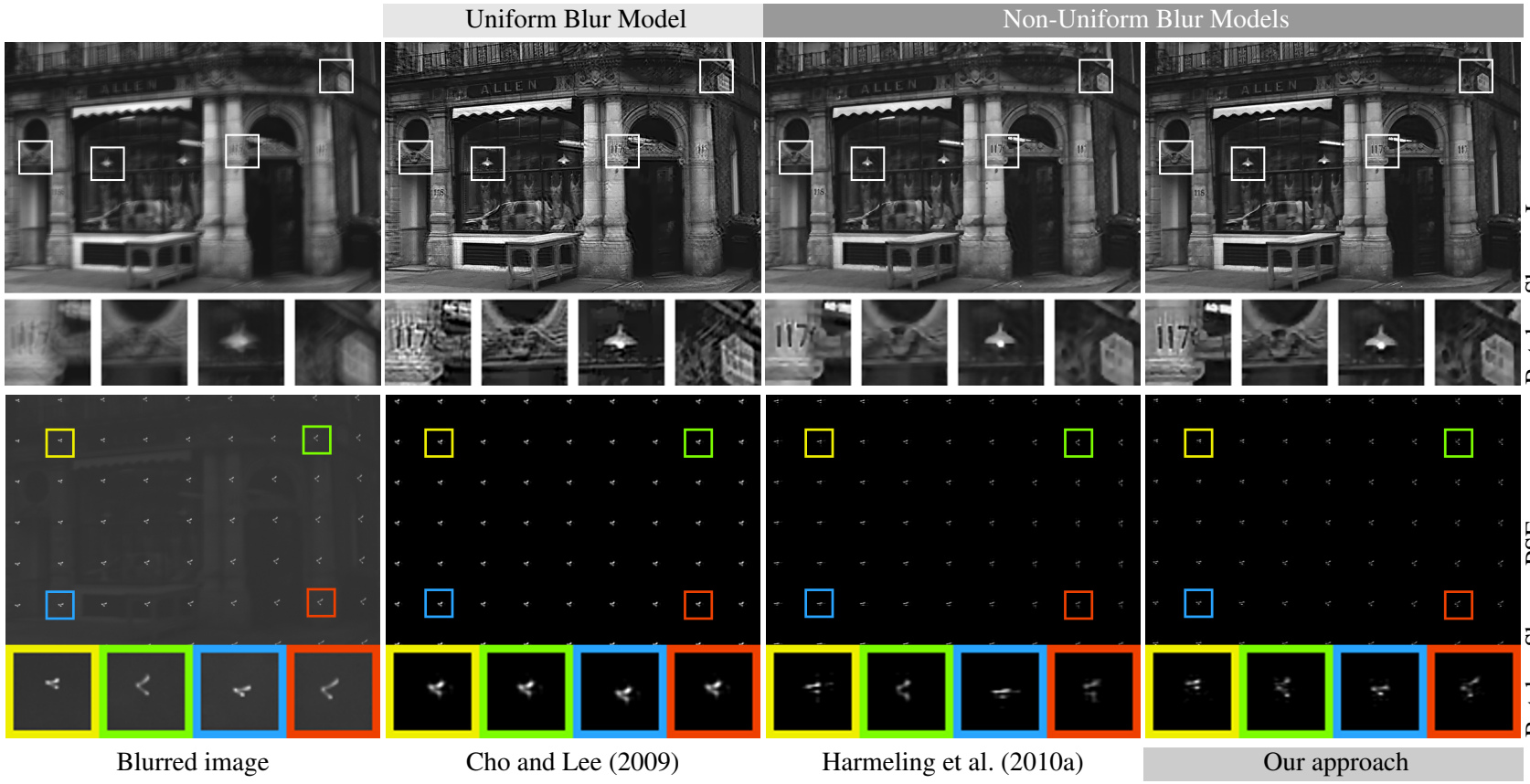



Figure 4.13: Controlled experiment: The left column shows the red and blue colour channels of a photo taken of a LCD screen with a handheld camera. While the red colour channel of the image that was displayed on the screen encoded a normal gray-scale image (top), the blue colour channel contained a point grid (bottom), which captured the motion blur during exposure and serves as a ground truth. Cho and Lee (2009) assume a uniform blur model which fails to describe the spatial variation of the PSF. In contrast, the EFF based method presented in Harmeling et al. (2010a) is able to do so. The approach proposed in this thesis even improves on the their result by constraining the EFF to allow for physically plausible motion blur only.

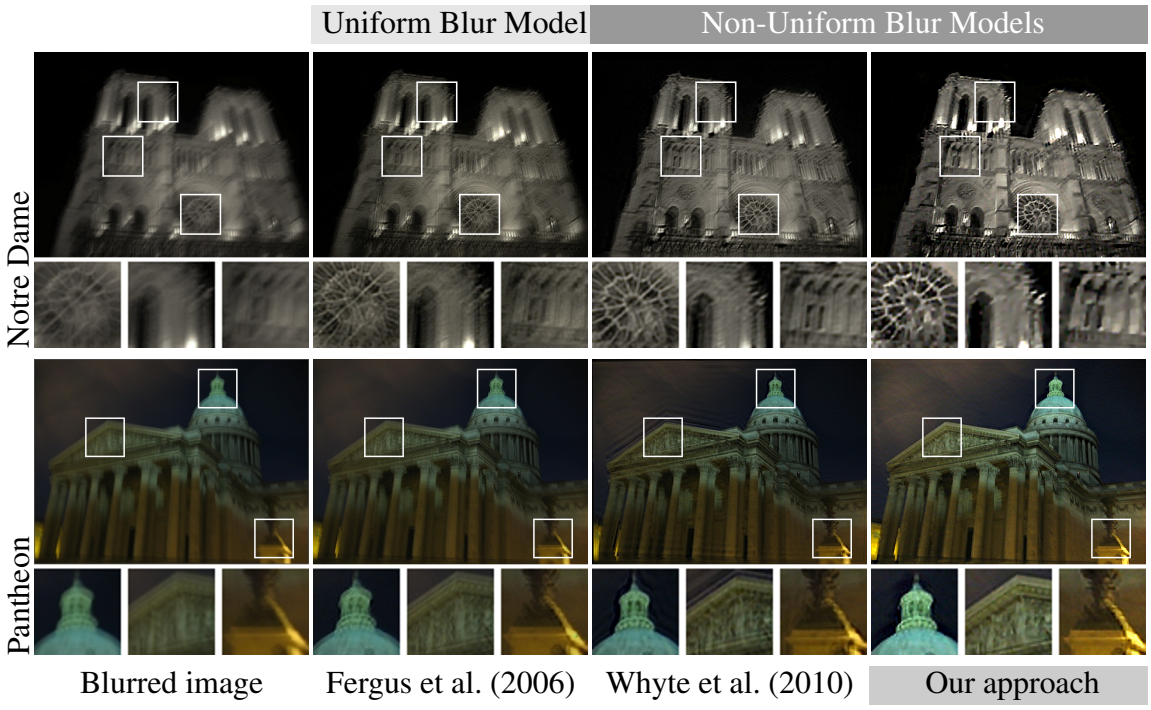



Figure 4.14: Comparison on real-world photos taken with a handheld camera. While Fergus et al. (2006) assume a uniform blur model and estimate an invariant PSF, Whyte et al. (2010) and our approach account for the spatial variation of the PSF. While Whyte et al. (2010) considers rotational camera motion, we take translations and in-plane rotations into account.

improvement in motion blur estimation can also be appraised by inspecting the PSF plots in Figures 4.11, 4.12 and 4.13.

#### 4.4.2.2 Comparison with Whyte et al. (2010)

The examples *Notre Dame* and *Pantheon* in Figure 4.14 show pictures with real camera shake taken from (Whyte et al., 2010). Results are shown for Fergus et al. (2006) who assume stationary blur and Whyte et al. (2010) who model the motion blur as PMPB caused by rotations only. The image obtained by Whyte et al. (2010) exhibits much more detail compared to Fergus et al. (2006) which suggests that the camera motion during exposure involved a significant amount of rotational motion. While Whyte et al. (2010) consider rotations (roll, pitch, yaw) for describing the motion blur, we took the basis of Gupta et al. (2010) comprising of translations in x- and y-direction and in-plane rotations. It equally well captures the motion blur which is verified by the good restoration quality of our approach.

Figure 4.15 is an interesting example, as Whyte et al. (2010) uses a noise/blurry image pair to estimate the motion blur (Yuan et al., 2008). In contrast, we are able to capture the blur blindly without using the noisy version of the image and recover a sharp image with comparative if not superior quality. For comparison, we also show the result of Cho and Lee (2009) which assume invariant motion blur. Although pursuing a similar strategy for kernel

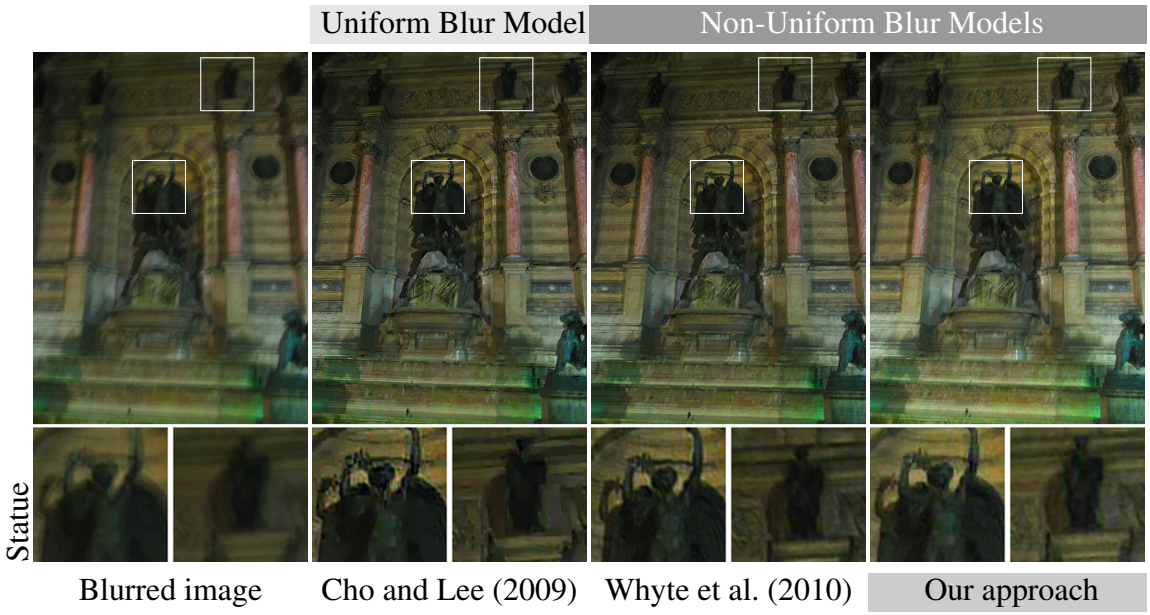



Figure 4.15: Comparison on noisy/blurry image pair (only blurry image is shown). While Whyte et al. (2010) uses both images for motion blur estimation, Cho and Lee (2009) and our approach use the blurry image only. See text for details.

estimation as Cho and Lee (2009), our deblurring algorithm is able to deal with non-uniform blur leading to a much better restoration result.

#### 4.4.2.3 Comparison with Gupta et al. (2010)

Figure 4.16 shows two real-world example images from (Gupta et al., 2010). To compare against a state-of-the-art deblurring method assuming uniform blur, we applied (Xu and Jia, 2010). In the case of the *Magazines* example, the uniform blur assumption of Xu and Jia (2010) is insufficient as their method fails to find a meaningful kernel. In contrast, the non-stationary PMPB model of Gupta et al. (2010) is able to capture and remove the blur such that the result reveals much more detail. Although using the same basis (in-plane rotations and translations) as Gupta et al. (2010), we are able to improve image quality even further, evident by less artifacts and clearly visible in the closeups.

The *Petrol Station* example image has been reported to be challenging due to the great variation of depth in the scene (Gupta et al., 2010). Indeed, despite yielding better restoration results compared to (Xu and Jia, 2010), both the method of Gupta et al. (2010) and our approach fail to fully remove the motion blur present in the up-left part of the image due to its increased depth compared to the foreground scene which is not covered by our model.

#### 4.4.2.4 Comparison with Joshi et al. (2010)

The *Coke* example in Figure 4.17 is another interesting example, as Joshi et al. (2010) uses data from inertial measurement sensors to determine the PSF. In contrast, we are able to estimate the blur *blindly* without exploiting the additional sensor data and recover a sharp

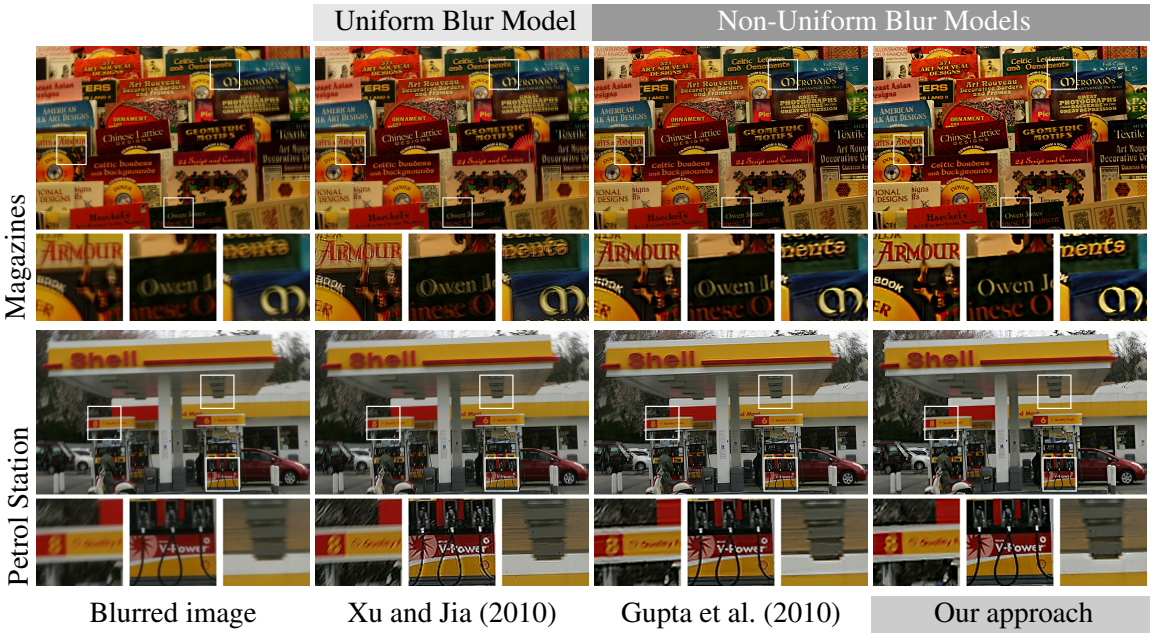



Figure 4.16: Comparison on a real-world photo taken with a handheld camera. While Xu and Jia (2010) considers translations only, both Gupta et al. (2010) and our approach additionally take in-plane rotations into account to capture the unknown motion blur. Note the great variation of depth in the *Petrol Station* example which limits restoration quality of all three presented approaches.

image with comparable if not superior quality. For comparison, we also show the result of Xu and Jia (2010) whose assumption of a invariant motion blur is again too restrictive to yield a good restoration result.

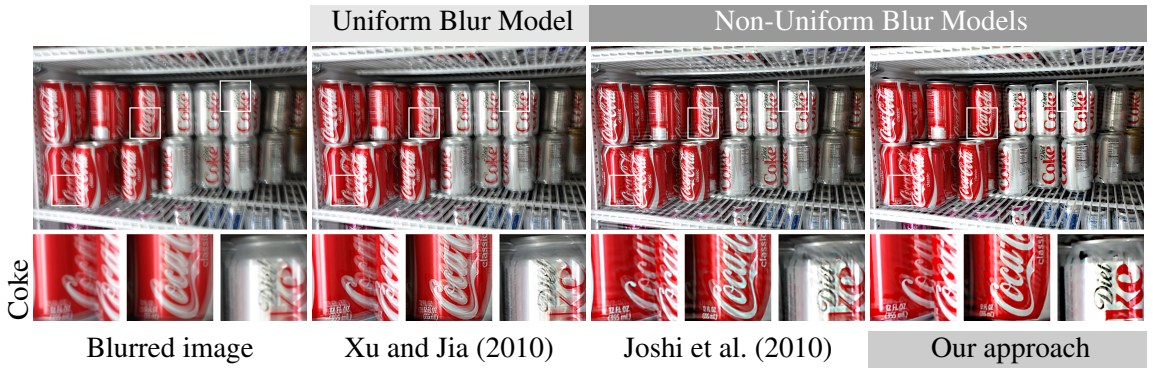



Figure 4.17: Comparison on real-world photos taken with a handheld camera. While Joshi et al. (2010) use additional information recorded by inertial measurement sensors at the time of image capture, Xu and Jia (2010) and our approach estimate the motion blur blindly.

## 4.5 Technical Details

### 4.5.1 Parameter Setting and Adjustment

**Parameter setting.** Our presented method is fairly automatic and necessitates little user intervention. Most of the presented results were obtained with the same or similar parameter settings. In particular, we set the regularisation constants $\lambda$ and $\eta$, in the objective (4.18) and (4.20) that is minimised for blur parameter estimation to 0.1 and 0.5, respectively. For non-blind image deconvolution, where we minimise objective (4.19) and (4.21) via direct deconvolution (4.23) we set the regularisation constant $\alpha$ to 0.5. For the final non-blind deconvolution we set the exponent $\alpha$ to 2/3 as it allows the use of an analytical solution for the w-subproblem (see (Krishnan and Fergus, 2009) for details). The regularisation parameter $\nu$ in the objective (4.24) ranged between $10^{-2}$ and $10^{-4}$ depending on the quality of the PSF estimation. The image pyramid is constructed by starting on the coarsest scale with flat filters of size $5 \times 5$ pixels and the correspondingly down-sampled version of the blurry image. For up- and down-sampling we employ a simple linear interpolation scheme. Both filter and image sizes are increased by a factor of $\sqrt{2}$ from one scale to the next. For the EFF we use an overlap of 0.5 and a Bartlett-Hanning window for interpolating between neighbouring filters. For all experiments we used about $9 \times 13$ filters with the exact ratio dependent on the image size ratio.

**Basis generation.** One important aspect to consider is the choice of basis for motion blur estimation. Under-sampling the space of possible camera poses will limit our ability to find an accurate representation of the underlying spatially varying PSF, however sampling it too finely causes unnecessary computations. In this context, an important issue is which degrees of freedom of the camera motion should be taken into account. This aspect is crucial as the search space increases exponentially with the number of considered dimensions. As mentioned above, we followed Gupta et al. (2010) who argued that translations and in-plane rotations are sufficient for the description of motion blur in most cases. The good restoration quality of our experiments confirms their findings. A free parameter of our model which we have to specify in advance, is the size of the local filters that parametrise the EFF. This is similar to the case of uniform blur models where the size of the PSF has to be set manually. The number of translations corresponds to the specified size of a local filter, while the sampling for in-plane rotations is chosen such that an angular increment corresponds approximately to a displacement of one pixel at the edge of the image. Since we are fundamentally limited by the resolution of our images, setting the resolution higher leads to redundant rotations, that are indistinguishable from their neighbours which will unnecessarily impede numerical optimisation. The frames of the PSF basis are precomputed at the beginning of each scale and are stored as sparse matrices to reduce memory demands.

**Manual parameter adjustment.** Finally, we mention a few parameters which were found to influence restoration quality and whose manual adjustment might improve the outcome of our algorithm. By default, the basis frames are generated by rotations around a pixel which is varied within a central image region as large as the size of a local filter. This

assumes that the camera has undergone rotations around a axis close to the optical axis only. However, in some cases such as the *Petrol Station* example it is obvious that this assumption does not hold. In such cases we set the centre of rotation roughly to the region which appeared visually sharpest. Another parameter to mention is the regularisation constant of the final non-blind deconvolution. Its value was chosen dependent on the quality of the estimated spatially varying PSF with smaller values in the case of poor motion blur estimation, giving more weight to the prior. Related is the regularisation parameter of the non-blind deconvolution (4.23) during motion blur estimation, which balances the trade-off between goodness of fit and the smoothness prior. For noisy images such as the *Notre Dame* example it prove beneficial to slightly increase its value to counter noise. One final issue worth mentioning concerns the regularisation parameter relevant for motion blur estimation. For images with few strong edges but much detail, it proved useful to put more weight on the prior terms.

### 4.5.2 Implementation and Run-time

Our algorithm has been prototyped in Matlab as it allows easy access and good inspection of intermediate results which eases debugging and studying the code behaviour. The typical run-time for a 1MP image is about 30 minutes on a single core of an Intel i5 processor with 2.4GHz.

However, the algorithm detailed above lends itself to parallelisation on a Graphics Processing Unit (GPU). We reimplemented all steps of the algorithm in PyCUDA (Klöckner et al., 2009), a Python wrapper to NVIDIA's CUDA API. To evaluate the speed-up, we

| Example | Size in pixels | | Processing time in seconds | | | |
|---|---|---|---|---|---|---|
| | | | GPU | | | CPU |
| | Image | Kernel | A | B | C | C |
| Elephant | $611 \times 441$ | $19 \times 19$ | 8.316 | 0.132 | 8.454 | 613 |
| Vintage Car | $441 \times 621$ | $19 \times 19$ | 8.306 | 0.038 | 8.354 | 673 |
| Butcher Shop | $401 \times 601$ | $25 \times 25$ | 13.638 | 0.036 | 13.686 | 724 |
| Notre Dame | $265 \times 354 \times 3$ | $21 \times 21$ | 7.388 | 0.082 | 7.474 | 521 |
| Pantheon | $274 \times 366 \times 3$ | $15 \times 15$ | 5.078 | 0.07 | 5.156 | 546 |
| Statue | $710 \times 523 \times 3$ | $21 \times 21$ | 12.434 | 0.174 | 12.618 | 783 |
| Magazines | $512 \times 768 \times 3$ | $17 \times 17$ | 10.3 | 0.172 | 10.482 | 767 |
| Petrol Station | $406 \times 679 \times 3$ | $17 \times 17$ | 8.194 | 0.132 | 8.338 | 599 |
| Coke | $749 \times 1123 \times 3$ | $21 \times 21$ | 12.776 | 0.27 | 13.07 | 1373 |

Table 4.1: Run-time of our Matlab and GPU implementation for several deblurring examples. A: kernel estimation. B: final deconvolution. C: total processing time.

compared the run time of our MATLAB implementation on a single core of an Intel Core i5 against our PyCUDA version on a NVIDIA Tesla C2050 with three gigabytes of memory, running on a 2.4Ghz Intel Xeon. Tab. 4.1 shows that deblurring real images of realistic sizes can be performed about $60\times$ faster on GPUs than on usual (single-core) processors and takes only about 30 seconds for a 2 Mega pixel image. However, note that for GPU computation

the image size is limited by the amount of available video memory. For example, with 1 GB RAM, the image size is restricted to $2000 \times 2000$ pixels. Due to implementation issues, our current GPU-based version is limited to images up to size $3000 \times 2000$ pixels.

## 4.6 Limitations and Perspectives

Although our presented approach considerably enlarges the regime where handheld photographs can be taken, there is a number of cases where our method breaks down and will fail to recover a high-quality image. In particular, current limitations include:

- **Saturation.** Our proposed method is not able to tackle the adverse effects of overexposure as our forward model does not account for saturated pixels. Image regions that exceed the dynamic range of the image sensor such as those shown in Figure 4.18 cause a non-linear pixel response. Hence, in these regions our linear model assumption is violated, a fact which impedes the inference process and adversely affects kernel estimation. Extending our model to account for saturated pixel regions similar to the approach we presented in Chapter 3, may be able to address this issue. Another approach for tackling the adverse effects of saturation that could be combined with our method has been proposed by Cho et al. (2011) recently.

- **Object motion.** Another major limitation of our approach is that it does not deal with moving or deformable objects. Our proposed image generation model only accounts for ego motion, *i.e*. motion of the camera that is more often than not unwanted in the form of camera shake. In contrast, moving objects in the scene such as shown in Figure 4.18 that also cause motion blur in the image are currently not modelled. A preprocessing step to separate moving objects as suggested in (Levin, 2006) might be one way of dealing with this problem.

- **Depth variation.** Out-of-focus blur and scenes with significant depth variation as shown in Figure 4.18 may also result in unsatisfying deblurring results. Both effects are not explicitly modelled by our approach. In particular, objects in different depths and partially occluding each other may cause different blurs despite lying in the focus range of the used lens. In this case, the scene can be segmented into different image regions each corresponding to a certain object that has a different distance to the camera. The PSF varies smoothly within these segments of the image, but between segments it may change abruptly. A more detailed discussion on the dependence of the PSF on depth variation within the scene can be found in (Harmeling et al., 2010a). Again, a preprocessing step to separate different depth layers may help to make the deconvolution more robust. For finding a meaningful segmentation, techniques such as those presented in (Loktyushin and Harmeling, 2011) could be applied.

- **Severe blur.** If the camera motion during exposure is too large, our method becomes computationally intractable as the search space of admissible solutions becomes vast and practically unmanageable by numerical optimisation techniques. Even restricting the space of possible camera trajectories to translations and in-plane rotations only, the

Saturated pixels | Out-of-focus blur | Objects in motion | Severe blur

Figure 4.18: Example images of current limitations of our proposed algorithm are shown.

search space grows cubically with PSF size. One solution to the problem might be the use of additional information from inertial measurement sensors such as accelerometers and gyroscopes. These sensors are nowadays standard electronic components and can be found on most mobile devices as they provide valuable motion and orientation information that is used for interactive control and navigation. In (Joshi et al., 2010), it has been demonstrated that such sensors are capable of recording the camera trajectory during exposure precisely enough to yield high-quality deblurring results. However, their presented system relies on a complicated and subtle calibration step that renders their approach unstable and vulnerable. As the sensor data is directly mapped to the PSF, even slight calibration or measurement imperfections lead to noticeable artifacts in the reconstructed image. One way to combine the strengths of both approaches, *i.e.* our BD approach and the one presented in Joshi et al. (2010), might be to use the inertial measurement sensor information to restrict effectively the search space of admissible solutions, thereby alleviating the optimisation problem especially for large motion blur.

- **Peculiar camera motion.** For certain types of motion such as linear uniform or stroboscopic camera motion, our kernel estimation fails to infer the corresponding PSF. Although these types of motion are quite peculiar in the sense that they rarely occur in real photographs, their study provides valuable insight into the capabilities of our kernel estimation procedure and why it fails sometimes to find a meaningful solution. As detailed in Section 4.3, at each scale kernel estimation is accomplished by iteratively estimating the blur parameters of the camera motion constrained EFF, based on the blurry as well as a predicted image. The predicted image is obtained via bilateral and shock-filtering. While bilateral filtering is applied to counter noise, shock-filtering aims at reverting the effect of blur by reestablishing smeared out edges. However, the latter will fail if an edge is linear uniformly blurred as it will appear as a constant area without leaving a working point for the shock filter. The same is true in the case of stroboscopic motion, where there are multiple sharp copies of the same scene overlaid within the same image as all copies are on equal footing and will be enhanced by the applied shock filter in the same way. More powerful priors might help to rule out the trivial solution and help to select a solution which we as humans would favour.

## 4.7 Conclusion and Outlook

In this chapter we proposed a single image blind deconvolution algorithm for removing non-uniform motion blur due to camera shake. By combining the efficiency of the EFF and the camera motion constraints of PMPB, we derived a fast forward model that outperforms recent approaches in computation time at a negligible loss in accuracy. To infer the unknown camera motion during exposure, we developed an efficient multi-hierarchical inference scheme by combining and adopting recent ideas of state-of-the-art techniques for uniform motion deblurring. The final image estimate is obtained by a non-blind deconvolution with a spatially varying PSF that corresponds to the estimated camera motion blur. To compensate for inaccuracies of the estimated PSF and to counter noise in the blurry image, we employ a regularisation term that promotes natural image statistics and helps to reduce restoration artifacts. Controlled experiments as well as a comprehensive comparison on a number of real-world examples demonstrated the extended capabilities and improved performance of our approach.

## 4.8 Acknowledgements

The work presented in this chapter was carried out in collaboration with Stefan Harmeling, Christian Schuler and Bernhard Schölkopf of the Max Planck Institute for Intelligent Systems. Christian Schuler assisted the GPU implementation of (Hirsch et al., 2011) and helped to take blurry images in the Boxenstop museum in Tübingen. Many fruitful discussions with Stefan Harmeling helped to clarify issues, amongst others a memorable whiteboard derivation of equation (4.13). Stefan Harmeling also assisted an early implementation of the algorithm presented in (Harmeling et al., 2010a) and the preparation of the manuscripts. Bernhard Schölkopf contributed in generously sharing ideas, giving scientific advice and in helping to prepare the manuscripts of the conference submissions. Except otherwise explicitly stated, all mathematical derivations, algorithmic implementations and experimental evaluations were performed by the author of this thesis.

This chapter is based on the following publications:

- *Space-Variant Single-Image Blind Deconvolution for Removing Camera Shake*, Stefan Harmeling, Michael Hirsch, and Bernhard Schölkopf. In Advances in Neural Information Processing Systems 23, Twenty-Fourth Annual Conference on Neural Information Processing Systems (NIPS 2010).
- *Fast Removal of Non-Uniform Camera-Shake*, Michael Hirsch, Christian Schuler, Stefan Harmeling, and Bernhard Schölkopf. In Proceedings of the 13th IEEE International Conference on Computer Vision (ICCV 2011).

The accompanying project webpage with additional material can be reached at
`http://webdav.is.mpg.de/pixel/fast_removal_of_camera_shake/`.

# Chapter 5

# A Blind Deconvolution Approach for Improving the Resolution of Cryo-EM Density Maps

Cryo-electron microscopy (cryo-EM) plays an increasingly prominent role in structure elucidation of macromolecular assemblies. Advances in experimental instrumentation and computational power have spawned numerous cryo-EM studies of large biomolecular complexes resulting in the reconstruction of three-dimensional density maps at intermediate and low resolution. In this resolution range, identification and interpretation of structural elements and modelling of biomolecular structure with atomic detail becomes problematic. In this paper, we present a novel algorithm that enhances the resolution of intermediate- and low-resolution density maps. Our underlying assumption is to model the low-resolution density map as a blurred and possibly noise-corrupted version of an unknown high-resolution map that we seek to recover by deconvolution. By exploiting the non-negativity of both the high-resolution map and blur kernel we derive multiplicative updates reminiscent of those used in non-negative matrix factorisation. Our framework allows for easy incorporation of additional prior knowledge such as smoothness and sparseness, on both the sharpened density map and the blur kernel. A probabilistic formulation enables us to derive updates for the hyper-parameters, therefore our approach has no parameter that needs adjustment. We apply the algorithm to simulated three-dimensional electron microscopic data. We show that our method provides better resolved density maps when compared with B-factor sharpening, especially in the presence of noise. Moreover, our method can use additional information provided by homologous structures, which helps to improve the resolution even further.

The outline of this chapter is as follows: we first give a brief introduction to both cryo-EM and blind deconvolution. We review state-of-the-art in both fields and motivate our approach. In Section 5.2 we develop a novel blind deconvolution algorithm by casting the problem into a set of coupled non-negative quadratic programs and derive multiplicative updates for both latent map and blur estimation. In Section 5.3 we discuss prior terms which are compliant with the multiplicative nature of our updates. We validate our approach on investigating its performance on simulated data in Section 5.4, followed by a presentation

of results on experimental density maps in Section 5.5. A summary and outlook in Section 5.6 concludes this chapter.

## 5.1 Introduction

Cryo-electron microscopy (cryo-EM) and low-resolution X-ray crystallography are emerging experimental techniques to elucidate the three-dimensional structure of large biomolecular complexes (Frank, 2002; Orlova and Saibil, 2004; Chiu et al., 2005; Brünger, 2005). A major drawback common to these methods is that the reconstructed density maps are only of intermediate or low resolution, typically in the nanometre range. In this resolution range, it becomes difficult to interpret the density maps unambiguously and to fit atomic models. A method to improve the quality of electron density maps has therefore the potential to broaden the scope of cryo-EM and low-resolution crystallography.

B-factor sharpening (DeLaBarre and Brunger, 2006; Rosenthal and Henderson, 2003; Fernández et al., 2008) is often advocated as a method for improving the resolution of density maps. The method operates in the frequency domain and applies a negative B-factor to the Fourier coefficients of the density map. This has the effect that high-frequency components encoding high-resolution features are amplified. B-factor sharpening has several limitations: First, the underlying model of the point spread function (PSF) is an isotropic Gaussian whose width is determined by the magnitude of the overall B-factor (the Fourier transform of a Gaussian is a Gaussian with inverted width). This assumption may be inappropriate for anisotropic data such as 2D crystals. Second, the method suffers from amplification of noise: Noise in density maps contributes high-frequency components, which are weighted up when applying a negative B-factor. Third, it is not possible to incorporate prior knowledge to regularise the recovered high-resolution density map. For example, the B-factor sharpened density map is not guaranteed to be non-negative.

In this chapter, we present a novel algorithm to sharpen electron density maps. The algorithm remedies some of the shortcomings of thermal factor sharpening. The underlying assumption is that low- to intermediate-resolution density maps can be viewed as distorted or "blurred" versions of high-resolution maps. Mathematically, this blurring process is modelled as a convolution

$$y = f * x \tag{5.1}$$

where $y$ denotes the observed blurry and noisy low-resolution map, $x$ the true high-resolution map, $f$ the linear shift-invariant blur kernel or PSF and $*$ the linear convolution operator. Our degradation model is illustrated in Figure 5.1 for a simulated density map of a monomer of the chaperonin complex GroEL-GroES-(ADP)7 (PDB-ID: 1AON). We propose Blind Deconvolution (BD) to sharpen electron density maps. BD aims to invert the blurring process and thereby recover the high-resolution map without any knowledge on the degradation or blur kernel. It does so by estimating the sharpened density map and the PSF simultaneously. In particular, we are interested in BD algorithms that do not assume a particular structural model and that are in this sense parameter-free. The recovered high-resolution map will be useful for density map interpretation and model fitting.

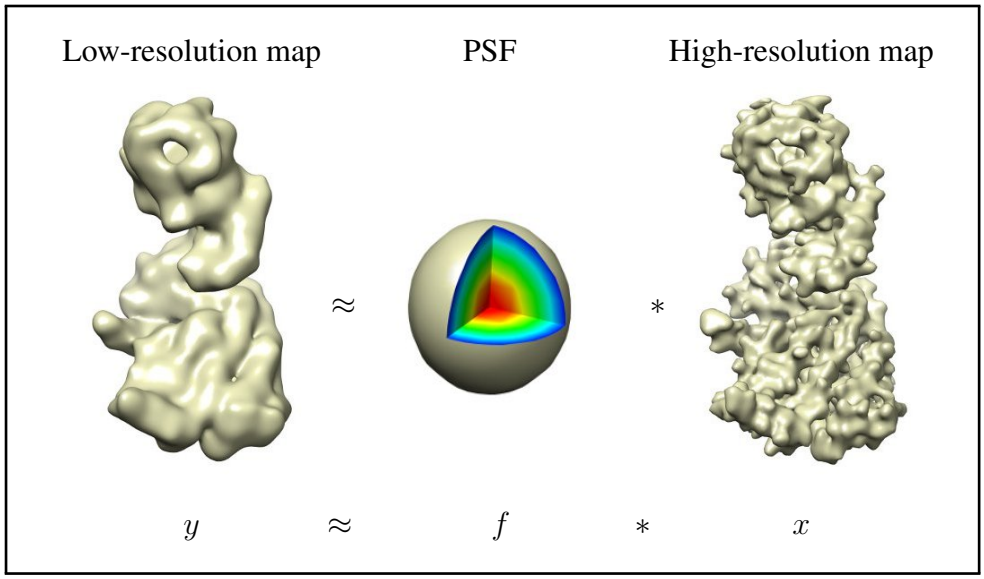



Figure 5.1: Degradation model: Low-to-intermediate resolution maps can be viewed as blurred versions of high-resolution maps. The point spread function (PSF) determines the form of blurring. Note that for illustration purposes the PSF is enlarged.

Blind Deconvolution is a severely ill-posed problem because there exists an infinite number of solutions and small perturbations in the data lead to large distortions in the estimated true map. The ill-posedness may be alleviated by confining the set of admissible maps to those which are physically plausible through the introduction of additional constraints. One such constraint is that electron density maps are inherently non-negative. We show that Non-negative Blind Deconvolution (NNBD) can be cast into a set of coupled quadratic programs that are solved using the multiplicative updates proposed in Sha et al. (2007). No learning rate has to be adjusted and convergence of the updates is guaranteed. By iterating between an update step for $x$ and $f$, we obtain an efficient BD algorithm that allows for straightforward incorporation of prior knowledge such as sparseness and smoothness of the true map and/or the PSF.

Blind deconvolution is a valuable tool in many image and signal processing applications such as computational photography, astronomy, microscopy, and medical imaging and thus has been treated in numerous publications. Many blind deconvolution algorithms have been proposed in various fields of research, for an overview confer (Kundur and Hatzinakos, 1996; Starck et al., 2002; Sarder and Nehorai, 2006; Levin et al., 2009). However, to our knowledge it has never been proposed in the field of cryo-EM.

## 5.2 Blind Deconvolution by Non-Negative Quadratic Programming

Our generative model underlying the image formation process is

$$y \approx f * x$$

where the degraded map $y$, the PSF $f$ and the true map $x$ are n-dimensional.[1] Assuming additive Gaussian noise with zero mean and variance $\tau^{-1}$, the likelihood of observing $y$ is given by

$$p(y|f,x,\tau) = Z(\tau)^{-1} \exp\Big\{-\frac{\tau}{2}\,\|y - f * x\|^2\Big\}$$

where $\|\cdot\|$ denotes the $L_2$-norm and $Z$ the normalising partition function, which depends only on the precision $\tau$. As a prior, we constrain $f$ and $x$ to be of finite size and to lie in the non-negative orthant: $p(x) \propto \chi(x \geq 0)$ and $p(f) \propto \chi(f \geq 0)$ where $\chi$ is the indicator function. Computation of the maximum a posteriori (MAP) estimate of $f$ and $x$ is equivalent to the non-negatively constrained problem of minimising the negative log-likelihood viewed as a function of the unknown parameters $f$ and $x$:

$$\min_{f\geq 0,\, x\geq 0} L(f,x) = \frac{1}{2}\,\|y - f * x\|^2 . \tag{5.2}$$

Here, the negative log-likelihood $L$ is expressed in units of $\tau$ and constants independent of $f$ and $x$ have been dropped. Because of the interdependence of $f$ and $x$ through the convolution, optimisations problem (5.2) is non-convex and a globally optimal solution cannot be found efficiently. Fortunately, the objective function $L(f,x)$ is sufficiently well-behaved as it is convex in each variable separately if the other is held fixed. This observation suggests a simple alternating descent scheme: instead of manumitting (5.2) directly we iteratively solve the minimisation's problems $\min_{f\geq 0} L(f)$ and $\min_{x\geq 0} L(x)$, where $L(f)$, $L(x)$ denotes $L(f,x)$ for fixed $f$, $x$, respectively. If we can ensure descent in each step, we will obtain a sequence of estimates $\{f^{(k)}, x^{(k)}\}$ that never increase the objective $L(f,x)$. Due to the symmetry of the convolution operation, $f * x = x * f$, we can restrict our exposition to the optimisation of $x$; equivalent results will hold for $f$.
Because convolution is a bilinear operation, the problem of epitomising $x$ can be written in matrix notation :

$$\min_{x\geq 0} L(x) = \frac{1}{2}\,\|y - f * x\|^2 = \frac{1}{2}\,x^T F^T F x - y^T F x + \frac{1}{2}\,y^T y, \tag{5.3}$$

where in this formulation $y$, $x$ and $f$ are zero-padded vectors stacked in lexicographical order and $F$ is a block-Toeplitz structured matrix. In the following we will use both notations interchangeably; the type of the involved quantities will be clear from the context. Minimising (5.3) is equivalent to solving a quadratic program with non-negativity constraint

$$\min_{x\geq 0}\ \frac{1}{2}\,x^T A\,x + b^T x \tag{5.4}$$

with $A = F^T F$ and $b = -F^T y$. Recently, a novel algorithm for solving Non-negative Quadratic Programs (NNQPs) based on multiplicative updates has been proposed Sha et al. (2007). In the derivation of the updates, only the positive semi-definiteness of $A$ is required. In particular, $A$ may have negative entries off-diagonal. The key idea is to decompose $A$

[1] The convolution is assumed to be non-circular and its value is taken only on its valid part, *i.e.* in the one-dimensional case, if $x \in \mathbb{R}^n$ and $f \in \mathbb{R}^m$, then $y$ is an element of $\mathbb{R}^{n-m+1}$. For discretized signals, $*$ reads $(f * x)_n = \sum_{i\in \mathrm{supp}(f)} f_i x_{n-i}$, where $\mathrm{supp}(f)$ denotes the support of $f$.

into its positive and negative part, *i.e.* $A = A^+ - A^-$, where $A^{\pm}_{ij} = (|A_{ij}| \pm A_{ij})/2$, and to construct an auxiliary function $G(x, x')$ for the objective (5.2) such that $\forall x, x' > 0$: $L(x) \leq G(x, x')$ and $L(x') = G(x', x')$. Because $G(x, x')$ is an upper bound on $L(x)$, minimisation with respect to $x$ yields an estimate $\hat{x} = \operatorname{argmin}_x G(x, x')$ which never increases the objective $L(x')$:

$$L(\hat{x}) \leq G(\hat{x}, x') \leq G(x', x') \leq L(x').$$

As shown in Sha et al. (2007) a valid auxiliary function for (5.4) is given by

$$G(x, x') = \frac{1}{2} \sum_i \frac{(A^+ x')_i}{x'_i} x_i^2 - \sum_i (A^- x')_i \, x'_i \log \frac{x_i}{x'_i} + b^T x - \frac{1}{2} \, x'^T A^- x'. \tag{5.5}$$

Minimisation of (5.5) with respect to its first argument yields the update:

$$x \leftarrow x \odot \frac{-b + \sqrt{b \odot b + 4 \, (A^+ x) \odot (A^- x)}}{2A^+ x}. \tag{5.6}$$

The symbol $\odot$ denotes voxel-wise multiplication, also division and square root are understood voxel-wise. For a non-negative observed map $y$ with $A^+ = F^T F$, $A^- = 0$ and $b = -F^T y$, update (5.6) reads

$$x \longleftarrow x \odot \frac{F^T y}{F^T F x}. \tag{5.7}$$

Contrary to previous approaches to NNQP (Johnston et al., 2000), no learning rate is involved that needs adjustment. Furthermore convergence to a global optimum is guaranteed. Note that as $f * x$ approaches $y$ the multiplicative factor in (5.7) tends to one. The update rules can be computed very efficiently using the Fast Fourier Transform (Press et al., 2007) due to the convolution theorem

$$Fx \equiv f * x = \mathcal{F}^{-1} \left\{ \mathcal{F}(f) \cdot \mathcal{F}(x) \right\}$$

and

$$F^T x \equiv f \star x = \mathcal{F}^{-1} \left\{ \mathcal{F}(f)^* \cdot \mathcal{F}(x) \right\},$$

where $\mathcal{F}$ denotes the discrete Fourier transform and $\star$ the n-way correlation between $f$ and $x$. Hence, we never have to compute matrices $F$ and $X$ explicitly. Because the objective is symmetric in $x$ and $f$, we obtain an equivalent update for $f$:

$$f \longleftarrow f \odot \frac{X^T y}{X^T X f}. \tag{5.8}$$

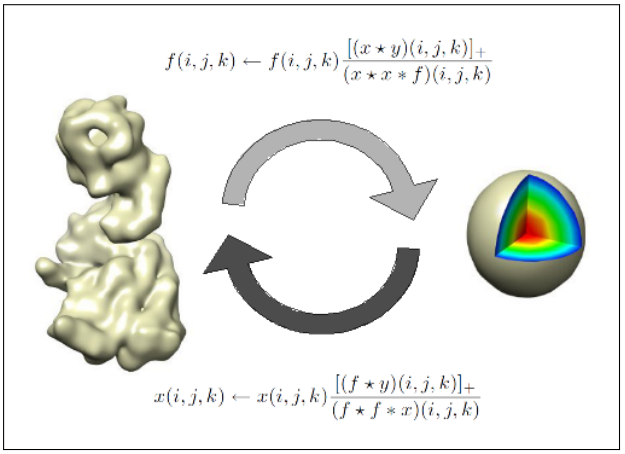
Coming back to a our original problem, namely solving (5.2) jointly in $x$ and $f$, we propose to iterate between update steps in $x$ and $f$. Cycling between (5.7) and (5.8) ensures that both $f$ and $x$ will remain in the non-negative orthant. Although multiplicative updates guarantee convergence to a global optimum in the case of NNQP, the proposed NNBD scheme only ensures convergence to a stationary point. Therefore, the solution might be sensitive to the initial values of $x$ and $f$. In our experiments, however, initialisation was never a problem: choosing flat maps for the initial $x$ and $f$ always led to good results. Algorithm (3) summarises our NNBD approach.

**Algorithm 3**: Non-negative Blind Deconvolution (NNBD)

**Input**: Degraded, blurry map $y$
**Output**: Sharp map $x$, blur kernel $f$

Initialisation of $f$ and $x$ with positive flat maps
**while** $\|y - f * x\|_F^2 > \epsilon$ **do**
  $f \leftarrow f \odot \frac{x \star y}{x \star x * f}$
  $x \leftarrow x \odot \frac{f \star y}{f \star f * x}$
**end**
**return**

Figure 5.2: Illustration of NNBD that alternatingly updates the latent density map and the PSF. The QR code below directs to a video showing the sharpening of a simulated noise-free density map by NNBD.




## 5.3 Incorporation of Prior Knowledge

In the absence of noise as well as in the case of high signal-to-noise ratios[2] (SNRs) our algorithm correctly decomposes a blurry observation into the true underlying map and the corresponding PSF.[3]

Figure 5.3 shows a simulated one-dimensional toy example, where $x$ is an equispaced sample of a Gaussian mixture model and $f$ is chosen such that it is irreducible.[4] The estimated map $\hat{x}$ and PSF $\hat{f}$ are close to the ground truth. However, Figure 5.3 shows that low SNRs raise difficulties in the reconstruction process and lead to noise-fitting and unfavourable solutions.

[2]Here, we define the signal-to-noise-ratio (SNR) of a signal as $\text{SNR}(\text{dB}) = 10 \log_{10} \frac{var(x)}{var(y - x*f)}$

[3]Note that this is true only up to an overall scaling factor, because for each estimate $\{\hat{f}, \hat{x}\}$ there exist infinitely many estimates $\{\frac{1}{\lambda}\hat{f}, \lambda\hat{x}\}$ with $\lambda \in \mathbb{R}^+$ that explain the observed data equally well. To rule these out, we fix the scale by normalising $f$. In addition to this scale invariance, the solution is also shift-invariant. Usually this effect can be corrected only by means of further prior knowledge.

[4]A signal $x$ is irreducible, if it cannot be decomposed into two or more nontrivial components $\{x_1, x_2, \ldots, x_n\}$ such that $x = x_1 * x_2 * \cdots * x_n$. Note that if either $f$ or $x$ is reducible, NNBD becomes inherently ill-posed, because $y = f * x$ cannot be decomposed unambiguously without employing additional prior knowledge.

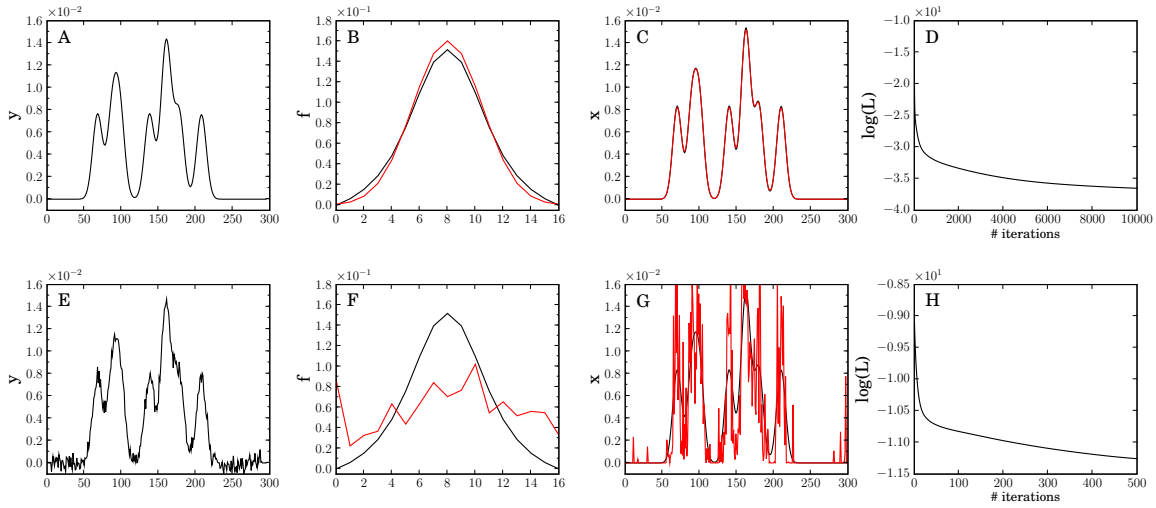



Figure 5.3: One-dimensional toy example. The top row shows the results of NNBD at SNR of 60 dB, the bottom row for SNR 20 dB. A, E: data $y$ used in NNBD. B, F: true (black) and estimated (red) PSF $f$. C, G: true (black), NNBD (red) estimate of the true signal $x$. D, H: negative log-likelihood (on logarithmic scale).

To further constrain the space of admissible solutions, additional knowledge about the unknown map and the PSF has to be utilised. This knowledge will be represented by non-uniform prior distributions $p(f|\theta)$ and $p(x|\theta)$ on $f$ and $x$, respectively, involving hyper-parameters $\theta$. With $p(\theta)$ denoting the prior of the hyper-parameters, the joint posterior is proportional to:

$$p(x, f, \theta|y) \propto p(y|f, x, \theta)\, p(x|\theta)\, p(f|\theta)\, p(\theta). \tag{5.9}$$

In the following, we describe prior distributions that are compatible with the multiplicative updates for $f$ and $x$ derived in the previous section. Again, because of the symmetry of (5.2) in $f$ and $x$, we will restrict ourselves to the incorporation of prior knowledge on the unknown map $x$.

Incorporating priors on $x$ introduces additional terms in (5.3) that have to be taken into account in the computation of the MAP estimate. In the derivation of the multiplicative update rule (5.6), we minimised the auxiliary function (5.5) defining an upper bound on $L(x)$. A close look reveals that all priors whose negative logarithm comprises terms that are either linear, quadratic, or logarithmic in $x$ can be incorporated into (5.5) and hence are compatible with the update (5.6). This includes the following priors:

- **Smoothness**: A desired property in many imaging applications is smoothness of the true map, which can be enforced by penalising the norm of its gradient $\|\nabla x\|$. The corresponding prior is

$$p(x|\lambda) \propto \exp\Big\{-\frac{\lambda}{2}\|\nabla x\|^2\Big\}. \tag{5.10}$$

  Note that $\|\nabla x\|^2$ can be rewritten as $x^T \Delta\, x$ where $\Delta x \equiv \nabla^T \nabla x = -\mathcal{L} * x$ is the negative Laplace operator, *i.e.* in the one-dimensional case $\mathcal{L} = (1, -2, 1)$.

- **Sparseness**: A further assumption commonly made is sparseness, which can be encoded in the exponential prior

$$p(x|\lambda) \propto \exp\{-\lambda \sum_i |x_i|\} = \exp\{-\lambda \mathbb{I}^T x\}, \tag{5.11}$$

  where the second equality holds for non-negative maps.

- **Orthogonality**: In some applications, it is useful to introduce a voxel-wise non-negative background $z$, which results in the model $y = f * x + z$. Such a background could, for example, account for the solvent in electron microscopic recordings or a homologous structure for model refinement (cf. Section 5.4.1). Usually, the background should be uncorrelated with the reconstructed map which can be enforced by penalising the overlap between $x$ and $z$, *i.e.*

$$p(x|\theta) \propto \exp\{-\lambda z^T x\}. \tag{5.12}$$

  We treat the background as a variable that we learn along with $f$ and $x$ using analogous multiplicative updates. In the following, we will refer to this regularisation term as orthogonality constraint. Of course, $z$ could be constant if such knowledge is available.

- **Entropy**: A reasonable assumption, especially for the form of the PSF, is that it exhibits a bump-like shape. This can be favoured by using the entropic prior

$$p(x|\lambda) \propto \exp\Big\{\lambda \sum_i \log x_i\Big\}. \tag{5.13}$$

  The Burg entropy $\sum_i \log x_i$ is compliant with the auxiliary function $G(x, x')$ and favours maximum entropy maps, *i.e.* constant maps. Entropy and sparseness/orthogonality can be combined into a single prior density: a voxel-wise Gamma distribution.

Table 5.1 summarises the presented prior distributions and the required modifications in (5.6).

### 5.3.1 Estimation of Hyper-Parameters

An important aspect is the estimation of the unknown hyper-parameters. Instead of resorting to heuristics or cross-validation, we use Bayesian inference to estimate the hyper-parameters $\theta$. For all hyper-priors introduced in the previous section, the Gamma distribution $G(\theta|\alpha, \beta)$ is a conjugate prior. The ideal approach to hyper-parameter estimation would be to calculate their marginal posterior distribution

$$p(\theta|y) = \int_{f\geq 0} \int_{x\geq 0} p(x, f, \theta|y)\ \mathrm{d}f\, \mathrm{d}x \tag{5.14}$$

and determine the mean or mode (Mackay, 1996). In our case, however, exact integration over $f$ and $x$ is infeasible. One would have to resort to computationally intensive methods

| **Prior** | $\boldsymbol{A}^+$ | $\boldsymbol{A}^-$ | $\boldsymbol{b}$ |
|---|---|---|---|
| Smoothness | $F^TF + \Delta^+$ | $\Delta^-$ | $-F^Ty$ |
| Sparseness | $F^TF$ | $0$ | $-F^Ty + \lambda\mathbb{I}$ |
| Orthogonality | $F^TF$ | $0$ | $-F^Ty + \lambda z$ |
| Entropy | $F^TF$ | $\lambda\,\mathrm{diag}\{x\}^{-2}$ | $-F^Ty$ |

Table 5.1: Modifications for the incorporation of prior knowledge in the update of the true map. $\Delta^+$ and $\Delta^-$ refer to the decomposition of the negative Laplacian $\Delta = \Delta^+ - \Delta^-$. $\mathrm{diag}\{x\}$ is a diagonal matrix with entries $x_i$.

like Markov chain Monte Carlo or alternatives such as variational (Molina et al., 2006) or approximate inference (Lin and Lee, 2005). Therefore we pursue the much simpler approach of computing the MAP estimate of the *joint* posterior, *i.e.*

$$\hat{\theta} = \mathrm{argmin}_\theta\, p(\hat{f}, \hat{x}, \theta|y), \tag{5.15}$$

where $\hat{f}$ and $\hat{x}$ denote the MAP estimate of the PSF and the true map, respectively. Although it has been argued that this approximation is crude and neglects valuable information (Levin et al., 2009), the joint MAP approach led to good results in our experiments. The estimates for the hyper-parameters $\hat{\theta}$ can be derived by solving (5.15) directly. The shape parameters $\alpha$ and $\beta$ of the Gamma hyper-prior are not estimated but set to fixed values $\alpha = 1$ and $\beta$ close to zero. According to Jin and Zou (2009) the sensitivity of the results on the shape parameters is negligible, which was confirmed by our experiments.

### 5.3.2 Discussion

Let us come back to the one-dimensional toy example at low SNR (cf. Figure 5.3). Figures 5.4 A-D show how enforcing smoothness of the signal using prior (5.10) prevents unfavourable noise-fitting and effectively helps us to recover the original signal and the PSF from the blurred and noisy observation. We further investigated the estimation of the regularisation parameter $\lambda$. We tested different fixed values for $\lambda$ and compared the reconstruction error of and the correlation with the true signal when applying our hierarchical Bayes approach. Figures 5.4 E and F show that the Bayes procedure yields a minimal reconstruction error and a maximal correlation for a wide range of fixed $\lambda$ values. The evolution of the regularisation parameter (Figure 5.4 G) reveals an important feature of our deconvolution algorithm. Starting at a small initial value, the regularisation parameter increases rapidly within a few iterations after which it gradually converges to a smaller optimal value. This finding may justify the heuristic regularisation scheme of Shan et al. (2008), which seems to be crucial for the success of their BD algorithm on natural images (Levin et al., 2009). Shan et al. propose to start the deconvolution with a large value of $\lambda$ – a conservative choice that puts

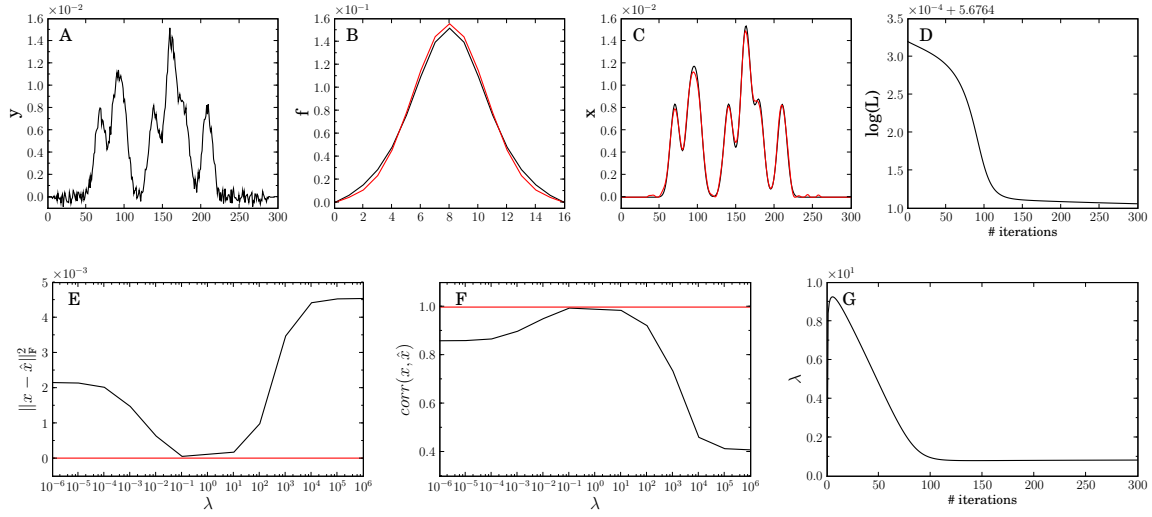



Figure 5.4: One-dimensional toy example. The top row shows the results of NNBD at a SNR of 20 dB. A: data $y$ used in NNBD. B: true (black) and estimated (red) PSF $f$. C: true (black), NNBD (red) estimate of the true signal $x$. D: negative log-likelihood (on logarithmic scale). The bottom row shows the absolute deviation (E)/correlation coefficient (F) of the reconstructed signal $\hat{x}$ from/and the true underlying signal $x$ for fixed values of the regularisation parameter $\lambda$ (black) and in the case of NNBD with additional hyper-parameter estimation (red) after 5000 iterations. G: Evolution of the hyper-parameter $\lambda$ with increasing number of iterations.

higher weight on the prior than on the data. As the deconvolution improves, the regularisation parameter is decreased to put more and more weight on the data. This is similar to simulated or deterministic annealing which aims to avoid trapping in sub-optimal local minima. The advantage of our approach is that, contrary to that of Shan et al. (2008), we do not need to choose a schedule for adjusting $\lambda$. Rather our update procedure automatically balances the influence of the data versus the importance of the prior.

## 5.4 Results on Simulated Density Maps

To evaluate the performance of our model and verify its validity we applied our algorithm to simulated three-dimensional density maps with a sampling of 1 Å/voxel. We used the program *pdb2mrc* from the EMAN software package (Ludtke et al., 1999) for density map simulation. First, we use Non-negative Blind Deconvolution to sharpen electron density maps. In the second application, we demonstrate the capabilities of our approach and the usefulness of the orthogonality prior by incorporating homologous structure information in the deconvolution.

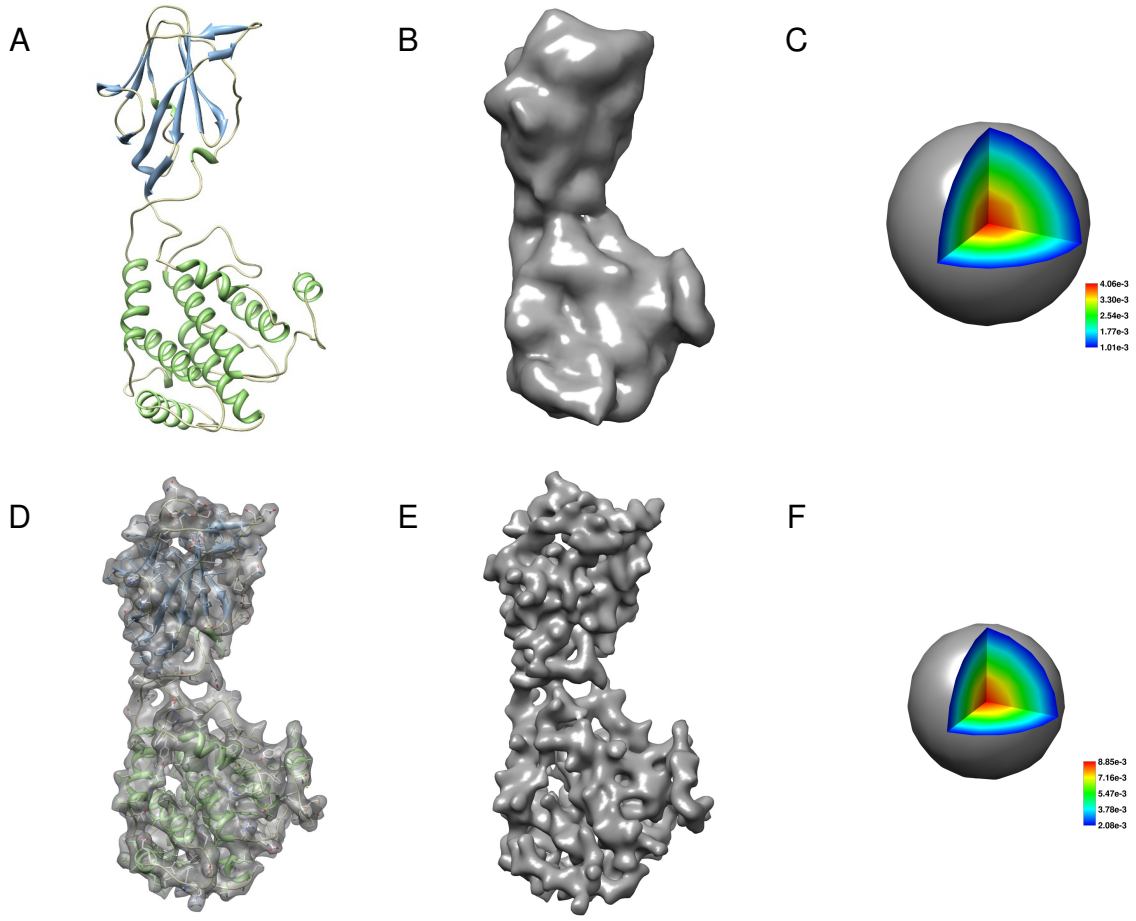



Figure 5.5: NNBD for the electron density map of the monomer of the bluetongue virus outer shell coat protein VP7 (PDB ID: 2BTV): Top row: A: molecular structure, B: simulated density map at 10 Å, C: point spread function. Middle row: D: NNBD reconstruction with molecular structure fitted into it, E: NNBD reconstruction, F: estimated point spread function.

### 5.4.1 Electron Density Maps of Proteins

For validation we used a monomer of the trimer of the bluetongue virus capsid protein VP7 (PDB ID: 2BTV) (Grimes et al., 1998). Figure 5.5 A shows the molecular structure, Figure 5.5 B the simulated electron density map at 10 Å resolution and Figure 5.5 C the corresponding PSF. Figures 5.5 D-F show the density map reconstructed with NNBD, the molecular structure fitted into it and the estimated PSF, respectively. The sharpened map reveals the nature of most secondary structure elements, whereas the original density map provides ambiguous secondary structure information. Also side chains become visible, which is important for modelling atomic details. To quantify the gain in resolution, we computed the correlation coefficient between the sharpened map and density maps simulated at higher resolutions. Figure 5.6 shows that the correlation coefficient is highest for a density map at a resolution of 6 Å. Hence, our algorithm is able to sharpen the original map and to improve

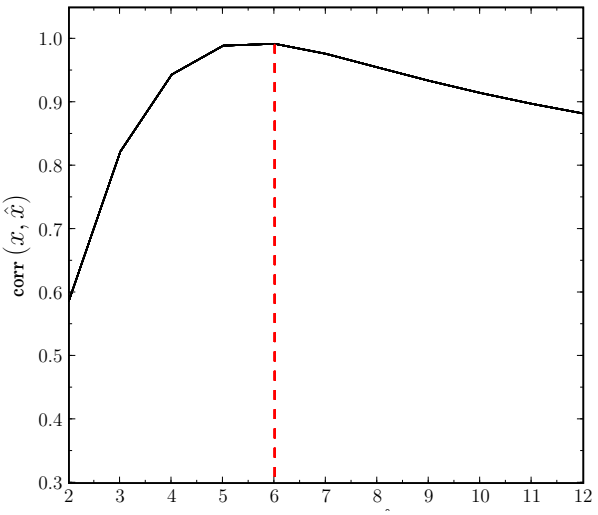



Figure 5.6: Correlation coefficient of the reconstructed density map shown in Figure 5.5 (E) with simulated density maps at various resolutions. The reconstructed density map peaks at 6 Å suggesting an improvement in resolution of 4 Å.

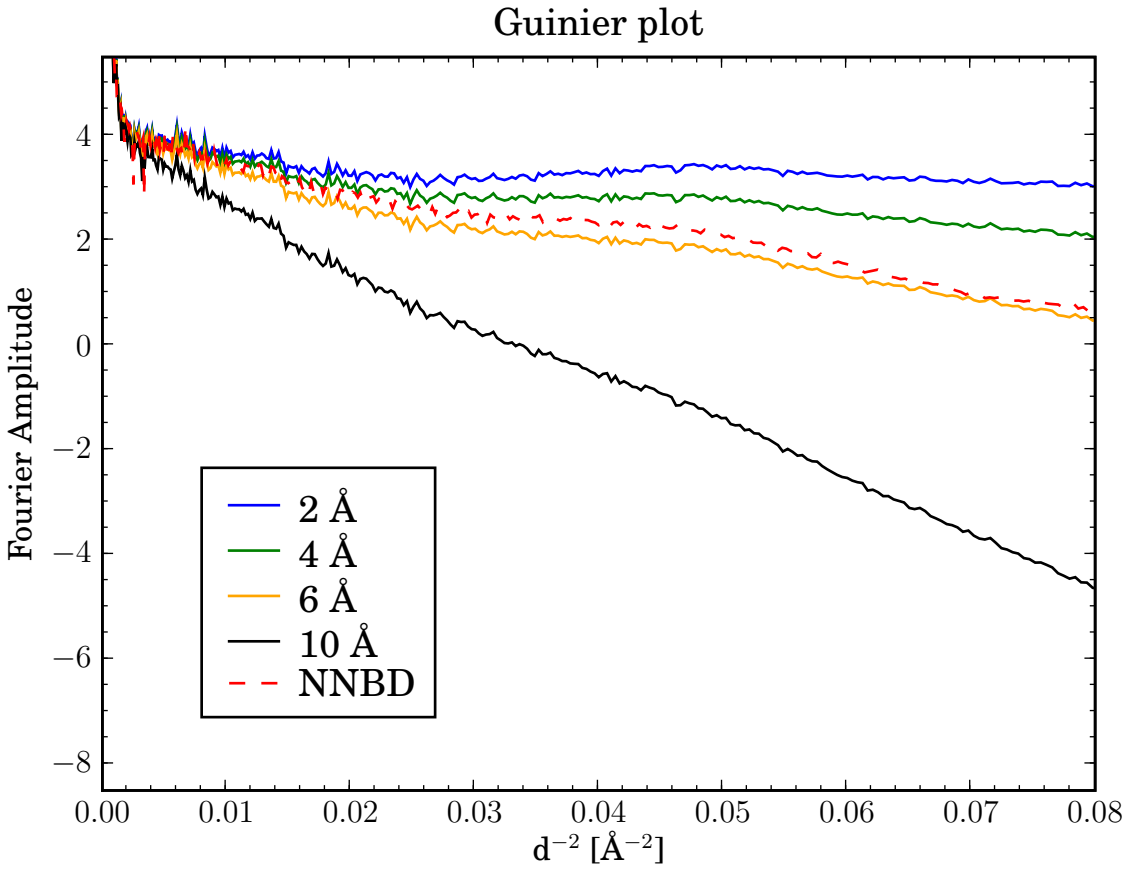



Figure 5.7: Guinier plot of the reconstructed density map shown in Figure 5.5 (E) with overlayed Guinier curves of simulated density maps at various resolutions. The plot suggests a resolution of about 6 Å which is in good agreement with the peak of the correlation coefficient in Figure 5.6.

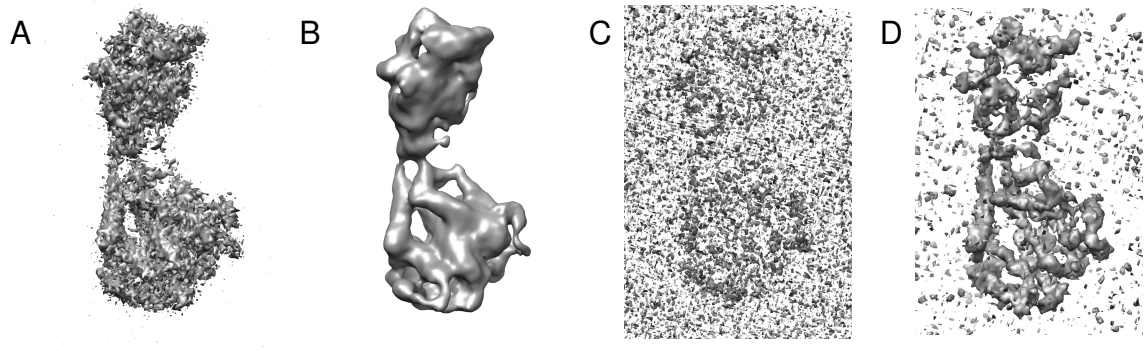



Figure 5.8: NNBD for the electron density map of the monomer of the bluetongue virus outer shell coat protein VP7 (PDB ID: 2BTV): A: simulated density map at SNR of 6 dB at 10 Å resolution, B: NNBD reconstruction, C: result of embfactor, D: median-filtered result of embfactor.

its resolution by almost a factor of two. Figures 5.5 C and F depict the true and estimated PSFs. The overall shape and functional form is determined correctly, however the estimated bandwidth appears to be smaller. This shrinkage of the PSF is largely due to the smoothness prior that downweights high-frequency components, which causes a loss of structural details but, at the same time, prevents amplification of noise. In this sense, underestimation of the bandwidth is conservative and should be viewed as a feature rather than a shortcoming.

Further insight is obtained by looking at the Guinier plot in Figure 5.7 showing the radially averaged power spectrum against the squared resolution. In physical terms, the Guinier plot quantifies the map's energy content at various spatial frequencies. Blurring has the effect that the Guinier plot drops off quite rapidly – convolution with a broad PSF acts as a low-pass filter that deletes all information above a certain cutoff frequency. The NNBD algorithm is able to recover high-frequency information to a large extent and lifts the Guinier curve above the curve of the simulated density map at a resolution of 6 Å (orange line in Figure 5.7).

To study the influence of noise, we corrupted the simulated density maps with Gaussian noise at different SNRs. We used the program *proc3d* from the EMAN software package (Ludtke et al., 1999) for noise corruption. Figure 5.8 A shows a noisy 10 Å-density map at a SNR of 6 dB. Figure 5.8 B shows the corresponding NNBD reconstruction using a smoothness prior (5.10). For comparison Figure 5.8 C shows the density map sharpened with *embfactor* (Fernández et al., 2008; Rosenthal and Henderson, 2003), the state-of-the-art method within the field.

### 5.4.2 Incorporating Homologous Structure Information

We now demonstrate how additional information from homologous structures can be incorporated to aid the deconvolution process and to detect secondary structure. We use the trimeric structure of the bluetongue virus capsid protein VP7 (PDB ID: 2BTV) as an example. Figures 5.9 A-C show the molecular structure, a top and side view of the simulated density of 2BTV at a resolution of 8 Å. The protein is made up of $\beta$-sheets and $\alpha$-helices in

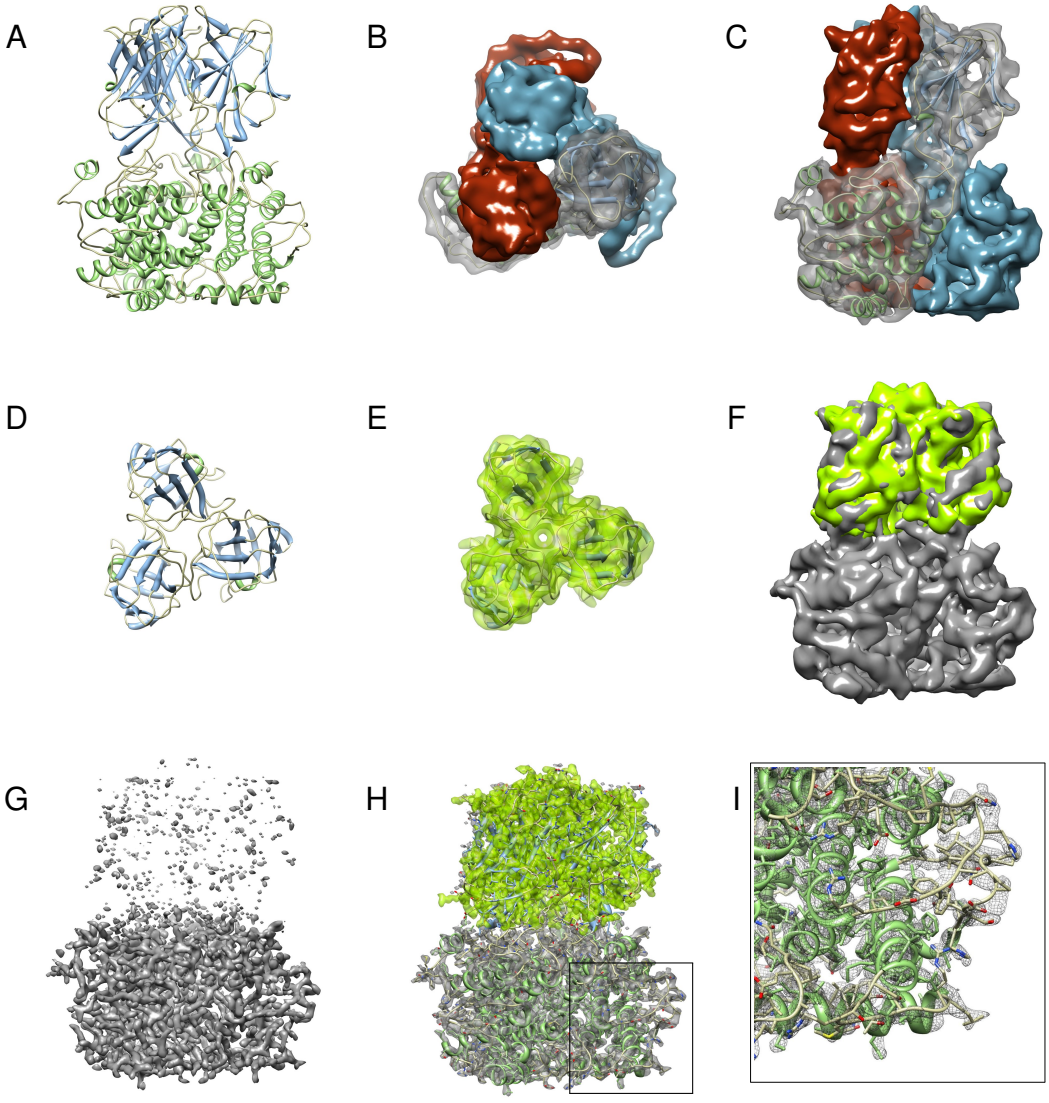



Figure 5.9: NNBD for the electron density map of the bluetongue virus capsid protein (PDB ID: 2BTV) using additional structural information from a homologous fold. Top row: A: molecular structure of trimer 2BTV, B: top view of simulated density map of 2BTV at 8 Å resolution, C: sideview. Middle row: D: molecular structure of the African horse sickness virus capsid protein (PDB ID: 1AHS), E: simulated density map of 1AHS at 8 Å resolution, F: density map of 1AHS fitted into the map of 2BTV by FOLDHUNTER. Bottom row: G: NNBD of 2BTV without density map of homologous fold, H: molecular structure of 2BTV fitted into the density map, I: closeup view of H. See text for details.

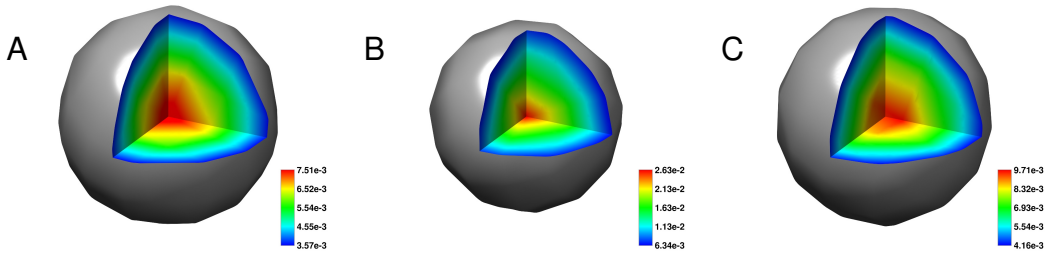



Figure 5.10: Comparison of true PSF (A) and the PSFs estimated by our approach without (B) and with homologous structure information (C). Note that (C) much closer resembles the true PSF as evident when comparing the intensity profiles. Hence, the incorporation of additional knowledge facilitates PSF estimation and improves accuracy.

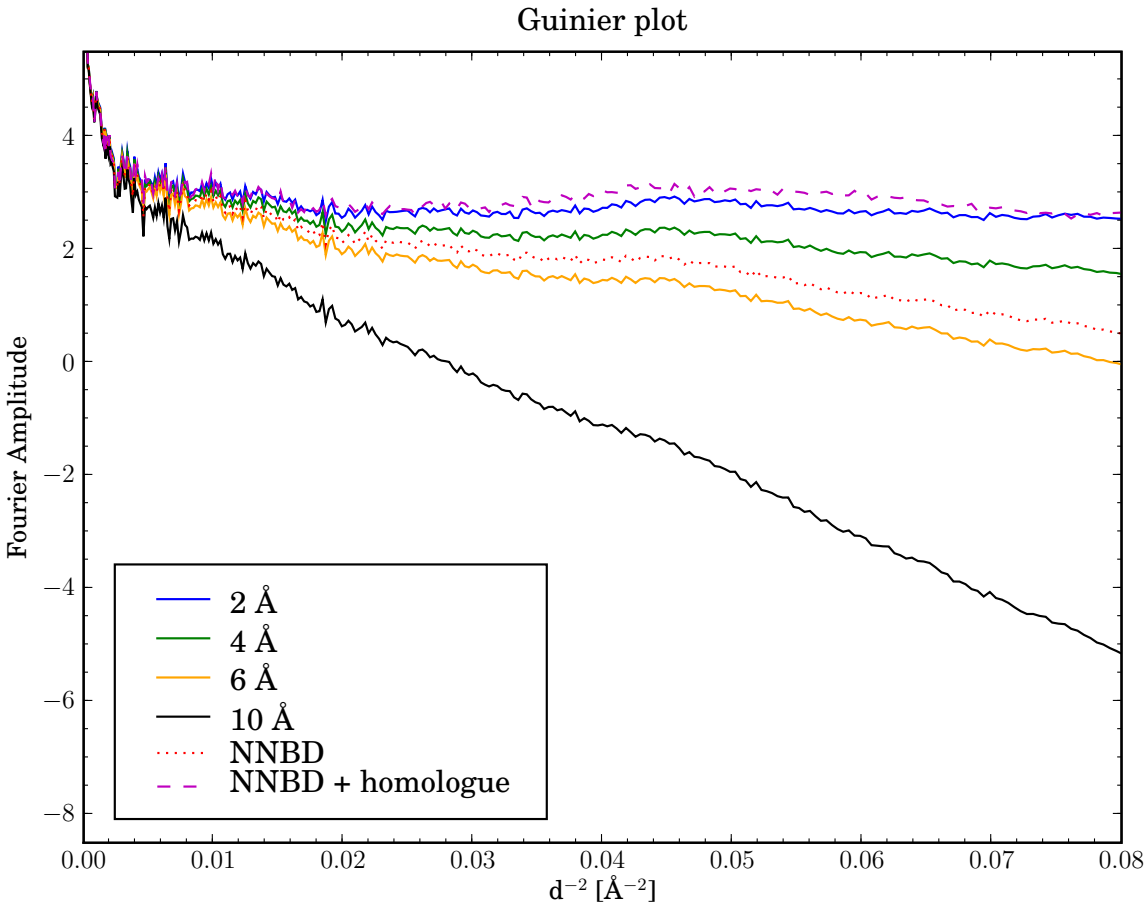



Figure 5.11: Guinier plot of reconstructed density maps with (magenta dashed line) and without homologous structure information (red dotted line). Incorporating additional knowledge in form of homologous structure information improves the resolution of the reconstructed density map from about 5 Å to 2 Å and therewith yields almost atomic resolution.

the upper and lower domains, respectively. The African horse sickness virus capsid protein (PDB ID: 1AHS) is a close structural homologue (RMSD: 1.4 Å) to the all-beta domain of 2BTV. Figures 5.9 D-F display the molecular structure, the simulated density at a resolution of 8 Å and the fit of 1AHS into 2BTV provided by FOLDHUNTER (Jiang et al., 2001). In B-factor sharpening, information from homologous folds is used to compute the optimal B-factor for density sharpening. In our blind deconvolution approach, we model the observed density map as being composed of the homologous structure simulated at a higher resolution and the remainder density of 2BTV. The density of the homologous fold is held fixed, only the missing density and the PSF are estimated during the deconvolution. As initial PSF, we use a Gaussian at 6 Å resolution corresponding to the resolution difference between the high-resolution density of 1AHS at 2 Å and the experimental density. During reconstruction, we apply the orthogonality constraint (5.12) to enforce that the 1AHS density and the unexplained region of 2BTV do not overlap. The result of NNBD is shown in Figs. 5.9 G-I. As clearly visible in the closeup (Figure 5.9 I), the sharpened density map reveals sidechains and information with almost atomic resolution. Figures 5.10 A-C compare the true PSF and the PSFs estimated by NNBD with and without homologous structure. As in the previous example, the width of the PSF is underestimated due to the smoothness prior. However, the additional structural information facilitates a more accurate estimation of the PSF (Figure 5.10 C) and thereby allows the restoration of a high-resolution density map (Figure 5.9 I). The Guinier plot in Figure 5.11 illustrates the improved recovery of high-frequency information and the increase in resolution.

## 5.5 Results on Experimental Density Maps

To test our proposed algorithm on real-world data, we downloaded a single particle reconstruction of GroES-ADP7-GroEL-ATP7 from the EMD database (EMD-ID: 1046), whose resolution was assessed as 23.5 Å by its authors (Ranson et al., 2001). Figure 5.12 shows a side and top view of both the original and the reconstructed density map after 1900 iterations. The estimated map reveals much more detail and exhibits structure that is not visible in the original density map. To assess the quality of the reconstruction we fitted the corresponding pdb structure which we downloaded from the PDBe[5] database (PDB-ID: 1gru) into the density map. The result is shown in Figure 5.13. Since the PDB structure was determined using the original density map at 23.5 Å, we do not expect a perfect fit. However, close inspection shows good agreement in places and even sidechains become visible, which suggests that a refinement of the proposed PDB structure with the help of the sharpened density map might be possible. For comparison we also show the results of B-factor sharpening, where we used the implementation of Fernández et al. (2008). The resulting density is of poor quality, which is probably due to unwanted noise amplification. Since the density map looks visually quite scattered, any meaningful structure determination becomes problematic, if not impossible.

Another example which demonstrates the usefulness of our method is illustrated in Fig-

[5] http://www.ebi.ac.uk/pdbe/

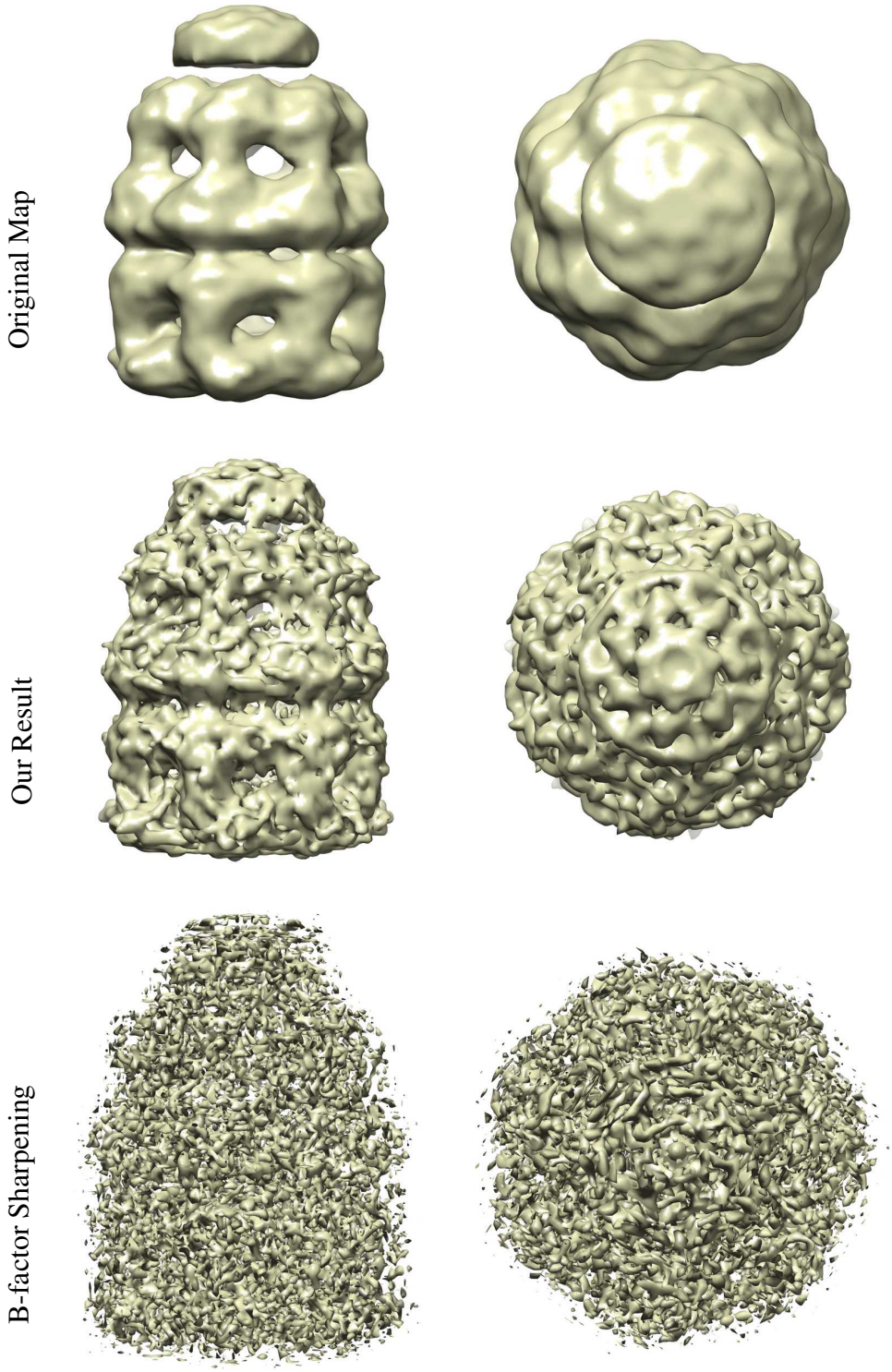



Figure 5.12: Sharpening of experimental density map: The top panel shows a side and top view of a single particle reconstruction of GroES-ADP7-GroEL-ATP7 at 23.5 Å downloaded from the EMD database (EMD-ID: 1046). The middle panel shows the result of our approach after 1900 iterations, which is clearly superior to the result of B-factor sharpening (shown in the bottom panel).

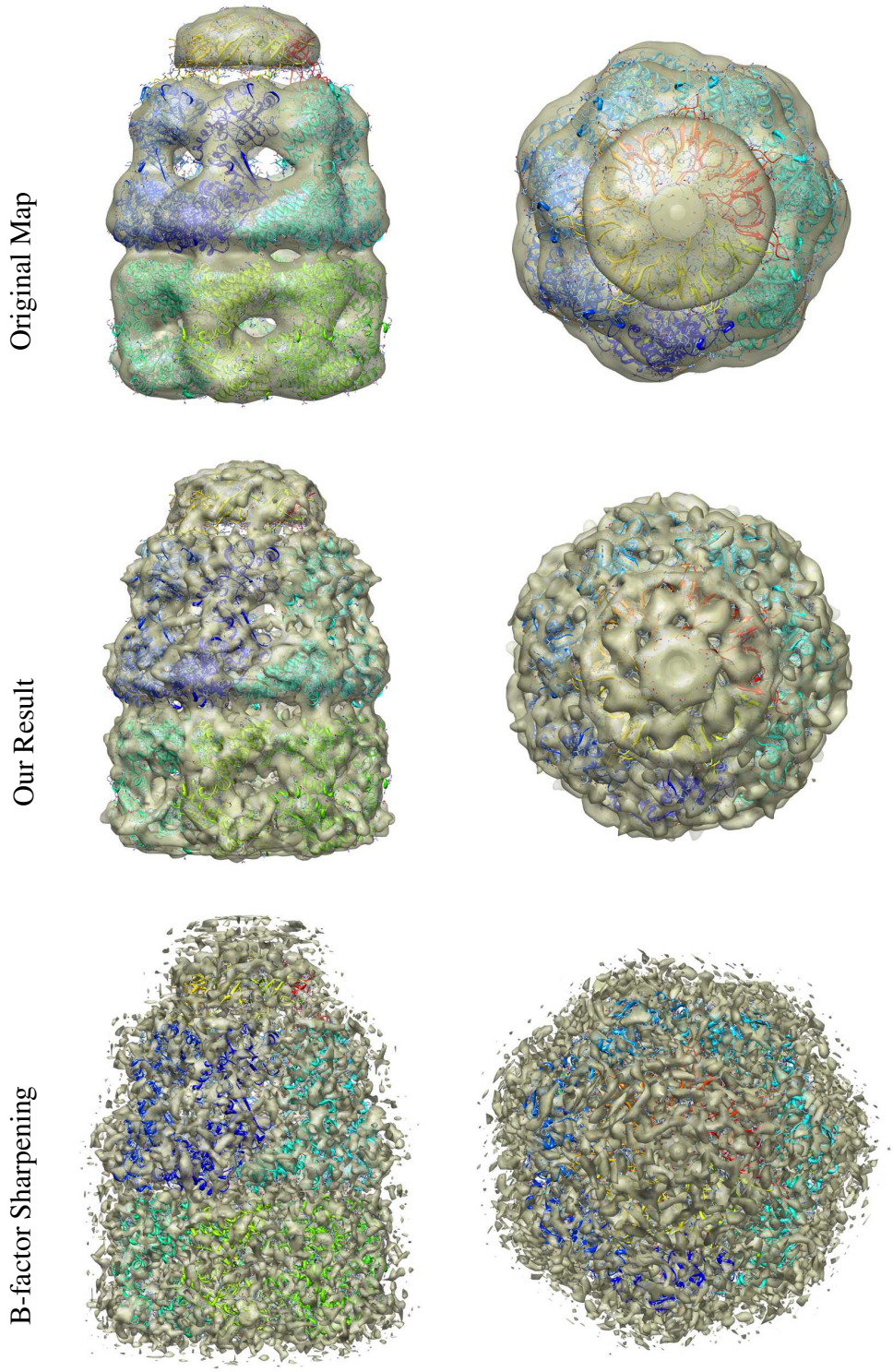



Figure 5.13: Sharpening of experimental density map: The top panel shows a side and top view of a single particle reconstruction of GroES-ADP7-GroEL-ATP7 at 23.5 Å downloaded from the EMD database (EMD-ID: 1046) with the proposed PDB structure of Ranson et al. (2001) fitted into it. The middle panel shows the result of our approach after 1900 iterations, which is clearly superior to the result of B-factor sharpening (shown in the bottom panel).

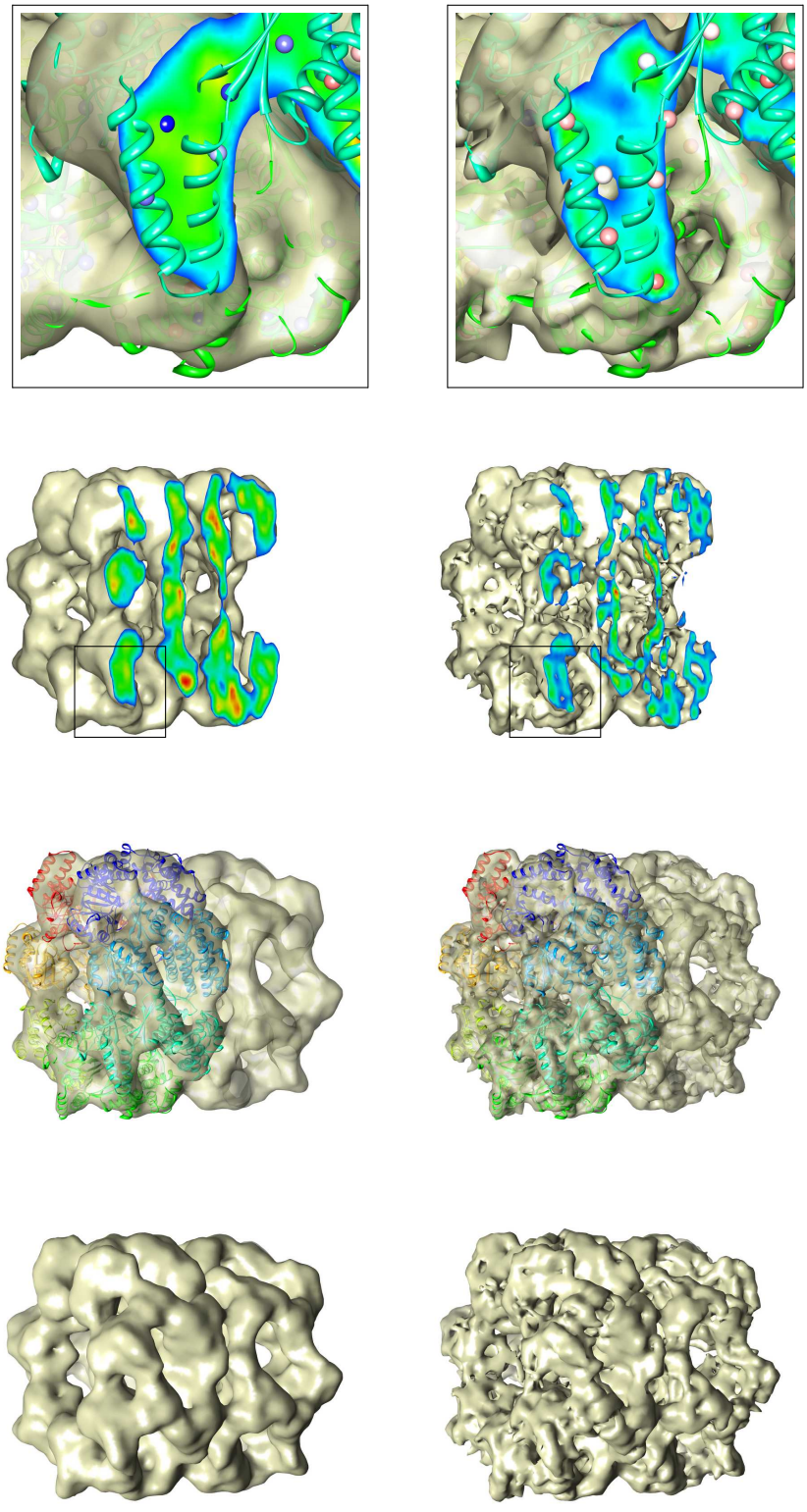



Figure 5.14: Sharpening of experimental density map and automatic secondary structure annotation via SSEHunter: The top panel shows a single particle reconstruction of GroEL at 11.5 Å downloaded from the EMD database (EMD-ID: 1080). The bottom panel shows the result of our algorithm after 150 iterations. The closeup reveals the increased information content, which helps to facilitate automatic secondary structure annotation. The depicted balls are automatically assigned pseudo-atoms whose colour encodes their probability for belonging to a helix (red) or a beta strand (blue) respectively.

ure 5.14. The top panel shows a single particle reconstruction of GroEL with a assessed resolution of 11.5 Å (Ludtke et al., 2001), which we downloaded from the EMD database (EMD-ID: 1080). The bottom panel shows the result of NNBD after 150 iterations. Beside the fitted PDB structure (PDB-ID:1OEL), we also show the intensity profile of a vertical slice. The aggregation of electron density and the sharper contouring is clearly visible. To assess the improvement in quality, we ran SSEHunter of Baker et al. (2007) for automatic secondary structure annotation on both the original and sharpened density maps. SSEHunter automatically assigns pseudo-atoms to the density to which a probability is assigned for belonging to a helix or a beta strand (red and blue balls in upright panel of Figure 5.14, respectively). The closeups in Figure 5.14 demonstrate how our proposed algorithm facilitates automatic structure determination by improving the resolution of electron density maps.

## 5.6 Conclusion and Outlook

In this chapter, we proposed a new method for improving the resolution of cryo-EM density maps by non-negative blind deconvolution. We provided an iterative algorithm for estimating simultaneously the sharpened density map and the blur kernel. We illustrated the generality of the proposed framework and showed that the derived updates allow for easy incorporation of prior knowledge such as smoothness and sparseness.The updates are multiplicative and do not require the adjustment of a learning rate, as opposed to previously proposed gradient descent techniques. In addition, the updates ensure the non-negativity of the sharp map and the PSF and guarantee convergence to a stationary point. A hierarchical Bayesian formulation also allowed us to derive update rules for the hyper-parameters, thus the method is fully parameter-free. The simplicity of the multiplicative updates allows for straightforward implementation. By employing the Fast Fourier Transform, we can reduce the computational complexity to large extent such that even medium and large sized problems (number of voxels $> 10^7$) can be tackled efficiently. Computation time is typically in the order of minutes to hours for large density maps ($> 400^3$) depending on the number of iterations one is willing to perform. Since our method allows the inspection of intermediate results, the user can decide when to stop either by visual inspection or by a user-set threshold of the monotonically decreasing cost function. We illustrated the performance and versatility of our algorithm by sharpening simulated electron density maps of the bluetongue virus capsid protein VP7 and by incorporating homologous structure information into the deconvolution process. Results on experimental density maps confirm that NNDB is a flexible and generic tool to improve the resolution of electron density maps.

## 5.7 Acknowledgements

The work presented in this chapter was carried out in collaboration with Michael Habeck and Bernhard Schölkopf of the Max Planck Institute for Intelligent Systems. Michael Habeck assisted the derivation of the algorithm, provided functionality for data handling and helped with the preparation of the conference submission. Bernhard Schölkopf contributed in giving scientific advice and in helping to prepare the manuscripts. Except otherwise explicitly

stated, all mathematical derivations, algorithmic implementations and experimental evaluations were performed by the author of this thesis.

This chapter is based on the following publications:

- *A new algorithm for improving the resolution of cryo-EM density maps*, Michael Hirsch, Bernhard Schölkopf and Michael Habeck. In 14th International Conference on Research in Computational Molecular Biology (RECOMB) 2010.
- *A Blind Deconvolution Approach for Improving the Resolution of Cryo-EM Density Maps*, Michael Hirsch, Bernhard Schölkopf and Michael Habeck. Journal of Computational Biology 18(3) 335-346, 2011.

# Chapter 6

# Conclusion and Outlook

In the last chapter, we summarise the contributions and findings of this thesis, discuss open problems and interesting future directions and conclude by giving a brief outline of some ongoing projects that are directly related to the work presented in this thesis.

## 6.1 Conclusive Summary

In this thesis, we developed and discussed an efficient and generic framework for the description and computation of spatially varying blur and presented novel solutions to challenging real-world applications in both scientific imaging as well as everyday photography.

### 6.1.1 Summary of Contributions

This thesis made contributions on three conceptual levels:

- **Theory.** In Chapter 2, we developed a mathematically sound and physically well-motivated derivation of our Efficient Filter Flow (EFF) framework, that allows to express and efficiently compute blur, which smoothly varies across the image plane. We derived our framework as a discrete approximation of the incoherent imaging equation, which extends the well-known and commonly employed invariant convolutional model, without, however, sacrificing its numerical amenities. It proved itself an essential and crucial ingredient in a number of challenging real-world applications (see Chapters 3 and 4), which demonstrated both the validity as well as the versatility of our approach.

- **Applications.** In Chapters 3, 4 and 5, we tackled a number of challenging real-world applications, for which we developed novel approaches which surpass current limitations and broaden the applicability of deconvolution methods in scientific imaging as well as everyday photography:

    - Chapter 3 addresses the problem of recovering a sharp latent image of a sequence of atmospherically blurred and noisy images. We developed a Blind

Deconvolution (BD) algorithm that processes the imagery in an online fashion, *i.e*. frame after frame, and is able to handle the effects of spatially varying blur, super-resolution and saturation correction. Encouraging results on simulated and real astronomical images demonstrated that our approach yields competitive, if not superior deblurring results when compared to existing approaches.

- In Chapter 4, we proposed a single image BD algorithm for removing non-uniform motion blur due to camera shake. By incorporating camera motion constraints into the EFF framework, we derived an algorithm that is not only substantially faster, but also leads to better deblurring results than existing approaches.
- Chapter 5 develops a novel BD algorithm for improving the resolution of three-dimensional density maps of macromolecular assemblies and large biomolecular complexes as obtained by cryo-electron microscopy. Results for both simulated and experimental density maps demonstrated that our proposed algorithm is a flexible and generic tool for facilitating structure elucidation of macromolecular assemblies.

- **Algorithms.** In Chapters 3, 4 and 5, we presented various BD algorithms that despite having been developed for a specific application, can be readily applied or easily adopted to other problem settings. The following section 6.2.2 will outline such a case, where we adopted the BD algorithm that has been developed for the sharpening of cryo-EM density maps (see Chapter 5), to an interesting problem in medical imaging. In this respect, the presented algorithms are noteworthy in their own right, as they might offer novel algorithmic solutions to similar inference problems.

### 6.1.2 Discussion and Outlook

Both benefits and shortcomings as well as limitations and possible extensions of the presented material have been discussed in great detail at the end of each chapter, hence we will not repeat them here. Instead, we will outline some problems that remain unsolved as well as interesting future directions:

- **Full Inference of EFF parameters.** In its current implementation, there are a couple of parameters within the EFF framework that necessitate manual adjustment. These parameters include the number of sampling sites which depends on the spatial correlation length of the point spread function (PSF), the support size of the local kernels and the interpolation scheme between them. As in some situations the spatial variance of the underlying PSF is position-dependent as *e.g*. it is often the case for lens aberrations, an approximation of the PSF by a non-uniform grid of sampling sites has the potential of improved accuracy and a speedup in computation time, which however comes at the cost of increased model complexity. As the EFF is linear in the parameter, which determines the interpolation scheme between neighbouring filters, an efficient inference scheme should be able to balance the trade-off between model complexity and goodness of approximation.

Particularly, in the context of astronomical imaging, some of the above-mentioned parameters are tightly connected to the statistical properties of the atmospheric turbulence. By incorporating this additional knowledge about the underlying physical properties, an inference scheme based on *e.g.* non-parametric Bayesian statistics could minimise human intervention and facilitate fully automatic image processing which is essential for the successful integration of the EFF framework into the data pipelines of future sky surveys.

- **Low Signal-To-Noise Ratio.** Many scientific imaging applications suffer from very low signal-to-noise-ratio (SNR). There are two ways of increasing the SNR for very faint signals: either increase the exposure time or the number of observations. There is a subtle trade-off between the two and Hasinoff et al. (2009) argues in the case of photography for the latter given a constraint time budget. However, all multi-frame deconvolution algorithms the author is aware of (incl. the online BD algorithm presented in Chapter 3) suffer from the presence of high noise and either lead to unwanted noise amplification or necessitate the use of additional regularisation terms. Often enough, simple averaging seems to be the method of choice due to a lack of efficient regularisation terms and robust deconvolution techniques. Evaluating and understanding the shortcomings of present deconvolution algorithms in the low SNR regime will help to improve current methods and to develop noise robust deconvolution techniques, thus broadening the applicability of deconvolution in scientific imaging.

- **Semi-blind deconvolution.** Commonly, non-blind and blind deconvolution are distinguished depending on whether the PSF or point response of the optical system is assumed to be known or unknown. However, in many situations partial information about the PSF is accessible by the use of additional measurement sensors, such as *e.g.* wavefront sensors or inertial measurement sensors. In other cases, the PSF describes the joint effect of several blurring sources, one of which can be analysed through additional measurement as *e.g.* in the case in astronomical imaging where blur is due to atmospheric turbulence and lens aberrations of the telescope optics. Only little work has been done along these lines which we refer to as *semi-blind* deconvolution. One interesting question that arises, is how to deal with uncertainties in the measured part of the PSF and how to optimally exploit the additional information available.

### 6.1.3 Conclusion

Overall, this thesis is witness of how various fields of research can mutually stimulate and enrich each other, often leading to fast progress and novel approaches to longstanding problems. In particular, based on the fundamental laws of physical optics, we developed a novel and generic framework for the description of spatially varying blur. By generalising the short-time Fourier transform commonly employed in the domain of audio and signal processing, we were able to devise an efficient implementation of the EFF. Being able to efficiently compute the gradients with respect to the model parameters enabled us the use of powerful gradient-based quasi-Newton methods from mathematical optimisation theory.

This in turn was crucial for its successful application to the various challenging real-world problems presented in this thesis.

The overall goal of this thesis was to take forward the state-of-the-art in the field and further the applicability of BD techniques in everyday photography as well as in scientific imaging. With the EFF, we presented for the first time a generic and versatile tool for modelling spatially varying blur which is a prevalent problem in many real-world applications. Through its efficient implementation and the provision of fast optimisation techniques for model parameter estimation, we hope to promote the use of more sophisticated but realistic imaging models and release researchers in digital image processing from having to make the frequently overly simplified assumption of an invariant PSF for the sake of computational tractability.

Our success in tackling various challenging real-world problems owes itself to a subtle trade-off between two integral parts of almost every computational approach: *modelling* and *inference*. The more accurate a model captures the underlying physics or dynamics of a problem, the more capable is a method in terms of quality and accuracy. On the other hand, the more complex a model is, the harder the inference of the underlying model parameters becomes, which might render a method computationally infeasible or intractable. The balancing between the two is often a delicate task, however at the same time also a key ingredient for the success of a computational method. One-sided viewpoints or the underestimation of one or the other, especially in fields where established models are rarely scrutinised, more often than not hinder or impede scientific progress. This thesis provides evidence for the importance of both *modelling* and *inference* in successfully approaching a challenging problem and suggests interdisciplinarity as an efficient catalyst.

## 6.2 Outlook and Perspectives

In this section, we will briefly outline some ongoing projects that are directly related to the work presented in this thesis and underpin the relevance and value of its findings.

### 6.2.1 Non-Stationary Correction of Optical Aberrations

This project directly builds on the EFF framework developed in Chapter 2, and is done in collaboration with Christian J. Schuler, Stefan Harmeling and Bernhard Schölkopf. First results have been published recently in (Schuler et al., 2011).

A direct application of the EFF is the modelling of optical aberrations, which describe departures of an optical system from an *ideal* optical system as described theoretically by paraxial optics: in an ideal optical system all light rays emitted by a point source converge to a single point in the focal plane, forming a clear and sharp image. Departures from this ideal behaviour are inevitable and the design of a lens is always a trade-off between various parameters, including price.

Manufacturers of photographic lenses attempt to minimise optical aberrations by combining several glass elements to so-called compound lenses comprising of as many as 15 elements or more, which renders high grade lenses probably the most expensive compo-

Figure 6.1: Self-made photographic lens with a 120mm focal length. Taken image without and with lens correction.

nents of high-end camera systems today.

Optical aberrations are commonly classified into:

- **Monochromatic aberrations** which cause blurring that varies across the image plane. This class of aberrations include:
  - **spherical aberration**, where the focal length is a function of the distance from the optical axis within a spherical lens
  - **coma** occurs in an oblique light bundle when the intersection of the rays is shifted with respect to its axis
  - **field curvature** occurs when the focal surface is non-planar
  - **astigmatism** occurs when the sagittal and tangential focal surfaces do not coincide (*i.e.*, the system is not rotationally symmetric for off axis light bundles)
- **Chromatic aberrations** caused by the wavelength dependence of the diffractive-index of most materials including glass.
- **Vignetting**, *i.e.* the intensity falls off towards the image corner due to absorption of oblique light bundles.

In (Schuler et al., 2011), we show that all of the above-mentioned optical aberrations can be well described by the EFF framework and even corrected for by digital image processing. By noting that aberrations of a linear optical system are fully described by their PSF, this does not come as a surprise, since in Chapter 2 the EFF framework was exactly derived as an efficient numerical approximation of a spatially varying PSF.

For a given lens/camera combination, the parameters of the EFF are determined via an automated calibration procedure that measures the PSF at a grid covering the image. We demonstrate that even for heavily degraded images, taken with a self-constructed lens consisting of a *single lens element* (see Figure 6.1), a reconstruction of a full-colour image, *i.e.*, all three colour channels at full resolution given a raw image is possible. Figure 6.2 shows degraded images taken with different lens/camera combinations that suffer from various monochromatic and chromatic aberrations. The result of our reconstruction significantly improves image quality. We believe that our work can have significant implications for the design of lenses by offering the possibility of drastically lowering the costs of current lenses.

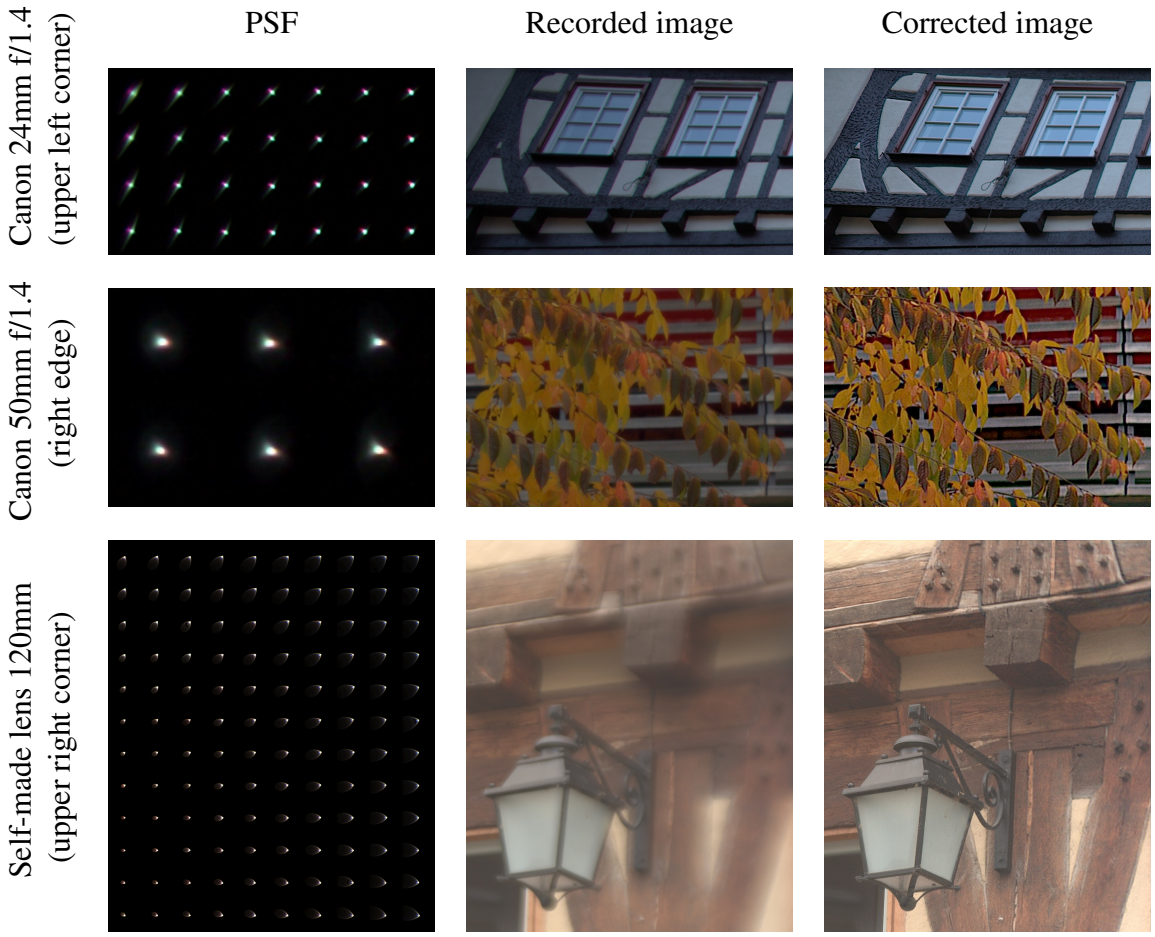



Figure 6.2: Comparison between original and corrected images and the respective PSFs. Only parts of the images are shown. For full resolution images please visit the project webpage linked below.

The accompanying project website contains many more images in full resolution and is readily accessible via:

### 6.2.2 A Blind Deconvolution Approach for Attenuation Map Prediction from MR Image Pairs

This project directly builds on the method developed in Chapter 5 and outlines an interesting application in medical imaging. The project is performed in collaboration with Matthias Hofmann, Frédéric Mantlik, Michael Habeck and Bernhard Schölkopf from the MPI for Intelligent Systems, Tübingen as well as Bernd Pichler from the Department of Preclinical Imaging and Radiopharmacy at the Eberhard-Karls-University Tübingen.

Recently, noninvasive hybrid imaging systems have been developed that combine Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI) (Judenhofer et al., 2008). PET visualises metabolic aspects of disease, while MRI produces high soft-tissue contrast and provides high-resolution anatomical information. Although it is therefore desirable to combine these two complementary modalities into one device, the development has posed many technical challenges. Prototypes from commercial manufacturers are now available, but one of the remaining problems for clinical adoption is that quantitative PET image reconstruction requires a so-called attenuation map that is typically obtained via an additional Computed Tomography (CT) scan.

Great efforts are made for circumventing the need of an additional CT scan. However, saving the patient from additional radiation exposure is possible only if one can generate the attenuation map from the information provided by MRI. This is a challenging problem because bone tissue contributes nearly no signal in standard MRI sequences and is therefore indistinguishable from air. In the attenuation image, bone and air are on opposite ends of the attenuation scale and it is therefore of particular importance that they can be distinguished reliably.

Recently, novel MRI sequences that use ultra short echo times (UTE) have been suggested as a promising approach. In UTE images, the signal is acquired within around 70$\mu$s of sending the excitation signal, which allows to yield signal even from solid structures such as cortical bone. In principle, the UTE image shows distinguishable responses for bone and air. However, as is visible in Figure 6.3, this comes at the price of lower image resolution compared to the typically used FLASH sequence protocol. The task is to harness the strengths from both images and to find a way to extract the available, though convoluted, information for attenuation map prediction.

Different methods such as (Keereman et al., 2008) have been proposed that make use of a pair of images, one acquired with an ultra short echo time and one with a normal echo time of around 1ms. These methods lack a principled approach and are based on heuristics. Through simple image arithmetics a pseudo CT image is generated that captures the essential features of the true CT image. However, the methods suffer from false bone prediction in some regions, *e.g.* the cerebellar hemisphere, and are sensitive to user set parameters.

In an ongoing project, we developed a novel approach that explicitly tackles the problem of resolution loss in the UTE image for the first time. We formulated the task as a BD problem and by exploiting the non-negativity we derived multiplicative updates for both the blur kernel and the sought-after pseudo CT image that are reminiscent of those presented in Chapter 5. Figure 6.3 illustrates that our method is able to predict bone with high accuracy using the MRI image pair only. The predicted bone image can be used to compute an improved attenuation map needed for PET reconstruction. Quantitative experiments will have to prove the performance and merit of our approach.

Although our method was developed specifically for attenuation and electron density estimation, it can be applied in the general case where two input images are provided, one image containing precise information about region boundaries and the other image containing intensity information about the region class.

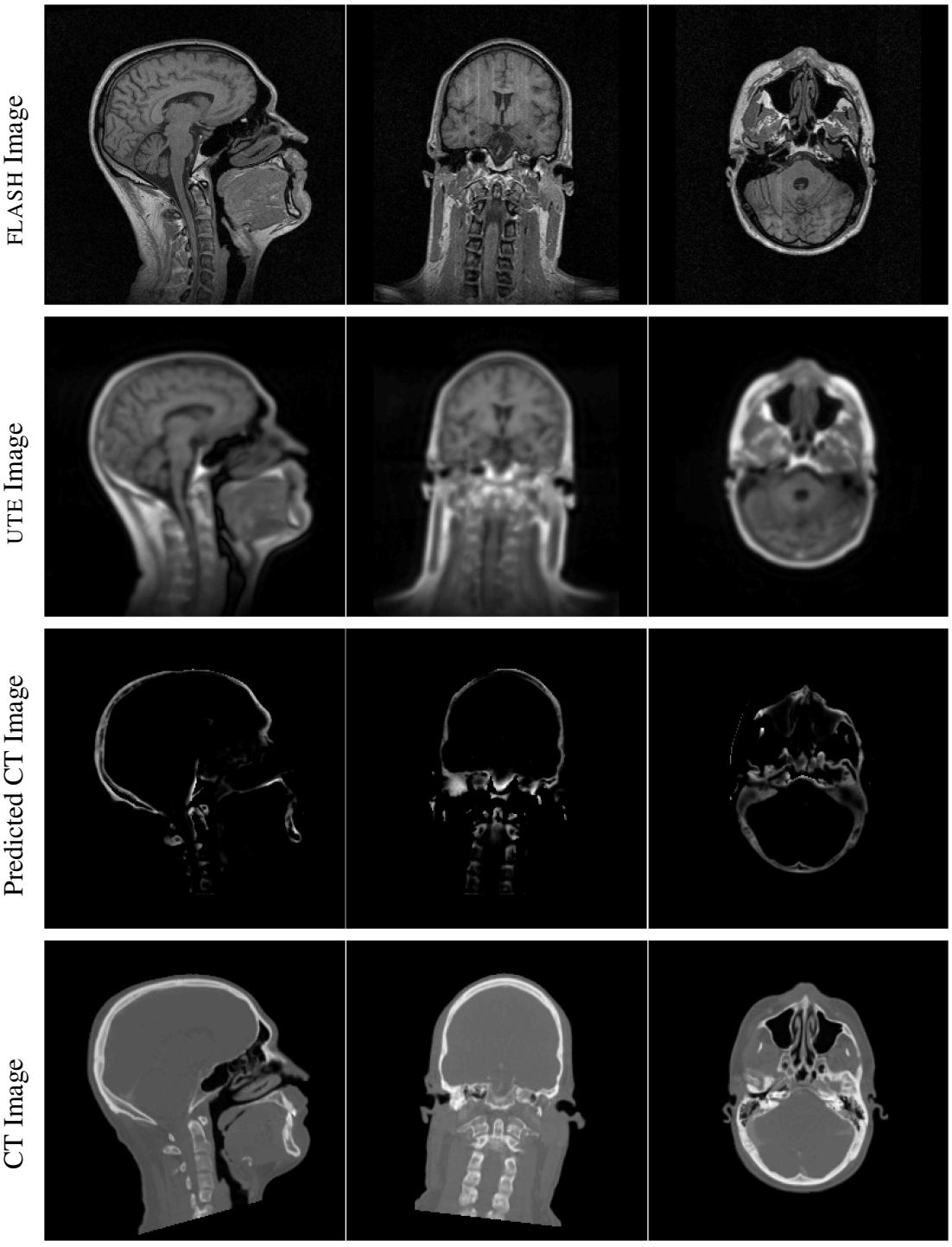



Figure 6.3: Mid-sagittal, coronal and transverse slice through human head. From top to bottom: FLASH, UTE, predicted bone and CT image. This figure is best viewed on screen, rather than in print.

# Appendix A

# Implementation Details

Although in sections 3.3 and 3.4 we only considered vectors, one-dimensional convolutions, and vector-norms, all results naturally generalize to two-dimensional images. However, efficiently implementing the resulting algorithms for two-dimensional images requires some care and handling of technical details.

## A.1 Convolution as Matrix-Vector Multiplication

We introduced $f * x$ as the convolution, which could be either circular or non-circular. Due to linearity and commutativity, we can also use matrix-vector notation to write

$$f * x = Fx = Xf. \tag{A.1}$$

The matrices $F$ and $X$ are given by

$$F = Z_y^\mathsf{T} \mathcal{F}^{-1} \operatorname{diag}(\mathcal{F} Z_f f) \mathcal{F} Z_x, \tag{A.2}$$

$$X = Z_y^\mathsf{T} \mathcal{F}^{-1} \operatorname{diag}(\mathcal{F} Z_x x) \mathcal{F} Z_f. \tag{A.3}$$

Matrix $\mathcal{F}$ is the discrete Fourier transform matrix, i.e., $\mathcal{F}x$ is the Fourier transform of $x$. The diagonal matrix $\operatorname{diag}(v)$ has vector $v$ along its diagonal, while $Z_x$, $Z_f$, and $Z_y$ are *zero-padding* matrices which ensure that $Z_x x$, $Z_f f$, and $Z_y y$ have the same length. Different choices of the matrices lead to different margin condition of the convolution.

For two-dimensional images and PSFs we have to consider two-dimensional Fourier transforms, which can be written as left- and right-multiplications with $\mathcal{F}$, and represented as a single matrix-vector multiplication using Kronecker products and the vectorization operator $\operatorname{vec}(x)$, which stacks columns of the two-dimensional image $x$ into a one-dimensional vector in lexicographical order; formally,

$$\operatorname{vec}(W x\, W) = (W \otimes W) \operatorname{vec}(x), \tag{A.4}$$

which follows from the identity (Horn and Johnson, 1991)

$$\operatorname{vec}(A\, B\, C^\mathsf{T}) = (C \otimes A) \operatorname{vec}(B). \tag{A.5}$$

The zero-padding operations for two-dimensional images can be written in a similar way.

## A.2 Resizing Matrices

The resizing matrix $D_n^m$ can be implemented efficiently using sparse matrices[1]. Resizing two-dimensional images can also be implemented by left- and right-multiplications: let $x$ be an $m \times n$ image, then $D_p^m \, x \, (D_q^n)^\mathsf{T}$ is an image of size $p \times q$. Using Eq. (A.5) we can write this operation as the matrix-vector product

$$\mathrm{vec}(D_p^m \, x \, (D_q^n)^\mathsf{T}) = (D_q^n \otimes D_p^m) \, \mathrm{vec}(x). \tag{A.6}$$

[1]Defining for instance in Octave, `D = kron(speye(m), ones(n,1)')*kron(speye(n), ones(m,1))/m;` the matrix-vector product `D*v` will be calculated efficiently.

# Bibliography


H.-M. Adorf. Towards HST restoration with a space-variant PSF, cosmic rays and other missing data. *The Restoration of HST Images and Spectra II*, pages 72–78, 1994.

H.M. Adorf, RN Hook, and LB Lucy. HST image restoration developments at the ST-ECF. *International Journal of Imaging Systems and Technology*, 6(4):339–349, 1995.

J. Allen. Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 25(3):235–238, 1977.

B. Anconelli, M. Bertero, P. Boccacci, M. Carbillet, and H. Lanteri. Reduction of boundary effects in multiple image deconvolution with an application to LBT LINC-NIRVANA. *Astronomy & Astrophysics*, 448:1217–1224, 2006.

G. R. Ayers and J. C. Dainty. Iterative blind deconvolution method and its applications. *Optics Letters*, 13:547–549, 1988.

M. L. Baker, T. Ju, and W. Chiu. Identification of secondary structure elements in intermediate-resolution density maps. *Structure*, 15:7–19, 2007.

J. Bardsley, S. Jefferies, J. Nagy, and R. Plemmons. Blind iterative restoration of images with spatially-varying blur. In *AMOS Technical Conference Proceedings*, 2005.

J. Bardsley, S. Jeffries, J. Nagy, and B. Plemmons. A computational method for the restoration of images with an unknown, spatially-varying blur. *Optics Express*, 14(5):1767–1782, 2006.

K.R. Barnes. *The Optical Transfer Function*. Hilger, London, 1971.

J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141, 1988.

Y. Biraud. A new approach for increasing the resolving power by data processing. *Astronomy & Astrophysics*, 1:124–127, 1969.

L. Bottou. Online algorithms and stochastic approximations. In David Saad, editor, *Online Learning and Neural Networks*, pages 9–42. Cambridge University Press, 1998.

A. T. Brünger. Low-resolution crystallography is coming of age. *Structure*, 13:171–172, 2005.

R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal of Scientific Computing*, 16(5):1190–1208, 1995.

W. Chiu, M. L. Baker, W. Jiang, M. Dougherty, and M. F. Schmid. Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure*, 13:363–372, 2005.

S. Cho and S. Lee. Fast motion deblurring. In *ACM Transactions on Graphics (SIGGRAPH ASIA)*, 2009.

S. Cho, J. Wang, and S. Lee. Handling outliers in non-blind image deconvolution. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011.

T. S. Cho, A. Levin, F. Durand, and W. T. Freeman. Motion blur removal with orthogonal parabolic exposures. In *IEEE International Conference in Computational Photography (ICCP)*, 2010.

M. E. Daube-Witherspoon and G. Muehllehner. An Iterative Image Space Reconstruction Algorthm Suitable for Volume ECT. *IEEE Transactions on Medical Imaging*, 5(2):61–66, 1986.

B. DeLaBarre and A. T. Brunger. Considerations for the refinement of low-resolution crystal structures. *Acta Crystallographica D*, 62:923–932, 2006.

L. Denis, Thiébaut E., and F. Soulez. Fast Model of Space-Variant Blurring and Its Applications to Deconvolution in Astronomy. In *IEEE International Conference on Image Processing (ICIP)*, 2011.

S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2):47–57, 2004.

R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *ACM Transactions on Graphics (SIGGRAPH)*, 2006.

J. J. Fernández, D. Luque, J. R. Castón, and J. L. Carrascosa. Sharpening high resolution information in single particle electron cryomicroscopy. *Journal of Structural Biology*, 164:170–175, 2008.

D. J. Field. What is the goal of sensory coding? *Neural computation*, 6(4):559–601, 1994.

J. Frank. Single-particle imaging of macromolecules by cryo-electron microscopy. *Annual Review of Biophysics and Biomolecular Structure*, 31:303–319, 2002.

D. L. Fried. Probability of getting a lucky short-exposure image through turbulence. *Journal of the Optical Society of America*, 86:1651–1657, 1978.

K. J. Friston, J. Ashburner, C. D. Frith, J. B. Poline, J. D. Heather, and R. S. J. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 2(1):1–25, 1995.

J. W. Goodman. *Introduction to Fourier optics*. McGraw-Hill, 2005.

J. M. Grimes, J. N. Burroughs, P. Gouet, J. M. Diprose, R. Malby, S. Ziéntara, P. P. C. Mertens, and D. I. Stuart. The atomic structure of the bluetongue virus core. *Nature*, 395: 470–478, October 1998.

A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2010.

C. M. Harding, R. A. Johnston, and R. G. Lane. Fast Simulation of a Kolmogorov Phase Screen. *Applied Optics*, 38:2161–2170, 1999.

S. Harmeling, M. Hirsch, S. Sra, and B. Schölkopf. Online blind image deconvolution for astronomy. In *Proceedings of the IEEE Conference on Computational Photography*, 2009.

S. Harmeling, M. Hirsch, and B. Schölkopf. Space-variant single-image blind deconvolution for removing camera shake. In *Advances in Neural Information Processing Systems (NIPS)*, 2010a.

S. Harmeling, S. Sra, M. Hirsch, and B. Schölkopf. Multiframe Blind Deconvolution, Super-resolution, and Saturation Correction via Incremental EM. In *IEEE International Conference on Image Processing (ICIP)*, 2010b.

S. W. Hasinoff. *Variable-Aperture Photography*. PhD thesis, University of Toronto, 2008.

S. W. Hasinoff, K. N. Kutulakos, F. Durand, and W. T. Freeman. Time-constrained photography. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2009.

E. Hecht. *Optics*. Pearson, 4th edition, 2003.

B. Hinman, J. Bernstein, and D. Staelin. Short-space Fourier transform image processing. *IEEE International Conference on ICASSP*, 9:166–169, 1984.

M. Hirsch, S. Sra, B. Schölkopf, and S. Harmeling. Efficient filter flow for space-variant multiframe blind deconvolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

M. Hirsch, S. Harmeling, S. Sra, and B. Schölkopf. Online multi-frame blind deconvolution with super-resolution and saturation correction. *Astronomy & Astrophysics*, 531:A9, 2011.

M. Hirsch, C. J. Schuler, S. Harmeling, and B. Schölkopf. Fast removal of non-uniform camera-shake. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011.

F. Hormuth, S. Hippler, W. Brandner, K. Wagner, and T. Henning. AstraLux: the Calar Alto lucky imaging camera. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7014 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, August 2008.

R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

W. Jiang, M. L. Baker, S. J. Ludtke, and W. Chiu. Bridging the Information Gap: Computational Tools for Intermediate Resolution Structure Interpretation. *Journal of Molecular Biology*, 308:1033–1044, 2001.

B. Jin and J. Zou. Augmented Tikhonov regularization. *Inverse Problems*, 25(2):5001, 2009.

R. A. Johnston, T. J. Connolly, and R. G. Lane. An improved method for deconvolving a positive image. *Optics Communications*, 181:267–278, 2000.

N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski. Image deblurring using inertial measurement sensors. In *ACM Transactions on Graphics (SIGGRAPH)*, 2010.

M. S. Judenhofer, H. F. Wehrl, D. F. Newport, C. Catana, M. Becker, S. Widmaier, A. Thielscher, M. Kneilling, K. Uludag, G. Reischl, H.-J. Machulla, R. E. Nutt, S. R. Cherry, C. D. Claussen, and B. J. Pichler. Simultaneous PET/MRI: A new perspective for functional and morphological imaging. *Nature Medicine*, 14:459–465, 2008.

V. Keereman, S. Vandenberghe, Y. De Deene, R. Luypaert, T. Broux, and I. Lemahieu. MR-based attenuation correction for PET using an Ultrashort Echo Time (UTE) sequence. In *IEEE Nuclear Science Symposium Conference Record, 2008. NSS'08*, pages 4656–4661, 2008.

A. Klöckner, N. Pinto, Y. Lee, B. C. Catanzaro, P. Ivanov, and A. Fasih. PyCUDA: GPU Run-Time Code Generation for High-Performance Computing. 2009. URL `http://arxiv.org/abs/0911.3456`.

K. T. Knox and B. J. Thompson. Recovery of images from atmospherically degraded short-exposure photographs. *Astrophysical Journal*, 193:L45–L48, 1974.

K. Königsberger. *Analysis 2*. Springer, 2002.

D. Krishnan and R. Fergus. Fast image deconvolution using hyper-Laplacian priors. In *Advances in Neural Information Processing Systems (NIPS)*, 2009.

D. Kundur and D. Hatzinakos. Blind image deconvolution. *Signal Processing Magazine*, 13(3):43–64, 1996.

H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Applications of Mathematics. Springer-Verlag, second edition, 2003.

A. Labeyrie. Attainment of diffraction limited resolution in large telescopes by fourier analysing speckle patterns in star images. *Astronomy & Astrophysics*, 6(1):85–87, 1970.

N. M. Law, C. D. Mackay, and J. E. Baldwin. Lucky imaging: high angular resolution imaging in the visible from the ground. *Astronomy & Astrophysics*, 446:739–745, February 2006.

A. Levin. Blind motion deblurring using image statistics. In *Advances in Neural Information Processing Systems (NIPS)*, 2006.

A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

B. Li, Z. Cao, N. Sang, and T. Zhang. Restoration of turbulence degraded images using improved multi-frame blind deconvolution algorithm. *Electronics Letters*, pages 1478–79, 2004.

Y. Lin and D. D. Lee. Bayesian regularization and nonnegative deconvolution for time delay estimation. In *Advances in Neural Information Processing Systems (NIPS)*, 2005.

M. G. Löfdahl. Multi-frame blind deconvolution with linear equality constraints. *Proceedings of the SPIE*, 4792:146–155, 2002.

A. W. Lohmann and D. P. Paris. Space-variant image formation. *Journal of the Optical Society of America*, 55(8), 1965.

A. W. Lohmann, G. Weigelt, and B. Wirnitzer. Speckle masking in astronomy - Triple correlation theory and applications. *Applied Optics*, 22:4028–4037, 1983.

A. Loktyushin and S. Harmeling. Automatic foreground-background refocusing. In *IEEE International Conference on Image Processing (ICIP)*, 2011.

L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745–754, 1974.

S. J. Ludtke, P. R. Baldwin, and W. Chiu. EMAN: semi-automated software for high-resolution single-particle reconstructions. *Journal of Structural Biology*, 128:82–97, 1999.

S. J. Ludtke, J. Jakana, J. L. Song, D. T. Chuang, and W. Chiu. A 11.5Å single particle reconstruction of GroEL using EMAN1. *Journal of Molecular Biology*, 314(2):253–262, 2001.

C. D. Mackay, R. N. Tubbs, R. Bell, D. J. Burt, P. Jerram, and I. Moody. Subelectron read noise at MHz pixel rates. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 4306, pages 289–298, 2001.

D. J. C. Mackay. Hyperparameters: Optimize, or integrate out? In *Maximum Entroy and Bayesian Methods*, pages 43–59, 1996.

G.F. Margrave. Theory of nonstationary linear filtering in the fourier domain with application to time-variant filtering. *Geophysics*, 63(1):244–259, 1998.

C. L. Matson, K. Borelli, S. Jefferies, C. C. Jr. Beckner, E. K. Hege, and M. Lloyd-Hart. Fast and optimal multiframe blind deconvolution algorithm for high-resolution ground-based imaging of space objects. *Applied Optics*, 48, 2008.

K. Mikurda and O. v. d. Lühe. High resolution solar speckle imaging with the extended knox–thompson algorithm. *Solar Physics*, 235(1):31–53, 2006.

J. Miskin and D. J. C. MacKay. Ensemble learning for blind image separation and deconvolution. In *Advances Independent Component Analysis*, 2000.

R. Molina, J. Mateos, and A. K. Katsaggelos. Blind deconvolution using a variational approach to parameter, image, and blur estimation. *IEEE Transactions on Image Processing*, 15:3715–3727, 2006.

J. H. Money and S. H. Kang. Total variation minimizing blind deconvolution with shock filter reference. *Image and Vision Computing*, 26(2):302–314, 2008.

J. G. Nagy and D. P. O'Leary. Fast iterative image restoration with a space-varying psf. *Advanced Signal Processing Algorithms: Architectures, and Implementations IV*, 3162: 388–399, 1997.

J. G. Nagy and D. P. O'Leary. Restoring images degraded by spatially variant blur. *SIAM Journal on Scientific Computing*, 19(4):1063–1082, 1998.

E. V. Orlova and H. R. Saibil. Structure determination of macromolecular assemblies by single-particle analysis of cryo-electron micrographs. *Curr. Opin. Struct. Biol.*, 14:584–590, 2004.

A. Oscoz, R. Rebolo, R. López, A. Pérez-Garrido, J. A. Pérez, S. Hildebrandt, L. F. Rodríguez, J. J. Piqueras, I. Villó, J. M. González, R. Barrena, G. Gómez, A. García, P. Montañés, A. Rosenberg, E. Cadavid, A. Calcines, A. Díaz-Sánchez, R. Kohley, Y. Martín, J. Peñate, and V. Sánchez. FastCam: a new lucky imaging instrument for medium-sized telescopes. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7014 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, August 2008.

S. Osher and L. I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal of Numerical Analysis*, 27(4):919–940, 1990.

E. Pantin, J. L. Starck, and F. Murtagh. *Deconvolution and Blind Deconvolution in Astronomy: Theory and Applications*. CRC Press, 2007.

S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction. *IEEE Signal Processing Magazine*, 2003.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes (3rd ed.): the art of scientific computing*. Cambridge University Press, 2007.

N. A. Ranson, G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich, and H. R. Saibil. Atp-bound states of groel captured by cryo-electron microscopy. *Cell*, 107 (7):869–879, 2001.

R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *ACM Transactions on Graphics (SIGGRAPH)*, 2006.

W. H. Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59, 1972.

P. B. Rosenthal and R. Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, 333:721–745, 2003.

P. Sarder and A. Nehorai. Deconvolution methods for 3-d fluorescence microscopy images. *Signal Processing Magazine, IEEE*, 23(3):32–45, 2006.

C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf. Non-stationary correction of optical aberrations. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011.

T. J. Schulz. Multiframe blind deconvolution of astronomical images. *Journal of the Optical Society of America*, 10:1064–1073, 1993.

F. Sha, Y. Lin, L. K. Saul, and D. D. Lee. Multiplicative updates for nonnegative quadratic programming. *Neural Computation*, 19(8):2004–2031, 2007.

Q. Shan, W. Xiong, and J. Jia. Rotational motion deblurring of a rigid object from a single image. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2007.

Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *ACM Transactions on Graphics (SIGGRAPH)*, 2008.

D. G. Sheppard, B. R. Hunt, and M. W. Marcellin. Iterative multiframe superresolution algorithms for atmospheric-turbulence-degraded imagery. *Journal of the Optical Society of America A*, 15:978–992, 1998.

M. Šorel and F. Šroubek. Space-variant deblurring using one blurred and one underexposed image. In *IEEE International Conference on Image Processing (ICIP)*, 2009.

F. Šroubek, G. Cristobál, and J. Flusser. A Unified Approach to Superresolution and Multichannel Blind Deconvolution. *IEEE Transactions on Image Processing*, 2007.

F. Šroubek, G. Cristobál, and J. Flusser. Simultaneous super-resolution and blind deconvolution. *Journal of Physics: Conference Series*, 124:2048 (8pp), 2008.

J. L. Starck, E. Pantin, and F. Murtagh. Deconvolution in Astronomy: A Review. *The Publications of the Astronomical Society of the Pacific*, 114:1051–1069, 2002.

C. Stelzer. Speckle1. http://www.tat.physik.uni-tuebingen.de/ stelzer/, 2009. version number 0.1.2.

C. Stelzer and H. Ruder. Suppressing anisoplanatism effects in speckle interferometry. *Astronomy & Astrophysics*, 475:771–774, November 2007.

T. G. Jr. Stockham. High-speed convolution and correlation. In *Proceedings of the April 26-28, 1966, Spring joint computer conference*, pages 229–233. ACM, 1966.

Y.-W. Tai, P. Tan, and M. S. Brown. Richardson-lucy deblurring for scenes under a projective motion path. *Technical Report of Korea Advanced Institue of Science and Technology (KAIST)*, 2009.

Y.-W. Tai, N. Kong, S. Lin, and S. Y. Shin. Coded exposure imaging for projective motion deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

Y.-W. Tai, P. Tan, and M. S. Brown. Richardson-lucy deblurring for scenes under a projective motion path. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2011.

M. Theusner. Avistack. http://www.avistack.de/index_en.html, 2009. Version number 1.80.

C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1998.

H.J. Trussell and S. Fogel. Identification and restoration of spatially variant motion blurs in sequential images. *Image Processing, IEEE Transactions on*, 1(1):123–126, 1992.

R.L. White. Restoration of images and spectra from the Hubble Space Telescope. In *Astronomical data analysis software and systems I*, volume 25, page 176, 1992.

O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2010.

L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Image deblurring with blurred/noisy image pairs. In *ACM Transactions on Graphics (SIGGRAPH)*, 2008.

J. Zhang, Q. Zhang, and G. He. Blind deconvolution of a noisy degraded image. *Applied. Optics*, 48(12):2350–2355, 2009.

C. Zhu, R. Byrd, and J. Nocedal. L-bfgs-b: Algorithm 778: L-bfgs-b, fortran routines for large scale bound constrained optimization. 23(4):550, 1997.

X. Zhu and P. Milanfar. Stabilizing and deblurring atmospheric turbulence. In *IEEE International Conference in Computational Photography (ICCP)*, 2011.

Y. V. Zhulina. Multiframe blind deconvolution of heavily blurred astronomical images. *Applied Optics*, 45(28):7342–7352, 2006.

# Notation

### Matrices & vectors

| | |
|---|---|
| $y, x, f$ | Lower case letters denote vectors |
| $X, F, A$ | Upper case letters denote matrices |
| $x_i$ | The i-th element of vector $x$ |
| $A_{ij}$ | The ij-th element of matrix $A$ |
| $\mathbb{I}$ | Identity matrix |
| $\mathrm{diag}(x)$ | Diagonal matrix with vector $x$ on diagonal |
| $\mathrm{vec}(A)$ | Vector by flattening matrix $A$ |
| $x^{\mathsf{T}}, A^{\mathsf{T}}$ | Vector or matrix transpose |
| $x^{\mathsf{H}}, A^{\mathsf{H}}$ | Conjugate transpose of a vector or matrix $X^{\mathsf{H}} = \bar{A}^{\mathsf{T}}$ |
| $\bar{x}, \bar{A}$ | Conjugate transpose of a complex vector or matrix |
| $A^{+}$ | Non-negative entries of matrix |
| $A^{-}$ | Negative entries of matrix |
| $A^{-1}$ | Matrix inverse |

### Operations & operators

| | |
|---|---|
| $\odot$ | Element-wise multiplication |
| $*$ | Convolution operator |
| $\star$ | Correlation operator |
| $\otimes$ | Kronecker product |
| $\diamond$ | Fast forward model operator |
| $\sum$ | Summation operator |
| $\int$ | Integral operator |
| $\iint$ | Double integral operator |
| $\nabla, \partial$ | Derivative operator |
| $\mathcal{L}$ | Laplacian operator, i.e. $\nabla^T \nabla$ |
| $\mathbb{R}$ | Real numbers |
| $\equiv$ | Defined to be equal to |
| $\approx$ | Approximately equal to |

# Abbreviations

| | |
|---|---|
| *BD* | Blind Deconvolution |
| *CCD* | Charged Coupled Device |
| *CPS* | Camera Pose Space |
| *cryo-EM* | Cryo-electron microscopy |
| *CT* | Computed Tomography |
| *DFT* | Discrete Fourier Transform |
| *EFF* | Efficient Filter Flow |
| *FFT* | Fast Fourier Transform |
| *GEM* | generalised expectation maximisation |
| *HST* | Hubble Space Telescope |
| *LST* | Linear Systems Theory |
| *MAP* | maximum a posteriori |
| *MDF* | Motion Density Function |
| *MFBD* | Multi-frame Blind Deconvolution |
| *MRB* | Motion Response Basis |
| *MRI* | Magnetic Resonance Imaging |
| *MTF* | modular transfer function |
| *MVM* | matrix vector multiplication |
| *NNBD* | Non-negative Blind Deconvolution |
| *NNQP* | Non-negative Quadratic Program |
| *OBD* | Online Blind Deconvolution |

| | |
|---|---|
| *OLA* | overlap-add |
| *OLS* | overlap-save |
| *PDB* | Protein Data Bank |
| *PET* | Positron Emission Tomography |
| *PMPB* | Projective Motion Path Blur |
| *PSF* | point spread function |
| *QR* | Quick Response |
| *SC* | saturation correction |
| *SGD* | stochastic gradient descent |
| *SNR* | signal-to-noise-ratio |
| *SR* | super-resolution |
| e.g. | for example (exempli gratia) |
| et al. | and others (et alias) |
| *i.i.d.* | independently and identically distributed |
| i.e. | that is (id est) |
| *w.r.t.* | with respect to |

# Index