TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY
Julkaisu 640 • Publication 640

Mejdi Trimeche

Super-Resolution Image Reconstruction Using Non-Linear Filtering Techniques

Tampere 2006

Tampereen teknillinen yliopisto. Julkaisu 640
Tampere University of Technology. Publication 640

Mejdi Trimeche

Super-Resolution Image Reconstruction Using Non-Linear Filtering Techniques

Thesis for the degree of Doctor of Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 8th of December 2006, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2006

ISBN 952-15-1688-7 (printed)
ISBN 952-15-1838-3 (PDF)
ISSN 1459-2045

Abstract

Super-resolution (SR) reconstruction is a filtering technique that aims to combine a sequence of under-sampled and degraded low-resolution images to produce an image at a higher resolution. The reconstruction attempts to take advantage of the additional spatio-temporal data available in the sequence of images portraying the same scene. The fundamental problem addressed in super-resolution is a typical example of an inverse problem, wherein multiple low-resolution (LR) images are used to solve for the original high-resolution (HR) image.

Super-resolution has already proved useful in many practical cases where multiple frames of the same scene can be obtained, including medical applications, satellite imaging and astronomical observatories. The application of super resolution filtering in consumer cameras and mobile devices shall be possible in the future, especially that the computational and memory resources in these devices are increasing all the time. For that goal, several research problems need to be investigated, i.e., precise modeling of the image capture process, fast filtering methods, accurate methods for motion estimation and optimal techniques for combining pixel values from the motion compensated images.

In this thesis, we investigate a number of topics related to the performance problems mentioned above. We develop novel solutions to improve the image quality captured by the sensors of a camera phone. Particularly, we present a framework for producing a high-resolution color image directly from a sequence of images captured by a CMOS sensor that is overlaid with a color filter array. In the proposed framework, we introduce a super-resolution algorithm that interpolates the subsampled color components and reduces the optical blurring. The results confirm that it is possible to improve the overall image quality by using few consecutive shots of the same scene.

Achieving accurate and fast registration of the input images is a critical step in super-resolution processing. Motivated by this basic requirement, we propose a novel recursive method for pixel-based motion estimation. We use recursive least mean square filtering (LMS) along different scanning directions to track the stationary shifts between a pair of LR images, which results in smooth estimates of the displacements at sub-pixel accuracy. The initial results indicate good performance, especially for tracking smooth global motion. One important advantage

of the proposed method is that it can be easily integrated into super-resolution algorithms thanks to its relative low computational complexity.

The overall performance of super-resolution is particularly degraded in the presence of motion outliers. Therefore, it is essential to develop methods to enhance the robustness of the fusion process. Towards this goal, we propose an integrated adaptive filtering method to reject the outlier image regions. The proposed approach consists in applying non-linear filtering techniques to improve the performance and robustness against motion outliers. In particular, we applied median filtering for robust fusion of the LR images, and we used generalized order statistic filters (OSF) for the enhancement of binary text images. Compared with conventional super-resolution algorithms, the proposed algorithms preserved well the fine details in the images, additionally, the result images exhibited less artifacts in the presence of registration errors. This confirms the advantage of using order statistic filtering in image super-resolution.

Preface

This thesis has been carried out while the author was working in Nokia Research Center with the Multimedia Laboratory in Tampere, Finland. The work was shaped while working in several projects within the Imaging Research Group; that's six years of close collaboration with a unique team of talented engineers and scientists.

First, I would like to thank my advisor Prof. Moncef Gabbouj from the Institute of Signal Processing in Tampere University of Technology. It was a great privilege and honor to study under his guidance.

I would like to thank the reviewers of the thesis Prof. Azedine Beghdadi and Prof. Edward J. Delp for their review and constructive feedback towards improving its content. I am grateful to Dr. Alessandro Foi and Dr. Faouzi Alaya Cheikh for their corrections and valuable suggestions in the initial drafts of the manuscript. I also thank Dr. Imed Hammouda for helping with proof reading.

The ideas in this thesis were developed through intensive collaboration with several colleagues. I would like to thank many people in NRC for sharing with me their knowledge, software and patience. I acknowledge the support of Jukka Yrjänäinen, Markku Vehvilainen, Ossi Kalevo and Petri Nenonen during the initial stages of the research. Also, I would like to thank my colleagues Sakari Alenius, Radu Bilcu, Jaana Taskinen, Jarno Nikkanen and Marius Tico for creating an inspiring atmosphere in Hermia 5. Many more researchers and managers in NRC have been a continuous source of support and inspiration.

This achievement would have been very challenging if it were not the PhD program that was created to encourage researchers from NRC to take the extra mile needed to finalize the thesis work. I gratefully thank Prof. Jukka Saarinen, head of the Multimedia Laboratory for nurturing the program together with Dr. Jari Kangas. Nokia Foundation is gratefully acknowledged for financial support.

During the last six years, Tampere University of Technology has been my second home. I would like to thank the professors of the Institute of Signal Processing for creating such an open and welcoming research environment, especially Prof. Jaakko Astola, Prof. Moncef Gabbouj and Prof. Karen Egiazarian. I thank Prof. Vladimir Katkivnik, Dr. Alessandro Foi and Dmitro Paliy for their active research partnership. Also, I wish to thank Elina Orava for her prompt and friendly assis-

tance with all administrative matters.

Special thanks go to the small Tunisian community in Tampere for sharing so many things. I am obliged to families Gabbouj, Alaya Cheikh, Hammouda, Bouazizi, Hamila and Maala. Also, I wish to thank Fehmi Chebil, Ali Hazmi and Monaem Lakhzouri for the pleasant moments we had together.

I am indebted to my family, especially to my mother Habiba, to my wife Narjess, to my brother Hamdi and to my sister Wided for their patience, love and support.

Finally, I dedicate this thesis to the memory of my father Abdellaziz Trimeche. Six years ago, the sudden onset of Asynchronous Lateral Sclerosis (ALS) has been a determining and sad moment for our family. It was a lost fight; however, his courage and patience towards his fate have inspired me a lot. As I write these words, I read through a moving letter he sent me when I started university eleven years ago; in that letter he urged me to seek excellence, perseverance and ethical correctness. I tried, and I shall keep trying.

Mejdi Trimeche
Tampere, December 2006

Contents

List of Publications

This dissertation is written based on the research published in the following articles.

M. Trimeche, R.C. Bilcu, and J. Yrjänäinen. Adaptive outlier rejection in image super-resolution. *Eurasip Journal on Applied Signal Processing, Special Issue on Image Super-resolution.* 2006(Article ID 38052):1–12, January 2006.

M. Trimeche, M. Tico, and M. Gabbouj. Dense optical flow field estimation using recusrsive LMS filtering. In *Proceeding of 14th European Signal Processing Conference, EUSIPCO 2006, Florence, Italy*, September 2006.

M. Trimeche and M. Vehvilainen. Super-resolution using image sequence in raw sensor domain. In *Second International Symposium on Communications, Control and Signal Processing (ISCCSP), Marrakech, Morocco*, March 2006.

M. Trimeche, D. Paliy, M. Vehvilainen, and V. Katkovnik. Multichannel image deblurring of raw color components. In *Proceedings of SPIE Conference on Computational Imaging, San Jose, USA*, volume 5674, pages 169–179, January 2005.

M. Trimeche and J. Yrjänäinen. A method for simultaneous outlier rejection in image super-resolution. In *Visual Content Processing and Representation*, volume 2849 of *Proceedings of 8th International Workshop on very low bitrate video coding (VLBV), Madrid, Spain*, pages 188–195. Springer LNCS, September 2003.

M. Trimeche and J. Yrjänäinen. Order filters in super-resolution reconstruction. In *Proceedings of SPIE Conference on Image Processing, San Jose, USA*, volume 5014, pages 190–200, January 2003.

M. Tico, M. Trimeche, and M. Vehvilainen. Motion blur identification based on differently exposed images. In *Proceeding of International Conference on Image Processing, ICIP 2006, Atlanta, USA*, October 2006.

A. Foi, S. Alenius, M. Trimeche, V. Katkovnik, and K. Egiazarian. A spatially adaptive Poissonian image deblurring. In *Proceedings of IEEE International Conference on Image Processing, ICIP 2005, Genova, Italy*, volume 1, pages 925–928, September 2005.

R.C. Bilcu, M. Trimeche, S. Alenius, and M. Vehvilainen. Regularized iterative image restoration in the blur space. *WSEAS Transactions on Communications*, 4(7):407–416, July 2005.

List of Abbreviations

1-D	-	One-Dimensional
2-D	-	Two-Dimensional
AWB	-	Automatic White Balancing
BMP	-	Bitmapped graphics format used in Microsoft operating systems
CBIR	-	Content Based Image Retrieval
CCD	-	Charge Coupled Devices
CFA	-	Color Filter Array
CFAI	-	Color Filter Array Interpolation
CFT	-	Continuous Fourier Transform
CMOS	-	Complementary Metal Oxide Semiconductor
DCT	-	Discrete Cosine Transform
DFT	-	Discrete Fourier Transform
EM	-	Expectation Maximization
FIR	-	Finite Impulse Response
HR	-	High Resolution
ICI	-	Intersection of Confidence Intervals
ISNR	-	Improvement in Signal to Noise Ratio
JPEG	-	Joint Picture Experts Group
LMS	-	Least Mean Squares
LPA	-	Local Polynomial Approximation
LR	-	Low Resolution
LS	-	Least Squares
MAD	-	Mean Absolute Deviations
MAP	-	Maximum *A-Posteriori*
MSE	-	Mean Square Error
ML	-	Maximum Likelihood
MRF	-	Markov Random Field
NLMS	-	Normalized Least Mean Square
OSF	-	Order Statistic Filters
PDF	-	Probability Density Function
POCS	-	Projection on Convex Sets
PSF	-	Point Spread Function

PSNR	-	Peak Signal to Noise Ratio
QVGA	-	Quarter Video Graphics Array, refers to 320×240 image resolution
RI	-	Regularized Inverse
RGB	-	Color space model defined using 3 primary colors: Red (R), Green (G) and Blue (B)
RGBG	-	Sampling pattern for Bayer sensor data. Even rows: Red-Green, odd rows: Blue-Green
RWI	-	Regularized Wiener Inverse
SNR	-	Signal to Noise Ratio
SR	-	Super-Resolution
VGA	-	Video Graphics Array, refers to 640×480 image resolution
YUV	-	Color space model defined using one luminance (Y) and two chrominance components (U,V)

List of Figures

Chapter 1

Introduction

Super resolution (SR) reconstruction refers to the process of combining a sequence of under-sampled and degraded low-resolution (LR) images in order to produce a single high-resolution (HR) image. The LR input images are assumed to portray slightly different views of the same scene. In broad sense, super-resolution techniques attempt to improve the spatial resolution by incorporating into the final result the additional new details that are revealed in each LR image. Conceptually, the processing allows to convert the temporal resolution into spatial resolution.

The basic assumption for super-resolution processing is that some LR images contain novel and non-redundant information about the scene details. This may be due to relative camera motion from one frame to another, possibly resulting from the combination of camera motion, moving objects in the scene, camera jitters, shaking, etc. In order to apply super-resolution, it is important to extract the relative displacement of the portrayed details at sub-pixel precision.

The fundamental problem that is addressed in super-resolution is a typical example of an *ill-posed* inverse problem wherein the original information (HR image) is estimated from the degraded observations (LR images). To solve for the inverse problem, explicit regularization strategies need to be incorporated in order to constrain the feasible solution space. The redundant information in the input LR images is inherently utilized in the solution to regularize the inverse problem and improve the final solution. Obviously, to obtain a meaningful solution of the inverse problem, it is critical to employ realistic modeling of the imaging process.

1.1 Super-resolution processing

Given a set of low-resolution images that result from the observation of the same scene from slightly different views, super-resolution algorithms produce a single high resolution image by fusing the input LR images such that the final HR image reproduces the scene with a better fidelity than any of the LR images. The central idea in super-resolution processing is to convert the temporal resolution

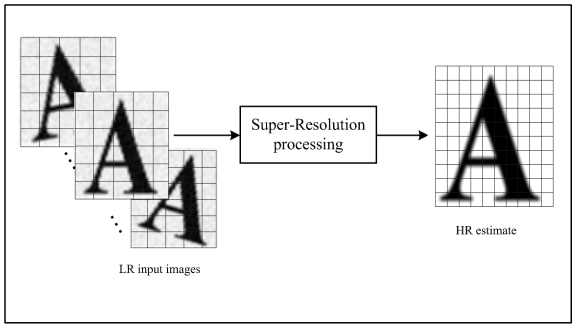


Figure 1.1: Illustration of super-resolution inverse problem: *Given a number of low resolution frames of the same scene, construct a single frame with an improved resolution.*

into spatial resolution. In broad sense, this approach can be used to perform any combination of the following image processing tasks:

- Interpolation
- Denoising
- Deblurring

Usually, super resolution methods consist of the following basic processing steps:

1. Motion estimation to determine the relative shifts between the LR images and register the pixels from all available LR images onto a common reference grid. This step is essential to enable motion compensated filtering.
2. Motion compensation and warping of the input LR images onto the reference grid. Note that the pixels of the LR images are usually non-uniformly distributed with respect to the reference grid.
3. Restoration of the LR images in order to reduce the artifacts due to blurring and sensor noise. The filtering is necessary to improve the perceived image quality.
4. Interpolation of the LR images with a predetermined zoom factor to obtain the desired HR size.
5. Fusing of the pixel values from the LR images. This temporal filtering operation is at the heart of all super-resolution algorithms, and complements the spatial filtering operations performed in the previous steps.

Fig. 1.2 illustrates the generic processing steps described above. It is important to note that in some algorithms, the order of the operations might be different. In the following chapter, different known approaches for super-resolution are presented and discussed in more detail.

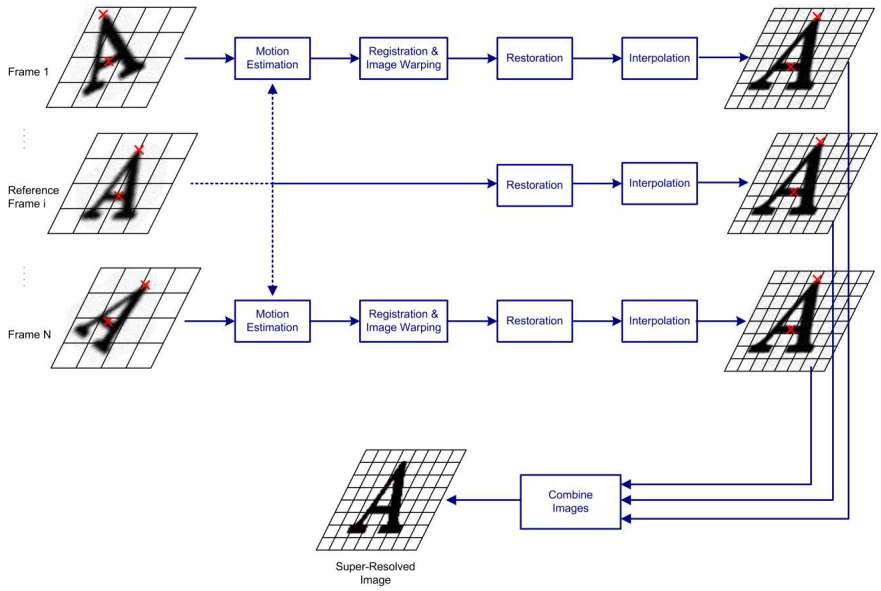


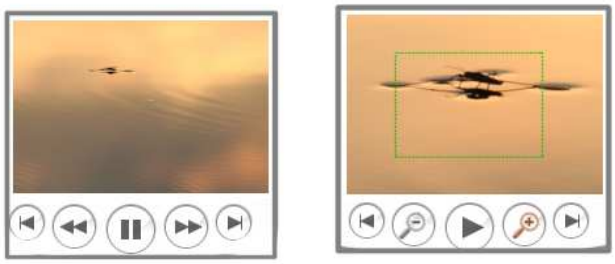
Figure 1.2: Schematic of an example algorithm. Several complex processing steps are integrated in super-resolution.

1.2 Applications

Super-resolution is a computationally intensive process. Nevertheless, the technique has already proved useful in many practical cases where multiple frames of the same scene can be obtained. Although most existing applications are still limited towards specialized imaging products, super-resolution is becoming a mainstream technique in image processing. Below, we list several industrial applications where this filtering technique could be used:

- Consumer photography
- Video cameras
- Surveillance applications (multisensor image fusion)
- Satellite and astronomical imaging
- Medical imaging (microscopy, X-ray, diffraction-limited tomography)
- Remote image sensing[1] (passive millimeter, infra-red, synthetic aperture radar)

[1]Super-resolution might be particularly useful in remote image sensing systems because the images are usually undersampled. For example, typical infrared devices have detector sizes in the order of 20-50 μm (in comparison with $3\mu m$ in CCD sensors). Lately, some are claiming designs for sensing systems with programmed motion of the sensor in order to apply SR (e.g. [11]).



(a) (b)

Figure 1.3: Example of a potential application of super-resolution for video playback. (a) User interface view while video is playing, (b) on pressing pause button, a zoom button appears in the toolbar. The adjacent frames are super-resolved to enhance the details in the region of interest.

Additionally, several novel usage scenarios in mainstream consumer applications can be easily conceived using super-resolution. The most intuitive example is that users capture several pictures of the same scene using a *burst* mode, and later post-process the images in order to enhance the resolution. Another example usage of super-resolution could be to enhance the zoom feature for video playback (see Fig. 1.3 for an illustrated user interface concept).

It is worth mentioning that the example applications above are not exhaustive. One may predict that in the future there will be more innovative applications, since the research in super-resolution has been lately very active. For example, super-resolution might be a key technology to achieve very high image quality by fusing the images captured with multi-sensor cameras (e.g. lenslet cameras [75]). Another potential application might be related to the emerging video devices that will be capable of capturing and processing video at a very high frame rate (e.g. [66]). The latter devices will raise the need for advanced processing methodologies that are capable of converting temporal and spatial video resolutions, as well as novel compression mechanisms that might be derived and linked to super-resolution video processing.

1.3 Super-resolution for consumer cameras

Recently, we have witnessed a revolution in digital photography. The quality and resolution of digital images have been constantly improving with the advances in sensor technologies, memory capacity, processing power and image processing techniques. Besides, there has been a significant reduction in manufacturing costs, which led to massive proliferation of consumer digital cameras. Indeed, it was estimated that more than 400 million cameras have been sold in 2005 [90].

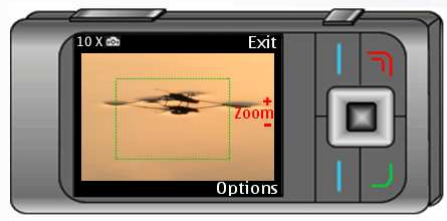


Figure 1.4: Potential application of super-resolution filtering in consumer cameras to enhance the quality of digital zooming, reduce noise and improve the dynamic range by processing multiple exposures of the same scene.

More than three quarters of these have been embedded in mobile phones, and it is predicted that main-stream use of camera-phones will soon erode low-end digital cameras. Nevertheless, due to constant pricing pressure and packaging limitations, there is still serious need for improvement in the imaging quality on camera phones. On the other hand, the computational and memory resources on mobile devices are increasing all the time, and it is already possible to consider the implementation of computationally intensive image processing algorithms such as super-resolution. This approach can help overcome the inherent hardware limitations of the integrated camera systems.

Super-resolution can be implemented in consumer cameras in various ways. For instance, the processing can be scheduled in off-line manner to combine the image sequence that is captured in video or in burst mode; the overall processing is invisible to the end-user. Super-resolution can be applied in video mode to enable the capture of still images without interrupting the video feed producing high resolution images, which can be used for example in automated video summarization, or for hard copy printing. Another different approach consists in applying super-resolution using embedded real time implementations, e.g., on hardware accelerators [20]. In this mode, the frames are continuously saved in a temporary buffer, and when the snap button is pressed, the latest frames are used to super-resolve the result image. This mode of operation can be used to enhance the performance of the *digital zoom* by panning and zooming into the region of interest (Fig. 1.4), without using mechanical parts to move the lens.

1.3.1 Factors impacting image quality

In general, the critical factors that limit the performance of the integrated cameras in consumer devices are the following:

– Spatial frequency response of optical apertures (combined effect of objective lens and sensor photodetectors).

– Optical system distortions, such as geometrical aberrations and vignetting.

– Subsampling of the different spectral components.

– Nonlinear response of photo-detectors, uneven color sensitivity.

– Photodetector noise, quantization errors.

– Reconstruction artifacts, simplistic filtering.

– Design constraints due to packaging and power consumption.

When designing a hardware camera system, several important criteria need to be considered. These criteria include the lens type, sensor type, pixel size, optical arrangement, interconnections, packaging, control electronics, power supply, clocking, dynamic range, exposure design, shuttering, etc. The final image resolution depends on the combination of all these design components.

Some argue that the main limitation in image resolution comes from the employed sensor technology, and put forward the argument that charge coupled device (CCD) technology is superior to its rival complementary metal oxide semiconductor (CMOS) technology. In reality, neither CMOS nor CCD technologies is categorically superior to the other [15] [74], especially with the ongoing maturing of the fabrication processes. Both CMOS and CCD chips sense light through similar mechanisms (Fig. 1.5), i.e., by exploiting the photoelectric effect that occurs when photons interact with crystallized silicon to promote electrons from the valence band into the conduction band.

So far, the most intuitive way to increase spatial resolution has been to reduce the pixel area. During the past few years, the pixel size has continued to shrink from the 10-20 microns pixels in the mid-1990s devices, to 2-3 microns sensors currently in the market. However, since the capacitance of semiconductor is proportional to the pixel area, there is a trade off between the pixel size and the associated light sensitivity. For this reason, larger pixels function better in low light situations, whereas smaller pixels require bright sunlight or a flash to obtain acceptable signal to noise levels. Besides the reduction in photon conversion efficiency, there are other fundamental optical limits which become increasingly important in the overall imaging process [47] [67], thus placing a practical lower limit on the pixel size. In other words, this means that the continuous reduction in pixel area is not a viable trend in consumer cameras.

The high cost for precision optics and sophisticated image sensors are an important concern in low-end consumer cameras. Increasing the pixel count by using a larger sensor area will continue to be an expensive option due to the fabrication processes on silicon. Additionally, this will result in increased power consumption, as well as an extra cost for the optical arrangement associated with an oversized sensor.

Based on the arguments mentioned above, the mere increase in pixel count will no longer be enough to improve the spatial resolution, or at least, we predict that

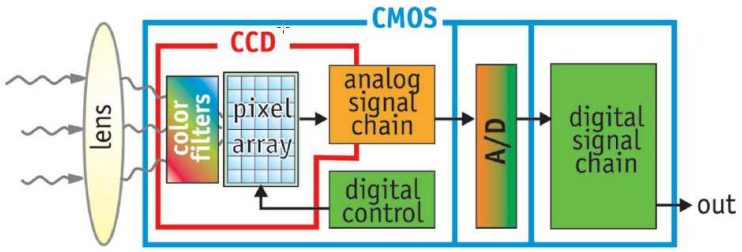


Figure 1.5: CCD versus CMOS sensor concept architectures in digital cameras. In CCD, the pixel's charge is transferred sequentially through a limited number of output nodes. The charge is converted to voltage, then buffered and sent off-chip as an analog signal, and the pixel's area is devoted to light capture. In a CMOS sensor, each pixel has its own charge-to-voltage conversion circuitry. Additionally, the pixel area may include amplifiers, noise-correction, and digitization circuits. Reprinted with permission from Albert Theuwissen [74].

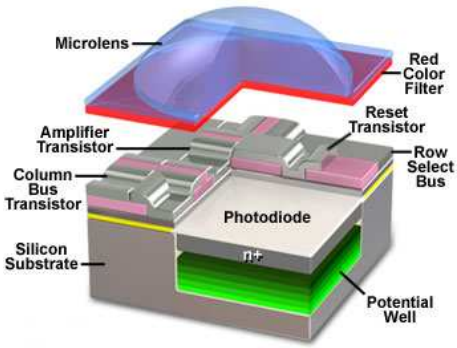


Figure 1.6: Anatomy of the active pixel area in CMOS sensors. Reprinted with permission from Michael W. Davidson [26].

there will be increasing need for sophisticated signal processing tools to follow the trend. In fact, the use of signal filtering techniques in image sensing is as old as digital imaging itself, but the employed techniques have been rather confined to simple and linear solutions. On the other hand, the research in the field of digital image and video processing has been very active lately, which resulted in a wealth of filtering solutions with confirmed results; however, these advanced solutions have been virtually unexploited in consumer camera applications.

1.3.2 Image processing to improve resolution

Super-resolution is an example of such processing techniques that may be successfully tailored for application in consumer cameras. In this context, multi-shot algorithms can be designed to correct the distortions and improve the specific resolution shortcomings by using a sequence of images captured consecutively. This type of processing may be very effective especially if we wish to target specific image degradations such as:

– Noise: especially in low-light capture conditions, or in the absence of a proper flashing mechanism. Multi-shot capture combined with proper motion compensated filtering can be efficient against visible noise artifacts.

– Dynamic range: multi-shot image capture with different exposures is an evident solution to overcome the problem of limited dynamic range and reduced sensitivity, which is due to increased analog gain values associated with the miniaturization of the pixel size.

– Optical blurring: especially for fixed focus cameras. In this context, multi-shot capture may help to regularize the inverse problem.

Nowadays, it is well accepted that the focus on the correction of these degradations is more actual problem than the increase in image size, especially for low-cost cameras. Although this means that the traditional application of super-resolution for interpolation may be left out, the application of multi-frame filtering can greatly benefit from the results in the field of super-resolution, since the basic approach is the same, i.e., motion compensated filtering. In this thesis, the focus is on the application of super-resolution algorithms for consumer cameras, and more specifically for non-dedicated imaging platforms such as camera-phones.

In general, super-resolution is usually considered as an attractive approach for image processing. However, before it is readily applicable in consumer cameras, several basic problems need to be investigated, this will also help understand the real potential of super-resolution. In the following Chapters, we elaborate more on these problems.

1.4 Beyond algorithms: link to the human visual system?

In [91], Schulz and Stevenson argued that the human visual system is capable of temporally integrating information in a video sequence, i.e., the perceived spatial resolution of a sequence appears much higher than the spatial resolution of an individual frame, however, the exact mechanism in the human visual system that performs this operation is yet to be discovered. In fact, there is significant knowledge to be gained from the discoveries about the mechanisms in early biological vision, which might be very useful in visual technology applications.

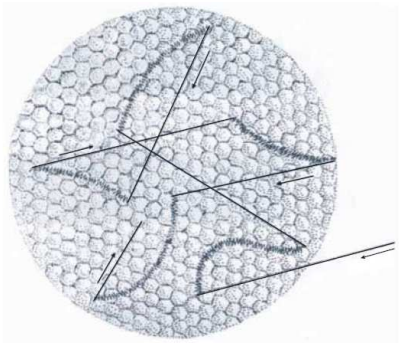


Figure 1.7: Diagram showing fixational eye movements projected on the retinal photoreceptors. High-frequency tremor is superimposed on slow drifts (curved lines). Microsaccades are fast jerk-like movements, which are believed to bring the image back towards the center of vision (straight lines), referenced in [77].

In the human eye, only the central part of the retina has a high concentration of color sensitive nerve cells, whereas the rest of the retina is mainly made up of monochrome nerve endings, which are especially good for motion detection. Knowing these facts, we are tempted to link the enhanced detail perception in video sequences to the role of saccadic eye motion. We raise the question whether this ambiguous eye motion is used by the human visual system to apply some sort of super-resolution processing. The rest of this section is not meant to provide any scientific evidence to answer this question, but rather to give a brief overview of a distinct feature in the human visual system, namely, the saccadic eye motion.

1.4.1 Saccadic eye motion

Our eyes perform different types of movements to accomplish essential early vision tasks; this is partly done through small, involuntary eye movements (saccade). According to Webster dictionary, a saccade is a small rapid jerky movement of the eye especially as it jumps from fixation on one point to another. Visual fixation refers to maintaining the gaze in a constant direction. Humans (and other animals with a fovea) constantly alternate saccades and visual fixations. For example, in reading, fixation refers to the human eye focusing upon an artifact of printed text such as a white space or a word [76]. Visual fixation is never perfectly steady, i.e., fixational eye movements occur involuntarily. Although the existence of fixational eye movements has been known and characterized since the 1950's, its exact role and importance are still debated [77].

1.4.2 Fixational eye movements

Fixational eye movements are usually classified into three types of motions: microsaccades, ocular drifts, and ocular tremor [77]. Microsaccades continuously jerk the center of gaze in straight lines by small, but resolvable distances. Drifts are irregular curvy motions that occur between microsaccades, and are characterized by low amplitude and relatively slower sweeps. Finally, eye tremor is made of extremely small oscillations that are superimposed on drifts. Tremor is characterized by constant, physiological, high frequency (peak 90Hz) and low amplitude vibrations.

Among these three categories of eye movements, microsaccades are the largest and easiest to characterize; although their role has remained a matter of controversy. Recent research [77] points to some evidence that the role of microsaccades is to counteract visual fading during fixation. In this theory, microsaccades continuously stimulate neurons in the early visual areas of the brain, which mostly respond to transient stimuli. This explains that even during periods of steady fixation, the visibility is maintained and the perception remains stable and continuous. It is argued in [37] that conceiving the eye as an electronic analog camera (with a simple lens system) does not correspond to the overwhelming evidence, which suggests that the photoreceptor cell is a differential rather than an integrating detector.

The exact role of the two other types of eye movements (drifts, tremor) is still unclear [76]. For instance, it was not until recently that some researchers were arguing that tremor is a useless feature that degrades vision. This belief has receded and nowadays there is agreement that tremors have a critical role in vision acuity, even though there is no solid evidence that fully explains the role of tremor eye motion. According to [37], our eyes employ an analytical signal processing channel for early vision, which is the primary determinant of the resolution performance and acuity. This process relies upon tremor as a fundamental mechanism and employs two dimensional correlation of the signals within the foveola.

Besides understanding its role in stabilizing our vision, there is a need to understand the precise mechanisms in which tremor motion is exploited. Does our visual system improve the visual acuity through these miniature shifts by applying a similar mechanism to super-resolution? Does the processing utilize the differential images to improve the acuity? How is that done? These are few of the questions about the human early vision, which hold no definite answers yet.

1.4.3 Dynamic theory of vision for hyperacuity

Due to the limited density of photoreceptors on the retina, normal visual acuity in humans is limited to about $1'$ of visual angle. This is imposed by the Nyquist sampling limit in relation to the number of photoreceptors. Surprisingly, it was found that the human visual system is capable of resolving certain stimuli (e.g. vernier

stimuli) at much higher resolutions (less than $5''$). This enhanced capacity to perceive details is called hyperactuity [38]. Several qualitative theories about visual perception have been proposed to explain this peculiar property of the human visual system [50]. Most notably, the so-called dynamic theory of vision, which claims that hyperacuity would require eye-micromovements (microtremor, microsaccades) to achieve this property. In this theory, small eye-movements would shift the photoreceptor grid across the stimulus leading to an enhanced discrimination capability when appropriate spatiotemporal integration is used. In [50], quantitative tests are reported to validate the theory under different experimental conditions. It is shown that eye micromovements indeed improve hyperacuity. Contrary to earlier assumptions, it is reported that eye micromovements have no effect in the central part of the retina, where optical blurring defines the limit for hyperacuity tasks; however, at above 5^o retinal eccentricity, eye micromovements clearly improve the acuity.

It is not our claim to make any conclusions about the theories explaining the role of eye movements in the human visual system. On the other hand, we believe that electronic image acquisition systems, as well as super-resolution processing techniques would greatly benefit when the roles of the different mechanisms in biological vision systems will be better understood.

1.5 Organization of the thesis

Following this introduction, we consider several topics of interest to improve the image resolution. We focus the discussion on the challenges for efficient use of image super-resolution in order to improve the imaging performance in portable camera devices.

Chapter 2 discusses in more detail the filtering methods employed in super-resolution. We formulate the main approaches for super-resolution that are known in the literature; and we present few example results to illustrate the possible improvement in image resolution when using this processing technique. This chapter provides the background knowledge to understand the theoretical and practical issues that limit the performance of super-resolution.

Chapter 3 is concerned with the restoration of a single observation of a degraded image. The goal is to lay down the basis for the extension towards the use of multi-frame image restoration in the following chapters. We present a novel algorithm for multichannel image deblurring of an image that is captured by a CMOS sensor in a camera phone. Our approach is distinct since we consider the application of the algorithm directly on the raw color image data, such that the restoration process is the first processing step in the image reconstruction pipeline. The proposed algorithm has shown to significantly reduce the optical blurring on camera-phone devices with fixed focus optics.

In Chapter 4, we present a framework for producing a high-resolution color

image directly from a sequence of images captured by a CMOS sensor that is overlaid with a color filter array. The proposed algorithm is based on iterative super-resolution that filters and interpolates the raw Bayer data from the sensor. We report experimental results using a synthetic image sequence, also using real data from CMOS sensors. The results exhibit significant improvement in quality when compared to demosaicing the color data using a single image.

Accurate and fast registration of the input images are critical in super-resolution processing. In Chapter 5, we propose a novel recursive method for pixel-based motion estimation. We use recursive LMS filtering along different scanning directions to track the stationary shifts between the LR images and produce smooth estimates of the displacements at sub-pixel accuracy. The initial results demonstrate the usability of the algorithm, especially when targeting video filtering applications that are based on motion-compensated filtering such as super-resolution.

The overall performance of super-resolution is particularly degraded in the presence of motion outliers; hence it is essential to develop methods to enhance the robustness of the fusing process. In Chapter 6, we propose an integrated adaptive filtering method to reject the outlier image regions. In the process of combining the gradient images due to each low-resolution image, we use adaptive FIR filtering. The coefficients of the FIR filter are updated using the LMS algorithm, which automatically isolates the outlier image regions by decreasing the corresponding coefficients. The adaptation criterion of the LMS estimator is the error between the median of the samples from the LR images, and the output of the FIR filter.

In Chapter 7, we investigate the use of order statistic filters in super-resolution. We propose to use signal dependent L-filters for the enhancement of binary text images. We incorporated a simple mechanism to select the most suitable data support to preserve the details along the edges. Additionally, we show that order statistic filtering, for instance median fusing, improves the robustness against motion outliers. Although this algorithm is developed in a heuristic manner, the experimental results demonstrate good performance of order statistic filters when used in image super-resolution.

1.6 Author's contribution

The author's contribution to the research in super-resolution image processing is presented in Chapters 3-7. These chapters cover several relevant topics necessary to understand the performance of super-resolution algorithms in consumer imaging. The distinct contributions of the thesis can be summarized in the following aspects:

- A unique approach to the topic by considering the practical application of super-resolution to the mobile imaging domain. This is visible across the entire thesis since we evaluated most of the proposed algorithms by using images taken with camera phones.

- An integrated adaptive filtering method to reject the outlier image regions in super-resolution (Chapter 6). This chapter is based on the work published in [103] and [112].

- A new method for dense motion estimation (Chapter 5). The method is based on recursive 1-D LMS filtering along different scanning directions. The algorithm is fast and successfully tracks the stationary shifts between a pair of images. The work in this chapter is published in [110].

- A super-resolution algorithm for demosaicing raw sensor data is presented (Chapter 4). The results of this algorithm have been recently published in [111].

- Novel approach for image deblurring by processing directly the raw color components, stemming from the observation that optical blur is different in the color channels (Chapter 3). This work was published in [108].

- We propose to use order statistic filters (OSF) in super-resolution (Chapter 7). The use of generalized OSF (L-filters) in super-resolution constitutes a novel and interesting approach because it can be further developed to target different models for noise and registration errors. This chapter is based on the work published in [113].

Other Work

Apart from the work presented in this thesis, the author was involved in other projects which are not presented here. One of these is concerned with the implementation of a document imaging system on camera-phones and the study of the relevant applications [104]. An example algorithm that was developed for this application is published in [19]. Another project dealt with the design of optimized JPEG quantization tables for specific camera models. Additionally, the author has been researching several topics in multimedia, for example digital rights management for mobile visual content [105], and the study of seamless transfer of multimedia content over wireless networks [109]. In earlier work, the author has been involved with research of content based indexing and retrieval (CBIR) system for images [107], and has been developing algorithms for shape similarity matching based on wavelet decomposition of the contour curves [106].

Chapter 2

Super-Resolution Techniques – An Overview

2.1 Introduction

This chapter discusses the filtering methods in image super-resolution. Section 2.2 gives a brief overview of the research developments in image super-resolution. In Section 2.3, we define a linear image formation model that relates the HR image to the LR observations. This enables to formulate image super-resolution in an inverse problem setting. In the same section, we give detailed examples of the blurring and nonuniform interpolation operations that are usually employed in super-resolution algorithms. In Section 2.4, different known approaches for super-resolution are presented and discussed. In the following Section 2.5, some example results are presented to illustrate the achievable enhancement in resolution using this processing approach. Next, in Section 2.6, we discuss the theoretical and practical issues that limit the performance of super-resolution; we focus the discussion on the challenges for efficient use of super-resolution in the mobile imaging context.

2.2 Related work

Extensive research literature exists on the topic of image super-resolution. The term "super-resolution" generally refers to the problems of recovering image spectrum beyond the diffraction limit through the use of signal processing techniques. It is worth of mentioning that the terminology is often used in different contexts, for instance in diffraction-limited applications, it refers to deconvolution of a single image. In this thesis, the focus is on super-resolution reconstruction, which consists in the process of creating a high resolution image from a sequence of low resolution images. This processing is also generically termed multi-frame restora-

tion [16]; and it implies the use of inter-frame motion information in processing the video data, i.e., using motion compensated filtering. In this sense, the common property to all super-resolution algorithms is that they combine both temporal and spatial filtering.

Early works in 1984 by Tsai and Huang [52] on super-resolution showed that the aliasing effects of the LR images can be reduced, or even completely removed if the relative sub-pixel displacement between the input images is exactly known. Their initial formulation of the problem in the frequency domain has attracted interest in the topic of image super-resolution. Later in 1987, Peleg *et al.* [87] formulated super-resolution in the pixel domain and proposed an iterative algorithm to minimize the error between the estimated HR image and the simulated LR images. The formulation of super-resolution in the pixel domain allowed to include arbitrary motion between LR images. In 1989, Stark and Oskoui [97] proposed a super-resolution algorithm to reduce sensor blurring due to large pixels, the algorithm was based on the method of projection onto convex sets (POCS). In 1992, Tekalp *et al.* [99] extended the POCS formulation to include sensor noise in the imaging model. In 1994, Cheeseman *et al.* [24] proposed a Bayesian statistical formulation of super-resolution, and applied the algorithm to restore astronomical images. Following these pioneering works, nowadays, there is a large number of competing approaches that propose to ameliorate the performance of the reconstruction process, while addressing different applications.

Several articles have surveyed the classic super-resolution methods. In 1998, Borman and Stevenson [17] published a review of different techniques that address the problem of super-resolution video restoration. Later, other review articles have followed, for example [85] and [32]. In 2001, Chaudhuri edited a book [23] containing a collection of articles relating different facets of super-resolution in imaging. Few special journal issues dedicated to the topic of super-resolution have followed recently, for example [59] and [83].

The objective comparison of the different super-resolution algorithms in the literature is a challenging task. The main difficulty stems from the complexity of the overall filtering process, which involves several detailed operations and a large number of different parameters that could bias the quality of the final result. Mostly, the precision of the estimated registration parameters can significantly impact the overall performance of the different algorithms. However, the common feature in super-resolution literature is that it is usually treated as an inverse problem, in the sense that the proposed algorithms attempt to solve the forward imaging process that relates the formation of a sequence of LR images from a single HR image scene.

2.3 Image formation model

In this section, the general model that relates the HR image to the LR observations is formulated. For tractability, the imaging model is usually assumed to be a linear one. The imaging model involves consecutively, geometric transformation, sensor blurring, spatial sub-sampling, and an additive noise term. In the continuous domain, the forward synthesis model can be described as follows: consider N observed LR images, we assume that these images are obtained as different views of a single continuous HR image. The i^{th} LR image can be expressed as:

$$g_i(x,y) = S \downarrow (h_i * f(\xi_i))(x,y) + \eta_i(x,y), \tag{2.1}$$

where g_i is the i^{th} observed LR image, f is the HR reference image, h_i the point spread function (PSF), ξ_i the geometric warping, $S \downarrow$ the down-sampling operator, η_i additive noise term, and $*$ denotes the convolution operator. If we assume that each LR image g_i is of equal size $(K \times L)$ and the down-sampling factor is S, then the HR image f has size $(SK \times SL)$.

After discretization, the model can be expressed as:

$$G_i = A_i F + \Upsilon_i. \tag{2.2}$$

where G_i, F and Υ_i correspond respectively to g_i, f and η_i in discrete domain, and are represented lexicographically column-wise into vectors. The matrix A_i combines successively the geometric transformation ξ_i, the convolution operator with the blurring parameters of h_i, and the down-sampling operator $S \downarrow$ [30]. If the down-sampling factor $S > 1$, then A_i is a sparse matrix with size $(KL \times S^2KL)$.

2.3.1 Problem statement

> *Given the observed set of LR images $\{G_i, i = 1 \cdots N\}$, solve for the HR image F according to the imaging model in (2.2).*

This type of problem is a typical example of an inverse problem, wherein the source of information (HR image) is estimated from the observed data (LR images). The linear formulation of the imaging model enables to formulate the super-resolution problem in a setting similar to a classic image restoration problem. The main difference is that we have several observations emanating from a single source data (when $N > 1$).

The inverse problem above is *ill-posed*. First, because the problem is likely to be under-determined due to insufficient number of LR images. Second, because the blurring (low-pass filtering) results in an ill-conditioned [12] matrix (A_i). In other words, this means that the forward imaging process involves an irrecoverable loss of information. Therefore, the information content of the solution, when it exists, is lower than that of the initial state. Since there is no direct solution

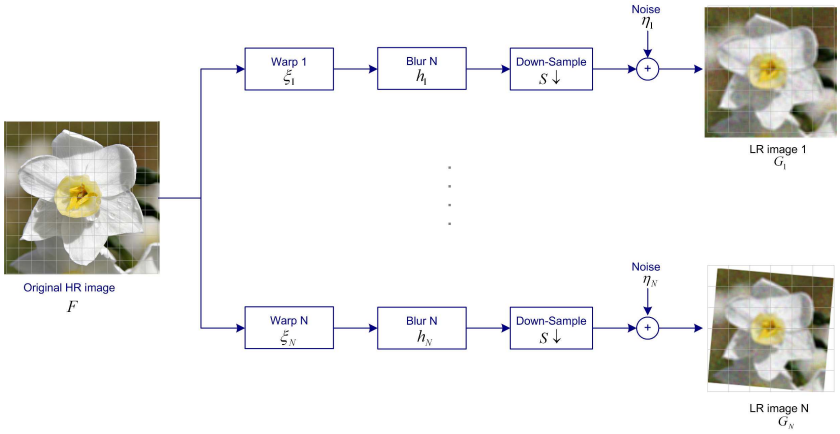


Figure 2.1: Illustration of the image formation model following the model in (2.1).

to the ill-posed problem, regularization procedures are necessary to stabilize the solution.

2.3.2 Simulation of the image formation model

As it will be discussed later, the simulation of the forward imaging model is usually required in super-resolution algorithms. Fig. 2.1 illustrates the model that is used for generating the sequence of LR images, the model is similar to the one described in [29].

In order to test various algorithms under controlled conditions, we simulated the forward imaging model described in (2.1). Given an original HR image, it is possible to generate a sequence of synthetic LR images by using random warps of the original image. For that, an 8-parameter projective geometry model is used; the corresponding parameters are saved for later use in the reconstruction experiments. A continuous Gaussian PSF is used as the blurring operator, which can be controlled through a single parameter. It is possible to specify any down-sampling parameter, and different types of additive noise models (Gaussian noise, impulsive, mixed). The obtained set of synthetic images are used as input to super resolution algorithms with the exact knowledge of motion and blur. This type of experimental data sets enables exact quantitative comparison of different super-resolution algorithms.

2.3.3 Point spread function

The Point-Spread Function (PSF) of an imaging system describes how the light energy from a point on the object plane is dispersed onto the sensor plane. The diagram in Fig. 2.2 shows the basic principle of blurring due to optics. Due to

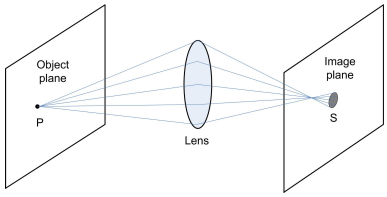


Figure 2.2: Illustration of optical blurring in imaging systems. Usually, the corresponding degradation is analytically simplified by space invariant linear convolution with a point spread function (PSF).

the optical system's diffraction and aberration pattern, each point P in the object plane is extended (spread) onto a region S in the image plane.

Since the image plane is sampled by the sensor, the PSF, h_i in (2.1), is assumed to incorporate the combined optical blurring and sampling effects of sensor. Usually, the corresponding degradation is simplified by assuming a space invariant linear convolution.

2.3.4 Nonuniform interpolation

All super-resolution algorithms need to implement at some stage nonuniform interpolation functions, sometimes referred to as projection functions. Due to arbitrary shifts between the LR images, the registered pixels are likely to be nonuniformly distributed over the reference grid. Thus, nonuniform interpolation is necessary to map those pixel values onto a uniformly spaced HR image (see Fig. 2.4 for an illustration). Even if the output size is the same as the input size, there is still need to perform this operation. Hence, careful and precise handling of the interpolation process is critical to achieve superior performance in super-resolution algorithms.

Besides pixel replication, the commonly used algorithms for image interpolation are bilinear, bicubic [63], and B-spline [114] interpolation. These methods employ a simple weighted sum operation to estimate pixel values at the interpolated image grid. Although these methods fail to effectively preserve edges and introduce additional blurring artifacts, they are simple and can be easily integrated in super-resolution algorithms when the interpolation step is intentionally designed as a stand-alone operation.

However, in most super-resolution algorithms, it is required to implement both the *synthesis* convolution and the *back-projection* convolution. The synthesis convolution is needed to simulate the forward imaging model, and accordingly generate the downsampled LR images from the HR image by incorporating the effect of the assumed PSF. On the other hand, the back-projection convolution is necessary to implement the inverse process in order to map the registered pixels from the LR image grid onto the HR image grid. Both operations can be integrated by assum-

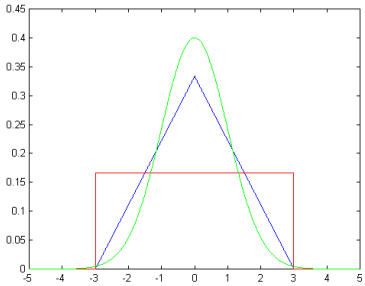


Figure 2.3: Different filters used in projection functions, plotted in 1-D.

ing a continuous interpolation filter that can be easily controlled through a single parameter (i.e., Gaussian interpolators). Although this choice of implementation limits the form of the PSF to a pre-defined parametric function, it allows significant flexibility in the implementation.

The projection functions are used to interpolate non-uniformly distributed pixels onto a rectangular reference grid. This means that we need to precisely calculate the distances between the central pixel position and the neighboring pixels. This procedure is needed to achieve efficient implementations of the projection functions, especially in the presence of significant rotations or perspective change between the LR images. Below, we describe in detail the back-projection function when considering the warping that is characterized by an 8-parameter perspective transformation P.

On the HR image grid, the pixel $f(m,n)$ is defined over the coordinate position (x,y) (center of the pixel). $f(m,n)$ is calculated as follows:

1. Initialize the HR pixel value $f(m,n) = 0$.
2. According to the transformation $(x',y') = P^{-1}(x,y)$, determine the coordinates of the projected pixel position onto the LR image grid.
3. Mark a rectangular window of size $(W \times W)$ around the coordinates defined by (x',y'). The pixels $(k = 1 \cdots W^2)$ inside this window will be used in the interpolation. The size of the window (W) depends on the desired precision, the employed filter (ψ), and the zoom factor.
4. For each pixel (g_k) inside the window:
   - 4.1. find the distance d_k between the center of that pixel and the point (x',y')
   - 4.2. using d_k, find the corresponding weight assigned by the filter $(\psi(d_k))$
   - 4.3. increment the HR pixel value by the LR pixel value times its corresponding weight: $f(x,y) = f(x,y) + \psi(d_k)g_k$
5. Normalize the HR pixel value by dividing it by the sum of the weights used: $f(x,y) = \frac{f(x,y)}{\sum_{k=1}^{W^2} \psi(d_k)}$

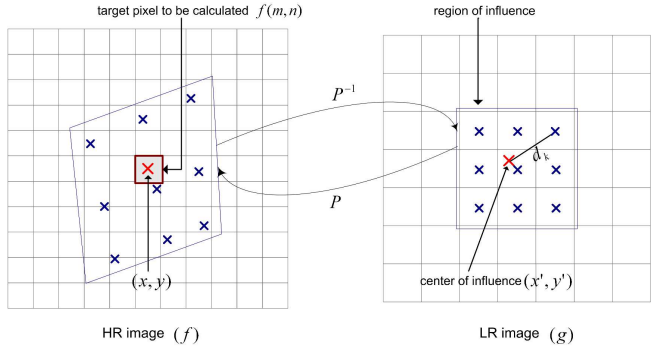


Figure 2.4: Dependency between LR and HR pixels in non-uniform interpolation.

Note that in this algorithm, the weights are assigned based on the distances d_k, which are calculated in the LR grid rather than on the HR grid. This approximation assumes that the tilt of the camera is small so that proportionally, the corresponding distances are quite close. Fig. 2.4 shows the dependency between the LR and HR pixels.

In our testing software, three different types of interpolation filters are used, these can be easily selected to generate the desired projection function. These are the zero order integrating function, triangular integrating function, and the parametric gaussian interpolator. These functions are continuous and truncated over a fixed support window. The window support depends on the extent of the assumed blurring.

2.4 Super-resolution algorithms: a review

In this section, we review some of the most referenced approaches for solving the super-resolution problem. In the following, the notation used is in accordance the image formation model in (2.2).

2.4.1 Iterated back-projection

Irani and Peleg [55] formulated the iterative back-projection (IPB) algorithm for super-resolution by utilizing a similar approach to that used in tomography. In Computer-Aided Tomography (CAT), the image of a 2-D object is reconstructed from its 1-D projections along many directions. In a similar fashion, the HR image $(\hat{F})$ is estimated by consecutively back projecting the error (difference) between simulated LR images via imaging model $(\hat{G}_i)$ and the observed LR images (G_i). Starting with an initial estimate $\hat{F}^0$ for the HR image, the back-projection process is repeated iteratively for each incoming LR image.

For the i^{th} inbound LR image, the basic update equation can be written as:

$$\begin{aligned} \hat{F}^i &= \hat{F}^{i-1} + H_{BP}(G_i - \hat{G}_i) \\ &= \hat{F}^{i-1} + H_{BP}(G_i - A_i\hat{F}^{i-1}) \end{aligned} \tag{2.3}$$

where H_{BP} is the back-projection filtering operator that performs the projection of the error image onto the HR estimate. In our notation, the matrix H_{BP} integrates the motion compensation and the interpolation filter h_{bp}, consecutively. Unlike the imaging blur due to h_{psf}, the back-projection filter (h_{bp}) may be chosen freely, for instance if we assume h_{bp} is Gaussian with parameter σ_{bp}, then the sharpness of the final result may be controlled by selecting a small value for σ_{bp}.

From a practical point of view, one advantage of this algorithm is that it can handle incoming LR images without the need of buffering, thus significantly lowering the memory use, while still producing competitive results. One difficulty with this filtering approach is the absence of a regularization step. This means that the algorithm may converge to several possible solutions, and keeps oscillating among some of these. Also, as the iterations go forward, the latest images may have more influence on the final result. The choice of the initial estimate does not significantly influence the performance of the algorithm in terms of speed of convergence or stability [56]. It may, however, influence which of the possible solutions is reached first. A good choice of initial estimate is the average of the motion-compensated LR images, which usually leads the algorithm to a smoother solution.

Fig. 2.5 shows a block diagram of an example algorithm based on iterative back-projection. Note that it is possible to integrate intermediate filtering steps. For example in [27], a Wiener filtering step is integrated prior to performing the back-projection in order to improve the deblurring and noise filtering performance. It is possible also to augment the derived algorithms with a few additional filtering steps such as additional iterations, regularization filters and simple checks to improve robustness against motion outliers.

2.4.2 Maximum *a-posteriori*

This approach (MAP) consists in solving the super-resolution problem by treating it as a statistical estimation problem (e.g. [21], [24], [45], [91]). The Bayesian formulation solves for the probability density function (PDF) of the original image by maximizing the *a-posteriori* conditional probability. Compared with the Maximum Likelihood (ML) solution, the MAP formulation provides for an easy method to integrate *a-priori* knowledge concerning the solution, which considerably helps to regularize the inverse problem.

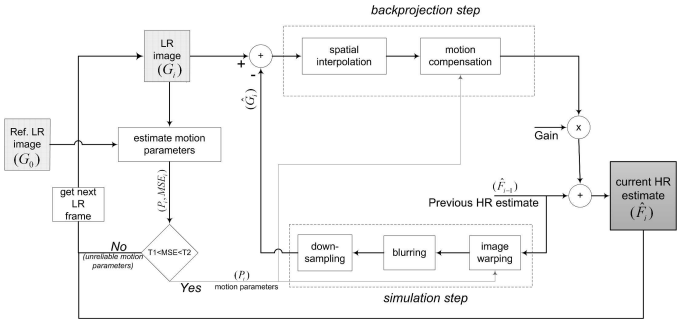


Figure 2.5: Block diagram of an example algorithm based on iterative back-projection

Problem formulation

The MAP estimator of $\hat{F}_{MAP}$ maximizes the *a-posteriori* PDF $Pr(F|G_1, \cdots, G_N)$ with respect to F. Applying Baye's rule, the MAP estimator can be formulated as:

$$\begin{aligned} \hat{F}_{MAP} &= \texttt{arg max} \ \{Pr(F|G_1, \cdots, G_N)\} \\ &= \texttt{arg max} \ \left\{ \frac{Pr(G_1, \cdots, G_N|F)Pr(F)}{Pr(G_1, \cdots, G_N)} \right\} \end{aligned} \tag{2.4}$$

Note that in the denominator, the joint probability function $Pr(G_1, \cdots, G_N)$ is independent of F. Further, since all the functions are positive, then if we take the log of (2.4), the MAP estimator can be written as:

$$\hat{F}_{MAP} = \texttt{arg max} \ \{\log[Pr(G_1, \cdots, G_N|F)] + \log[Pr(F)]\} \tag{2.5}$$

In the equation above, the first term is the *log-likelihood* function, and the second term is *a-prior* density of the assumed solution. If we drop the *a-prior* term, the problem formulation is equivalent to that of the *Maximum Likelihood* (ML). On the other hand, the *a-prior* density function enables to incorporate regularization by biasing the set of possible solutions towards the assumed prior model.

According to the imaging model in (2.2), the likelihood function is completely specified by the joint PDF of the noise $(Pr_{(\Upsilon)})$, i.e.:

$$Pr(G_1, \cdots, G_N|F) = Pr_{(\Upsilon)} \{(G_1 - A_1F), \cdots, (G_N - A_NF)\} \tag{2.6}$$

Typically, it is assumed that the noise is identically and independently distributed in the different LR images G_i. This simplifies the joint density function $Pr_{(\Upsilon)}$ into a multiplication of the probability distribution of each term. Further, if we assume a zero-mean Gaussian noise model with variance σ^2, then the conditional PDF above reduces into the following analytic expression:

$$Pr(G_1, \cdots, G_N|F) = \frac{1}{(2\pi)^{M/2}\sigma^M} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \|G_i - A_iF\|^2 \right\} \tag{2.7}$$

where $M = S^2KL$ is the number of pixels on the HR image.

Prior model

In order to solve for the MAP estimator in equation (2.4), we need to define the prior probability $Pr(F)$ for the data model, which is the distinctly Bayesian contribution. The selected statistical image model should reflect the characteristics of the random process from which F is considered as a realization. A general model for the prior distribution $p(u)$ is usually defined as a Markov random field (MRF) that is characterized by its Gibbs distribution

$$Pr(F(k)) = \frac{1}{Z}\exp\{-\Omega(k)\} = \frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}}\varphi_c(k)\right\} \tag{2.8}$$

where Z in the normalizing term is called the partition function in physics terminology, which encodes the statistical properties of a system in thermodynamic equilibrium. Ω is called the energy function and has the form $\Omega(F) = \exp\left\{-\sum_{c\in\mathcal{C}}\varphi_c(F)\right\}$, where $\mathcal{C}$ denotes the set of cliques for the MRF, and φ_c is a potential function defined on a clique.

Below, we list some typical prior models that are usually used in image restoration tasks:

- $\Omega = ||F||_2^2$ (L_2 norm) white noise prior
- $\Omega = ||\nabla^2 F||_2^2$ (L_2 norm of Laplacian) smoothness prior
- $\Omega = ||F||_1$ (L_1 norm) impulsive data prior
- $\Omega = |\nabla F|$ (magnitude of gradient) also known as total variation (TV) prior

Using the Gibbs prior model, the overall MAP estimator is obtained by minimizing the following cost function

$$\hat{F}_{MAP} = \texttt{arg min}\ \left\{\sum_{i=1}^{N}||G_i - A_iF||^2 + \lambda\Omega(F)\right\} \tag{2.9}$$

where λ is the regularization parameter which controls the balance between the influence of the Gibbs prior term and that of the likelihood term. Note that the ML estimator is a special case of MAP estimation with no prior term ($\lambda = 0$). However, due to the ill-posedness of SR inverse problems, the regularization term is needed to avoid the divergence of the solution. On the other hand, a too large λ will emphasize the prior model on the expense of the fidelity to the observed data.

In practice, the minimization of the cost function in (2.9) is solved through iterative conjugate descent techniques. The gradient function E is obtained by

differentiating with respect to F and is obtained as follows:

$$E(F) = -2\sum_{i=1}^{N} A_i^T(G_i - A_iF) + \lambda\frac{\partial\Omega}{\partial F} \tag{2.10}$$

and the HR estimate is iteratively computed as follows:

$$\hat{F}^{k+1} = \hat{F}^k + \mu^k E(\hat{F}^k) \tag{2.11}$$

where μ^k is the step size. μ^k is usually calculated through the steepest descent method, and is given by

$$\mu^k = \frac{N||E(\hat{F}^k)||^2}{\sum_{i=1}^{N}||A_iE(\hat{F}^k)||^2} \tag{2.12}$$

The MAP paradigm for solving the SR problem has been proposed in several different settings. In [91], Schultz and Stevenson argued that Gaussian image priors are not effective for image data, instead they proposed a discontinuity preserving model using the Huber-Markov Gibbs prior, resulting in a constrained optimization problem with a unique minimum. Cheeseman *et al.* [24] applied MAP super-resolution restoration techniques to astronomical images. Their formulation assumes Gaussian noise and utilizes prior terms which lead to a linear system of equations, and which are solved using Gauss-Jacobi methods. Hardie *et al.* [45] proposed a similar MAP formulation under the assumption of Gaussian noise and a Gaussian MRF prior model. Additionally they considered the simultaneous estimation of both the HR image and the motion parameters, and they provided with a detailed algorithm to the resulting iterative optimization problem.

2.4.3 Projection on convex sets

The method of projection onto convex sets (POCS) defines the feasible solution space as the region of intersection of a collection of convex constraint sets, which represent the space containing the possible HR images that can satisfy the image formation model. The convex sets C_k may represent constraints such as fidelity to the observed data, positivity, bounded energy, smoothness and so on. The estimated HR image is restricted to lie in the intersection space of these constraint sets $C_{\hat{F}} = \cap_{k=1}^{m} C_k$, and the solution is obtained recursively as follows:

$$\hat{F}^{n+1} = \mathcal{P}_m\mathcal{P}_{m-1}\cdots\mathcal{P}_1\hat{F}^n \tag{2.13}$$

where $\hat{F}^0$ is an initial estimate of the HR image, $\mathcal{P}_k$ is the projection operator that projects an arbitrary image onto the convex set C_k, $(k = 1\cdots m)$. Fig. 2.6 illustrates this approach in solving the super-resolution problem.

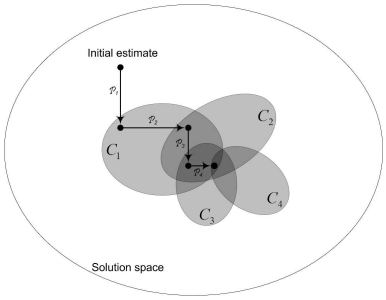


Figure 2.6: Illustration of the projection onto convex sets (POCS) approach

One constraint that is commonly used in super-resolution is the consistency with observed data. It is usually imposed using the following constraint set

$$C_{G_i} = \left\{ \hat{F} : |G_i(u,v) - (A_i \hat{F})(u,v)| \leq T_i(u,v) \right\} \tag{2.14}$$

which is defined for each pixel position (u, v) on the LR image grid. $T_i(u, v)$ is a threshold that may be freely determined based on the assumed noise power, the precision of motion parameters, the number of iterations, etc.
In addition to data consistency constraints, the range values in the solution image may be bounded with the set C_B,

$$C_B = \left\{ \hat{F} : v_{min} \leq \hat{F}(u,v) \leq v_{max} \right\} \tag{2.15}$$

Bounded energy is another common constraint that can help regularize the projection functions. Also, another useful constrain is to limit the solution to lie within some distance from a reference image F_{ref}, which can be for instance the interpolated reference image that is used in motion estimation. This may be imposed using the following constraint set,

$$C_r = \left\{ \hat{F} : ||\hat{F} - F_{ref}|| \leq \epsilon_r \right\} \tag{2.16}$$

Additional constraints on the solution may be defined in a similar manner. The inclusion of prior knowledge in this fashion constrains the solution space thus enabling robust performance in the presence of noise, inconsistent data or missing data. One important advantage of the POCS approach is that it provides a convenient framework for regularizing the inverse problem by incorporating simple methods to include a-priori information.

Initially, Stark and Oskoui [97] proposed the POCS method in super-resolution to reduce sensor blurring due to large pixels, although the image formation model they assumed did not include a noise term. Tekalp *et al.* [99] extended the POCS formulation to include sensor noise, and later Patti [86] *et al.* further incorporated the motion blur occurring during the aperture time of the camera, this has been

done by considering a spatially varying point spread function in the image formation model. They also described a detailed implementation of the POCS-based reconstruction algorithm.

The main difficulty with the POCS approach lies in defining proper projection functions that ensure to reach the intersection of the defined constraint sets. The associated implementations usually require considerable computational cost, and a large number of iterations to achieve convergence [16]. In practice, better performance is usually achieved with heuristic algorithms that combine both the statistical inference methods (i.e., ML or MAP) and POCS approach, since they combine the speed of convergence of gradient based optimization techniques with POCS-based intuitive methods for inclusion of simple constraint rules to avoid the divergence of the solutions.

2.4.4 Other approaches

Nonuniform interpolation approach

This approach is the most intuitive method to implement super-resolution. The pixels from all the LR images are registered directly onto a HR image grid, and a non-uniform interpolation technique is employed to interpolate in all missing pixel positions. Later, it is possible to perform image restoration stage in order to reduce noise or blur artifacts. For example, Ur and Gross [115] proposed an algorithm that performs nonuniform interpolation, which is followed by a deblurring step. Nguyen and Milanfar [84] proposed a wavelet-based algorithm that is a combination of interpolation and restoration operations. They exploit the interlacing structure of the sampling grid in SR and derive a computationally efficient wavelet interpolation for interlaced image data.

In general, the advantage of this type of approach is the simplicity in implementation and the lower computational requirements, which makes real-time applications possible. Additionally, it may be easier to integrate heuristic approaches that improve the performance of the overall process. On the other hand, the main difficulty with this approach lies within its unpredictable performance due to the limited degradation models, also the optimality of the whole reconstruction algorithm is not guaranteed since the interpolation and restoration steps are implemented independently of each other.

Frequency domain processing

Early works on super-resolution by Tsai and Huang [52] published in 1984 showed that it is the aliasing effects in the LR images that enable the recovery of the high-resolution (HR) fused image. They derived the equations that describe the relationship between LR images and the original HR image by exploiting the relative motion between LR images. Based on the shifting property of the Fourier trans-

form, they formulated an elegant system of equations relating the aliased DFT coefficients of the observed LR images to the CFT of the unknown image. This system of equations is solved for the frequency-domain coefficients of the original scene, which is then recovered using the inverse DFT. This formulation was the first to propose super-resolution by using sub-pixel shifts to improve the spatial resolution of the images. However, this approach is limited to global translational motion model, for this reason, Fourier domain approach is considered nowadays useless in practical applications.

More recently, Altunbasak *et al.* [3], [40] proposed a method to perform super-resolution in compressed domain. They incorporated the compression artifacts in the image formation model, particularly the quantization effect. They proposed an algorithm that is based on the technique of projection of convex sets (POCS), and which operates directly on the DCT coefficients. Although it is quite challenging to precisely model the compression artifacts and the implementation might be quite complex, this approach is interesting for practical video applications, since more often video data is presented exclusively in compressed domain.

Recognition based approaches

Motivated with the observation that smoothness priors usually lead to overly smooth HR estimates, Baker and Kanade [8] proposed to use recognition-based priors when the target magnification factor is large. The face hallucination algorithm they propose is based on the recognition of generic local features, which they use to predict an image prior that replaces the traditional smoothness priors. They formulated their solution as a MAP, and integrated the learnt models into the prior term. They claim that the recognition process provides additional source of information for super-resolution, which leads to the enhanced results of face images.

With the target application of face recognition, Gunturk and Batur [42] proposed to transfer the super-resolution reconstruction from pixel domain to a lower dimensional Eigen face space. The reconstruction algorithm no longer tries to obtain a visually improved high-quality image, but instead directly constructs the information required by the recognition system. Such an approach has the advantage of a significant decrease in the computational complexity as well as an improved performance.

The idea of integrating sophisticated recognition-based priors in the super-resolution process is conceptually interesting. It also softens the borders between different research areas, which might bring some promising results in future. However, the main difficulty lies in defining robust and generic methods to exploit the prior information into the super-resolution process. Another drawback is that the processing is closely tied with a limited scope of image models and applications.

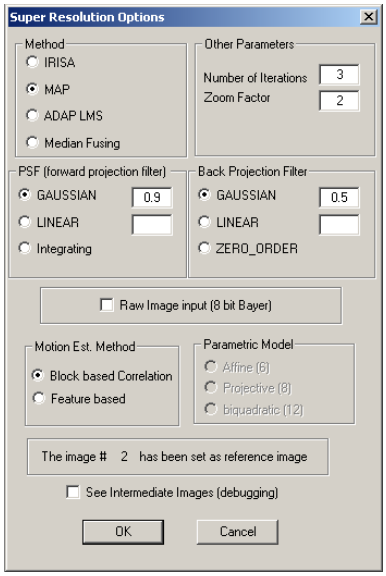


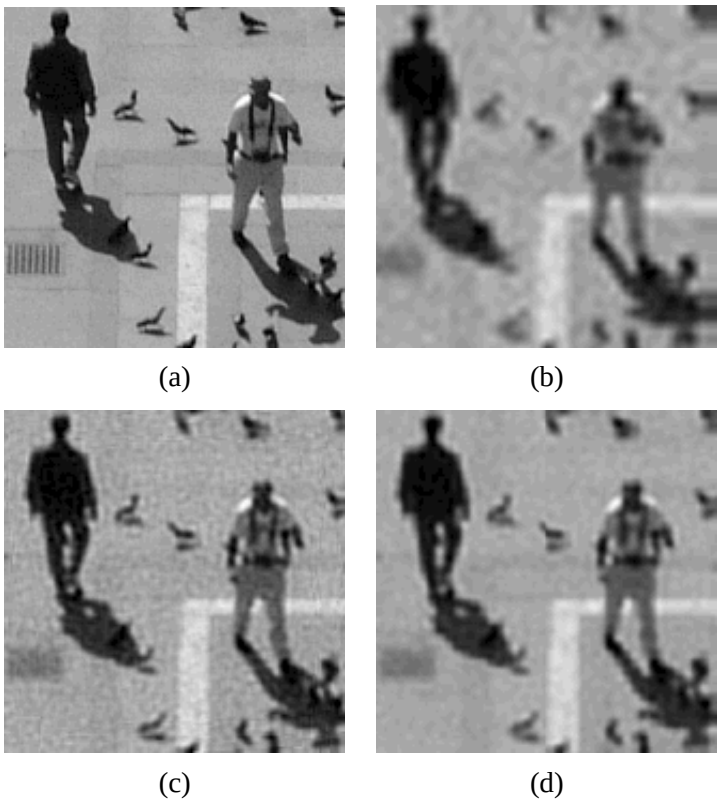
Figure 2.7: Screen shot of user interface options for testing super-resolution algorithms.

2.5 Example results

In this section, we show some example results in order to illustrate the possible resolution enhancement with super-resolution algorithms. We used two different methods, the back-projection and the MAP algorithms, which are discussed in the previous section. We present two experiment settings to demonstrate the performance of SR algorithms. In the first experimental setting, a synthetic image sequence is used. The LR images are generated from a single HR image with the exact knowledge of the registration parameters and the image formation model. In the second experiment, the algorithms are tested on real world sequences obtained from camera hardware.

2.5.1 Testing software

Most of the algorithms that are described in this thesis have been implemented in a Windows based application, which is used as a development and testing environment for a variety of different image processing algorithms. Additionally, we frequently use Matlab© to test some algorithms in a quick fashion. Fig. 2.7 shows a screen shot of the user interface options that are related to the testing of different super-resolution algorithms. The user interface makes it easy to select different generic parameters such as the zoom factor, the algorithm used, the number of iterations, the PSF, the back-projection filters, the motion estimation algorithm, etc.



(a) (b)

(c) (d)

Figure 2.8: Example of super-resolution on noisy LR sequences ($\sigma_\eta^2 = 30$). Target zoom factor 4, 16 input images used. (a) Original image. (b) Reference frame zoomed by 4 using bicubic interpolation, $SNR = 1.30$. (c) Result using back-projection algorithm, 4 iterations, $SNR = 9.11$. (d) Result using MAP method (smooth prior), 4 iterations, $SNR = 10.35$.

2.5.2 Example results with synthetic image sequences

In Fig. 2.8 and Fig. 2.9, the results of super-resolution on sequences of synthetic test images are shown. The input images are generated from a single HR image according to the imaging model described in equation (2.2). The original HR image was randomly warped using an 8-parameter projective model. The registration parameters were saved for the reconstruction experiments. This enables us to compare the images using an objective metric. We used the Signal to Noise Ratio (SNR), which is defined as follows:

$$SNR = 10 \log \frac{||F||^2}{||\hat{F} - F||^2} \tag{2.17}$$

In the first experiment, we used a continuous Gaussian PSF $(\sigma_{psf} = 0.6)$

(a) (b)

(c) (d)

Figure 2.9: Example of super-resolution on noisy LR sequences ($\sigma_\eta^2 = 15$). Target zoom factor 3, 16 input images used, only luminance component (Y) is processed. (a) Original image. (b) Reference frame zoomed by 3 using pixel replication $SNR = 13.76$. (c) Result using back-projection algorithm, 1 iteration, $SNR = 16.79$. (d) Result using MAP method (smooth prior), 3 iterations, $SNR = 17.19$. Note that visual fidelity of the original images may be altered due to resizing and dithering operations used in the printing process.

as the blurring operator and we down-sampled the images by 4 to obtain the 16 LR input images. All the images were contaminated with additive zero-mean Gaussian noise ($\sigma_\eta^2 = 30$). Fig. 2.8 (b) shows the reference frame zoomed by 4 using bicubic interpolation, $SNR = 1.30$. Fig. 2.8 (c) shows the result obtained using the back-projection algorithm after 4 iterations, $SNR = 9.11$. Fig. 2.8 (d) shows the result using MAP method (with a smooth prior) after 4 iterations, $SNR = 10.35$. The visual results, as well as the SNR numbers confirm the improvement in quality when using super-resolution algorithms compared to a single interpolated image. Also, comparing the performance of iterated back-projection against MAP when using similar parameters and computational load, we can remark the superiority of the MAP estimator, which is more robust to

(a) (b)

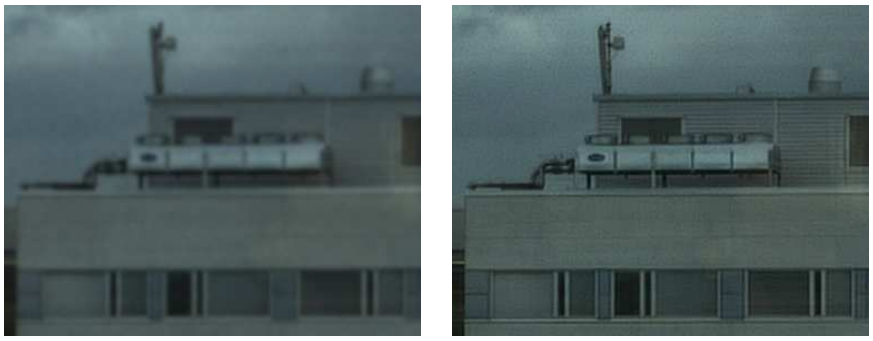
Figure 2.10: Example results using input sequence from a digital camera. 16 JPEG compressed images are used. (a) Reference frame zoomed by 3 using bicubic interpolation. (b) Super-resolution result using iterated back-projection technique, 1 iteration, real time operation.

noise. This is due to the smoothness prior used in MAP, which corresponds well with the original image structure, and ultimately leads to improved regularization behavior of the MAP estimation.

In the second experiment, the original HR image is a color image which was blurred with a Gaussian PSF ($\sigma_{psf} = 0.6$) and down-sampled by 3 to obtain 16 LR input images. The images were randomly warped and contaminated with additive zero-mean Gaussian noise ($\sigma_\eta^2 = 15$). Fig. 2.9 (b) shows the reference frame zoomed to original size using pixel replication, the corresponding signal to noise ratio of the luminance component was found to be $SNR = 13.76$. Fig. 2.9 (c) shows the result obtained using the back-projection algorithm after 1 iteration, $SNR = 16.79$. Fig. 2.9 (d) shows the result using MAP method (with a smooth prior) after 3 iterations, $SNR = 17.19$. In (c) and (d), we can notice that the details are sharper, whereas the noise artifacts are less visible, especially in (d). The super-resolution processing was applied only to the luminance component (Y), whereas the chrominance components have been interpolated from the reference image. It is worth noting that even if we processed in the RGB domain (which takes 2 times more processing time), the visual quality of the obtained results is similar to the images shown in (c) and (d). This is due to the human eye being less sensitive to the chrominance than to the luminance spatial resolution.

2.5.3 Example results with camera sequences

The true indication on whether an algorithm performs adequately is to test it on an image sequence captured with a digital camera. In this section we report two related experiments. In the first experiment, an image sequence of 16 JPEG compressed LR frames (160x120 pixels) is used. The images are captured with a low-



(a) (b)

Figure 2.11: Example results using input sequence from digital camera (Micron board MI-SOC1310). 6 cropped frames from an uncompressed video sequence are used. (a) Reference frame zoomed by 2 using bicubic interpolation. (b) Super-resolution result using MAP technique, 5 iterations.

end consumer camera (Casio QV-3EX). This sequence was used to obtain a super-resolved image of size (480x360). The image in Fig. 2.10 (a) shows the zoomed reference frame using bicubic interpolation (zoom factor = 3), and the image in Fig. 2.10 (b) shows the super-resolved image obtained using a single iteration of the back-projection algorithm, which runs in real-time on a PC. Visual inspection of the images reveals significant improvement in the details of the image; note the details which are completely invisible in the reference image, and which have been reconstructed from different images in the sequence. In this experiment, we used only one iteration of the back-projection algorithm; it is remarkable that a single iteration produces significant improvement. In this sense, limiting the number of iterations works as a regularization mechanism in the back-projection algorithm, and prevents the edges from overshooting and the noise amplification that may be due to inaccurate assumptions in the assumed image formation model. In general, this example illustrates remarkable performance for super-resolution reconstruction, especially that JPEG compression, which eliminates much of the high-frequency content, has not been considered in the image formation model.

In the second experiment, we used a short image sequence of 6 raw images (RGB, BMP format), which were captured with a test camera board (Micron board MI-SOC1310). The image sequence was used to obtain a super-resolved image with a target zoom-factor of 2. The image in Fig. 2.11 (a) shows the zoomed reference frame using pixel replication, and the image in Fig. 2.11 (b) shows the super-resolved image obtained using 5 iterations of the MAP iterative algorithm. Comparing the images, we could observe significant improvement in the details of the image, as well as a reduction in visible noise artifacts. This second example illustrates the usefulness of super-resolution algorithms when used for restoration

of uncompressed images. The algorithm parameters in this case are not optimized; instead, they are roughly estimated since the exact image formation model is not known. This is typical in the case of real hardware experiments. In this respect, we noted that the MAP estimation is less sensitive to modeling deviations (noise variance, PSF) than the iterated back-projection technique, which is more prone to divergence when using the same parameters. This is due to the smooth regularization step that is associated with the MAP estimator, which attenuates the effect due to deviations from the real acquisition model.

2.6 Factors limiting the performance of super-resolution

Recently, there has been some criticism towards the overall efficiency of the super-resolution process. The skepticism is fueled by the fast advances in sensor technology. In fact, some argue that the most direct and cost effective solution to increase spatial resolution is to reduce the pixel size by sensor manufacturing techniques. However, as the pixel size decreases, the number of photons incident on the pixel array decreases, generating shot noise that significantly decreases the signal to noise ratio (SNR). In practice, this performance problem will be most noticeable in low-light conditions, where the noise becomes a major problem. In fact, besides the reduction in photon conversion efficiency, there are other fundamental optical limits which become increasingly important in the overall imaging process, which place a practical lower limit on pixel size ([47], [67]). Therefore, in these applications, super-resolution processing might be the solution to overcome the future limitations of sensing technologies.

Besides digital cameras, there are several applications that might benefit from super-resolution. In the previous chapter, we listed some potential applications, although we are convinced that there are many more applications that were not considered. So the real questions that we should try to address are not about the applicability of super-resolution, but rather about the performance. In the following, we discuss the performance limits of super-resolution. First, we consider the theoretical bounds that limit the performance of super-resolution algorithms. Next, we raise some practical issues that limit the performance of the derived algorithms.

2.6.1 Necessity of aliasing

In frequency domain setting, super-resolution processing restores the DFT samples at a finer resolution and extrapolates the frequency content such that the restored spectrum is wider than any of the observed LR images. The extrapolation is the operation that is most specific to super-resolution since it explains the recovery of the lost information during the sampling process. In [53], Hunt explained the recovery of the information beyond the diffraction limit cut-off because the sinc

function (due to rectangular spatial sampling) is infinite in extent, which means that there will be components of the spectrum portion above the diffraction limit that will be mirrored into the spectrum below the diffraction cut-off. This means that it is the presence of the aliased high frequencies that make super-resolution feasible. In practice, this means that the images need to be under-sampled at the sensor level, without undergoing excessive low pass filtering due to optical blurring or motion blur.

2.6.2 Ill-posedness

The fundamental problem that is addressed in super-resolution is a typical example of an *ill-posed* inverse problem. This means that explicit regularization strategies need to be employed in order to achieve meaningful solutions. In practice, regularization is incorporated in the solution as terms that express a-priori assumptions about the structure of the imaged scene.

Most, if not all super-resolution algorithms are based on the reconstruction constraints which assume that the LR images are generated from a single HR image. In [8], Baker and Kanade derived some analytical results which show that the reconstruction constraints provide less and less useful information as the magnification factor increases. Most noticeable is their remark that even if the reconstruction constraints are invertible (which generally are *not*), the condition number grows at least as fast as quadratic function of the target magnification factors, which roughly indicates that *ill-posedness* of the inverse problem is growing exponentially with the magnification. This means that the overall estimate of the HR solution, especially when combined with smoothness priors, is more and more irrelevant at large magnification factors. In one experiment, Baker showed that for a magnification factor of 16, the smoothness prior provides more information than the reconstruction constraints. This illustrates how fast the inverse problem can become *ill-conditioned*.

2.6.3 Simplistic modeling

Most of the proposed methods for super-resolution in the literature suffer from simplistic assumptions. Besides the simplification in the image formation model, usually sub-pixel motion is assumed to be exactly known. In practice, the displacement between the consecutive frames has to be separately estimated, which makes the super-resolution reconstruction a compound process that heavily depends on the precision and implementation details of the motion estimation. Additionally, motion blur[1], which has a substantial degrading effect on the performance of super resolution [11] is usually skipped in the image formation model.

[1] In [101], we proposed a new method for the identification of motion blur based on the availability of two differently exposed images the same scene.

In fact, few publications dissect in detail the implementation of the different algorithms required in super-resolution. Rather, most publications concentrate on the simulated inverse problem and the associated regularization strategies. We reckon that there is need to investigate and understand the problems that are posed when considering super-resolution as an entire process, which integrates the motion estimation process, and which tackles the dependent problems such as motion outliers, motion blur and internal camera settings, etc.

Nonlinearities

The most evident simplification lies in the image formation model itself, which is usually assumed linear for tractability of the solution. Usually the employed models do not cater for the different processes that happen during the sensing of the scene, which result typically in nonlinearities of the model. For instance, besides the model recently presented in [43], which considers the limited dynamic range and the non-linear sensor response in the super-resolution model, most of the existing literature assume linear sensing models and constant camera exposure time. Still, there are several additional physical and processing parameters which are usually skipped in the modeling, e.g., uneven sampling of the different colors, uneven color response, different gain levels used with each picture and the resulting signal dependent amplification of the noise, different optical aberrations such as vignetting or geometrical distortions, etc.

Even when adequate linearization of these processes is corrected by applying point-wise or spatial processing on the captured images, some additional enhancement algorithms are usually applied inside the camera to reduce the noise, or to improve the sharpness or the color contrast. The resulting process is extremely difficult to capture using a simple model, this means that the linear models used in the formulations of image super-resolution are approximative at best.

Noise modeling

Usually Gaussian noise models are used in the modeling of the image formation process. It is well accepted that this is an over-simplification since in reality the sensor's noise is due to a combination of several sources, e.g., shot-noise, photonic noise, dark current noise, dark signal level, thermal noise. The processing on sensor itself might complicate the noise modeling by introducing errors due to fixed-pattern noise, photon-response non-uniformity, amplifier noise, circuit noise, pixel cross-talk, correlated double sampling, quantization noise, chromatic conversions, etc. [25].

Most of the research literature available on the detailed sensor's noise analysis is developed within the electronics community, and typically comes to conclusions and models which are useful only for the purpose of electronic hardware design and integration. These fragmented models are particularly inadequate for image

processing applications since accurate pixelwise knowledge of the noise model is required in order to properly restore the image details.

In general, signal dependant noise models need to be considered in order to improve the fidelity of the reconstruction process in super-resolution. For example, recently Foi *et al.* [36] investigated the effect of precise Poissonian noise modeling on the performance of image deblurring. It was found the assumption of signal dependent noise, which is closer to reality, significantly improves the performance of the image restoration process, especially for images captured with sensors having small pixel size. In a further development [35], he proposed a deconvolution technique for observations corrupted by signal-dependent noise. The deblurring is performed in transform-domain and is applied on varying size blocks. The results demonstrate good performance of the proposed method, which can be easily combined with other transform-domain processing.

2.6.4 Algorithmic performance

Precise motion estimation

One critical requirement to achieve good performance in image super-resolution is the availability of accurate registration parameters. In fact, sub-pixel precision in the motion field is needed to achieve the desired improvement. In real-life electronic imaging applications, the motion occurring between frames is not known exactly, since precise control over the data acquisition process is rarely available. Thus, motion estimates must be computed to determine pixel displacements between frames. Towards achieving practical implementations of super-resolution, the problems of sub-pixel image registration and outlier robustness need to be investigated in more detail.

The employed algorithm for motion estimation needs to compromise the following properties: precision of registration, noise robustness, locality of motion estimates, robustness to motion outliers and reasonable computational complexity. It is well accepted that motion estimation is the most challenging in super-resolution because it is affected by aliasing and the degradation in the image formation process, which are precisely the factors that super-resolution proposes to solve for.

It is well known that an accurate estimation of the motion blur parameters is non-trivial problem and requires strong assumptions about the camera motion during integration [101]. It was shown in [11] that even when an accurate estimate of the motion blur parameters is available, motion blur has a significant influence on the super-resolution result.

The overall performance of super-resolution algorithms is particularly degraded in the presence of persistent outliers, for which registration has failed. The artifacts caused by an incorrectly registered image are visually much more disturbing than the intrinsic poor spatial resolution in a single input image. To enhance the ro-

bustness of the processing against this problem, super-resolution algorithms need to integrate adaptive filtering strategies in order to reject the outlier image regions.

Computational requirements

Super-resolution is a computationally intensive process. Since the initial problem is numerically ill-posed, most solutions require iterative processing in order to reach an acceptable solution. Additionally, this filtering technique requires several processing stages (motion estimation, restoration, interpolation) that are usually inter-dependent. This makes the overall implementation quite complex, requiring large memory to store the intermediate results, as well as considerable computational resources to calculate the final result.

When the targeted filtering is to run in mobile imaging devices, it is desirable to have real time operation. This might be quite challenging to achieve, especially that portable devices are currently constrained with limited memory, computational and power resources. Additionally, if the number of input images is large, or the output image size is large, the processing delay needed to perform the overall filtering might be inexorably long.

On the other hand, the computational power is increasing all the time, and the rate of improvement is even faster for portable devices. User's expectations of image and video quality is also rising all the time. This means that the opportunity for integrating heavy processing techniques, such as super-resolution will be possible in the future. In this goal, super-resolution algorithms can be introduced at first by scaling down the processing, for instance by reducing the number of iterations, thus favoring the approach of acceptable quality improvement in real time operation at the expense of best possible quality but slow operation.

Chapter 3

Image Deblurring

3.1 Introduction

This chapter is concerned with the restoration of a single degraded image. The objective is to lay down the basis for the extension towards multi-frame image restoration in the following chapters. The problem of image restoration is usually considered a classic topic, however, in this work, we present practical solutions in order to reduce the optical blur in the images captured by a camera phone. We use a novel approach which consists in processing the raw color components captured by the sensor.

In the following, we present a novel multichannel image restoration algorithm. The goal is to develop practical solutions to reduce optical blur from noisy observations produced by the sensor of a camera phone. Iterative deconvolution is used separately for each color channel. Additionally, we use a modified iterative Landweber algorithm which includes an adaptive denoising filter for optimal regularization of the inverse problem. The denoising method is based on the local polynomial approximation (LPA), which operates on windows that are selected by the rule of intersection of confidence intervals (ICI) [60]. In order to avoid false coloring due to independent component filtering in the RGB space, we integrated a novel saturation control mechanism that smoothly attenuates the high-pass filtering near saturated regions. It is shown by simulations that the proposed filtering is robust with respect to errors in point-spread function and approximated noise models.

In Section 3.3, we present the proposed processing paradigm and we describe the image acquisition model. In Section 3.4, we briefly describe the Landweber iterative restoration and we introduce the modified iterative solution which includes an adaptive denoising technique. The analysis of the proposed algorithm shows that the proposed regularization enhances the robustness of the restoration towards modeling errors of the point-spread function (PSF) and of the noise parameters. In Section 3.5, we discuss the practical issues relating to the efficient implementation

of the deconvolution algorithm. Finally, experimental results with the proposed technique on images taken with a fixed-focus camera are shown. The results indicate that the proposed processing produces significant improvement in perceived image resolution.

3.2 Related work

Image restoration research begun with the early space programs in the 1960s. Considering the cost and impact related to acquiring the images from space crafts, the degradation of the data was not a negligible problem. Hence, there was a critical need for processing techniques that could correct and revert the unwanted effects due to suboptimal systems, mechanical vibrations, motion, etc. [10]. Although nowadays this approach towards image processing is still limited for high end applications, such as astronomy and medical imaging; its use in consumer electronics starts to take off, especially that more processing power is available in these devices, and that users are more aware of image quality on non-dedicated imaging platforms such as camera phones.

Image restoration usually requires the knowledge of the degradation process in order to solve the consequent inverse problem. The inverse problem is generally ill-posed [12], i.e., if the direct solution is considered, a small perturbation in the input can result in an unbounded output. The direct inverse methods such as the regularized inverse (RI) and regularized wiener inverse (RWI) deconvolution techniques [61] are effective, but sensitive to modeling errors. On the other hand, the iterative methods are more robust to modeling errors ([13], [73]), hence more interesting for practical implementations. Several algorithms have been proposed to solve the ill-posed inverse problem by introducing a regularization step that suppresses over-amplification of the solution. For example, a directional adaptive regularization technique was proposed to reduce the ringing artifacts and the over-smoothing in the iterative restoration process [71]. Another regularization method [79] suggests the use of spatially adaptive intensity bounds in the framework of gradient projection method. The local bounds were shown to offer a flexible method for constraining the restoration problem.

When the degradation function is unknown, the restoration problem becomes a blind image deconvolution. A basic survey of different techniques can be found in [69]. Blind image deconvolution is usually performed using two main approaches. First approach is to identify the impulse response of degradation, and then restore the original image using some equalization technique. An example of this approach can be found in [95], where a mutually referenced equalizers approach is applied to a set of observations of the same image undergoing different degradation channels. The second approach to blind deconvolution is to jointly identify the degradation function and simultaneously restore the original image. The associated techniques usually assume regularization terms based on the available prior

information that model various statistical properties of the image, and relate these into constrains over the estimated image. A good example of this approach can be found in [96], where stochastic approach is developed using a multi-channel framework, and a maximum *a posteriori* (MAP) solution jointly estimates the blurring functions and the original image. Another interesting approach was proposed in [54] where the image blur is assumed to be a multivariate Gaussian function, and the deblurring is performed by extrapolation in the blur parameters. This technique for restoration in the blur space is based on the compositional properties of the Gaussian PSF. In [14], we proposed an extension of this approach by using an iterative algorithm in the blur domain. The proposed algorithm uses a simple stopping rule that finds the optimum number of iterations and a regularization mechanism to control edge overshooting. Simulations performed on artificially degraded images have shown good performance of the proposed method, and most notable is that restoration does not necessitate prior knowledge of the blur parameters.

The specific problem of restoring noisy and blurred color images has been investigated in the literature since the mid-eighties. Several algorithms ([62], [80], [100]) have been proposed to restore the color images by utilizing the inter-channel correlation between the different color components. Nevertheless, most of literature approaches the problem as a post-processing correction, i.e., the processing is applied after the image is captured, processed, and stored. Our approach (discussed next) is inherently different: we consider the application of the image restoration algorithm directly (and separately) on the raw color image data, so that the deblurring and denoising are at the first step of the image reconstruction chain. Applying the image restoration as a pre-processing step ensures that the linear modeling of the problem holds best. Also, the choice of implementing the deblurring at this level of the image formation chain benefits the following cascaded operations from improved resolution and contrast. Example of the following typical processing steps include automatic white balance (AWB), and color filter array interpolation (CFAI) (typically non-linear operations). A similar processing paradigm was proposed earlier [81] in order to reduce color cross-talk and to decorrelate the different color components, although the processing was carried out after color conversion, which may introduce the cross-talk itself. The restoration was proposed without consideration of the difference in the blur of the different color channels. In this work, we use separate processing of the raw RGB color components measured by the camera sensor, and each color channel is separately deblurred according to its corresponding optical blur. In fact, the optical blur in each color channel is different, since the focal length depends on the wavelength of the incoming light [47].

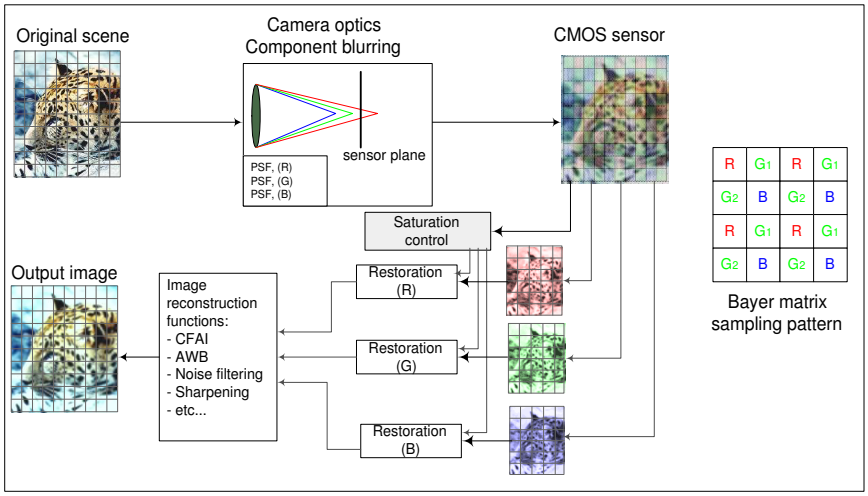


Figure 3.1: Block diagram of the proposed restoration system. The color channels are restored according to the corresponding component blur. The restoration algorithm is applied as the first operation in the image reconstruction chain to minimize non-linearities in the image formation model.

3.3 Multichannel image deblurring of raw color components

3.3.1 Imaging model

Fig. 3.1 depicts the block diagram of the proposed multi-channel restoration within the image reconstruction chain. In the imaging model, the incoming light is blurred by the camera optics, and the image data is measured by a sensor through the Bayer sampling pattern. The optical blurring and the noise sensitivity of each color channel can be different. In fact, by implementing the restoration directly on the raw color components, we are aiming to avoid nonlinearities produced by the cascaded image reconstruction functions such auto-white balancing (AWB), color filter array interpolation (CFAI), distortion corrections, denoising, sharpening, etc. [58].

If we assume a linear response at the sensor and a linear space-invariant blur in each color channel, then the observed image can be modeled as:

$$g_i(x) = (h_i * f_i)(x) + \eta_i(x), \; i = 1, ..., 4 \tag{3.1}$$

where g_i is the measured color component image, f_i is the original color component, h_i is the corresponding PSF in that component, and η_i is an i.i.d. additive Gaussian noise term. g_i, f_i and η_i are defined over the array $x = (m, n)$ of pixels spanning the entire image area. In equation (3.1), $*$ is the discrete convolution operator. The index $i = \{1, 2, 3, 4\}$ denotes respectively the data corresponding to

the *Red*, *Green1*, *Green2*, and *Blue* color channels, those are measured according to the Bayer matrix sampling pattern. Note that each of these images is quarter size of the final output image. The restoration problem can be stated as recovering the original image f_i from its degraded observation g_i.

3.4 Iterative restoration

3.4.1 Generalized Landweber method

Iterative methods have shown to be an attractive alternative for implementing the inverse solution of image deblurring, especially when the blurring parameters can exhibit some modeling errors. The standard *Landweber method* ([57], [73]) to solve for f_i from the observations g_i in equation (3.1) is given by the following iterative process:

$$f_i^{(0)} = 0, \tag{3.2}$$

$$f_i^{(k+1)} = f_i^{(k)} + \mu_i \cdot h_i^T * (g_i - h_i * f_i^{(k)}),\ k = 0, 1, ...,\ i = 1, ..., 4 \tag{3.3}$$

where μ_i is the update parameter, $h_i^T(t) = h_i(-t)$. If the image formation model (3.1) is noise-free, $\eta_i(t) = 0$, the iterative process described above is converging [57]. However, the slow convergence [73] is a significant drawback. The problem of choosing μ_i is one of balancing the stability against the rate of convergence, i.e., a large μ_i ensures a quick convergence but also increases a risk of instability.

Another aspect of the Landweber method in equation (3.3) is the fact that it is designed to solve a problem $g_i(t) = (h_i * f_i)(t)$. As a result, the obtained solution is sub-optimal in presence of noise. We propose to use the following modifications in order to incorporate a noise filtering stage and to enhance convergence:

$$f_i^{(0)} = 0, \tag{3.4}$$

$$f_i^{(k+1)} = \widetilde{f}_i^{(k)} + \mu_i \cdot d_i * h_i^T * (g_i - h_i * \widetilde{f}_i^{(k)}), \tag{3.5}$$

$$\widetilde{f}_i^{(k+1)} = \Phi\{f_i^{(k+1)}\},\ k = 0, 1, ...,\ i = 1, ..., 4 \tag{3.6}$$

where d_i is an impulse response of a special linear filter that can be used to accelerate the *convergence* of the solution. $\Phi\{\cdot\}$ is an intermediate filtering operator that is intended to enhance the robustness of the solution. It can be considered as a separate regularization step. It is interesting to note that in the context of *expectation-maximization* (EM) methods [34], in the iterative process described above, the E-step coincides with equation (3.5), and M-step corresponds to filtering stage in equation (3.6). The operator $\Phi\{\cdot\}$ can be, for example, a simple averaging filter, or any other sophisticated filter that takes into consideration the local signal statistics. In our algorithm, we chose to plug-in an adaptive denoising

filter in order to preserve the image details from over-smoothing. The adaptive filter is based on the polynomial approximation of neighboring pixels from dynamically selected windows. The windows are selected by the rule of intersection of confidence intervals in order to ensure statistical homogeneity of the data in the localized windows. Detailed explanation and results of this filtering approach (LPA-ICI) can be found in the following references ([60], [61]). This adaptive denoising technique plays an important role in our proposed solution because it preserves image details and ensures also efficient noise removal, which is difficult to achieve using filters operating on fixed data support.

3.4.2 Convergence

To study the convergence and the sensitivity of the proposed iterative deconvolution technique, the analysis is done in the Fourier domain. The image formation model in equation (3.1) can be written in the frequency domain as follows:

$$G_i(\omega) = H_i(\omega)F_i(\omega) + \Gamma_i(\omega),\ i = 1, ..., 4 \tag{3.7}$$

where $F_i(\omega) = \mathcal{F}\{f_i(x)\}$ is the Fourier transform of f_i, $G_i(\omega) = \mathcal{F}\{g_i(x)\}$, $H_i(\omega) = \mathcal{F}\{h_i(x)\}$, and $\Gamma_i(\omega) = \mathcal{F}\{\eta_i(x)\}$.

Now, consider the equations (3.4-3.5) in the frequency domain:

$$\begin{aligned} F_i^{(0)} &= 0, \\ F_i^{(k+1)} &= F_i^{(k)} + \mu_i D_i H_i^*(G_i - H_i F_i^{(k)}) \end{aligned} \tag{3.8}$$

where, $D_i(\omega) = \mathcal{F}\{d_i(x)\}$, and H_i^* is the complex conjugate of H_i. If we rewrite equation (3.8) in the following form:

$$F_i^{(k+1)} = (1 - \mu_i D_i |H_i|^2)F_i^{(k)} + \mu_i D_i H_i^* G_i, \tag{3.9}$$

then the error between the original signal and the signal estimate in the frequency domain can be expressed as:

$$E_i^{(k+1)} = q_i E_i^{(k)} - \mu_i D_i H_i^* \Gamma_i \tag{3.10}$$

$$q_i = 1 - \mu_i D_i |H_i|^2, \tag{3.11}$$

where $E_i^{(k)} = \mathcal{F}\{e_i^{(k)}\}$, and $e_i^{(k)} = f_i - f_i^{(k)}$ is the error between the estimated and original images.

The idea behind using the operator D_i is to accelerate the iterative process (3.8), while at the same time ensure that the overall solution converges. As it can be inferred from equation (3.11), the rate of convergence of the iterative process (3.8) is characterized by the variable $q_i(\omega) = 1 - \mu_i D_i(\omega)|H_i(\omega)|^2$. Further, the

convergence takes a place only if $|q_i(\omega)| < 1$. This gives us the conditions that the update parameter μ_i has to satisfy:

$$0 < \mu_i D_i(\omega)|H_i(\omega)|^2 < 2, \text{ for all } \omega, \tag{3.12}$$

where it is assumed that $D_i(\omega)$ is real and positive everywhere on its frequency support. Note that for the pure Landweber method in equation (3.3), $D_i = 1$.
The fastest convergence happens when factor q_i (3.11) is minimal. If we introduce the following variable:

$$\lambda_{i,\max} = \max_{\omega} D_i(\omega)|H_i(\omega)|^2, \tag{3.13}$$

then, according to equation (3.12), the solution converges if μ_i is selected as follows:

$$\hat{\mu_i} = \frac{2}{\lambda_{i,\max} + \epsilon}, \tag{3.14}$$

where $\epsilon > 0$. When considering frequency domain implementations of the iterative process in (3.8), we propose to use the following realization of D_i in order to accelerate convergence:

$$D_i = \frac{1}{|H_i|^2 + r_i^2}, \tag{3.15}$$

where r_i^2 is a regularization parameter. This can be motivated by the fact that this choice of D_i allows us to approach the behavior of the pseudo-inverse filter at each iteration, and hence significantly accelerate the convergence to the solution. In fact, it can be seen from equation (3.9), that the pseudo-inverse filter can be a particular case of this realization when $\mu_i = 1$.

3.4.3 Simulation results with LPA-ICI regularization

To study the properties of the proposed method, we used the *Cameraman* test image in Fig. 3.2 (a). The image was corrupted by the following Gaussian PSF blur

$$h_i(m,n) = \frac{1}{2\pi\sigma_{psf}^2} \exp(-\frac{m^2+n^2}{2\sigma_{psf}^2}). \tag{3.16}$$

We used $\sigma_{psf} = 1$. We further degraded the blurred image with an additive white Gaussian noise, $\sigma_\eta^2 = 40$ (Fig. 3.2 (b)). We compared the restoration results obtained with the standard Landweber method (Fig. 3.2 (c)) against the proposed method with the integrated LPA-ICI filtering (Fig. 3.2 (d)). It can be seen from the images and from the improvement in signal to noise ratio $(ISNR)$ values that the proposed denoising step significantly enhances the performance of the restoration process. It is worth mentioning that the images shown above correspond to the results of the implementation in frequency domain.

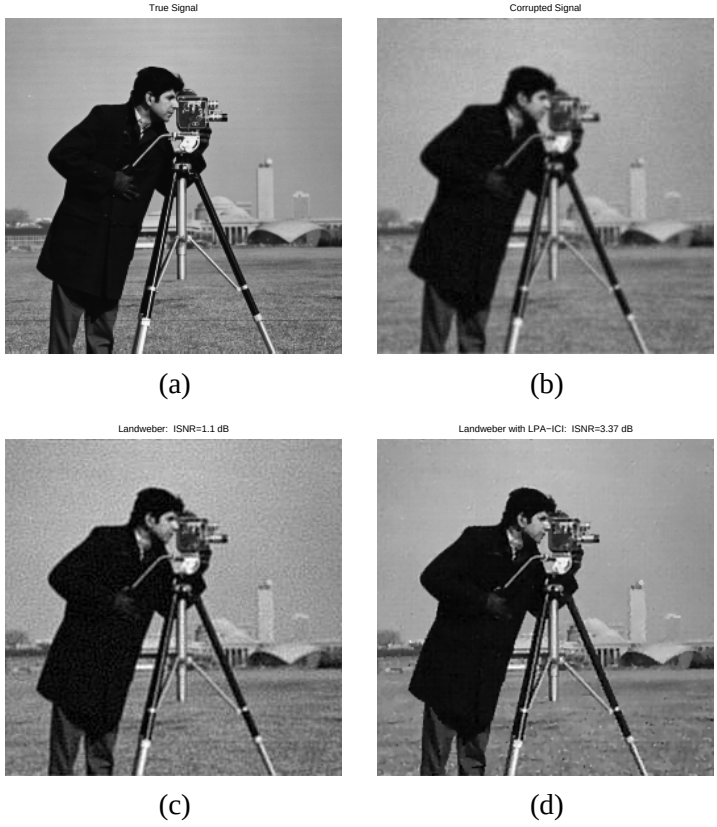
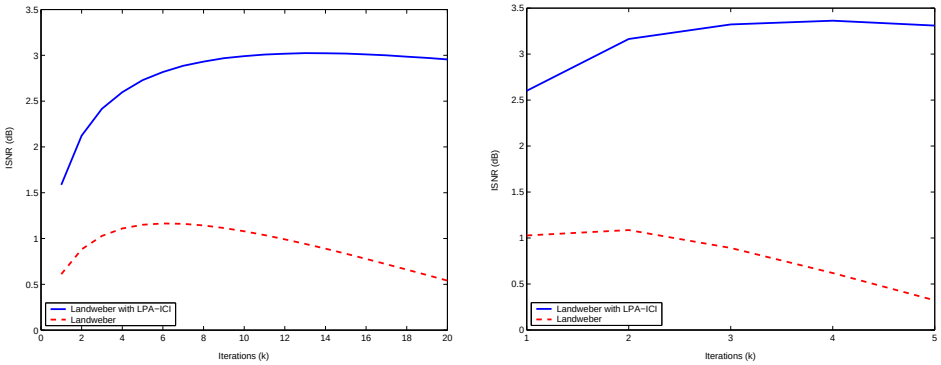


Figure 3.2: (a) Original *Cameraman* test image. (b) Blurred and noisy image, Gaussian PSF ($\sigma_{psf} = 1$), Gaussian additive noise ($\sigma_{\eta}^2 = 40$). (c) Restoration result with the standard iterative Landweber method after 7 iterations. (d) Proposed Landweber method with $LPA - ICI$ after 4 iterations.

Fig. 3.3 (a) shows the improvement in terms of signal to noise ratio ($ISNR$) across the iterations for the restored images in Fig. 3.2 (c) and Fig. 3.2 (d). The $ISNR$ corresponding to the standard Landweber technique (dashed line) takes its maximum and then tends down, due to the noise amplification. The curve corresponding to the proposed technique with LPA-ICI denoising (solid line) clearly improves the stability of the solution. In fact, this adaptive filtering stage acts as a regularization for the inverse solution, while also enhancing the overall quality of restored images.

In Fig. 3.3 (b), we integrated the acceleration operator D_i in equation (3.15) into the frequency domain implementation of the iterative restoration. It can be seen that the number of iterations that is needed to achieve similar performance is significantly smaller. This result confirms the potential of integrating an accelerating spatial operator in the Landweber process in general.



(a) (b)

Figure 3.3: $ISNR$ (in dB) vs. number of iterations (k). (a) Iterative restoration *without* acceleration. (b) Iterative restoration *with* acceleration. In both Figures, the Landweber technique with LPA-ICI denoising (*solid line*) is compared with the standard Landweber technique without denoising (*dashed line*).

3.4.4 Sensitivity to PSF errors

In practice, it is rarely possible to have precise estimates for the point spread function (PSF). Therefore, it is essential to have restoration algorithms that are robust against deviations in PSF parameters.

In order to test the robustness of the proposed solution, we run the algorithms when the exact PSF h_i is known, and when we deliberately introduced different amounts of errors Δh_i into the assumed PSF ($\hat{h_i}$). The corrupted PSF that is actually used in the restoration model can be expressed as follows:

$$\hat{h_i} = h_i + \Delta h_i,\ i = 1, .., 4. \tag{3.17}$$

In our experiments, we used Gaussian PSF $\hat{h_i}$ with parameter $\sigma_{psf} = 1 \pm \tau$, where $\tau \in \{0, 0.1, ..., 0.6\}$ is the assumed estimation error. In Fig. 3.4, we compared the proposed technique (solid line) with the standard Landweber method (dashed line). It is clear from the $ISNR$ curves that the proposed solution is more robust against possible errors in PSF, since the performance was consistently better than the standard Landweber method for all the values of τ that were used.

3.5 Practical considerations

3.5.1 Blur identification

Typically, one of the most difficult practical problems to be solved when restoring images is usually not the choice of a restoration algorithm itself, but rather finding a good PSF. The problem is that the PSF changes as a function of the wavelength

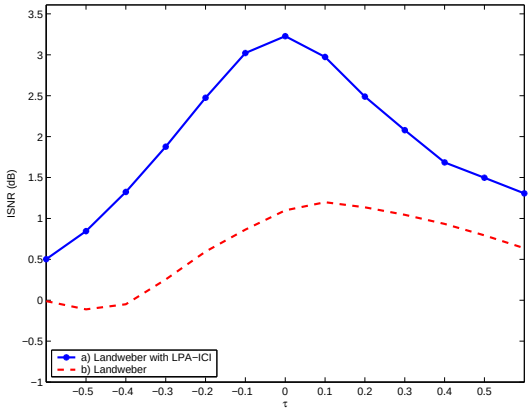
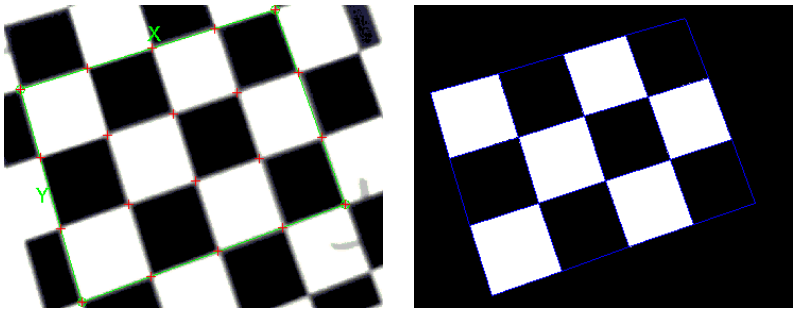


Figure 3.4: Simulation of the sensitivity of the iterative deblurring methods to possible errors in PSF estimates $(\hat{h}_i)$. Gaussian PSF with parameter $\sigma_{psf} = 1 \pm \tau$ is used, where τ is an error that is deliberately introduced.

and the distance of the imaged target with respect to the camera. In the following, we simplify the problem by assuming out-of-focus close range imaging. We further assume space invariant blurring. In order to enable the application to the images captured with a camera phone, we developed a simple PSF estimation technique and used it to find the blurring corresponding to each color component. The procedure is described next.

Given a blurred raw image corresponding to one color component of a checkerboard pattern, the four outer corner points are located manually, and a first rough estimate of the corner positions is calculated. Next, the exact locations (at subpixel accuracy) are recalculated again by refining the search within a square window of 10x10 pixels. The algorithms for corner detection are based on the implementation of the camera calibration Matlab toolbox developed by Heikkilä ([48], [49]). Using the detected corner points, we reconstruct an approximation for the original grid image by averaging the central parts of each square and by asserting a constant luminance value to those squares. Fig. 3.5 (a) shows an example of a test image, and Fig. 3.5 (b) shows the corresponding segmented and reconstructed grid image. Now that we obtained the blurred image (Fig. 3.5 (a)), and the assumed original input (Fig. 3.5 (b)), the blur can be inferred by applying an inverse filtering method.

In the experiments, the PSF estimates are obtained using simple pseudo-inverse filtering (in Fourier domain). Since the pseudo-inverse technique is quite sensitive to noise, we further imposed a low pass cut-off frequency to limit the noise propagation. We repeated the procedure with several images (more than 10, with different orientations of the checkerboard images), and we finally averaged all of them to obtain smooth and reliable estimates. Fig. 3.5 (c) shows an example of the



(a) (b)

Figure 3.5: Procedure to estimate the PSF. (a) From the captured raw image corresponding to each color channel; the corners of the checker-board are located at sub-pixel accuracy. (b) The corner locations are used to reconstruct the sharp pattern of the original checkerboard images.

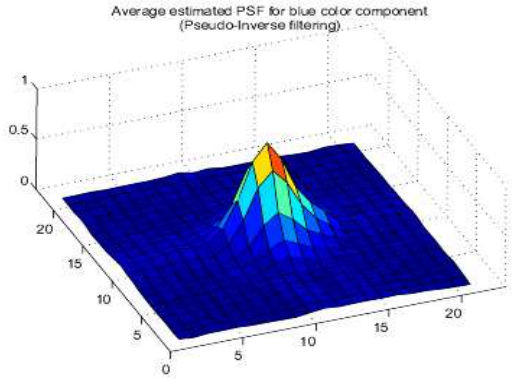


Figure 3.6: An example of the estimated PSF for the blue color channel using raw data from Nokia 6600 camera phone. 10 images are used in the calibration process, all captured at close range ($\sim$ 10cm).

estimated PSF for the blue color channel with a truncated support of 21x21 pixels. In our experiments, it was also confirmed that the 3 color components exhibited slightly different blurring characteristics; interestingly, the blue color component, although it was least contrasted, was the sharpest component.

3.5.2 Implementation of restoration

In order to realize practical real time implementations of the restoration algorithm in equations (3.4-3.6), some approximations in the algorithm have been considered. We have implemented a simplified integer implementation of the algorithm. This causes a loss of precision in the normalized PSF. Because the linear convolu-

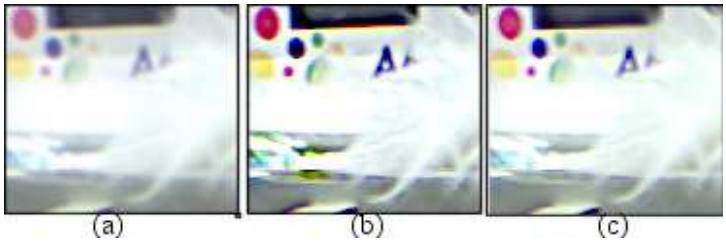


Figure 3.7: Effect of the proposed saturation control mechanism to avoid false coloring in restored images. (a) Original blurred image (b) Restored image (4 iterations) *without* saturation control; remark the green false coloring. (c) Restored image (4 iterations) *with* saturation control. Remark the green false coloring has disappeared. The same reconstruction chain was used in all 3 images.

tion is proportional to the square of the size of the filter support, we have truncated the PSF support to a window of 9x9, which contains most of the information about the defocus degradation, and we considered an approximation to the formal solution in equation (3.4-3.6). We assumed that the first approximation in (3.4) is the observed image itself. So, the iterative approximate solution can be compactly expressed as:

$$f_i^{(0)} = g_i, \tag{3.18}$$

$$f_i^{(k+1)} = \widetilde{f}_i^{(k)} + \mu_{i,adap} \cdot \widetilde{h}_i * (g_i - h_i * \widetilde{f}_i^{(k)}), \tag{3.19}$$

$$\widetilde{f}_i^{(k+1)} = \Phi\{f_i^{(k+1)}\}, i = 1, ..., 4 \tag{3.20}$$

where $\mu_{i,adap}$ is a pixel-wise step size parameter that is designed to avoid false coloring. It is derived in the following section.

3.5.3 Saturation control

In the literature, the formulation of the image acquisition process is invariably assumed to be a linear one (equation 3.1). It is true that this assumption simplifies the inverse problem and allows to develop compact and attractive solutions. However, in practice, the sensor electronics introduce nonlinearities in the acquisition model, of which the saturation effect is one of the most serious. In fact, due to the sensitivity difference of the three color channels, and fuzzy exposure controls, pixel saturation can happen incoherently in each of the color channels. The separate channel restoration near those saturated areas can result in over-amplification of that color component alone, thus creating artificial color mismatch and false coloring near those regions. To avoid this, we propose a novel mechanism that smoothly regulates the restoration process near saturated regions. The saturation control is integrated in the iterative solution of equation (3.20). The idea is to

spatially adapt the update parameter μ_i so as to limit the restoration effect near saturated areas. The adaptive update parameter is given as follows:

$$\mu_{i,adap}(m,n) = \beta_{sat}(m,n)\mu_i, \tag{3.21}$$

where μ_i is the global step-size as discussed earlier, and β_{sat} is the local saturation control that modulates the step size. β_{sat} is obtained using the following algorithm:

1. for each color channel image g_i, $i = \{1 \dots 4\}$,
2. consider the values of the window $(w \times w)$ surrounding pixel $g_i(m,n)$,
   2.1. count the number of saturated pixels $S_i(m,n)$ in that window,
3. the saturation control parameter β_{sat} is calculated as follows:

$$\beta_{sat}(m,n) = max\{0, [w^2 - \sum_{i=1}^{4} S_i(m,n)]/w^2\}.$$

β_{sat} varies between 0 and 1 depending on the number of saturated pixels in any of the color channels. This means for example that if the pixels of a certain color channels are saturated, then β_{sat} will be zero, and no restoration will be performed in that window. Fig. 3.7 shows the effect of the proposed modification to avoid false coloring in the restored images. It can be seen from the examples that the proposed procedure effectively suppresses color mismatch near saturated areas.

3.5.4 Image reconstruction chain

It is important to remind that our processing framework (in Fig. 3.1) proposes to apply restoration of each of the color components as the first spatial filtering operation in the image reconstruction chain. Therefore, an example implementation of the overall imaging chain includes the following cascaded operations:

- Deblurring of color components *(proposed)*
- Pedestal noise removal
- Automatic White Balance
- Color Filter Array Interpolation
- Color gamut conversion
- Geometrical correction and vignetting elimination
- Noise reduction (optional)
- Sharpening (optional)

(a) (b)

Figure 3.8: (a) Test image taken with a Nokia-6600 camera phone and reconstructed with the default processing chain. (b) Final image processed with the proposed deblurring of the raw data after 4 iterations, and reconstructed with the same chain.

It is evident that the final image quality depends on the combined result of all these operations. The implementations corresponding to these individual operations are usually non-linear. As discussed earlier, the use of restoration as the first operation in the reconstruction chain ensures the best fidelity to the assumed linear model. The following algorithms, especially the color filter interpolation (CFAI) and the noise reduction algorithms, can act as additional regularization steps to prevent over-amplification due to excessive restoration.

Fig. 3.8 shows the final result that is obtained when the proposed multichannel restoration algorithm is integrated in the reconstruction chain of a real camera system. The processing was carried out on the raw pictures captured with a Nokia 6600 camera-phone. As it can be seen from the images, there is an evident improvement in detail resolution and in color contrast. We have also tested with a large set of real scene images, and the visual quality of the result images was consistently better than the images obtained with the default reconstruction chain.

3.6 Conclusions

In this chapter, we presented an integrated filtering method that reduces the optical blur. The filtering is an integral part of the camera image reconstruction chain. We implemented a modified iterative Landweber algorithm which uses an adaptive window denoising method to regularize the inverse filtering process. The proposed deblurring method makes use of the estimated PSF that characterizes the optical blurring in each of the three color components. To avoid false coloring that may happen due to independent component filtering in RGB space, we integrated a novel saturation control mechanism, which smoothly attenuates the

(a) (b)

Figure 3.9: (a) Test image taken with a Nokia-6600 camera phone and reconstructed with the default processing chain. (b) Final image processed with the proposed deblurring of the raw data after 3 iterations, and reconstructed with the same chain.

iterative restoration near saturated regions. The experimental results have shown that proposed method for image restoration attenuates the blurring due to fixed-focus cameras integrated in mobile devices. In general, the results demonstrate the potential of considering image deblurring as an integral part of image reconstruction chains, especially when implemented on hardware.

Chapter 4

Super-Resolution from Sensor Data

4.1 Introduction

Despite the continuous improvement in the performance of imaging sensors, there are still several physical and practical constraints that limit the final image quality. Therefore, signal processing techniques are widely used to improve the quality of the sensed raw data. In this chapter, we present a framework for producing a high-resolution color image directly from a sequence of images captured by a CMOS sensor overlaid with a color filter array. The algorithm attempts to utilize the additional temporal resolution in order to improve the demosaicing of the color components and filter the noisy and blurred image data. The method is based on iterative super-resolution that performs separately the filtering of the individual color image planes. We present experimental results using synthetic image sequence as well as real data from CMOS sensors.

4.1.1 Spatial resolution in image sensors

Lately, the industry trend has been focusing on reducing the pixel size in order to improve the spatial resolution. This approach leads to reduced sensitivity of the individual pixels and amplification of the noise levels because CMOS sensor performance is limited by low quantum efficiency and by dark current non-uniformity [15]. In fact, regardless of the sensor manufacturing technology, there is a fundamental trade-off between spatial sampling (number of pixels), pixel size, and temporal sampling.

4.1.2 Color plane interpolation

In most consumer products, the integrated cameras consist of a single imaging sensor that uses Color Filter Arrays (CFA) to sample different spectral components. The most common sampling pattern is called the Bayer-matrix (Fig. 4.2), which consists of color filter elements arranged such that green component is 50% of the total number of pixels, whereas each of the red and blue components represent 25% of the total number of pixels. At each pixel location, the missing colors must be interpolated from neighboring samples. This color plane interpolation is known as demosaicing, and it is one of the important tasks in the image reconstruction (formation) pipeline.

There has been significant work related to demosaicing of raw color images. A good review of the common techniques can be found in [44]. Besides the traditional interpolation techniques such as nearest-neighbor replication, bilinear interpolation, and cubic spline interpolation; several algorithms have been proposed to enhance the performance of demosaicing. For example, in [1], [58] and [65], inter-channel correlation is exploited in various ways to perform edge-directed interpolation and to prevent color mismatch across the edges. In [41], a technique using alternating projections onto convex sets is proposed. The performance was superior to earlier techniques [41] because the projections across the color planes provided for an efficient way to exploit inter-channel correlation. In all of the algorithms cited above, only the spatial correlation of the neighboring pixels is used, without exploiting the additional information that is available in the adjacent frames.

Super-resolution can be used to combine a short sequence of raw video output from the sensor to reconstruct a high-resolution image. The idea is to exploit the non-redundant samples that fall on the Bayer sampling lattice in order to perform demosaicing of the color data, and possibly increase the original image size, with improved color fidelity. However, to date, almost all super-resolution methods have been designed to increase the resolution of a single monochromatic channel (luminance component), and it was not until recently that the use of multi-frame processing was considered in the problem of demosaicing ([31], [39], [93]). In [39] and [93], it is argued that although a two-pass algorithm (demosaicing followed by super-resolution) improves the overall resolution, this approach results in blurring effects and artifacts similar to those observed in demosaiced images. Also, they have shown that it is possible to obtain precise motion using block matching from aliased raw images. However, the overall method is complicated, and involves among other processing steps edge directed interpolation. Therefore, it is not straight-forward to separate the distinct contribution of super-resolution in the final demosaiced result. In [31], a method based on MAP estimation is proposed to jointly perform demosaicing and super-resolution, but the paper does not address the problem of estimating motion between the images, and only presents results with synthetic data sets.

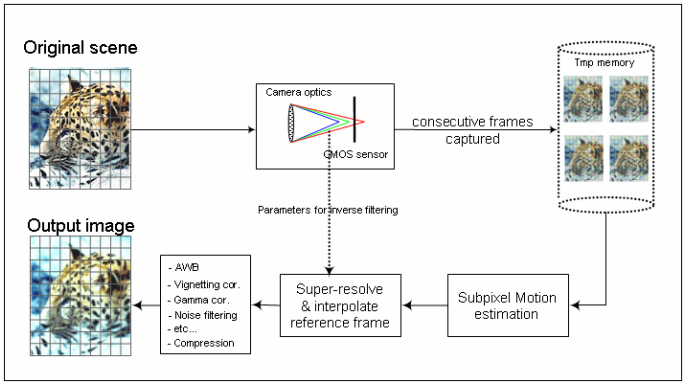


Figure 4.1: Integrated image formation (reconstruction) model using the proposed super-resolution filtering.

4.1.3 Our approach

In the following, we consider the application of super-resolution directly on a sequence of raw images. The idea is to enhance the interpolation of the different color planes by utilizing the data from neighboring image frames. One benefit from applying super-resolution as a pre-processing step (before cascading into the following operations) is that it ensures that the linear modeling of the problem holds best, at least conceptually. In earlier work [108], we have demonstrated that by applying deblurring directly on the raw color components, we were capable of producing superior results, especially after implementing simple mechanisms to avoid color mismatch.

In Section 4.2, the image formation model is defined, and in Section 4.3, the solution is derived by applying super-resolution on the color components, also, some implementation issues are highlighted. Experimental results of the proposed algorithm with synthetic image set as well as with real sensor data are presented in Section 4.5. Finally, conclusions are drawn in Section 8.

4.2 Image formation model

In order to produce the final image from the sensor data, the default processing pipeline includes several cascaded operations such as automatic white balance (AWB), gamma correction, contrast enhancement, color filter array interpolation (CFAI), denoising, sharpening. Most of the filtering that is involved in these operations is non-linear. Hence, our proposal to apply super-resolution as the first step of the reconstruction pipeline (before cascading into the following operations) ensures that the linear modeling of the problem holds best. The modified reconstruction chain uses few raw images to perform enhanced interpolation of the

Bayer color data, which replaces the conventional step of CFAI filtering. Fig. 4.1 depicts the block diagram of the proposed super-resolution reconstruction chain.

In the assumed model, the incoming light is blurred by the camera optics, and the image data is measured by a sensor through the Bayer sampling pattern. The optical blurring and the noise sensitivity of each color channel can be different. We assume linear sensor response, as well as linear space-invariant blur for each color channel. Now, consider N observed raw LR images $(g_i\ , i = 1 \cdots N)$ that are captured by the sensor in video mode, such that the images are adjacent observations of a static scene corresponding to a HR color image (f). Further, we assume that each subsampled color component image is an independent realization of the imaged scene, i.e., for the i^{th} LR image g_i, we capture separately 4 color channels $g_{i(c)}$, where the index $c = \{1, 2, 3, 4\}$ denotes respectively the data of the *Green1*, *Red*, *Blue*, and *Green2* color channels; as measured according to the Bayer sampling pattern (Fig. 4.2). Using the assumptions above, the image formation model can be written as:

$$\begin{array}{rcl} g_{i(1)}(x,y) & = & S\downarrow \left(h_{(1)} * f_G\left(\xi_i\right)\right)(x,y) + \eta_{(1)}(x,y) \\ g_{i(2)}(x,y) & = & S\downarrow \left(h_{(2)} * f_R\left(\xi_i\right)\right)(x,y) + \eta_{(2)}(x,y) \\ g_{i(3)}(x,y) & = & S\downarrow \left(h_{(3)} * f_B\left(\xi_i\right)\right)(x,y) + \eta_{(3)}(x,y) \\ g_{i(4)}(x,y) & = & S\downarrow \left(h_{(4)} * f_G\left(\xi_i\right)\right)(x,y) + \eta_{(4)}(x,y) \end{array} \tag{4.1}$$

where $f = (f_R, f_G, f_B)$ is the HR reference image corresponding to the imaged scene in the RGB domain. $h_{(c)}$ denotes the point spread function, or the PSF due to optical blurring in each color channel, $*$ denotes the convolution operator, and $S\downarrow$ the down-sampling operator. Note that in equation (4.1) each color component is subsampled at a different offset due to the specific pattern of the Bayer sampling matrix. ξ_i is the mapping function corresponding to the geometric warping due to the scene displacement in each of the LR images relative to the HR image f, while $\eta_{(c)}$ is an additive noise term that is associated separately with each color channel.

After discretization, the model can be expressed in matrix form as follows:

$$\begin{array}{rcl} \overline{g}_{i(1)} & = & \mathbf{A}_{i(1)}\overline{f}_G + \overline{\eta}_{(1)} \\ \overline{g}_{i(2)} & = & \mathbf{A}_{i(2)}\overline{f}_R + \overline{\eta}_{(2)} \\ \overline{g}_{i(3)} & = & \mathbf{A}_{i(3)}\overline{f}_B + \overline{\eta}_{(3)} \\ \overline{g}_{i(4)} & = & \mathbf{A}_{i(4)}\overline{f}_G + \overline{\eta}_{(4)} \end{array} \tag{4.2}$$

The matrix $\mathbf{A}_{i(c)}$ combines successively, the geometric transformation ξ_i, the convolution operator with the blurring parameters of $h_{(c)}$, and the down-sampling operator $S\downarrow$ over the Bayer grid. For notation convenience, we integrate the RGB correspondence in matrices $\mathbf{A}_{i(c)}$, and we express the image model using the following equation:

$$\overline{g}_{i(c)} = \mathbf{A}_{i(c)}\overline{f} + \overline{\eta}_{(c)} \tag{4.3}$$

where $\overline{g}_{i(c)}$, $\overline{f}$ and $\overline{\eta}_{(c)}$ are lexicographically ordered.

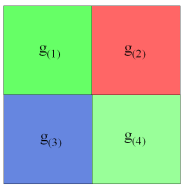


Figure 4.2: Bayer matrix sampling pattern

Referring to equation (4.3), the size of each of the LR images $\overline{g}_{i(c)}$ is quarter that of the sensed raw image. The separation of the raw image data into 4 individual subsampled images introduces significant aliasing, since for each color plane we are skipping the pixels from the next color component without low-pass filtering. In principle, super-resolution algorithms exploit the aliasing of the image data in order to reconstruct high frequency information. In the following section, we present the solution for interpolating each color component and combining the information from adjacent frames.

4.3 Super-resolution from raw sensor data

In this section, we describe the algorithm that performs simultaneously the demosaicing of the color components while fusing the data from the LR frames. The HR image is in RGB domain, while the individual LR images are assumed to be the monochrome components after performing the subsampling according to the Bayer pattern. The problem of super-resolution reconstruction can be described as estimating for each color channel the best HR image $\overline{f} = \left(\overline{f}_R, \overline{f}_G, \overline{f}_B\right)$, which when appropriately projected and down-sampled by the model in (4.2) will generate the closest estimates of the LR images $\overline{g}_{i(c)},\ c = \{1 \cdots 4\}$.

4.3.1 Cost function

For each observation $\overline{g}_{i(c)}$, we associate the following cost function:

$$\epsilon_{i(c)} = ||\widehat{g}_i - \overline{g}_i||^2 = ||\mathbf{A}_{i(c)}\overline{f} - \overline{g}_{i(c)}||^2, \tag{4.4}$$

where $\widehat{g}_{i(c)}$ is the simulated LR image through the forward imaging model. If all LR images $(i = 1 \cdots N)$ are assumed to contribute equally to the cost function, then the overall cost function is the following:

$$\epsilon_{(c)} = \sum_{i=1}^{N} \epsilon_{i(c)} = \sum_{i=1}^{N} ||\mathbf{A}_{i(c)}\overline{f} - \overline{g}_{i(c)}||^2 \tag{4.5}$$

Further, if Gaussian noise is considered in the imaging model $(\overline{\eta}_{(c)})$, then minimizing the cost function in (4.5) (least squares solution) is equivalent to the maximum likelihood solution.

4.3.2 Iterative super-resolution

In order to minimize the error functional in (4.5), the method of iterative gradient descent is used. This technique seeks to converge $\epsilon_{(c)}$ towards a local minimum by following the trajectory of the negative gradient; i.e., at iteration n, the high-resolution image is updated as follows:

$$\overline{f}^{n+1} = \overline{f}^{n} + \mu^n \overline{r}^n_{(c)}, \tag{4.6}$$

where μ^n is the step-size, and $\overline{r}^n_{(c)}$ is the residual gradient due to the LR color images (c).
The residual gradient $\overline{r}^n_{(c)}$ is computed as follows:

$$\overline{r}^n_{(c)} = \sum_{i=1}^{N} \mathbf{W}_{i(c)} \left(\overline{g}_{i(c)} - \mathbf{A}_{i(c)} \overline{f}^n \right). \tag{4.7}$$

The matrix $\mathbf{W}_{i(c)}$ corresponds to $\mathbf{A}^{(-1)}_{i(c)}$, i.e., the inverse process of the image formation. In practice, $\mathbf{W}_{i(c)}$ combines successively the up-sampling and the inverse geometric warp (ξ_i^{-1}) such as to map the i^{th} LR image grid onto the HR grid.

In the update equation (4.6), the same step size μ^n is used for all color channels $(c = 1, 2, 3, 4)$); this means that all the color channels are iteratively minimized at the same speed to minimize consequent false coloring. For example the step size is calculated using the green component $(\overline{f}_G)$ using the method of steepest descent [12] as follows:

$$\mu^n = \frac{1}{N} \sum_{i=1}^{N} \frac{||\overline{g}_{i(1)} - \mathbf{A}_{i(1)} \overline{f}^n_G||^2}{||\mathbf{A}_{i(1)} \overline{r}^n_{i(1)}||^2}. \tag{4.8}$$

4.4 Implementation

In the following, we highlight some implementation issues that are needed to realize a practical implementation of the proposed algorithm. We recall that super-resolution shall be considered as a processing framework that involves several operations such as motion estimation, interpolation and demosaicing. In the specific context of super-resolution of raw images, we deal with multiple data channels corresponding to different spectral components (colors) as well as different temporal samples (images), this makes the alignment of all these data channels a challenging task in the implementation stage. Next, we show how we deal with this problem when implementing the different operations needed in super-resolution.

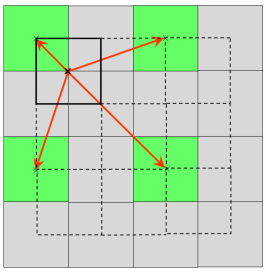


Figure 4.3: Pixel projection from interpolated RGB domain (dashed lines) onto a single raw color component (green 1). Note the uneven spacing that is used in the pixel projections.

4.4.1 Motion estimation

One critical aspect to achieve efficient implementations of image super-resolution is the need for accurate sub-pixel registration of the input images. The problem of estimating sub-pixel motion from raw data has been investigated in detail in [93]. In our implementation, we used the subsampled components to estimate global projective motion parameters. The motion was estimated for each of the color channels separately, and the resulting motion parameters were refined using a simple averaging operation.

Because the overall performance of super-resolution algorithms is particularly degraded in the presence of persistent outliers, we included a simple mechanism in the motion estimation process that asserts the confidence of the obtained estimates; i.e., if the mean square error (MSE) between the reference frame and the motion-compensated LR image is larger than a given threshold, then we skip that frame throughout the entire reconstruction process.

4.4.2 Initialization of iterative super-resolution

It is well known that the iterative Least Squares solution, as in equation (4.6), is prone to divergence especially when the number of input images is limited. If this happens, annoying artifacts start to appear when over-iterated. This is due to the absence of a proper regularization term. To avoid this, we use a smooth initial estimate of the HR image and we limit the number of iterations, especially when we know that the input sequence is noisy. The initial HR estimate is obtained by demosaicing the reference frame (which is used in the motion estimation process) by applying simple bilinear CFAI, and then interpolating to the desired zoom factor, for instance by using bilinear interpolation in the RGB domain.

4.4.3 Projection functions

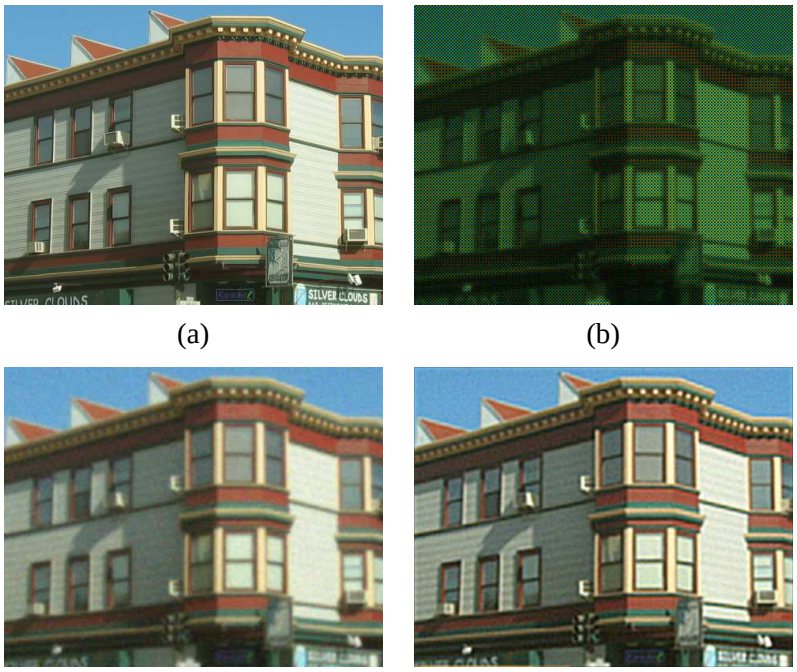
When implementing the image synthesis and the inverse process defined respectively by $\mathbf{A}_{i(c)}$ and $\mathbf{W}_{i(c)}$, we used a process similar to that described in [21]. In the synthesis process, or the forward-projection, we warp the HR image as point samples and convolve with a continuous form of the point-spread function (PSF), and we downsample at the required positions on the Bayer pattern. We assumed the PSF can be approximated with a Gaussian function, so that we can easily integrate the blurring as a single parameter in the convolution process. This operation is rather delicate to implement, Fig. 4.3 shows an example of the assumed positioning of the HR image grid (in dashed line) with respect to the LR image grid. The corresponding half pixel shifts need to be integrated in the motion parameters of each LR image. The inverse mapping, or the back-projection process $(\mathbf{W}_{i(c)})$ is handled in a similar manner: instead of naive downsampling, we interpolate to the required Bayer pattern positions. The region of influence affected by the back-projected pixel is determined by the interpolation filter (also Gaussian). The smaller the variance of the filter, the sharper the resulting HR image is. However, this also means that more LR samples (N) are needed to avoid amplification of the noise and annoying pixelized effect in the solution. A smoother interpolation filter will make a compromise between the number of input images, the noise level and the sharpness of the result.

4.4.4 Processing the green channel

Another problem that we need to take in consideration is the fact that for each LR image, we have 2 sub-images corresponding to the green spectral component, $\overline{g}_{i(1)}$ and $\overline{g}_{i(4)}$, which correspond to a single channel in the HR image $(\overline{f}_G)$. This can be handled in many ways, for example by averaging the corresponding back-projected components, i.e., the residual gradients $\overline{r}^n_{(1)}$ and $\overline{r}^n_{(4)}$ due to the green spectral component, this procedure has been used in the experiments presented in the following section. Alternatively, it is possible to use one of the channels for the regularization of the iterative solution, meaning that while $\overline{g}_{i(1)}$, $\overline{g}_{i(2)}$ and $\overline{g}_{i(3)}$ components of the Bayer image are used as reconstruction constraints of the G, R and B color components respectively, $\overline{g}_{i(4)}$ may be used as smoothness constraint of the final solution, especially if we assume a MAP iterative implementation, in this case $\overline{g}_{i(4)}$ can be used to calculate a non-redundant prior distribution of the HR image.

4.5 Experimental results

In this section, we present experiments on synthetic and on real sensor data. First, we tested the algorithm on a sequence of synthetic test images. The images, 6 in

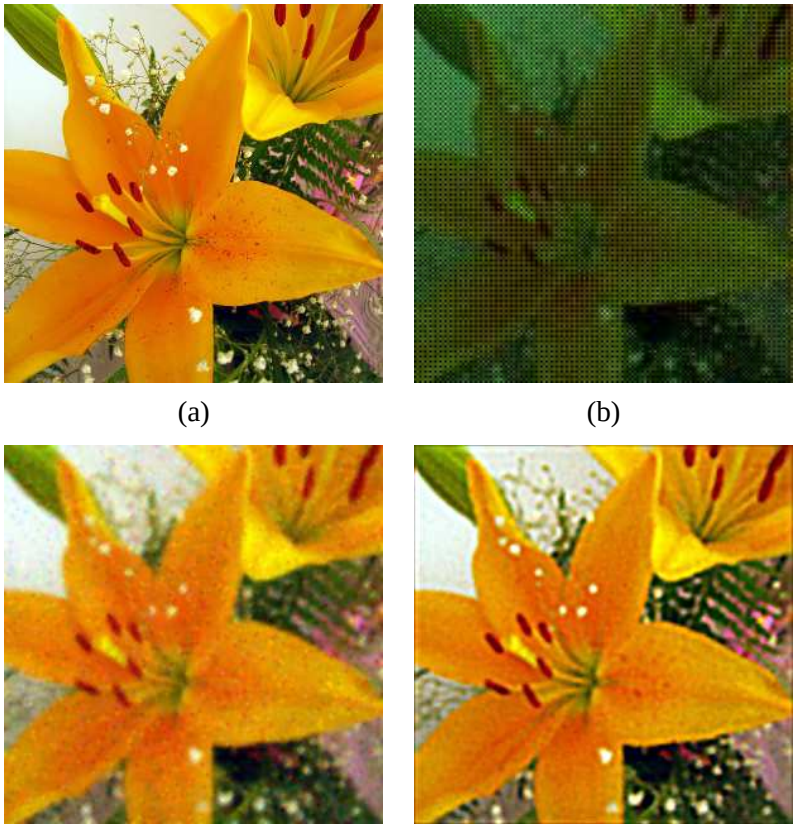


(a) (b)

(c) (d)

Figure 4.4: (a) Original HR image. (b) Example LR image obtained according to model in equation (4.1), Gaussian PSF ($\sigma^2_{psf} = 1.5$), zoom factor 2, additive Gaussian noise ($\sigma^2 = 20$). (c) Image obtained using bilinear CFAI interpolation and bicubic interpolation ($SNR_R = 9.44$, $SNR_G = 10.77$, $SNR_B = 10.45$, $SNR_Y = 10.19$). (d) Image obtained using the proposed algorithm, 2 iterations, ($SNR_R = 10.88$, $SNR_G = 12.19$, $SNR_B = 11.68$, $SNR_Y = 11.50$).

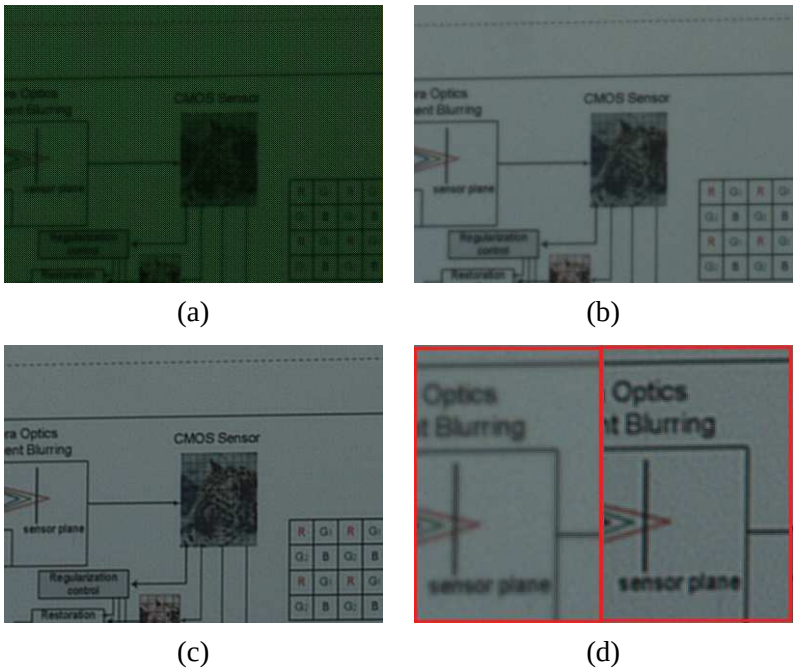
total, were generated from a single HR image according to the imaging model described in equation (4.1). The original HR image was randomly warped using an 8 parameter projective model. We used a continuous Gaussian PSF ($\sigma^2_{psf} = 1.5$) as the blurring operator, and we down-sampled the images by 2 to obtain the 6 LR images. We added to all input images a zero-mean Gaussian noise with variance $\sigma^2_\eta = 20$. Fig. 4.4 (c) shows the image that is reconstructed using bilinear CFAI interpolation and interpolated to target size (by 2) using bicubic interpolation. Fig. 4.4 (d) shows the result image that is obtained using the proposed algorithm after 2 iterations. For the images in (c) and (d), we calculated the signal to noise ratio with respect to the original image in (a) corresponding to the RGB color channels (SNR_R, SNR_G and SNR_B), as well as for the luminance component (SNR_Y). These are shown in the caption of Fig. 4.4. In terms of SNR, the proposed super-resolution algorithm enables an improvement of about $1.5db$ over equivalent traditional processing which consists in applying separately the CFAI interpolation



(a) (b)

(c) (d)

Figure 4.5: (a) Original HR image. (b) Example LR image obtained according to model in equation (4.1), Gaussian PSF ($\sigma^2_{psf} = 2.5$), zoom factor 2, additive mixed noise (Gaussian noise $\sigma^2_\eta = 20$, impulsive noise $p = 0.06$). (c) Image obtained using bilinear CFAI interpolation and bicubic interpolation ($SNR_R = 9.5$, $SNR_G = 9.43$, $SNR_B = 10.12$, $SNR_Y = 9.21$). (d) Image obtained by applying super algorithm (median fusing), 4 iterations, ($SNR_R = 10.92$, $SNR_G = 11.22$, $SNR_B = 12.16$, $SNR_Y = 11.19$).



(a) (b)

(c) (d)

Figure 4.6: (a) Example of raw data captured (4 images) with Micron test camera board (MI SOC1310). (b) Image obtained using bilinear CFAI interpolation of reference image. (c) Image obtained by applying proposed algorithm, zoom factor 1, 3 iterations. (d) Close-up comparison between zoomed portions of the images shown in (b) and (c).

(in this case bilinear) and resizing operation (bicubic interpolation). In addition, close visual inspection of the images in (c) and (d) confirms the good performance of the proposed algorithm seeing that some fine details of the original image have been recovered in the super-resolved image shown in (d), whereas these details are not visible when using a single interpolated image (c). The perceived image quality is also improved, since the contrast is enhanced and the colored artifacts due to noise are reduced thanks to the integrated filtering approach used in the algorithm. It is worth mentioning that the algorithm is relatively fast, for example in Fig. 4.4, the target image size was 320×240, and in this setting, the resulting images were computed in real-time on an ordinary PC.

In Fig. 4.5, we perform a similar experiment as above with different burring and noise parameters. 4 LR images were generated from a single HR image by random warp. We used a continuous Gaussian PSF ($\sigma^2_{psf} = 2.5$) and down-sampling factor 2. A mixture of Gaussian noise ($\sigma^2_\eta = 20$) and impulsive noise ($p = 0.06$) is added to all LR images. The aim is to emulate random noise that may appear in CMOS sensed images. Fig. 4.5 (c) shows the image that is re-

constructed using bilinear CFAI interpolation and interpolated to target size (by 2) using bicubic interpolation. Fig. 4.5 (d) shows the result image that is obtained using the proposed algorithm after 4 iterations. One simple modification with respect to earlier results is that in equation (4.7) we used the median filter instead of averaging filter, the idea is to account for impulsive noise present in the LR images. In terms of SNR, the super-resolved image is about $2db$ better than the image obtained using equivalent traditional processing, in particular the super-resolved image has less false coloring artifacts that are due to impulsive noise in the individual color components. This result is interesting because it shows that it is relatively easy to account for other types of spatial noise (in this case impulsive) by modifying the fusing step of the super-resolution algorithm, this result may be further developed in future work to account for other types of noise distributions.

In Fig. 4.6, we show the performance of the proposed method using real sensor data without prior knowledge of the distortion or motion that happened between the LR images. The proposed algorithm is applied on a set of 4 images captured by a CMOS camera board (Micron SOC1310). The images were taken slightly out-of-focus, on purpose to simulate a fixed-focus optical system. In Fig. 4.6 (d), we show a zoomed portion of the image to compare the results obtained by applying simple bilinear interpolation against the images obtained using the proposed algorithm. Although the parameters that were used in our algorithm were tuned without accurate knowledge of the forward imaging model; the obtained results were good, i.e., from the visual inspection it is clear that the details became sharper, the noise was decreased and the contrast in the result image was better. This confirms the usability of the proposed solution for practical application in camera systems.

4.6 Conclusions

In this chapter, we presented a super-resolution algorithm that takes a sequence of raw color images, and produces a demosaiced color image in the RGB domain. The proposed algorithm was compared with Bilinear CFAI interpolation, and it showed an outstanding performance both visually and in terms of signal to noise ratio; and at a reasonable computational complexity. The results confirm that this approach for super-resolving raw sensor data is promising, and is capable of producing superior results. In future work, we shall validate our method against an extensive set of known demosaicing algorithms to confirm that the use of multi-shot mode is capable to improve significantly the color and detail resolution. Another interesting direction is to explore integrated fusing methods that may be more appropriate to use in different noise environments of the sensors. Also, we shall consider the development of fast motion estimation techniques that will be suitable for the application of super-resolution on raw image data.

Chapter 5

Motion Estimation

5.1 Introduction

The estimation of motion in image sequences is often required in various video applications. In the literature, there is an enormous variety of different approaches, which we can not exhaustively survey here. Instead, we provide in this chapter an introductory discussion on the challenging problem of motion estimation (Section 5.2), we focus on the specific context of image super-resolution. In the following Sections (5.3–5.8), we present a novel filtering method that is based on a recursive LMS filtering scheme to match the intensity values. The adapted filter coefficients are used to calculate a dense motion field between two adjacent video frames. The employed filtering takes advantage of the localized correlation of image data in adjacent frames, and produces refined estimates of the displacements at sub-pixel accuracy. The proposed method for pixel-based motion estimation is well suited for the estimation of small displacements within video frames, and can be applied in several applications such as super-resolution, video stabilization and denoising of video sequences.

5.2 Image registration

Image registration establishes the correspondences between the pixel positions from a target frame with respect to a reference frame. This operation is fundamental to many image and video processing tasks, and most notably in video compression systems. Although the understanding of the issues involved in motion estimation has significantly increased in the last decade, we are still far from generic, robust and real-time algorithms [98]. Fundamentally, the problem is challenging because of its ill-posedness, the aperture problem, and the occlusion that happens when 3-D moving objects past each other are projected on a 2-D plane.

In this thesis, we are mostly concerned with motion estimation techniques that can be used for super-resolution. In this context, the major requirements for image

registration are the following:

– Accuracy: sub-pixel precision is required in super-resolution

– Robustness towards image degradations [1]

– Robustness to outliers represented by moving objects in the scene

– Robustness to changes in illumination

– Reasonable computational complexity [2]

It is well known that satisfying these requirements together is not straightforward. In principle, motion estimation aims to reveal the 3-D motion trajectory from the observed image intensities, which means that we need to constrain the problem and assume certain imaging and scene models. The most common assumption is that the image intensity along a motion trajectory remains constant. This is usually known as the brightness constancy assumption. Additionally, some other constraints may be assumed, such as the constant gradient assumption, or the smoothness of the motion field, etc.

5.2.1 Motion field representation

Model	**Transformation** $\begin{bmatrix} x \\ y \end{bmatrix} \Rightarrow \begin{bmatrix} x' \\ y' \end{bmatrix}$	**Camera motion**	**Number of parameters**
Translational	$\begin{cases} x' = x + t_x \\ y' = y + t_y \end{cases}$	2D translation	2
Euclidean	$\begin{cases} x' = x\cos(\alpha) - y\sin(\alpha) + t_x \\ y' = x\sin(\alpha) + y\cos(\alpha) + t_y \end{cases}$	translation, rotation	3
Affine	$\begin{cases} x' = ax + by + t_x \\ y' = cx + dy + t_y \end{cases}$	translation, rotation, skew	6
Projective	$\begin{cases} x' = (ax + by + t_x)/(gx + hy + 1) \\ y' = (cx + dy + t_y)/(gx + hy + 1) \end{cases}$	translation, rotation, skew, perspective	9

Motion field representations can be divided into two broad categories: non-parametric and parametric models. In the first model, the motion field is represented at each point (or point samples) by a vector indicating the estimated

[1] e.g. noise, blur caused by camera motion during the exposure time, or by out-of-focus optical system.

[2] image registration is usually the most computationally complex part of a multi-shot image processing application. This requirement is important if real-time operation with large image sizes is desired.

displacement at that point. In the parametric representation, only few parameters corresponding to a pre-determined geometric model are needed to represent the motion in a given image area. The table above displays some common linear parametric motion models and their corresponding geometric transformations. Additionally, if the imaging system suffers from optical distortions, higher order polynomial models can be used to integrate the corresponding geometric distortions into the parametric model.

The motion fields, in particular the parametric representations, are usually represented on various partitions of the image plane. The most frequently employed regions of support for motion are the following: global models, block-based models, triangular or hierarchical block-based meshes, and single point support (to represent dense motion fields).

For super-resolution processing, one particular advantage of the parametric motion field representation is its capability to accurately register camera translation, rotation, zooming, or projection on a static planar scene. On the other hand, a major problem (with global registration techniques) is that they are limited to the assumed parametric model, and completely fail in the presence of local outliers. Such outliers may be for instance due to moving objects inside the scene, or due to the presence of repetitive textures and noisy areas. In those cases, non-parametric registration techniques (dense optical flow field) may be more suitable to handle the moving objects. However, this usually comes at the expense of precision loss and additional memory to store the motion field.

5.2.2 Common approaches for motion estimation

Two different approaches for motion estimation have been separately developed; i.e., image based discrete motion estimation (block matching), and gradient based techniques, or optical flow estimation.

Block-based methods

This method for discrete motion estimation establishes the correspondences by measuring similarity using blocks or masks (block-matching). The technique is simple, but nevertheless the most widely used because of its ubiquitous application in video coding schemes such as H.261, H.263, MPEG-1, MPEG-2, and MPEG-4, AVC. The motion of each block (typically 16x16 pixels) is represented by a two-parameter vector. Sub-pixel accuracy is often achieved by interpolating between the pixel intensity values. There are several variations of this basic estimation approach using different combinations of minimization strategies, search techniques, interpolation procedures, hierarchical derivations, smoothing methods, etc.

In general, the advantages of block-matching are simplicity and reliability for discrete large motion. Originally, this approach has been mainly developed to

improve compression performance in video coding applications. As a result, the motion vectors do not necessarily reflect real motion, suffer from discontinuities along the block boundaries, and they generally fail to register detailed motion. Due to the shortcomings listed above, the direct application of block-based motion estimation in filtering applications is not considered a viable option.

Optical flow methods

These methods aim to obtain a velocity field by computing the spatial and temporal image derivatives. Since the initial formulation of the optical flow equation for motion estimation by Horn and Schunck [51] in 1981, several variations based on this approach have been described in the literature. The reported algorithms usually consist in deriving a dense flow field by applying a variety of techniques to compute the gradient images and regularization strategies to solve for the local parametric motion at each pixel position. This approach has been very popular in applications considering different sorts of computer vision tasks.

In general, the advantage of this approach is the capability to accurately estimate of the elastic displacement between frames, i.e., the adequate handling of piecewise motion. However, the computational complexity associated with optical flow estimation techniques is usually prohibitively expensive (see Table 1 in [9]), and this makes it unsuitable for direct use into video filtering applications.

Motion estimation for filtering applications

Based on the arguments mentioned above, and mainly because they have been designed for different applications, one may conclude that the traditional approaches for motion estimation are not very suitable for motion-compensated filtering applications, such as super-resolution. For instance, in video filtering applications, we can afford to skip areas that may not be accurately registered, but we may not afford blocking artifacts in the final result. This raises the need for research in new approaches for motion estimation, which may provide a compromise between computational complexity, ease of integration and accuracy.

5.3 Dense optical flow field estimation using recursive LMS filtering

Introduction

In the following, we develop a novel motion estimation algorithm that can be used in video filtering applications. The method is inspired from the analogy of audio echo-cancellation, which successfully applies adaptive filtering to match the delayed audio components. Similarly, we attempt to use a 2-D LMS filter to match the intensity values and calculate the displacement between two adjacent video

frames. The proposed algorithm adapts the corresponding set of FIR coefficients. The peak value in the resulted coefficient distribution points to the localized displacement that happens between two successive frames.

The adaptive filtering takes advantage of the localized correlation of image data in adjacent frames, and produces refined estimates of the displacements at sub-pixel accuracy. One particular advantage is that the proposed method is flexible and well suited for the estimation of small displacements within video frames, which makes it useful in video filtering applications such as super-resolution [103], video stabilization [102], and denoising of video sequences.

Adaptive filters

An adaptive filter has an adaptation mechanism that is meant to monitor the environment and vary the filter transfer function accordingly. The algorithm starts from a set of initial conditions, that may correspond to complete ignorance about the environment, and based on the actual observed signals, it tries to find the optimum filter design. In principle, adaptive filters derive all their information from a given realization of a stochastic process, i.e., one sequence of time samples, thus canceling the need for a-priori assumptions.

A wealth of adaptive filtering algorithms have been developed in the literature, and can be found from classic books on the topic (e.g. [46], [89]). Rather than solving optimally for an unique solution, different types of adaptive algorithms are usually considered as a set of tools that enable to filter non-stationary signals in a given way, and each alternative algorithm offers desirable features of its own.

LMS filters

One of the simplest adaptive filters is the well known Least Mean Squares (LMS) algorithm. It is usually used to determine the filter coefficients that relate to producing the least mean squares of the error signal, i.e., the difference between the desired and the observed signals. The LMS uses the FIR filter input vector as a correction vector modulated by a step size parameter, which controls the speed of the adaptation. LMS filters do not make a priori assumptions about the statistics of the signal data, and this enables a robust performance against various types of noise and outliers signal. This filter is particularly attractive due to its simplicity and its low computational complexity. However to date, LMS filters have not been widely used in image processing applications.

LMS filtering for image and video processing

Earlier, 1-D LMS filtering has been extended to the 2-D case, and has been applied in few example image processing tasks, e.g., image enhancement [94]. In [6], a 2-D block diagonal LMS algorithm was developed for image processing

applications. It was claimed that it is possible to preserve the local correlation information of the pixels in both directions when utilizing a 2-D diagonal scanning pattern. In [92], a two-dimensional recursive least squares (LS) filtering scheme was introduced. The filter was tuned to remove the mismatching effects in a stereo image pair, and the weights of the filter were computed using a block-based LS method. In the context of block motion estimation, it was suggested in [82] that the estimation of motion vectors based on the spatio-temporal neighborhood information is an effective solution to reduce the effects of uneven error surface. In [64], an adaptive matching scan was employed to reduce the amount of computations needed to perform the full-search block-matching algorithm. However, it is somehow surprising that none has yet proposed to use LMS algorithms for direct estimation of the optical flow in image pairs.

In the following Section (5.4), we present the observation model which assumes small displacement between two successive frames. In Section 5.5, we introduce the LMS matching filter, and then the procedure that we used to extract the motion from the adapted coefficient distribution. In Section 5.6, we discuss the effect of the scanning from one direction, and we propose a method to scan from different directions and to combine the final result. Experimental results are presented in Section 5.7 using synthetic test set to illustrate the capability of the proposed algorithm to detect motion as well as motion outliers. Finally, in Section 5.8, we discuss the advantages of the proposed method in the context of video filtering and we draw some conclusions.

5.4 Observation model

Consider two successive frames of a video sequence, a reference image I, and a template image T, which we would like to register with respect to I. Both images have the same size (X, Y). The images are ordered lexicographically into vectors, such that $I(k)$ and $T(k)$ denote the intensity values on the grid position k $(1 \leq k \leq XY)$. We want to estimate the displacement field $D(k) = [u(k), v(k)]$, which establishes the correspondence between $I(k)$ and $T(k)$. We assume that the relative displacement $D(k)$ is constrained, such that

$$\begin{cases} -s \leq u(k) \leq s \\ -s \leq v(k) \leq s \end{cases} \tag{5.1}$$

In order to solve for the pixel-based motion estimation problem, the following cost function may be considered

$$C(k) = [T(k) - \tilde{I}(k + D(k))]^2 \tag{5.2}$$

$\tilde{I}$ denotes the estimated intensity value of the reference image after performing the motion compensation. Note that the displacement $D(k)$ need not be integer

valued. In equation (5.2), we chose the simple quadratic functional of the registration error for tractability of the formulation, especially in case of Gaussian additive noise.

The main hypothesis in our formulation is that the pixel value $I(k)$ in the reference image can be expressed as an estimate using a linear filter combination of the window around the central pixel location $T(k)$ in the template image. That is:

$$I(k) = w(k)^{'} * T_w(k) + \eta(k) \tag{5.3}$$

where $T_w(k)$ is a matrix of windowed pixel values from the template image with size $S = (2s+1)^2$ and centered around the pixel position k. $w(k)$ corresponds to the modulating coefficient matrix, $\eta(k)$ is an additive noise term. For notation convenience, the matrices $T_w(k)$ and $w(k)$ are ordered lexicographically into column vectors, and $'$ denotes the transpose operation.

The model in (5.3) tells that each pixel value in the reference image can be estimated with a linear model of a window that contains the possible shifted pixels in the template image. In this setting, the motion estimation problem can be mapped into the simpler problem of linear system identification, i.e., we have the desired signal $I(k)$, the input data $T_w(k)$, and we would like to estimate $w(k)$ according to the linear model in (5.3). The goal is to minimize the cost function in (5.2) by limiting the motion search within the bounds expressed in (5.1).

5.5 2-D LMS adaptive pixel matching

The 2-D LMS filter is essentially an extension of its 1-D counterpart. In our solution, it takes the two dimensional window $T_w(k)$ as input data and the desired response to be matched is the intensity value in the reference image $I(k)$. In order to solve for the weight array $w(k)$, we apply the standard LMS recursion [46]. The recursion is applied along a pre-determined scanning path of the image grid (indexed by n), as follows

$$\begin{cases} e(n) &= I(k) - w(n-1)^{'} T_w(k) \\ w(n) &= w(n-1) + \mu(n) T_w(k) e(n) \end{cases} \tag{5.4}$$

where $\mu(n)$ is a positive step size parameter, $e(n)$ is the output estimation error, n refers to the iteration number, and k denotes the current pixel position that we are filtering. Note that if the indexing of the pixels k is the same as the indexing of the scanning path, then n and k are identical. $w(n-1)$ refers to the coefficient values that were estimated in the previous pixel position following the employed scanning direction (see the following section for discussion). Fig. 5.1 shows an illustration of this basic filtering process.

Like its 1-D counterpart, the 2-D adaptive filter does not assume any knowledge about the cross correlation functions [94]. The filter approximates their values by using instantaneous estimates at each pixel position according to the step

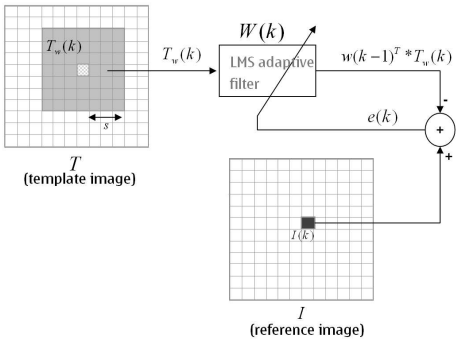


Figure 5.1: Illustration of the LMS filtering that is used to calculate the optical flow between the frames.

size μ. For LMS filters, there is a well-studied trade-off between stability and speed of convergence, i.e., a small enough step size $\mu(n)$ will result in slow convergence; whereas a large step size may result in unstable solutions. Alternatively, there are several modifications of the standard LMS algorithm that offer simpler stability requirements, for example, the normalized LMS (NLMS). The NLMS algorithm is obtained by substituting in equation (5.4) the following step size:

$$\mu(n) = \frac{\mu}{\varepsilon + \|T_w(k)\|^2} \tag{5.5}$$

where ε is a small positive constant. In this form, the filter is also called ε-NLMS [2], and the stability condition is given by:

$$\mu < \frac{2}{3} \tag{5.6}$$

The choice of the step size parameter is critical in tuning the proper performance of the overall algorithm. In general, the motion can be assumed locally stationary, such that a small step size μ can be used to favor smooth and slowly varying motion field rather than a spiky and fast changing motion field.

Determining the motion from the adapted filter coefficients

The function of the 2-D LMS filter is to match the pixels in a search window on the template image to the central pixel in the reference image. This matching is done through the smooth modulation of the filter coefficient matrix. Fig. 5.2 shows an example plot of the coefficient values, which peak at the position of the displacement between the corresponding images.

In order to obtain the displacement vector $D(k)$ from the adapted coefficient distribution $w(k)$, we apply a simple filtering operation, which first finds the cluster of neighboring coefficients that contains the global maximum coefficient value

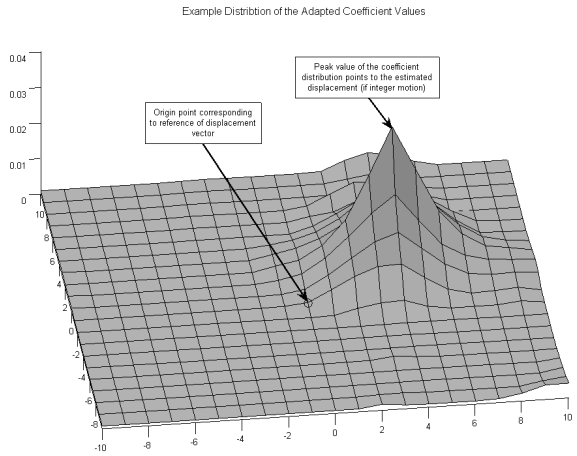


Figure 5.2: Example distribution of adapted coefficient values. The peak value points to the displacement that happened between the two frames at pixel location k.

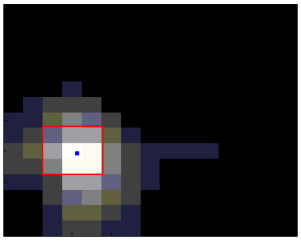


Figure 5.3: Obtaining the displacement from the coefficient distribution.

(Fig. 5.3). Then, the center of mass of this cluster is calculated over the support window. The resulting displacement in the x and y directions constitute the horizontal and vertical components of $D(k)$ at sub-pixel accuracy. A simple intermediate check is inserted to assert whether motion can be confirmed from the obtained coefficient distribution. This filtering operation is described next in more detail:

1. Find the 3×3 window support, over which the sum of neighboring coefficients is maximum.
2. Check that the sum is larger than a pre-determined threshold (confidence in estimation process). If not, assert an empty pointer.
3. Calculate the center of mass over the obtained 3×3 support window. The vector from the origin to the resulting position is the estimated motion vector.

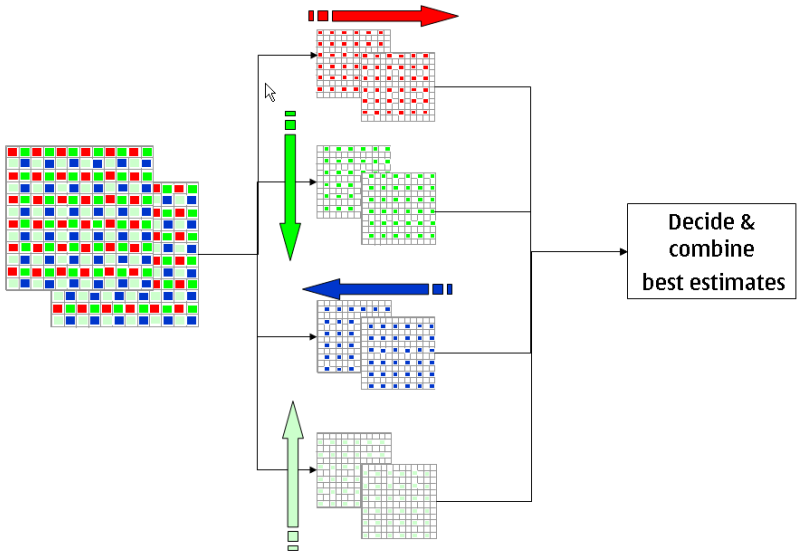


Figure 5.4: Implementation alternative of the proposed motion estimation technique for raw Bayer data. The scanning is performed from 4 different directions separately for each subsampled color component. The final result is obtained by fusing the resulting motion fields in a robust manner.

One important remark relates to the last step where the motion vector is estimated as the center of mass of the maximum coefficient distribution. This is done by calculating the weighted average of the pre-selected 3×3 grid positions with respect to the corresponding adapted FIR coefficients. In our experiments, this operation is performed in floating point, which means that the obtained displacement is inherently at sub-pixel precision. Alternatively, the operations can be easily converted to include desired the subpixel integer precision, this modification is useful if we would like to further reduce the computational complexity of the algorithm.

5.6 Scanning direction

The proposed filtering method is based on recursive scanning of the 2D image grid. As a consequence, the employed scanning pattern impacts the coefficient adaptation, especially if we favor stable adaptation by using a small step size μ. This means that the overall estimation process is spatially causal with respect to the employed scan method. In case the motion is global stationary and constrained, which may be for instance due to camera shaking with respect to a fixed scene, even the simplest of scanning patterns, e.g., raster scan, is sufficient to correctly estimate the stationary displacement.

5.6.1 Multiple scanning directions

On the other hand, if we want to detect arbitrary and localized motion, it may not be possible to estimate the corresponding motion field by utilizing a single scanning direction. Instead, the scanning can be performed in four different directions and obtain the displacement field independently for each scanning direction. The final motion field is computed by combining the resulted motion fields obtained from the different directions. The combining of the displacement vectors can be performed by selecting the vector that minimizes the corresponding error value at each pixel location (error images due to LMS adaptation are stored temporarily in the memory). Another elegant method is to apply a component-wise scalar median filter (or vector median [4]) for the obtained displacement vectors; this allows to obtain the consolidated motion through a voting process and enhance the performance of the estimation process against outliers.

5.6.2 Scanning method for raw Bayer data

In the previous chapter, we proposed to use super-resolution for raw sensor data in RGBG domain (Bayer image data). Interestingly, since the raw data inherently has four separate color components, which are assumed to undergo the same motion; we can apply the multiple scanning scheme described above in 4 different directions by treating each subsampled color component as a separate data source. This may be done at no extra computational cost. The final motion field can be obtained by fusing the motion field obtained from the different directions, as described above. This scheme (illustrated in Fig. 5.4) will be combined in future work with the method proposed in the previous chapter in order to develop an integrated method for super-resolving raw sensor data.

5.6.3 Enhanced scanning patterns

Additionally, instead of the basic raster scan, space-filling curves [78] can be used to traverse the image plane while adapting the LMS coefficients. The typical space filling patterns (e.g., Peano and Hilbert curves [78]) are defined over grid areas that are powers of 2. Fig. 5.5 shows an example of the Hilbert scanning pattern for a rectangular window of 16x16. This mode of scanning through the pixels, though more complicated, has the important advantage of staying localized within areas of stationary shifts before moving to another area. This scanning mode typically results in superior performance of the overall estimation process, especially in the presence of localized motion or other random outliers. The pattern in Fig. 5.5 can be easily mirrored and traversed from four directions as discussed previously.

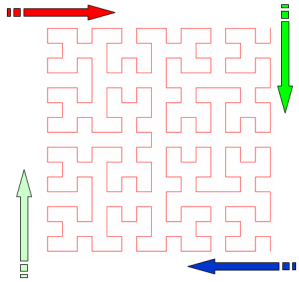


Figure 5.5: Possibility to employ more elaborate scanning patterns. In this example, we propose to use mirrored Hilbert scanning patterns to traverse the image plane from 4 different directions.

5.7 Experimental results

In this section, we briefly show the performance of the proposed approach for motion estimation. We present three different experiments to simulate practical situations that may arise in video filtering applications.

In the first experiment illustrated in Fig. 5.6, a template image is generated from an original reference image by simulating a global translation ($D = [0, 3.5]$). Zero-mean Gaussian noise ($\sigma_\eta^2 = 40$) is added to both the template and the reference image. We tested our algorithm with the following parameters ($s = 7, \mu = 0.02$)[3]. In this experiment, simple raster scanning of the image was employed to adapt the LMS coefficients. The obtained sampled motion field is displayed in Fig. 5.6; the red arrows display the estimated displacement vectors, whereas the blue points show areas where the algorithm cannot resolve constrained motion with certainty; these areas generally contain little image details, which confuse the LMS adaptation. The algorithm was successful in determining the global translational motion, e.g., the motion vector that was estimated in the middle of the image was $D_{est} = [0.048, 3.513]$. In fact, since the step size μ was small, the overall performance was robust against noise, meanwhile the coefficient adaptation was capable to track the stationary shift, because the same shift is consistently confirmed in the areas that contain contrasted details.

In the second experiment, a template image is generated by warping the reference image with an affine transformation. The test image is meant to simulate the geometric skew that is due to camera rotation with respect to an axis, which is parallel to the image plane. Fig. 5.7 shows the estimated motion field that was obtained with the following algorithm parameters ($s = 15, \mu = 0.02$). The filtering was performed using a single scanning direction (raster scan). In the border

[3]The extent of the search (s) can be set according to the video type and the target filtering application. When we know the motion is constrained, smaller search windows ensure faster operation and improved precision.

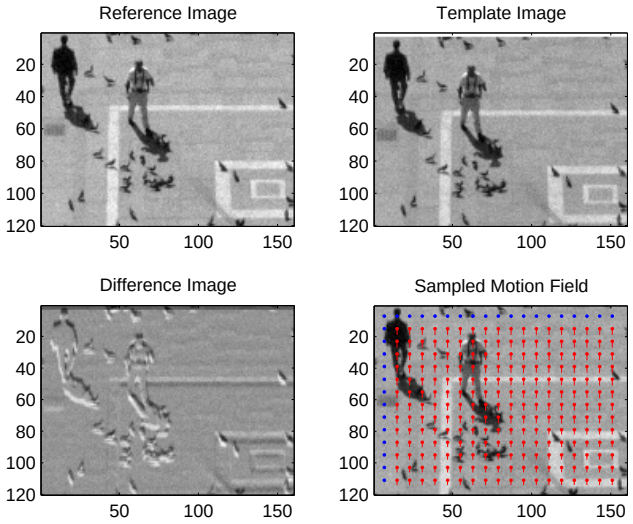


Figure 5.6: Example of the estimated motion field that is obtained by scanning in a single direction (horizontal). Global translation and noisy input images (Gaussian noise, $\sigma_\eta^2 = 40$). Blue points represent the pixel positions where the algorithm cannot resolve any motion with certainty.

area, the algorithm did not detect motion due to the absence of any image details that fall inside the search window. In the rest of the image area, the algorithm was capable to track the smoothly varying motion field.

In the third experiment, a template image is generated by translating the reference image and by inserting an outlier area in the middle of the image. This experiment is designed to simulate the performance of the algorithm in the presence of combined motion. The following algorithm parameters were used $(s = 10, \mu = 0.02)$. In this setting, we used the block Hilbert scanning to traverse the image plane from 4 different directions, and we finally fused the obtained displacement vector components using a median filter. Fig. 5.8 displays the estimated motion field. The blue points show areas where the algorithm cannot resolve for motion with certainty, which corresponds well to the outlier area that was in the template image. This experiment reveals that the use of multiple scanning from different directions and the subsequent voting process through the median selector adds robustness to the motion estimation. We reckon that this result is useful in video filtering, since the detected outlier points can be left out from the filtering process to avoid unwanted artifacts.

In all our experiments, the obtained dense motion field is smooth and spatially correlated, which reflects well the real motion that happened in the video frames. The complexity of the algorithm is $O((2s+1)^2 N)$. Comparing to the complexity

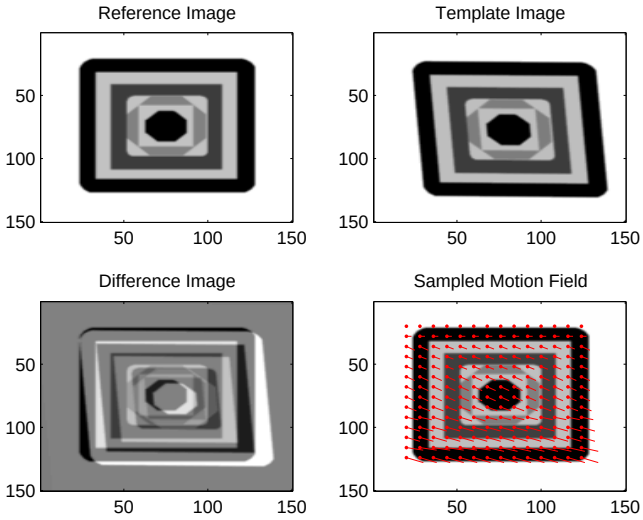


Figure 5.7: Example of the estimated motion field that is obtained by scanning in a single direction (horizontal). The template image was obtained from the reference image by an affine geometric transformation.

of optical flow methods (table 1 in [9]), the proposed method is much simpler and faster, thus enabling real-time operation of motion compensated filtering.

5.8 Conclusions

In this chapter, we presented a novel recursive method for pixel-based motion estimation. The proposed algorithm employs 2-D LMS filter to adapt a window of coefficients so that we can match the pixel value in the reference frame. The peak value in the resulted coefficient distribution points to the displacement between the frames at each pixel position. The recursive LMS filtering along the scanning direction enables to track the stationary shifts that happen between the reference and template frames, and inherently produces smooth estimates of the displacements, directly at sub-pixel accuracy. We also proposed variations of the initial algorithm such as the use of multiple scanning directions and patterns in order to track complicated motion in the scene.

Experimental results have demonstrated that the overall performance was robust against Gaussian noise. Also, the algorithm was capable to accurately track smooth affine motion, even when using a single scanning direction. When using multiple scanning directions, it was possible to single out outlier regions which correspond for example to moving or disappearing objects in the scene.

One important advantage of the proposed method is its simplicity and relative

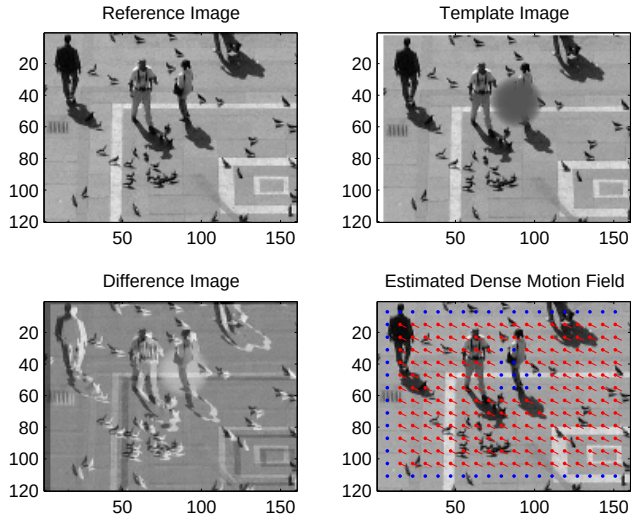


Figure 5.8: Example of the estimated motion field in the presence of outliers. Blue points represent points where the algorithm cannot resolve motion with certainty. The result was obtained by combining the motion estimates from 4 directions (using median operator). The algorithm detected the outlier region (blue points in the center) and isolated it from the smooth motion field in the rest of the image.

low computational complexity. The initial results demonstrate the usability of the algorithm, especially when targeting motion-compensated filtering such as video denoising, video stabilization and super-resolution processing. In future work, we may investigate different variations of the algorithm in order to enhance the basic motion tracking performance, and derive simple rules for LMS step size adaptation. Also of interest, is the research of different extensions of the algorithm in order to cover complex motion patterns.

Chapter 6

Robust Fusion in Super-Resolution

6.1 Introduction

As we have discussed in the previous chapter, one critical aspect in image super-resolution is the need for accurate sub-pixel registration of the input images. The overall performance is particularly degraded in the presence of persistent outliers. To enhance the robustness of super-resolution algorithms, we propose in this chapter an integrated adaptive filtering method to reject the outlier image regions. In the process of combining the gradient images due to each low-resolution image, we use adaptive FIR filtering. The coefficients of the FIR filter are updated using the LMS algorithm, which automatically isolates the outlier image regions by decreasing the corresponding coefficients. The adaptation criterion of the LMS estimator is the error between the median of the samples from the LR images, and the output of the FIR filter. Through simulated experiments on synthetic images and on real camera images, we show that the proposed technique performs well in the presence of motion outliers. This relatively simple and fast mechanism enables to add robustness in practical applications of image super-resolution, while still effective against Gaussian noise in the image formation model.

6.2 Related work

Super-resolution is considered to be one of the most promising techniques that can help overcome the limitations due to optics and sensor resolution [85]. The technique consists in combining a set of low-resolution images portraying slightly different views of the same scene, in order to reconstruct a high-resolution image of that scene. The idea is to increase the information content in the final image by exploiting the additional spatio-temporal information that is available in each of

the LR images.

In practice, the quality of the super-resolved images depends heavily on the accuracy of the motion estimation; in fact, sub-pixel precision in the motion field is needed to achieve the desired improvement. Global parametric motion estimation using affine or projective models can provide accurate enough registration, which positively impacts the over-all performance of the SR algorithms. If the images exhibit optical distortions, higher order polynomial models can be used to obtain better pixel correspondence within the LR images. One major problem with global registration techniques is that they are limited to the assumed parametric model, and more importantly, they completely fail in the presence of local outliers. For example, such outliers may be due to moving objects inside the scene, or due to the presence of repetitive textures or localized noisy areas. In those cases, the super-resolved image can exhibit severe artifacts. Local registration techniques such as optical flow are capable of handling moving objects, although their performance suffers from from lack of precision [7], and the result is not completely prone to outliers. For these reasons, robustness towards registration errors is a critical requirement in super-resolution, especially if we target to realize commercial applications. Moreover, if we consider current mobile devices, we can afford only a limited number of LR frames in the memory buffer; so it's useful to consider optimized algorithms that reject localized outliers, but that are able to exploit the rest of the image areas to improve the final resolution.

Registration error

Several solutions have been proposed to handle registration errors by solving it as part of the regularization of the solution (e.g. [45], [70], [91]). In [45] and [91], motion error noise is incorporated as *a priori* information within the smoothness prior, and the result image is obtained as the MAP solution. In [70], a regularization functional is plugged in a constrained least squares setting and solved by iterative gradient descent. This approach for handling the registration error as part of the regularization certainly helps towards the conditioning of the ill-posed inverse problem. However, it is argued in [8] that for large magnification factors, and regardless of the number of LR images used, regularization suppresses useful high-frequency information, and ultimately leads to smooth results. Note that in most of the literature, localized motion outliers are not properly handled in the model. Further, it is implicitly assumed that the extra resolution content is equally distributed among all the LR images, and usually, the result is obtained by averaging the contributions from all LR images, which propagates the outlier pixels from any of the LR images into the final HR image.

Robust fusion

In [33], it was shown through simulations that in the presence of small errors due to motion estimation or due to inconsistent pixel areas in the consecutive frames, the combined noise is better modeled with a Laplacian distribution rather than a Gaussian distribution. So, if this is taken into consideration, the mixed noise model is best handled through the minimization of the L_p $(1 \leq p \leq 2)$ norm. Specifically, if the L_1 norm is considered, the pixelwise median minimizes the corresponding cost function, and when used together with the bilateral prior regularization [33], the solution was robust towards errors, and still preserved details near sharp edges. In the context of super-resolution reconstruction, the median filter was used earlier [117] in the fusing process of the gradient images. It was shown that together with a bias detection procedure, it is possible to increase resolution even for those regions that contained outlier objects. However, it is well known that the median operator is not optimal for filtering Gaussian noise. Also, the median tends to consistently eliminate those measurements that significantly deviate from the majority, and which may contain most of the novel high-frequency information. So at least in principle, there is a delicate trade-off between outlier rejection performance, noise removal capability, and the capability to reconstruct aliased high frequencies. One possible approach is to consider studying, instead of the mean or median filters, the α-trimmed mean or $\{r, s\}$-trimmed mean[1] in the fusing process. The generalized class of order statistics filters, or L-filters [5] constitute a suitable filtering framework to derive the desired balance between the different trade-offs that are involved in the fusing process of the LR images. We have used this approach [113] to super-resolve text images by emphasizing either the maximum or minimum values to enhance the contrast near character edges.

Our approach

In order to efficiently handle localized outliers, in the following we propose to use an adaptive FIR scheme that automatically reduces the contribution of the outliers, and averages the rest of the pixels. As the scanning progresses over the image grid, the weights associated with each LR image are adapted using an LMS estimator. The median estimator is used as an adaptation criterion that tunes the FIR coefficients to reject consistent outliers. Our approach is different in that we use the median estimator as an intermediate step in the adaptation process, and this inherently eliminates the need for a bias detection procedure [117], making the overall algorithm more robust to Gaussian noise in the image formation model.

The rest of the chapter is organized as follows: in Section 6.3 we present the assumed imaging model. In Section 6.4, the general framework of the iterative

[1] These filters are effective against impulsive noise that may be due to registration errors, and are relatively easy to tune.

super-resolution is presented. Next, we review briefly the existing fusing techniques, and we explain the issues that need to be addressed in order to tune the SR algorithm for robustness against outlier regions. In Section 6.6, we introduce our approach that uses an adaptive FIR filter to combine the gradient images. In Section 6.7, we show the experimental results, and Section 6.8 concludes the topic.

6.3 Imaging model

In this section, the general model that relates the HR image to the LR observations is formulated. The degradation process involves consecutively, geometric transformation, sensor blurring, spatial sub-sampling, and an additive noise term. In continuous domain, the forward synthesis model can be described as follows: consider N observed LR images, we assume that these images are obtained as different views of a single continuous HR image. Following a similar notation as in [22], the i^{th} LR image can be expressed as:

$$g_i(x,y) = S \downarrow (h_i * f(\xi_i))(x,y) + \eta_i(x,y), \tag{6.1}$$

where g_i is the i^{th} observed LR image, f is the HR reference image, h_i the point spread function (PSF), ξ_i the geometric warping, $S \downarrow$ the down-sampling operator, η_i additive noise term, and $*$ denote the convolution operator. The overall degradation process is illustrated in Fig. 6.1.

After discretization, the model can be expressed in matrix form as follows:

$$\overline{g}_i = \mathbf{A}_i \overline{f} + \overline{\eta}_i. \tag{6.2}$$

The matrix $\mathbf{A}_i$ combines successively, the geometric transformation ξ_i, the convolution operator with the blurring parameters of h_i, and the down-sampling operator $S \downarrow$ [30]. Note that in equation (6.2) $\overline{g}_i$, $\overline{f}$ and $\overline{\eta}_i$ are lexicographically ordered.

6.4 Iterative super-resolution

The problem of super-resolution reconstruction is formulated as estimating the best HR image, which when appropriately warped and down-sampled by the model in (6.2) will generate the closest estimates of the LR images $\overline{g}_i$. If we assume that $\overline{\eta}_i$ is Gaussian white noise, then the least squares solution maximizes the likelihood that each LR image is the result of an observation of the original HR image. In other words, for each observation $\overline{g}_i$, the corresponding solution is a high-resolution image $\overline{f}$, which minimizes the following cost function:

$$\epsilon_i = ||\widehat{g}_i - \overline{g}_i||^2 = ||\mathbf{A}_i \overline{f} - \overline{g}_i||^2, \tag{6.3}$$

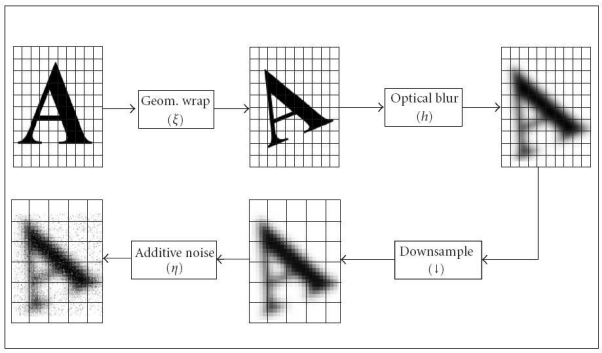


Figure 6.1: An illustration of the image degradation process following the model in (6.2).

with $\widehat{g}_i$ being the simulated LR image through the forward imaging model.

In order to minimize the error functional in (6.3), the method of iterative gradient descent is commonly employed. This optimization technique seeks to converge ϵ_i towards a local minimum, following the trajectory defined by the negative gradient. That is, at iteration n, the high-resolution image according to observation $\overline{g}_i$, is updated as:

$$\overline{f}^{n+1} = \overline{f}^{n} + \mu_i^n \overline{r}_i^n, \tag{6.4}$$

μ_i^n and $\overline{r}_i^n$ are respectively, the step-size and the residual gradient at iteration n. The residual gradient $\overline{r}_i^n$ is computed as follows:

$$\overline{r}_i^n = \mathbf{W}_i \left(\overline{g}_i - \mathbf{A}_i \overline{f}^n \right). \tag{6.5}$$

The matrix $\mathbf{W}_i$ combines successively the up-sampling, and the inverse geometric warp ξ_i^{-1}. The step size μ_i^n that achieves the steepest descent is given by [12]:

$$\mu_i^n = \frac{||\overline{g}_i - \mathbf{A}_i \overline{f}^n||^2}{||\mathbf{A}_i \overline{r}_i^n||^2}. \tag{6.6}$$

In equation (6.4), each scaled gradient term, $\overline{p}_i = \mu_i^n \overline{r}_i^n$, corresponds to the update image that verifies the reconstruction constraint for the i^{th} observation $\overline{g}_i$. We define $\mathbf{z}_k$ as the data vector that points to the values from all the gradient images at pixel position k, $\mathbf{z}_k = \{p_i(k), i = 1 \ldots, N\}$. In the process of SR reconstruction, we need to perform a temporal filtering operation that combines the observations in $\mathbf{z}_k$. For convenience of notation, we denote this filtering operator Φ. For each pixel k on the HR image grid, the resulting update value y_k is given as

$$y_k = \Phi(\mathbf{z}_k), \tag{6.7}$$

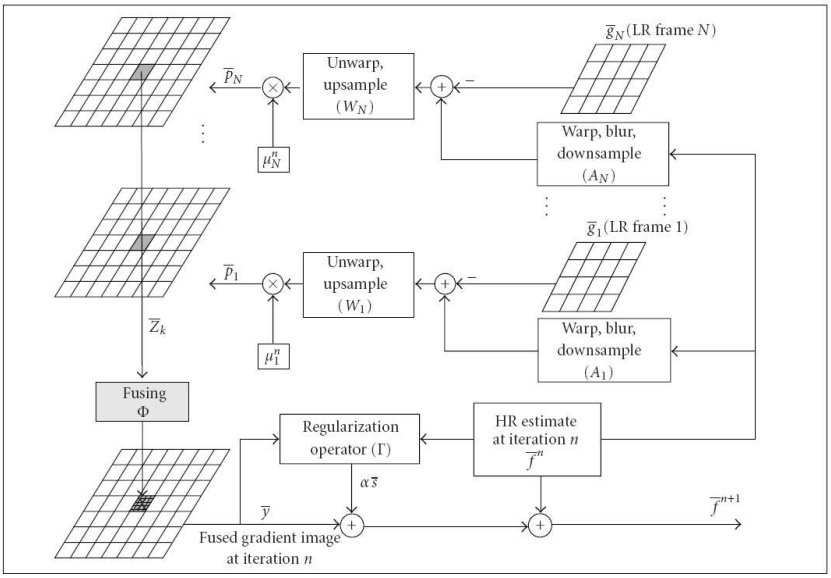


Figure 6.2: Generic block diagram of the iterative super-resolution process. The gradient images are combined using a filtering operator Φ that can be modulated depending on the application.

where Φ is a generic filtering operator that performs the fusing of the pixels from all the available gradient images. Fig. 6.2 depicts an illustration of the iterative SR algorithm that we considered. Note that so far our formulation does not assume a proper regularization term for the solution. Certainly, super-resolution is an ill-posed inverse problem, so regularization is necessary to obtain a stable solution. In the literature, there has been significant effort to formulate suitable prior models for iterative super-resolution (e.g. [32], [45], [70]). These solutions can be implemented in the iterative setting of Fig. 6.2 by assuming a generic filter Γ that operates on the previous SR estimate $\overline{f}^n$, or on the fused gradient image. If we denote s_k as the contribution that is due to the regularization process at pixel k, then at iteration n, the final output at each pixel k is updated as follows:

$$f_k^{n+1} = f_k^n + y_k + \mu^n \alpha s_k, \tag{6.8}$$

where α is the regularization parameter that controls the conditioning of the solution. In the rest of the chapter, and in our experiments, we omitted the use of a regularization operator, i.e., we assumed $s_k = 0$. We focus the discussion on the efficient implementation of the fusing process Φ in the presence of motion outliers.

6.5 Fusing the gradient images

Ideally, the fusing process defined by operator Φ will retain the novel information from each LR frame, filter out the noise due to the image formation process, and of course reject the motion outliers. Thus, at least in principle, we shall consider all the observations independently, and design a filtering mechanism that adapts itself to instantly recognize and reject the outliers, while constantly adjusting its behavior according to the non-stationary noise distribution of the input images.

One straightforward implementation of the fusing process would be to select Φ as the mean filter. In this case, if Gaussian noise is assumed in the imaging model, the algorithm is equivalent to the maximum likelihood solution. However, the solution is not robust against outliers. Another possibility is to select the median filter, which would be efficient against impulsive errors in $\mathbf{z}_k$. This idea was used earlier in iterative super-resolution [117], and was shown to improve the robustness against motion outliers. In fact, the median minimizes the L_1 cost function [33] which corresponds to the Laplacian distribution of the combined noise. However, in the case when the errors have a mixed distribution, for instance, Gaussian and impulsive, the class of $trimmed$ mean filters might have better performance. Note that the filters discussed above can be derived as special cases of the generalized L-filters[2] which operate on the sorted data vector $\mathbf{z}_{(k)}$.

When we consider error modeling due to motion estimation, it is difficult in practice to assume a stationary distribution. This is especially true when dealing with local outliers, for example due to moving objects inside the scene. More difficult is the case when the user tilts the camera, resulting in a significant perspective change. This situation is quite challenging for most motion estimation techniques, which may register parts of the image correctly, but may completely fail in some other regions. It is beneficial to use an adaptive fusing strategy that is capable of automatically isolating localized outliers. In the following section, we introduce our approach which is based on spatially adaptive FIR filtering of the gradient images. It is shown that this technique enables the overall process to deal adequately with outliers.

6.6 Our approach

6.6.1 Outlier rejection by adaptive FIR filtering

In equation (6.7), the fusing operator Φ is implemented as a weighted mean filter, i.e., at each iteration, the update value y_k is calculated as the output of an FIR

[2]For example, the median filter is a special case of the L-filters, which can be obtained by selecting all the coefficients to be zero, except for the center coefficient that has unity value.

filter, as follows:

$$y_k = \sum_{i=1}^{N} a_i p_i(k) = \mathbf{a}^T \mathbf{z}_k, \tag{6.9}$$

where $\mathbf{a}$ is the FIR coefficient vector. The filter coefficients relate the contribution that each LR image brings into the fused image. In most conventional techniques, it is generally implied that all the LR images contribute equally to the total gradient image, i.e., $a_i = \frac{1}{N}$, $i = 1, \ldots, N$. However in the presence of outliers, the computed solution may be corrupted by the consistent presence of large projection errors coming from the same frames.

To take into account the presence of outlier regions at the fusing stage, we introduce an adaptation mechanism that modulates the weights associated with each input image. The coefficients of the FIR filter are varying with the pixel location k, that is in equation (6.9), we use $\mathbf{a}_k$ instead of $\mathbf{a}$.

6.6.2 Coefficient adaptation

For its simplicity and computational efficiency, we chose to use the least mean squared (LMS) estimator to adapt the filter coefficients. The coefficients are updated progressively according to a pre-determined scanning pattern across the selected image region ($k = 1 \ldots L$). Our proposed method for spatially adapting the FIR coefficients, and simultaneously computing the update value is described below:

1. Initialization: $k = 1$, $a_0 = \left[\frac{1}{N}, \ldots, \frac{1}{N}\right]$
2. While scanning inside the image ($k \leq L$)
   2.1. Filtering: $y_k = \mathbf{a}_{k-1}^T \mathbf{z}_k$
   2.2. Error computation: $e_k = d_k - y_k = median(\mathbf{z}_k) - y_k$
   2.3. Coefficient update: $\mathbf{a}_k = \mathbf{a}_{k-1} + \lambda e_k \mathbf{z}_k$
   2.4. Move to next pixel location ($k + 1$)
3. Update the HR estimate according to equation (6.8)

In the LMS coefficient adaptation shown above, λ is the step-size parameter. We set the desired response of the LMS estimator (d_k) to be the median of all the errors. In this setting, the median is used to point out those frames that consistently present error values that deviate from the majority. For example, if scanning progresses through an area where the i^{th} LR image contains an outlier region, then pixel after pixel, the error with respect to the median is going to be large, and the coefficient bias due to $\lambda e_k \mathbf{z}_k(i)$ is going to adjust the corresponding FIR coefficient $\mathbf{a}_k(i)$. Fig. 6.3 depicts an illustration of the proposed filtering method.

When combined with a suitable step-size, the LMS estimator gathers reliable statistics from the immediate pixel neighborhood. The resulting FIR coefficients tend to stabilize, rejecting the outlier contribution, while still averaging the rest of

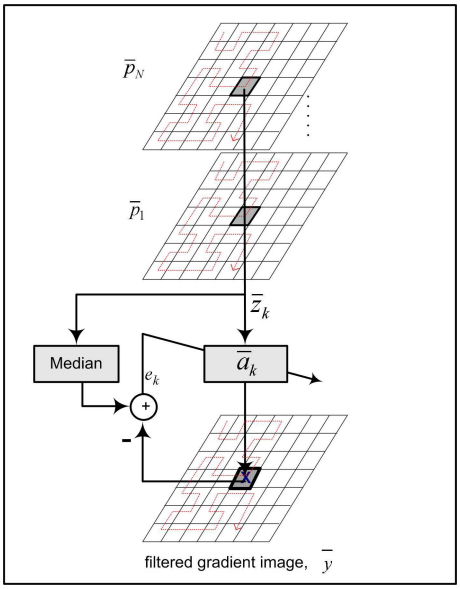


Figure 6.3: Block diagram of the proposed fusing method. The gradient images are combined with a spatially varying FIR filter. The coefficients of the FIR are chosen with an LMS estimator that is tuned to reject outliers.

the error values. Given a sufficient set of samples, the median can approximate the mean quite well [5], although with a reduced set of LR images (fewer samples), the result can be biased, and that's why we chose to set it only as an intermediate step for the coefficient adaptation. The experiments in the following section confirm that this fusing scheme is also efficient to filter the Gaussian noise assumed in the image formation model.

Note that the desired response of the LMS estimation (d_k) can be changed to modulate the performance of the super-resolution process. In this case, the median estimator is used to tune the algorithm for robustness against local outliers. Other functions might be studied and plugged in d_k to obtain a specific property of the fusing process. For example, to speed up the reconstruction property for all the input images, we can set $d_k = 0$. In this case, since we are fusing gradient images, the algorithm will favor the contribution of those LR images that consistently present most of the novel information.

6.6.3 Stability of LMS adaptation

Although its simplicity and good adaptation performance, the LMS has also some sensible points that must be addressed. First issue is the initialization of the step-size λ. It is well known that the value of λ provides a trade-off between speed

of convergence and quality of adaptation. If its value is large, the convergence is fast, but at the expense of an increased adaptation error. On the contrary, a small step-size provides good adaptation performance, but the transient time is increased.

The problem of stability and adaptation speed for the LMS estimator is well studied in the literature [46]. Several modified solutions have been proposed to solve the problem for 1D signals. To ensure the stability of the LMS estimator, the step-size must be bounded[3]:

$$0 < \lambda < \frac{2}{3tr\left[\mathbf{R}\right]}. \tag{6.10}$$

where $\mathbf{R} = E\left\{\mathbf{z}_k \mathbf{z}_k^T\right\}$ is the cross-correlation matrix of the input vector, $E\{\}$ denotes the expectation operator, and $tr\left[\mathbf{R}\right]$ is the sum of the diagonal elements of matrix $\mathbf{R}$.

The above stability criterion is valid and easy to implement when the input sequence is stationary. However, for non-stationary inputs, as it is often the case with image data, the cross-correlation matrix $\mathbf{R}$ changes when scanning through the image. As a consequence, the stability interval in (6.10) is not fixed throughout the entire image. To overcome this difficulty, the simplest solution consists in selecting a small value of λ, such that it is always within the stability bounds for all the pixel locations. However, such a small step-size will significantly slow down the convergence. Moreover, although in some parts of the image, a small step-size will be beneficial to avoid fast and unnecessary variations in the the FIR coefficients, a larger value of λ will be required in regions containing outliers.

To overcome those difficulties, and to simplify the setup of the algorithm, we have implemented the normalized LMS (NLMS). The gradient step factor is normalized by the energy of the data vector. In our case, λ_k is modified depending on the pixel location, and is given by the following equation:

$$\lambda_k = \frac{\gamma}{||\mathbf{z}_k||^2}, \tag{6.11}$$

where $||\mathbf{z}_k||$ is the Euclidean norm of vector $\mathbf{z}_k$, and γ denotes the step size parameter of the NLMS estimator. In this setup, the stability condition of (6.10) becomes:

$$0 < \gamma < \frac{2}{3}. \tag{6.12}$$

As it can be seen from equation (6.11), the algorithm maintains a step size value that is inversely proportional to the input power. As a result, the normalized algorithm converges faster within fewer samples in many cases. To overcome the

[3] for several applications, relaxed boundary conditions may be used for λ. However, the stability condition in equation (6.10) has been shown to ensure stability for a wider class of input statistics, including non-stationary signals.

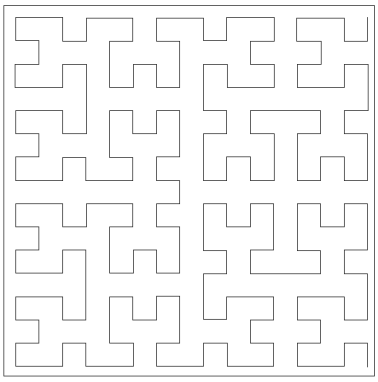


Figure 6.4: Hilbert scanning pattern is used to maximize efficient adaptation of the FIR coefficients.

possible numerical problems when $||\mathbf{z}_k||^2$ is very close to zero, the step-size of the Normalized LMS in equation (6.11) is usually modified as follows [46]:

$$\lambda_k = \frac{\gamma}{c + ||\mathbf{z}_k||^2}, \tag{6.13}$$

with $c > 0$. Note that the stability interval of γ remains unchanged, and is the same as in (6.12). In equation (6.13), the constant c can be used to prevent very large changes of the step size. If we use a relatively large value of $c >$, the speed of coefficient adaptation is decreased, but on the other hand, this improves the robustness of the employed NLMS adaptation against fast changing edges and other local image details that are present in the gradient images.

6.6.4 Scanning pattern

To better handle outlier regions, especially those due to moving objects, the proposed fusing algorithm is most efficient when the coefficient adaptation procedure stays localized around the 2-D outlier patterns. Ideally, we would like the scanning path to satisfy the following constraints:

- cover the entire image area,
- pass through each point only once,
- stay in the highly correlated image areas as long as possible.

By default, if we use the simple raster scan over the entire HR image, we fail to satisfy condition 3. One immediate solution is to divide the image into areas of equal size, and to apply the filtering in these areas independently, with careful

handling of the borders. Instead of the raster scan, space-filling curves can be used to traverse the image plane during the filtering process. These curves have been successfully used in several other applications such as image coding [78]. This mode of scanning through the pixels, though more complicated, has the important advantage of staying localized within areas of similar frequencies before moving to another area. Fig. 6.4 shows the Hilbert scanning pattern for a rectangular window of 16×16. Notice that the filtering following the Hilbert path will stay longer in regions having 2-D correlation than the one following the raster scan. In our implementations, we tested the Hilbert space filling curves of 64x64, as well as 16×16. It was clear to us that applying this type of scanning pattern significantly enhanced the coefficient adaptation, and allowed to use smaller values of λ, thus resulting in better stability of the LMS estimator. It is worth of mentioning that these scanning patterns are easily integrated in the overall implementation using pre-defined look up tables.

The typical space filling patterns (such as Peano, Hilbert [78]) are defined over grid areas that are powers of 2. To confine with this restriction, the image area is devided into smaller tiles that are powers of 2. This option is rather a limitation to the performance of the LMS estimator. Moreover, if the tiles happen inside an outlier area, some artifacts might appear at the borders of the tiles, and may get amplified with the iterations. To avoid these artifacts, one immediate solution is to slow down the LMS adaptation by decreasing λ. Another solution is to smooth the coefficients at the borders of adjacent tiles, but this procedure makes the overall implementation rather cumbersome. Better solution would be to apply space filling curves that are defined over arbitrary sized images, for example the scanning technique that is proposed in [88] provides an elegant method for preserving two-dimensional continuity.

To further enhance the stability of the LMS estimator, the adapted FIR coefficients are saved in between successive iterations of the super-resolution algorithm. These are used to initialize the input coefficients at the beginning of each scanning block. In fact, in the presence of consistent outliers, the coefficients tend to stabilize quickly after scanning through a small part of the image (see Fig. 6.6), and the outlier regions can be pointed out, since their corresponding coefficients are much smaller than the rest. The detected outlier regions can be thrown away when processing the following iterations to reduce the computational complexity of the overall algorithm.

6.7 Simulation results

In this section, we show the performance of the proposed technique. First, we tested the algorithm on a sequence of synthetic test images. The images, 5 in total, were generated from a single HR image according to the imaging model described in equation (6.1). The original HR image was randomly warped using

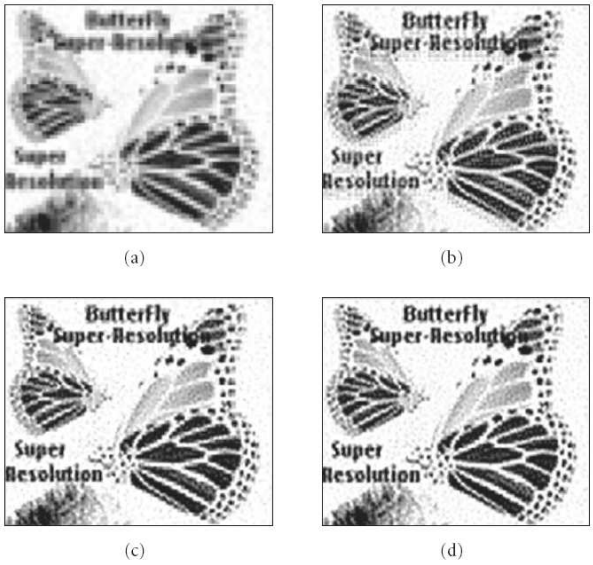


Figure 6.5: 5 noisy LR were synthetically generated by random warp and down-sampling by 2, additive Gaussian noise ($\sigma_\eta^2 = 40$); 1 outlier image. (a) Reference LR image, $SNR = 11.85$. (b) SR result with mean fusing (ML solution) after 10 iterations, $SNR = 14.12$. (c) Iterative median fusing after 10 iterations, $SNR = 15.32$. (d) SR using adaptive FIR filtering after 10 iterations, $SNR = 15.99$.

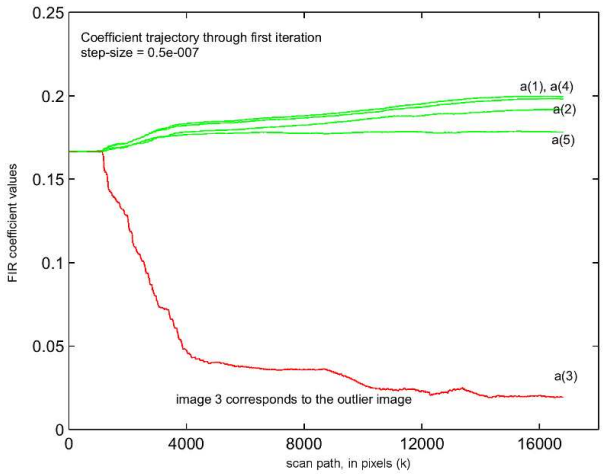


Figure 6.6: Adaptation of the filter coefficients during the first iteration corresponding to the image shown in Fig. 6.5 (d). The coefficient $a(3)$ reflecting the contribution of the outlier image is automatically decreased.

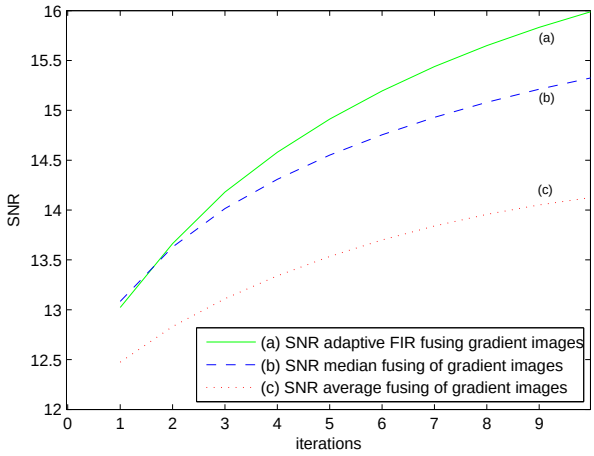


Figure 6.7: SNR comparison across the first 10 iterations for the super-resolved images shown in Fig. 6.5. SNR curves for (a) proposed adaptive solution, (b) median fusing of the gradient images, and (c) average fusing of the gradient images.

an 8 parameter projective model. The registration parameters were saved for the reconstruction experiments. We used a continuous Gaussian PSF ($\sigma^2_{psf} = 0.5$) as the blurring operator, and we down-sampled the images by 2 to obtain the 5 LR images. All the images were contaminated with additive Gaussian noise ($\sigma^2_\eta = 40$). Out of the 5 obtained images, we singled out one image, and we introduced a deliberate error in its registration parameter corresponding to a translation error of 1.5 pixels on the LR image grid.

We ran the algorithm on the resulting set of images. Fig. 6.6 shows the trajectory of the adapted coefficients through the first iteration. In this experiment, the LMS step-size is fixed to a small value $\gamma = 5 \cdot 10^{-7}$. Although the step size is relatively small, the LMS estimator successfully singles out the outlier image (third image) by decreasing its corresponding FIR coefficient $a(3)$, after scanning through a small part of the image.

We compared the results of iterative super-resolution obtained using the proposed fusing process against the mean and median filters. For the three compared techniques, the same step size μ^n_i in the update equation (6.4) is used. Fig. 6.5 shows the result images; both our fusing technique and the median fusing successfully singled out the outlier image and improved the robustness of the overall SR process. Compared to median fusing, the proposed filtering has shown better robustness towards noise, and was able to reconstruct finer character details. Fig. 6.7 shows the corresponding SNR values across the iterations. The SNR num-

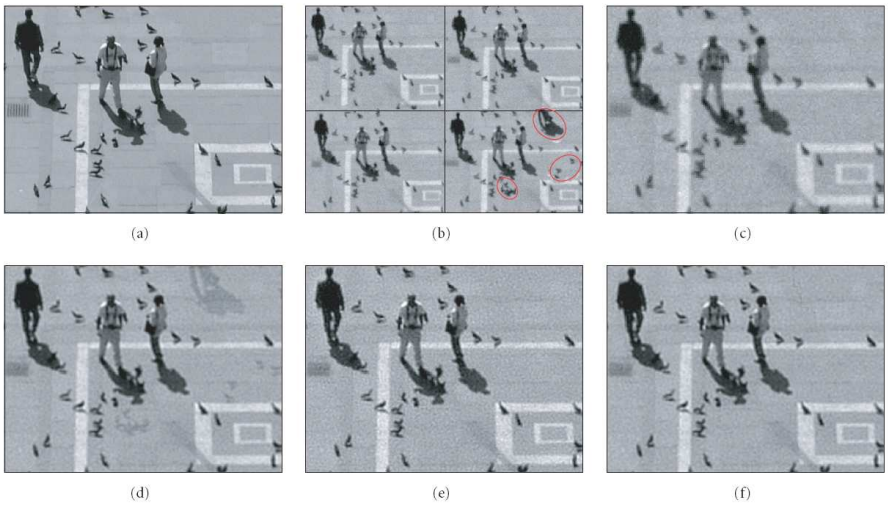


Figure 6.8: (a) Original HR image. (b) The set of LR images used in the experiment: 4 noisy LR were synthetically generated from the original HR image. The last image was generated from the same image with artificial objects inserted. All images were shifted, downsampled by 2 and contaminated with additive Gaussian noise $(\sigma_\eta^2 = 40)$. (c) Interpolated reference image (pixel replication), $SNR = 8.6$. (d) SR result using iterative mean fusing after 4 iterations, $SNR = 11.4$. Remark the shaded outlier regions. (e) SR result using iterative median fusing after 4 iterations, $SNR = 11.3$. (f) SR using adaptive FIR filtering after 4 iterations, $SNR = 12.1$.

ber confirm that the proposed filtering scheme consistently performs better than the mean and median filters. It is worth mentioning that the intermediate result was truncated in between iterations, which helped to constrain the solution and achieve steadier convergence for this set of almost binary images. Note that in all the experiments, we have not used a regularization operator, because we are mainly interested to isolate the effect of the fusing strategy. We assume that it would be possible to enhance the final result, when we assert some prior knowledge about the image content in the regularization step.

In Fig. 6.8, we repeated the same experiment. We generated 4 LR resolution images with the same parameters described above, but in this setting, we selected the last LR frame, and we inserted several outlier objects. Fig. 6.10 shows the SNR values across the iterations for the three fusing techniques. The convergence of the SR algorithm is fast during the first 4 iterations of the steepest descent (SD), but in the following iterations, the SNR starts to oscillate without significant improvement. This example illustrates the need for a regularization step in order to ensure the convergence of the solution. Early abortion of the iterations is the only

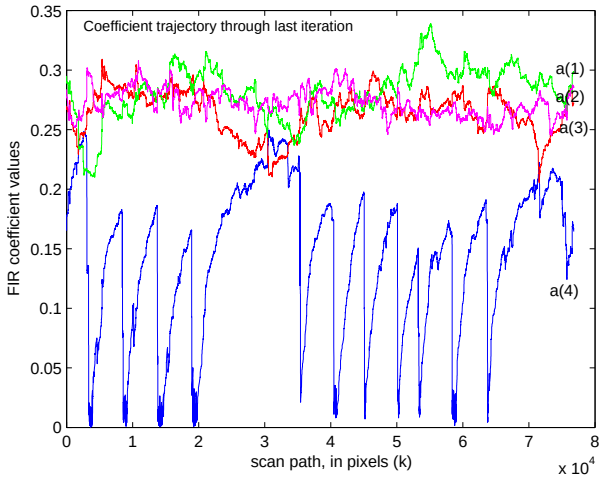


Figure 6.9: Adaptation of the filter coefficients during the fourth and last iteration corresponding to the result in Fig. 6.8 (f). The coefficient $a(4)$ reflecting the contribution of the last LR image is automatically decreased when inside an outlier region, when the scanning steps outside the outlier area, the coefficient increases again. 16×16 Hilbert scanning is used in this example.

available option to avoid over-amplified edges. In Fig. 6.8, we show the results after 4 iterations, again, both the median and the proposed solution eliminated the outlier areas, whereas the mean failed. Better SNR performance, as well as better visual result was obtained with our fusing method (Fig. 6.8 (f)). Fig. 6.9 shows the trajectory of the adapted coefficients through the last iteration. The coefficient $a(4)$ reflecting the contribution of the last LR image is automatically decreased when stepping inside an outlier area. When the scanning steps outside the outlier area, the coefficient increases again. The other coefficients corresponding to the non-outlier images are kept around the same level. As indicated in Fig. 6.9, basically our method operates as a weighted mean filter, except for the detected outlier areas. So, compared to median fusing, an improved performance against Gaussian noise is predictable. In Fig. 6.8 (f), it can be noticed some artifacts near the borders of the Hilbert scanning blocks that contain outlier regions. These are due to the fast and abrupt change of the coefficient values on the borders of the sub-areas that were used for the scanning. To reduce this effect, some implementation enhancements can be designed, such as the use of larger scanning areas, or the smoothing of the coefficients near adjacent blocks.

Fig. 6.11 shows the super-resolved images obtained using 5 LR scenery images taken with a cameraphone (Nokia 6600). To register the pixels on the reference HR grid, we used hierarchical block matching in the central parts of the image,

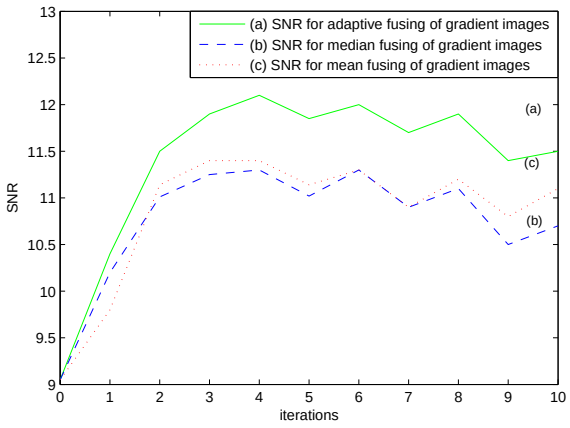


Figure 6.10: SNR comparison across the first 10 iterations for the super-resolved images shown in Fig. 6.8. SNR curves for (a) proposed adaptive solution, (b) median fusing of the gradient images, and (c) average fusing of the gradient images.

followed by the estimation of the global projective motion parameters. In one of images, the registration failed due to a significant perspective change. Fig. 6.11 (a) shows the interpolated reference frame (pixel replication). Fig. 6.11 (b) shows the result when simple mean fusing is used; note the picture of a ghost car that does not belong to the original scene. Fig. 6.11 (c) and (d) show respectively the results after 5 iterations, when fusing with the median and with the proposed technique. For both images, the sharpness of the scene detail is significantly enhanced and the outlier region in the bottom of the image is successfully eliminated. In this specific set of input images, the clouds were particularly difficult to register, because they were deformed from one shot to the next. In fact, for the corresponding area, the only information that needs to be considered is the one that comes from the reference frame. This specific example illustrates the inadequacy of the median filter to fuse this kind of fuzzy regions (Fig. 6.11 (c)). Since the input samples do not constitute a reliable majority to obtain a correct vote, the median filter picks borders randomly from any one of the input images. The proposed filtering does not solve the problem completely, but at least it prevents the formation of excessive artifacts in those regions (clouds in Fig. 6.11 (d)). The reason is that similar FIR coefficients are employed when filtering adjacent pixels, unless a clear outlier frame is consistently voted after scanning through several consecutive pixels, which is not the case in this example. Note that Zomet et. al. [117] have tackled this problem and proposed to use a bias detection procedure in conjunction with the median. The detection procedure outputs a binary mask indicating

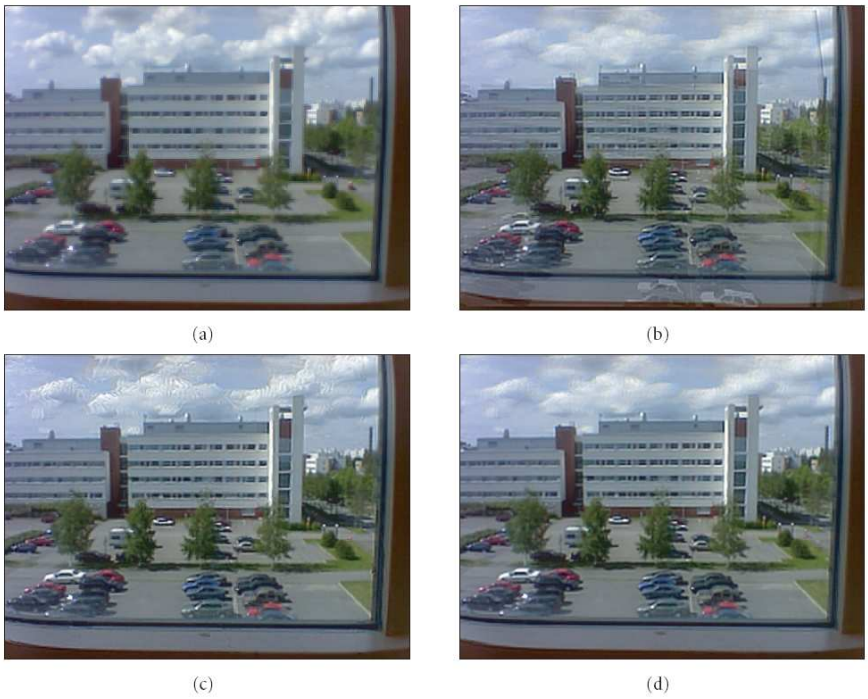


Figure 6.11: The super resolved images using the proposed algorithm, 5 LR images were used. The global motion estimation failed to register at least one frame. (a) Interpolated reference frame, zoom factor 2; (b) result using mean fusing, (c) result using median fusing, and (d) super-resolved image using the proposed method.

where to perform the filtering. Although it is unclear how the thresholds and the windows would be selected.

Fig. 6.12 shows a similar example depicting the performance of the proposed algorithm on real image scenes. We used 5 LR images that were cropped from VGA pictures imaged at close range (the images are JPEG compressed at 90%). The last frame contained an outlier object. Again, note that the median fusing (c) and our technique (d) successfully wiped out the outlier object from the reconstructed scene. Looking more closely, we can notice that the result image of the proposed filtering method has less noise artifacts, especially on smooth areas.

6.8 Conclusions

The overall performance of super-resolution algorithms is particularly degraded in the presence of persistent outliers, for which registration has failed. The artifacts

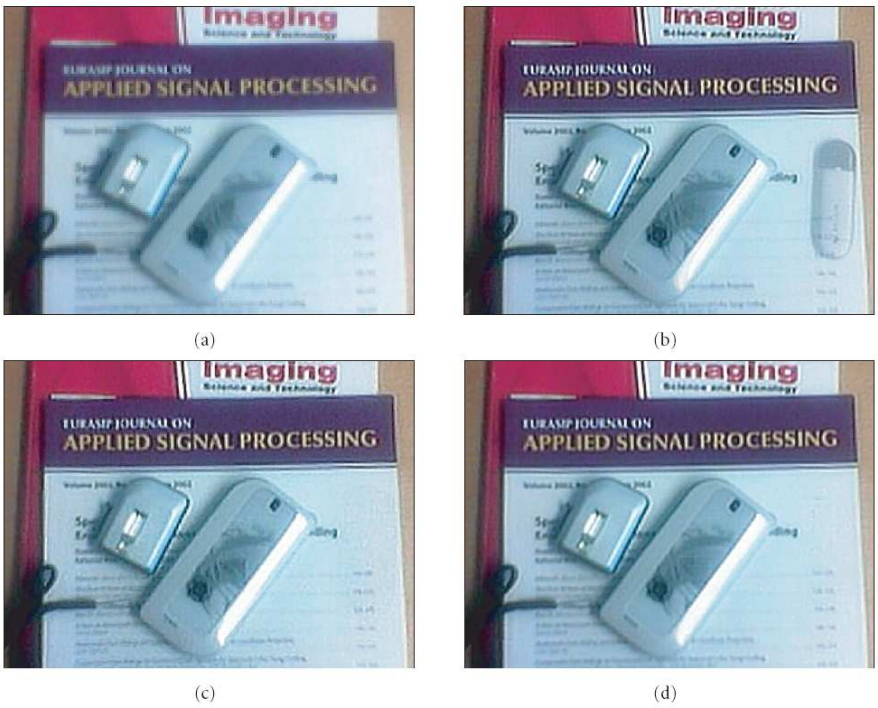


Figure 6.12: The super resolved images using the proposed algorithm. 5 LR images were cropped from VGA images taken with a cameraphone (Nokia 9500). One outlier object appears in the last frame. (a) Zero order interpolated reference frame, zoom factor 2; (b) result using mean fusing, (c) result using median fusing, and (d) super-resolved image using the proposed method.

caused by an incorrectly registered image are visually much more disturbing than the intrinsic poor spatial resolution in a single input image. To enhance the robustness of the processing against this problem, super-resolution algorithms need to integrate adaptive filtering strategies in order to reject the outlier image regions.

In this chapter, we have proposed to use adaptive FIR filtering of the gradient images in iterative super-resolution. The FIR coefficients are adapted using an LMS estimator that is tuned to detect motion outliers. The algorithm performs adequately in the presence of Gaussian noise, and is capable of automatically isolating outlier regions, which are due to registration errors. The proposed method is useful to enhance the robustness of super-resolution in practical applications.

The topic of outlier rejection in super-resolution has been somehow underexplored in comparison to the wealth of solutions in the literature that propose to solve for the inverse problem by assuming exact registration parameters. Possible future research in the topic may include the development of systematic filtering

approach for the fusing step and the improvement of the performance in the presence of different types of noise and motion outliers. The results may be useful in several more applications which propose to employ multi-frame fusion such as high dynamic range image capture, panoramic scene reconstruction, multiple focused imaging, etc.

Chapter 7

Order Statistic Filters in Super-Resolution

7.1 Introduction

In this chapter, we investigate the use of order statistic filters in the iterative process of super-resolution reconstruction. At each iteration, order statistic filters are used to filter and fuse the error images. We use a signal dependent L-filters structure that tunes its coefficients to achieve maximum noise suppression in homogeneous regions. We incorporate a simple mechanism to select the most suitable data support, which preserves the details along the edges. The filter switches to use the orientation that is most likely to preserve the image edges depending on the variance of the pixels across different directional masks. Experimental results show the improvement obtained on sequences of noisy text images, when the motion is exactly known.

In another setting, we show that the integration of an order statistic filtering stage into the iterative process of super-resolution improves the robustness towards motion errors and image outliers. We simulate the filtering by assuming inaccurate registration of a sequence of synthetic LR images. The results demonstrate that the proposed filtering is robust towards random motion errors, and produces visually acceptable results.

7.2 Related work

We recall that super-resolution algorithms consist of the following basic processing steps:

1. Motion estimation to determine the relative shifts between the LR images and register the pixels from all available LR images onto a common reference grid. This step is essential to enable motion compensated filtering.

2. Motion compensation and warping of the input LR images onto the reference grid. Note that the pixels of the LR images are usually non-uniformly distributed with respect to the reference grid.

3. Restoration of the LR images in order to reduce the artifacts due to blurring and sensor noise. The filtering is necessary to improve the perceived image quality.

4. Interpolation of the LR images with predetermined zoom factor to target HR size.

5. Fusing of the pixel values from all the LR images. This temporal filtering operation is at the heart of all super-resolution algorithms, and compliments the spatial filtering operations performed in the previous steps.

In the following, we are mostly concerned with the last step, which deals with the fusing of the pixels from the registered LR images. We consider the application of the generic order statistic filters (L-filters) to enhance the images. The filtering is applied in a maximum likelihood (ML) iterative estimation framework, and can be considered as a heuristic method to improve the robustness of the estimation process and to help regularize the inverse problem.

In super-resolution, the ML estimate of the HR image minimizes the mean square error (MSE) between the LR images and the simulated ones through the assumed imaging model. In [22], it is illustrated through simulations that the poor conditioning of the problem makes the result extremely sensitive to even small amounts of noise in the input images, the main reason is that super-resolution is an ill-posed inverse problem. A common way to regularize the problem is to impose a smoothness prior to the result in a maximum a-posteriori (MAP) framework (e.g. [24], [45]). However, for large magnification factors, the smoothness priors lead to overly smooth results with little high-frequency content, and this regardless of the number of images used [8]. Such smooth images may not correspond to user expectations, especially if the target images contain well-defined details, such as text, geometric shapes, etc. To address this problem, Schultz and Stevenson [91] proposed to incorporate in a Bayesian framework a discontinuity-preserving prior through a Huber penalty function on edge response. Experimental results in [21] have confirmed the suitability of this kind of total variation priors as opposed to simplistic smoothing ones, especially for the enhancement of text images. In the following, we target a similar application for super-resolving text images, and we use order statistic filters (OSF) in order to encourage a piecewise continuity along the edges of the HR image. The data support of the filtering is selected by choosing the orientation that minimizes the variance of the error from the motion-compensated gradient images. The mechanism for the selective directional filtering is similar to that presented by Li and Orchard [72] for image interpolation. At each pixel location, the local covariance is estimated in several directions of the

gradient images, and the orientation yielding the minimum covariance is selected to assert a possible step edge in that direction.

It is well known that the quality of the super-resolved images depends heavily on the accuracy of the calculated motion between the LR frames. In fact, the presence of registration errors usually leads to the propagation of a signal dependent noise from the motion-compensated LR images. On the other hand, it is well known that the generic class of *L*-filters is effective against signal-dependent noise as well as being robust against impulsive noise. A good overview of order statistic filters can be found from the following references ([5], [68], [116]). In the context of SR reconstruction, the median filter was used earlier [117] in a temporal filtering scheme of the gradient images, and it was shown that the procedure increases the resolution even for regions with outliers. This property is particularly interesting for practical implementations, so in the following, we also apply the proposed order statistic filtering to reduce the artifacts due to motion error. We assumed that the forward imaging model and the motion parameters of the LR images are exactly known, and we simulated the inaccurate registration by introducing a random error into the motion coefficients. The experimental results show the robustness of the proposed technique.

The rest of the chapter is organized as follows: in the next Section (7.3), we present the image degradation model used for solving the super-resolution problem. Section 7.4 presents the gradient-based formulation of the maximum likelihood solution. In the same section, we present the fusing technique used in our proposed method and we describe the orientation selection procedure. In Section 7.6, we explain our approach, which uses *L*-filters to enhance the image edges. In Section 7.7, we present the experimental results obtained when applying the method on a synthetic sequence of LR text image. We present also experimental results when random motion error is introduced. Conclusions and possible extensions are discussed in Section 7.8.

7.3 Model used

We use the same image formation model as in the previous chapter. Consider N observed LR images $\overline{g}_i, i = \{1 \dots N\}$, we assume that these images are produced from a single HR image $\overline{f}$. The i^{th} LR image can be expressed as:

$$\overline{g}_i = \mathbf{A}_i \overline{f} + \overline{\eta}_i. \tag{7.1}$$

Where $\mathbf{A}_i$ corresponds to the image formation process, which involves consecutively the geometric transformation mapping the HR image grid onto the observed LR image grid, sensor blurring and spatial sub-sampling. $\overline{\eta}_i$ is an additive noise term. Note that in equation (7.1) the images $\overline{g}_i$, $\overline{f}$ and $\overline{\eta}_i$ are lexicographically ordered into vectors, meaning for example that if the LR images have size $(K \times L)$,

then the pixel position (i, j) on the LR image grid is indexed as $\overline{g}_i(m)$ where $m = j + iL$. This indexing is used throughout the rest of the chapter to simplify the notation.

7.4 Maximum likelihood estimation

Assuming that for all images, the additive image noise is independently distributed Gaussian with zero mean and variance σ_η^2, the probability that the LR image $\bar{g}_i$ is generated by the HR image $\bar{f}$ is given by [22]:

$$\Pr\left(\bar{g}_i \mid \bar{f}\right) = \prod_{\forall m} \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{\left(\hat{\bar{g}}(m) - \bar{g}(m)\right)^2}{2\sigma_\eta^2}\right) \tag{7.2}$$

where $\hat{\bar{g}}$ is the simulated low resolution image given an estimate of $\bar{f}$. The associated negative likelihood function is:

$$L\left(\bar{g}_i\right) = -\sum_{\forall m} \left(\hat{\bar{g}}(m) - \bar{g}(m)\right)^2 = -\|A_i\bar{f} - \bar{g}_i\|^2 \tag{7.3}$$

If we further assume that all the images are equally likely to be the realizations of the same statistical process, then the maximum likelihood estimate, $\bar{f}_{ML}$, is obtained by maximizing the summation of the function in (7.3) over all observed images.

$$\bar{f}_{ML} = \arg\max_i \left(\sum_{i=1}^{N} L\left(\bar{g}_i\right)\right) = \arg\max_i \left(\sum_{i=1}^{N} \|A_i\bar{f} - \bar{g}_i\|^2\right) \tag{7.4}$$

The solution above coincides with the total least squares formulation. If we use the same method of recursive gradient projections to maximize the likelihood function in (7.4), we get a solution similar to that in equation (6.4), i.e., at iteration n, the HR estimate is updated as:

$$\bar{f}_{ML}^{n+1} = \bar{f}_{ML}^{n} + \frac{1}{N}\sum_{i=1}^{N} \bar{p}_i \tag{7.5}$$

where $\bar{p}_i$ is the scaled gradient image that is due to each LR image and is calculated as described in (6.5) and (6.6).

7.5 Gradient fusing process

In the maximum likelihood solution above, all LR images are assumed to contribute equally to the total gradient update image. In the presence of outliers, the

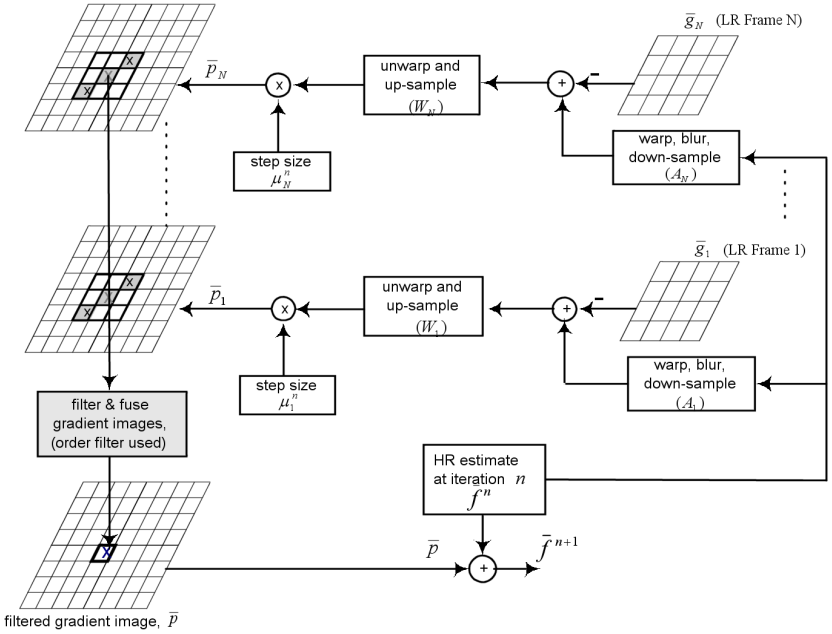


Figure 7.1: An illustration of the proposed iterative SR method.

computed solution may be corrupted by the consistent presence of large projection errors across the iterations. Also, in regions of high frequencies, large error values may be due to aliasing [117], which provides the main source of true resolution enhancement in the process. For these reasons, we believe that the use of a non-intuitive filtering stage is necessary to produce a reliable gradient update. Consequently, the update equation at iteration n is given as follows:

$$\bar{f}^{n+1} = \bar{f}^n + \mathrm{T}\left(\bar{p}_1, \bar{p}_2 \cdots \bar{p}_N\right) \tag{7.6}$$

where $\mathrm{T}\{\}$ denotes a generic filtering operator that fuses the gradient images and produces a single update image. Ideally, the fusing takes into account the LR observations independently, eliminates possible outliers, restores the aliased high frequencies and adjusts its behavior according to the local error content. A schematic of the overall SR framework is shown in Fig. 7.1.

Note that if T is assumed to be the mean operator, then the update equation in (7.6) is equivalent to the iterative solution of the maximum likelihood estimation in (7.5). Alternatively, if T is selected to be the pixel wise median operator, then the solution is equivalent to the super-resolution algorithm proposed in [117]. In the following, we propose to ameliorate the fusing process by considering also the spatial neighborhood around the central pixel location. The goal is to detect possible edge features in the gradient images, and perform the filtering along the

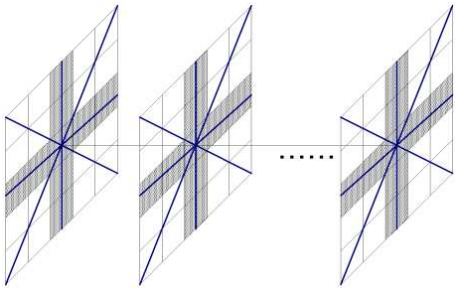


Figure 7.2: Possible orientations used in the experiments $(q = 2)$. This figure illustrates the data support that is used as input to the filtering operation in (7.7). The input samples are collected from all the gradient images.

corresponding data support, in such a way to achieve edge preservation, as well as maximum noise suppression in homogeneous regions.

Mask selection procedure

The main motivation to use an orientation adaptive scheme comes from the fundamental property of an ideal step edge; i.e., the intensity field evolves more slowly along the edge orientation than across the edge orientation [72]. At each pixel location, the filter considers four candidate masks: horizontal, vertical, cross 1 and cross 2, and this for all the data from the error images $(\bar{p}_1, \bar{p}_2 \cdots \bar{p}_N)$ (see Fig. 7.2). We assume that the filter mask that presents the minimum error variance is most likely to align an image feature in that direction. The values coming from that orientation will be used in the filtering step to update the central pixel location.

Along each orientation, we form the possible windows of dimensions $2q + 1$ around the central pixel k, as follows:

$$W(k) = \begin{bmatrix} p_1(k-p) & \cdots & p_1(k) & \cdots & p_1(k+p) \\ \vdots & & \vdots & & \vdots \\ p_N(k-p) & \cdots & p_N(k) & \cdots & p_N(k+p) \end{bmatrix} \tag{7.7}$$

If we arrange the samples in lexicographical order (row-wise) and combine the windowed samples from all the gradient images into vector **w**; we obtain the following vector of size $M \times 1$, where $M = N(2q + 1)$:

$$\begin{aligned} \mathbf{w}(k) &= [w_1(k), w_2(k), \cdots, w_M(k)]^T \\ &= [p_1(k-q), \cdots, p_1(k+q), \cdots, p_N(k-q), \cdots p_N(k+q)]^T \end{aligned} \tag{7.8}$$

The choice of the window sizes defined by q depends on the target interpolation factor and the noise level in the LR images.

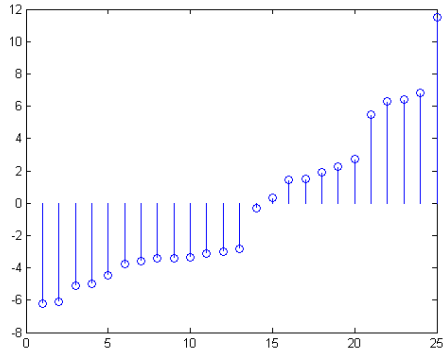


Figure 7.3: Example distribution of the ordered error pixels ($\mathbf{w}_L(k)$) within a rectangular filter mask.

An unbiased estimator for the variance of the mask σ^2_{mask} can be calculated as:

$$\hat{\sigma}^2_{mask} = \frac{1}{M-1}\sum_{i=1}^{M}(w_i(k) - mean(\mathbf{w}(k)))^2 \tag{7.9}$$

The mask that has the minimum variance (σ^2_{mask}) is selected as a data support for the fusing stage, since it is most likely to align an image feature along that direction.

7.6 Order statistic filtering: an enhancement process

Now that we selected the window $w(k)$ that indicates a possible edge feature, we propose to filter it using an order statistic filter. The median filter is well known for its ability to remove impulsive noise while preserving edges, however, the only tunable parameter is its sliding window shape or size. Hence it is more flexible to use *L*-filters, which output is defined as a linear combination of the order statistics (sorted samples). *L*-filters are parameterized by a set of N weights, which allow the filter to reach various profiles, including the mean or median behavior [68].

At each pixel position k, the corresponding ordered input vector $\mathbf{w}_L(k)$ denotes the following data:

$$\mathbf{w}_L(k) = \left[w(k)_{(1)}, w(k)_{(2)}, \cdots, w(k)_{(M)}\right]^T \tag{7.10}$$

where $w(k)_{(1)} \leq w(k)_{(2)} \leq \cdots \leq w(k)_{(M)}$ are the ordered error values coming from the window $w(k)$. The vector $\mathbf{w}_L(k)$ is commonly referred to as the *order statistics vector* [68]. Fig. 7.3 shows an example distribution of the ordered error samples.

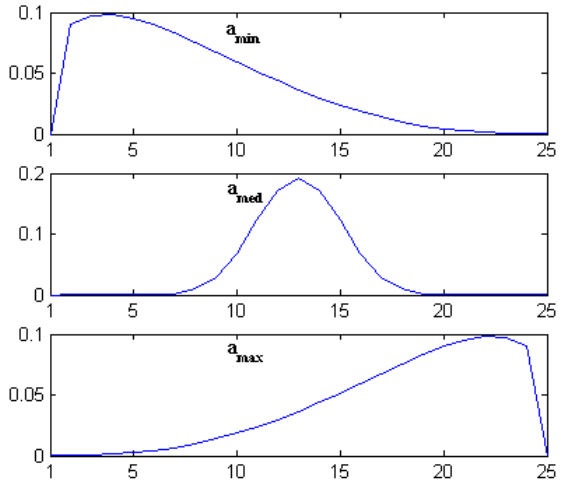


Figure 7.4: L- filters used in the experiments. The x-axis depicts the ordered index of the pixel over the employed filter mask, y-axis shows the corresponding weight.

Let $p(k)$ be the output of the generic L-filter that linearly weights the ordered statistic vector of the observation errors $\mathbf{w}_L(k)$; it is given by [18]:

$$p(k) = \mathbf{a}^T \mathbf{w}_L = \sum_{i=1}^{M} a_i w_{(i)}. \tag{7.11}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$ is the vector of coefficients of the L-filter. Note that the formulation presented above is non-restrictive. If we assume equal values of the L-filter coefficients and we further limit the window at the central pixel values $(q = 0)$, then we end up averaging the gradient images, which is equivalent to the maximum likelihood (ML) solution.

L-filters used

In the proposed algorithm, three different L-filters $\mathbf{a}_{min}$, $\mathbf{a}_{med}$ and $\mathbf{a}_{max}$ are used. Since we are targeting the application for the enhancement of text images, i.e. we know in advance that the original image data is binary; the idea is to use biased distributions ($\mathbf{a}_{min}$ and $\mathbf{a}_{max}$) in order to favor the extremes of the input samples. Additionally, we use a symmetric L-filter structure ($\mathbf{a}_{med}$) when we detect large deviations in the input samples, which indicates the presence of evident outliers.

The employed filters were derived using the *generalized beta distribution* [28], the coefficients of $\mathbf{a}_{min}$, $\mathbf{a}_{med}$ and $\mathbf{a}_{max}$ were defined by sampling the following beta functions $\mathrm{B}(\alpha = 2, \beta = 5)$, $\mathrm{B}(\alpha = 5, \beta = 5)$ and $\mathrm{B}(\alpha = 5, \beta = 2)$, respectively. There is no specific reason why we chose to use this particular distribution,

however, looking at distribution of the coefficients in Fig. 7.4, we can remark that it is possible to filter out extreme outliers. Additionally, it was easy to tune the coefficients by modulating only two parameters, and yet obtain symmetric distributions with different curve shapes.

Signal dependent filtering

At each pixel location, we evaluate the variance of the error values in the employed mask (σ^2_{mask}). If the variance is larger than a pre-defined threshold value, then we use $\mathbf{a}_{med}$ to smooth the data in that window. Otherwise, if the variance is below the threshold value, then we use either $\mathbf{a}_{min}$ or $\mathbf{a}_{max}$ in order to enhance the image features. In that case, the decision on which filter to use ($\mathbf{a}_{min}$ or $\mathbf{a}_{max}$) is simply taken by thresholding the median of the error values. The binary nature of text images allows us to bias the projected error towards negative errors (using $\mathbf{a}_{min}$) or positive errors (using $\mathbf{a}_{max}$) in order to enhance the image contrast. We recall that this filtering procedure is plugged in an iterative SR algorithm, so if there is an error in the filter selection process, it is likely to be corrected in the following iterations.
The proposed fusing algorithm is defined as follows:

```
if (σ²_mask ≥ S_σ)
    p(k) = a_med w_L
else if (median(w_L) < −S_m)
    p(k) = a_min w_L
else if (median(w_L) > S_m)
    p(k) = a_max w_L
else
    p(k) = a_med w_L
```

The threshold values S_σ and S_m are preset positive threshold values that are determined empirically.

7.7 Experimental results

In this section, we show the performance of the proposed technique on a sequence of synthetic LR images. The images, nine in total, were synthetically generated from a single HR text image according to the imaging model described in Section 7.3. The LR images were randomly warped using a projective model (eight parameters). The registration parameters are exactly known. We used a continuous Gaussian PSF ($\sigma^2_{psf} = 0.5$), as the blurring operator and we down-sampled the images by three to obtain the LR images. All the images were contaminated with an additive Gaussian white noise $\sigma^2_\eta = 40$. The HR estimate $\hat{f}^0$ is initialized as the average of all registered LR images, this ensures that we start with a smooth image

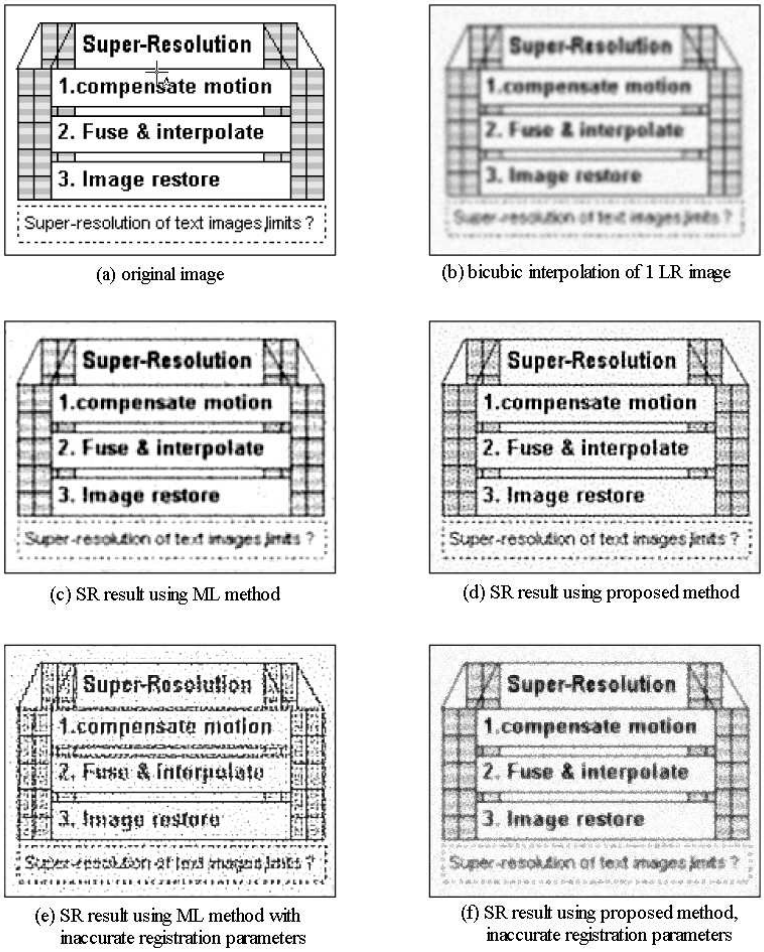


Figure 7.5: Super-resolution at zoom factor 3, 9 LR images used. (a) Original HR text image. (b) One LR frame interpolated using bicubic resample, $SNR = 10.31$. (c) Result after 10 iterations of gradient averaging (ML solution), $SNR = 15.19$. (d) Result using proposed SR filtering technique, $SNR = 15.52$. For (c) and (d), exact motion coefficients were used. (e) result using ML, when random uniform error is used to corrupt the registration parameters, $SNR = 13.58$, (f) same data as in (e), super-resolved using the proposed method, $SNR = 13.98$.

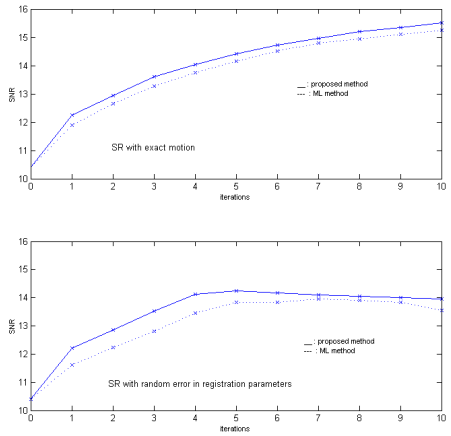


Figure 7.6: SNR comparison across the first 10 iterations for the result images shown in Fig. 7.5.

which is approximately equidistant to all LR observations, hence encouraging the solution to converge to a local minimum.

We compared the proposed SR technique against the constrained ML solution, which is obtained by averaging the gradient images as specified in (7.6). For both methods, we constrained the HR estimate in between the iterations to be in the range $[0, 255]$. The same step sizes were used in both methods. In our algorithm, the orientation selection operation was performed through all the iterations. We set the values of both S_σ and S_m to 25 and 20, respectively.

In Fig. 7.5, we show the resulting images obtained after 10 iterations. The signal to noise ratio (SNR) is calculated for each image and is indicated in the caption. Fig. 7.5 (c) shows the result using ML estimation, and Fig. 7.5 (d) shows the result using the proposed filtering. It can be noticed that the visual quality of the images obtained with our algorithm is better than the result of the ML solution; this is most visible in homogeneous regions. Straight edges and details corresponding to lines are also adequately enhanced due to the use of the orientation selective filtering. As predicted, the contrast is also enhanced in our result thanks to the use of the filters that emphasized extreme values.

To check for the robustness of the filtering towards registration errors, we added a uniformly distributed random error to the motion coefficients corresponding to translation. The error causes the registration to vary in both directions in the interval $[0 - 2]$ pixels. In this setting, Fig. 7.5 (e) shows the result when applying ML estimation. We can notice in this case the presence of annoying artifacts that are due to dislocated pixels from the mis-registered images. In Fig. 7.5 (f), the result obtained with the proposed method looks relatively better; this is because

the spatial filtering limits the propagation of errors in the iterative process and enhances the stability, and the median selector isolating the outlier pixels from the mis-registered images. These observations are confirmed by the corresponding SNR values indicated in the caption of Fig. 7.5.

In Fig. 7.6, the SNR values are plotted for both techniques across all the iterations. The plots show the convergence of the proposed method in terms of SNR, and the small improvement it presents over the ML solution across iterations. It is worth mentioning that both iterative methods are clipping the intermediate images between 0 and 255, which helps the convergence of the solutions. It should also be said that the use of clipping and zero-order interpolation in the iterative process complicates the task of identifying the separate contribution of each step in the SR procedure, especially for binary text images.

7.8 Conclusions

In this chapter, we proposed a novel filtering method of the gradient images in iterative super-resolution. The filtering is aimed at the enhancement of text images by incorporating a prior that assumes the edges to be piecewise constant along a finite number of orientations. We used order statistic filtering to encourage piecewise constancy together with an orientation selection mechanism that is based on the variance of the error images along certain orientations.

We tested the filtering scheme and found that it improves resolution and preserves the edges. The proposed method performs better than the ML solution. The integrated spatial filtering step reduces the signal dependent error that is due to mis-registration.

In the described algorithm, the added computational complexity and the dependence of the filtering on threshold values may overshadow the improvement in quality. In future work, we shall concentrate on the refinement of the method to reduce its computational complexity and to improve the orientation selection mechanism. Also, we may consider the generic application of L-filters in super-resolution such as to minimize modified cost functions.

Chapter 8

Conclusions

Super-resolution reconstruction consists in the process of creating a high resolution image from a sequence of low resolution images. In the context of low-end consumer imaging products, this filtering technique has the potential to make significant economies of scale since it proposes to overcome the intrinsic hardware limitations by using the available computational resources.

Recently, super-resolution has been a very active research area. However, we are still far from generic, robust and real-time algorithms. In the absence of precise imaging models that relate the LR observations to the original scene, the understanding of the potential and limits of super-resolution remains a challenging task. Alternatively, we focused on the development of fast and robust algorithms that are applicable for consumer imaging devices.

In this dissertation, we addressed several aspects of the resolution enhancement problem. First, we presented an integrated filtering method that reduces the optical blur in a single image. The proposed filtering is an integral part of the camera image reconstruction chain, and makes use of the estimated PSF that characterizes the optical blurring in each of the RGB color components. This unique approach, which consists in filtering the raw sensor data was extended to produce high-resolution color image from a sequence of images captured by a CMOS sensor that is overlaid with a color filter array. The experimental results have confirmed that this approach is promising, and is capable of producing superior resolution from sensor data.

Second, we considered the problem of pixel registration in the input images. We proposed a novel recursive algorithm for pixel-based motion estimation. We used recursive LMS filtering along different scanning directions to track the stationary shifts between the LR images, and produce smooth estimates of the displacements at sub-pixel accuracy. The initial results demonstrated the usability of the algorithm, especially when targeting video filtering applications that are based on motion-compensated filtering such as video denoising, video stabilization and super-resolution processing.

Third, we investigated the problem of robust fusion of the motion compensated images. We proposed an integrated adaptive filtering method, which uses the median estimator as an intermediate filtering step to reject the outliers that are due to registration errors. Experiments have shown that the algorithm performs adequately in the presence of Gaussian noise, and is capable of automatically isolating outlier regions.

Finally, we applied non-linear filtering techniques to improve the performance and robustness of super-resolution. In the last chapter, we used generalized order statistic filters for the enhancement of binary text images. We tested the proposed filtering scheme and found that it improves resolution and adequately preserves the edges. Except for the added computational complexity, order-statistic filters present a considerable advantage in super-resolution, since they are efficient against signal dependent error, and are capable to reject outliers that are due to registration errors.

It is well acknowledged that super-resolution methodology is useful since it allows to combine the temporal and spatial filtering of the video content. Therefore, we believe that the incremental development of the research in this area will be beneficial for the development of future multimedia systems. In future work, we will try to further build on the results obtained so far, and develop integrated algorithms with reduced computational complexity. The ultimate goal is to develop scalable algorithms that can be utilized to improve the performance of digital cameras in mobile devices.

Bibliography

[1] J. Jr. Adam. Design of practical color filter array interpolation algorithms for digital cameras. In *Proc. SPIE*, volume 3028, pages 117–125, February 1997.

[2] T. Al-Naffouri, A. Sayed, and T. Kailath. On the selection of optimal nonlinearities for stochastic gradientadaptive algorithms. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1:464–467, September 2000.

[3] Y. Altunbasak, A. Patti, and R. Mersereau. Super-resolution still and video reconstruction from mpeg coded video. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(4):217–227, April 2002.

[4] J. Astola, P. Haavisto, and Y. Neuvo. Vector median filters. *Proceedings of the IEEE*, 78(4):678–689, April 1990.

[5] J. Astola and P. Kuosmanen. *Fundamentals of Nonlinear Digital Filtering*. CRC Press, New York, 1997.

[6] M. Azimi-Sadjadi and P. Hongye. Two-dimensional block diagonal LMS adaptive filtering. *IEEE Transactions on Signal Processing*, 42(9):2420–2429, September 1994.

[7] S. Baker and T. Kanade. Super-resolution optical flow. Technical Report CMU-RI-TR-99-36, Robotics Institute, Carnegie Mellon University, October 1999.

[8] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, September 2002.

[9] S. Baker and I. Matthews. Lukas-kanade 20 years on: a unifying framework. *International Journal of Computer Vision*, 56(3):221–255, March 2004.

[10] M. Banham and A. Katsaggelos. Digital image restoration. *IEEE Signal Processing Magazine*, 14(2):21–41, March 1997.

[11] M. Ben-Ezra, A. Zomet, and S. Nayar. Jitter camera: High resolution video from a low resolution detector. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 135–142, June 2004.

[12] M. Bertero and P. Boccacci. *Inverse Problems in Imaging*. IOP Publishing, 1998. pp. 138.

[13] D. Biggs and M. Andrews. Iterative blind deconvolution of extended objects. In *International Conference on Image Processing, IEEE Proceedings*, volume 2, pages 454–457, October 1997.

[14] R.C. Bilcu, M. Trimeche, S. Alenius, and M. Vehvilainen. Regularized iterative image restoration in the blur space. *WSEAS Transactions on Communications*, 4(7):407–416, July 2005.

[15] A. Blanksby and M. Loinaz. Performance analysis of a color CMOS photogate image sensor. *IEEE Transactions on Electron Devices*, 47(1):55–64, January 2000.

[16] S. Borman. *Topics in Multiframe Superresolution Restoration*. Phd, University of Notre Dame, Indiana, April 2004.

[17] S. Borman and R. Stevenson. Super-resolution from image sequences - a review. *Midwest Symposium on Systems and Circuits*, pages 374–378, 1998.

[18] A. Bovik, T. Huang, and D. Munson. A generalization of median filtering using linear combinations of order statistics. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 31(6):1342–1350, December 1983.

[19] A. Burian, M. Vehvilainen, M. Trimeche, and J. Saarinen. Document image binarization using camera device in mobile phones. In *Proceedings of IEEE International Conference on Image Processing, ICIP 2005, Genova, Italy*, volume 2, pages 546–548, September 2005.

[20] G. Callico, S. Lopeza, R. Llopisb, R. Sethuramanc, A. Nuneza, J. Lopeza, M. Marreroa, and R. Sarmientoa. Practical considerations for real-time super-resolution implementation techniques over video coding platforms. In *Proccedings SPIE*, volume 5489, pages 628–637, 2005.

[21] D. Capel and A. Zisserman. Super-resolution enhancement of text image sequences. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, volume 1, pages 1600–1605, 2000.

[22] D. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 627–634, 2001.

[23] S. Chaudhuri, editor. *Super-Resolution Imaging*. Kluwer Academic Publishers, 2001.

[24] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson. Super-resolved surface reconstruction from multiple images. Tech. Rep. FIA-94-12, NASA Ames Research Center, Moffett Field, CA, December 1994.

[25] R. Costantini and S. Süsstrunk. Virtual sensor design. In *Proceedings IS&T/SPIE Electronic Imaging 2004: Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, volume 5301, pages 408–419, January 5301.

[26] M. W. Davidson. *Molecular Expressions*. http://microscopy.fsu.edu [Online; accessed 3-June-2006], 2003.

[27] F. Dekeyser, P. Bouthemy, P. Perez, and E. Payot. Super-resolution from noisy image sequences exploiting a 2D parametric motion model. In *Proceedings of the International Conference on Pattern Recognition*, volume 3, pages 350–353, 2000.

[28] E. Dougherty. *Random Processes for Image and Signal Processing*. SPIE/IEEE Series on Imaging Science & Engineering, 1999.

[29] M. Elad and A. Feuer. Super-resolution reconstruction of image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):817–834, September 1999.

[30] M. Elad and A. Feuer. Superresolution restoration of an image sequence: adaptive filtering approach. *IEEE Transactions on Image Processing*, 8(3):387–395, March 1999.

[31] S. Farsiu, M. Elad, and P. Milanfar. Multi-frame demosaicing and super-resolution from under-sampled color images. In *Proceedings of the SPIE Conference on Computational Imaging*, volume 5299, pages 18–22, January 2004.

[32] 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, August 2004.

[33] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super-resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, October 2004.

[34] M. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, August 2003.

[35] A. Foi, S. Alenius, M. Trimeche, and V. Katkovnik. Adaptive-size block transforms for poissonian image deblurring. In *Proceedings of International TICSP Workshop on Spectral Methods and Multirate Signal Processing, SMMSP 2006, Florence, Italy*, September 2006.

[36] A. Foi, S. Alenius, M. Trimeche, V. Katkovnik, and K. Egiazarian. A spatially adaptive poissonian image deblurring. In *Proceedings of IEEE International Conference on Image Processing, ICIP 2005, Genova, Italy*, volume 1, pages 925–928, September 2005.

[37] J. Fulton. *Processes in Animal Vision*. http://www.4colorvision.com [Online; accessed 3-March-2006], 2003.

[38] S. Geisler. Physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 1(7):775–782, July 1984.

[39] T. Gotoh and M. Okutomi. Direct super-resolution and registration using raw CFA images. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR04)*, volume 2, pages 600–607, 2004.

[40] B. Gunturk, Y. Altunbasak, and R. Mersereau. Super-resolution reconstruction of compressed video using transform-domain statistics. *IEEE Transactions on Image Processing*, 33(1):33–43, January 2004.

[41] B. Gunturk, Y. Altunbasak, and R. M. Mersereau. Color plane interpolation using alternating projections. *IEEE Transactions on Image Processing*, 11(9):997–1013, September 2002.

[42] B. Gunturk, A. Batur, Y. Altunbasak, M. Hayes, and R. Mersereau. Eigenface-domain super-resolution for face recognition. *IEEE Transactions on Image Processing*, 12(5):597–606, May 2003.

[43] B. Gunturk and M. Gevrekci. High-resolution image reconstruction from multiple differently exposed images. *IEEE Signal Processing Letters*, 13(4):197–200, April 2006.

[44] B. Gunturk, J. Glotzbach, Y. Altunbasak, R. Schafer, and R. Mersereau. Demosaicking: color filter array interpolation. *IEEE Signal Processing Magazine*, 22(1):44–54, January 2005.

[45] R.C. Hardie, K. Bernard, and E. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 6(1621–1632), 1997.

[46] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 3 edition, 1996.

[47] E. Hecht. *Optics (4th edition)*. Addison Wesley, 2002.

[48] J. Heikkilä. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077, October 2000.

[49] J. Heikkila. Matlab toolbox for camera calibration, October 2000. http://www.ee.oulu.fi/ jth/calibr [Online; accessed 2-April-2006].

[50] H. Henning and F. Wörgötter. Eye micro-movements improve stimulus detection beyond the nyquist limit in the peripheral retina. In *Advances in Neural Information Processing Systems*, volume 16, Denver, Colorado. MIT Press.

[51] B. Horn and B. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, August 1981.

[52] T. Huang and R. Tsai. Multi-frame image restoration and registration. *Advances in Computer Vision and Image Processing*, 1:317–339, 1984.

[53] B. Hunt. Super-resolution of imagery: understanding the basis for recovery of spatial frequencies beyond the diffraction limit. In *Proceedings of Information, Decision and Control IDC 99*, pages 243–248, February 1999.

[54] J. Immerkaer. Use of blur-space for debluring and edge-preserving noise smoothing. *IEEE Transactions on Image Processing*, 10(6):837–840, June 2001.

[55] M. Irani and S. Peleg. Super resolution from image sequences. In *Proceedings of IEEE conference on Pattern Recognition*, volume 2, pages 115–120, June 1990.

[56] M. Irani and S. Peleg. Improving resolution by image restoration. *Computer Vision, Graphics and Image Processing*, 53:231–239, 1991.

[57] M. Jiang and G. Wang. Convergence studies on iterative algorithms for image reconstruction. *IEEE Transactions on Medical Imaging*, 22(5):569–579, May 2003.

[58] O. Kalevo and H. Rantanen. Sharpening methods for images captured through Bayer matrix. In IS&T-SPIE's Electronic Imaging Science and Technology conference, editors, *Sensors, Cameras, and Applications for Digital Photography V*, volume 5017, pages 286–297, January 2003.

[59] M. Kang and S. Chaudhuri, editors. *Special Issue on "Super-resolution Image Reconstruction"*, volume 20. IEEE Magazine on Signal Processing, May 2003.

[60] V. Katkovnik. A new method for varying adaptive bandwidth selection. *IEEE Transactions on Signal Processing*, 47(9):2567–2571, September 1999.

[61] V. Katkovnik, K. Egiazarian, and J.Astola. A spatially adaptive nonparametric regression image deblurring. *IEEE Transactions on Image Processing*, 14(10):1469–1478, 2005.

[62] A. Katsaggelos and J. Paik. Iterative color image restoration algorithms. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pages 1028–1031, April 1988.

[63] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(6):1153–1160, 1981.

[64] J. Kim and T. Choi. A fast full-search motion-estimation algorithm using representative pixels and adaptive matching scan. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(7):1040–1048, October 2000.

[65] R. Kimmel. Demosaicing: Image reconstruction from CCD samples. *IEEE Transactions on Image Processing*, 8:1221–1228, 1999.

[66] S. Kleinfelder, S. Lim, X. Liu, and A.E. Gamal. A $10,000$ frames/s CMOS digital pixel sensor. *IEEE Journal of Solid-State Circuits*, 36(12):2049–2059, December 2001.

[67] T. Komatsu, K. Aizawa, T. Igarashi, and T. Saito. Signal-processing based method for acquiring very high resolution image with multiple cameras and

its theoretical analysis. *IEEE Proceedings on Communications, Speech and Vision*, 140(1):19–24, February 1993.

[68] C. Kotropoulous and I. Pitas *(editors)*. *Nonlinear Model-Based Image/Video Processing and analysis*. John Wiley & Sons, April 2001. pp. 121–213.

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

[70] E. Lee and M. Kang. Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration. *IEEE Transactions on Image Processing*, 12(7):826–837, July 2003.

[71] S. Lee, N. Cho, and J. Park. Directional regularization for constrained iterative image restoration. *Electronic Letters*, 39(23):1642–1643, November 2003.

[72] X. Li and M. Orchard. New edge-directed interpolation. *IEEE Transactions on Image Processing*, 2(10):1521–1527, October 1997.

[73] L. Liang and Y. Xu. Adaptive Landweber method to deblur images. *IEEE Signal Processing Letters*, 10(5):129–132, May 2003.

[74] David Littwiller. CMOS vs. CCD: Maturing technologies, maturing markets. *Photonics Spectra*, August 2005.

[75] J.P. Makinen, J. Aikio, V.P. Putila, K. Keranen, P. Karioja, T. Kolehmainen, and J. Haavisto. Prototyping of miniature plastic imaging lens. *Telescopes and Instrumentation for Solar Astrophysics, Proceeding of SPIE*, 5178:89–100, 2004.

[76] S. Martinez-Conde, S. Macknik, and D. Hubel. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3):229–240, March 2004.

[77] S. Martinez-Conde, S. Macknik, X. Troncoso, and T. Dyar. Microsaccades counteract visual fading during fixation. *Neuron*, 49:297–305, January 2006.

[78] N. Max. Visualizing hilbert curves. In *Proceedings of IEEE Visualization*, pages 447–450, October 1998.

[79] K. May, T. Stathaki, and A. Katsaggelos. Spatially adative intensity bounds for image restoration. *Eurasip Journal on Applied Signal Processing*, 12:1167–1180, December 2003.

[80] R. Molina, J. Mateos, A. Katsaggelos, and M. Vega. A general multichannel image restoration method using compound models. In *IEEE International Conference on Pattern Recognition (ICPR2002)*, volume 3, pages 835–838, 2002.

[81] W. Na, J. Paik, and C. Lee. An image restoration system for a single-CCD color camcorder. *IEEE Transactions On Consumer Electronics*, 41(3):563–572, August 1995.

[82] K. R. Namuduri. Motion estimation using spatio-temporal contextual information. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(8):1111–1115, August 2004.

[83] M. Ng, T. Chan, M. Kang, and P. Milanfar, editors. *Special Issue on "Super-Resolution Imaging: Analysis, Algorithms, and Applications"*, volume 2006. Eurasip Journal on Applied Signal Processing, January 2006.

[84] N. Nguyen and P. Milanfar. An efficient wavelet-based algorithm for image superresolution. In *International Conference on Image Processing, IEEE Proceedings*, volume 2, pages 351–354, September 2000.

[85] S. Park, M. Park, and M. Kang. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, May 2003.

[86] A. Patti, I. Sezan, and M. Tekalp. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE Transactions on Image Processing*, 6(8):1064–1076, August 1997.

[87] S. Peleg, D. Keren, and L. Schweitzer. Improving image resolution using subpixel motion. *Pattern Recognition Letters*, 5:223–226, 1987.

[88] A. Perez, S. Kamata, and E. Kawaguchi. Peano scanning of arbitrary size images. In *Proceeding IEEE International Conference on Pattern Recognition, ICPR*, volume 3, pages 565–568, Spetember 1992.

[89] J. Proakis, C. Rader, F. Ling, M. Moonen, I. Proudler, and C. Nikias. *Algorithms for Statistical Signal Processing*. Prentice-Hall, 2002.

[90] Mobile Imaging Report. The camera-phone phenomenon: A new deal in imaging. Research report, Future Image, (http://www.futureimage.com), May 2004.

[91] R. Schultz and R. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Transactions on Image Processing*, 5(6):996–1011, June 1996.

[92] S. H. Seo, M. Azimi-Sadjadi, and B. Tia. A least-squares-based 2-d filtering scheme for stereo image compression. *IEEE Transactions on Image Processing*, 9(11):1967–1972, November 2000.

[93] M. Shimizu, T. Yano, and M. Okutomi. Super-resolution under image deformation. In *Proceedings of the IEEE Conference on Pattern Recognition*, volume 3, pages 586–589, August 2004.

[94] T. Soni, B. Rao, J. Zeidler, and W. Ku. Enhancement of images using the 2-d lms adaptive algorithm. In *IEEE Conference on Acoustics, Speech and Signal Processing*, volume 4, May 1991.

[95] W. Souidene, K. Abed-Meraim, and A. Beghdadi. Deterministic techniques for multichannel blind image deconvolution. In *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications*, volume 1, pages 439–442, August 2005.

[96] F. Sroubek and J. Flusser. Multichannel blind deconvolution of spatially misaligned images. *IEEE Transactions on Image Processing*, 14(7):874–883, July 2005.

[97] H. Stark and P. Oskoui. High-resolution image recovery from image-plane arrays, using convex projections. *Journal of the Optical Society of America A*, 6(11):1715–1726, November 1989.

[98] C. Stiller and J. Konrad. Estimating motion in image sequences. *IEEE Signal Processing Magazine*, 16(4):70–91, July 1999.

[99] A. Tekalp, M. Ozkan, and M. Sezan. High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, volume III*, pages 169–172, 1992.

[100] M. Tekalp and G. Pavlovic. Space-variant and color image restoration using kalman filtering. In *Proceedings of IEEE Symposium on Circuits and Systems*, volume 1, pages 8–11, 1989.

[101] M. Tico, M. Trimeche, and M. Vehvilainen. Motion blur identification based on differently exposed images. In *Proceeding of International Conference on Image Processing, ICIP06, Atlanta, USA*, October 2006.

[102] M. Tico and M. Vehvilainen. Constraint translational and rotational motion filtering for video stabilization. In *IEEE International Conference on Image Processing (ICIP05)*, volume 3, pages 569–572, September 2005.

[103] M. Trimeche, R.C. Bilcu, and J. Yrjänäinen. Adaptive outlier rejection in image super-resolution. *Eurasip Journal on Applied Signal Processing (JASP)*, 2006(Article ID 38052):1–12, January 2006.

[104] M. Trimeche, A. Burian, J. Kangas, and M. Vehviläinen. Enhanced mobile imaging for document applications. In *Proceeding of 13th International Conference on Systems, Signals and Image Processing (IWSSIP 2006), Budapest, Hungary*, September 2006.

[105] M. Trimeche and F. Chebil. Digital rights management for visual content in mobile applications. In *Proceedings of the International Symposium on Control, Communications and Signal Processing ISCCSP'04, Hammamet, Tunisia*, March 2004.

[106] M. Trimeche, F. Alaya Cheikh, and M.Gabbouj. Similarity retrieval of occluded shapes using wavelet-based shape feature. In *Proceedings SPIE conference on Internet Multimedia Management Systems, Boston, USA*, volume 4210, pages 281–289, November 2000.

[107] M. Trimeche, F. Alaya Cheikh, M.Gabbouj, and B.Cramariuc. Content-based description of images for retrieval in large databases: Muvis. In *Proceedings of European Signal Processing Conference, EUSIPCO 2000, Tampere, Finland*, September 2000.

[108] M. Trimeche, D. Paliy, M. Vehvilainen, and V. Katkovnik. Multichannel image deblurring of raw color components. In *Proceedings of SPIE Conference on Computational Imaging, San Jose, USA*, volume 5674, pages 169–179, January 2005.

[109] M. Trimeche, R. Suomela, A. Aaltonen, G. Lorho, T. Aarnio, T. Dossaji, and S. Tuoriniemi. Enhancing end-user experience in a multi-device ecosystem. In *Proceedings of Fourth International Conference on Mobile and Ubiquitous Multimedia, Christchurch, New Zealand*, volume 1, pages 19–25, December 2005.

[110] M. Trimeche, M. Tico, and M. Gabbouj. Dense optical flow field estimation using recusrsive LMS filtering. In *Proceeding of 14th European Signal Processing Conference, Eusipco 2006, Florence, Italy*, September 2006.

[111] M. Trimeche and M. Vehvilainen. Super-resolution using image sequence in raw sensor domain. In *Second International Symposium on Communications, Control and Signal Processing ISCCSP06, Marrakech, Morocco*, March 2006.

[112] M. Trimeche and J. Yrjänäinen. A method for simultaneous outlier rejection in image super-resolution. In *Visual Content Processing and Representation*, volume 2849 of *Proceedings of 8th International Workshop on very low bitrate video coding (VLBV), Madrid, Spain*, pages 188–195. Springer LNCS, September 2003.

[113] M. Trimeche and J. Yrjänäinen. Order filters in super-resolution reconstruction. In *Proceedings of SPIE Conference on Image Processing, San Jose, USA*, volume 5014, pages 190–200, January 2003.

[114] M. Unser, A. Aldroubi, and M. Eden. Fast B-spline transforms for continuous image representation and interpolation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:277–285, 1991.

[115] H. Ur and D. Gross. Improved resolution from subpixel shifted pictures. *Graphical Models and Image Processing*, 54:181–186, March 1992.

[116] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo. Weighted median filters: a tutorial. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 43(3):157–192, March 1996.

[117] A. Zomet, A. Rav-Acha, and S. Peleg. Robust super-resolution. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 645–650, 2001.

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O. Box 527
FIN-33101 Tampere, Finland