# Regularized Super-Resolution of Multi-View Images

by

Mustafa H. Fanaswala, B.Sc.

A thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Master of Applied Science in Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Systems and Computer Engineering
Carleton University
Ottawa, Ontario
August, 2009

The undersigned hereby recommends to the
Faculty of Graduate Studies and Research
acceptance of the thesis

Regularized Super-resolution of Multi-view Images

submitted by Mustafa H. Fanaswala, B.Sc.

in partial fulfillment of the requirements for the degree of
Master of Applied Science in Electrical Engineering

---

Professor Richard M. Dansereau, Thesis Co-Supervisor

---

Professor Eric Dubois, Thesis Co-Supervisor

---

Professor Howard Schwartz, Chair,
Department of Systems and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Systems and Computer Engineering
Carleton University
August, 2009

# Abstract

Super-resolution is the process of obtaining either a higher resolution still image or a sequence of higher resolution images from a corresponding sequence of low resolution images of a particular scene. It extends the performance limits of single image interpolation by leveraging the unique information present in the multiple albeit slightly different images. These multi-image techniques seek to recover frequency content beyond that present in any of the individual observed images and are hence termed 'super-resolution' algorithms. In its simplest form, a super-resolution algorithm aims to align the mutually shifted low resolution images on a higher resolution grid. The alignment process requires precise knowledge of the displacement occurring in the scene, which is estimated using the low resolution images. This allows a formulation of multiple observed data constraints that can be used together with knowledge about the imaging process to estimate the high resolution image. Although super-resolution algorithms have been shown to perform well in synthetic scenarios, many of the modeling assumptions break down in real world imaging conditions. Super-resolution performance is then heavily dependent on how the forward imaging model is constructed, which is a recurring theme of this thesis. We constrain ourselves to a specific imaging device and examine the effects of super-resolution when the characteristics of the camera are uniquely identified. Displacement estimation has been identified as a major factor in the performance of super-resolution and the choice of displacement models for different scenes is examined in the thesis. The estimation of the high resolution image is carried out using regularization-based methods (both algebraic and stochastic). The thesis also addresses artifacts arising from inaccurate displacement estimates either due to inconsistent displacement models and/or occlusions occurring in the scene. A complete system is built in this thesis and the results obtained show significant improvement over single image bi-cubic interpolation.

# Acknowledgements

I am ever grateful to God, for his blessings and benevolence that have helped me through all that I have endeavored to do. I would like to thank my supervisors, Dr. Richard Dansereau and Dr. Eric Dubois for their invaluable guidance and patience during my often incomprehensible ramblings. I would particularly like to thank Dr. Richard Dansereau for his constant support and many suggestions to refine my work. Professor Dansereau took an active interest in my course work, even helping me choose project topics and suggesting resources for the various courses that I took. I am also extremely thankful to Dr. Eric Dubois for first of all, agreeing to co-supervise me late into my degree and for being the ideal sounding board for my various hare-brained schemes. The freedom afforded to me and Professor Dubois' infinite patience is greatly appreciated. I hope that some of his acumen and wisdom has rubbed off on me. It was a pleasure working with them and more importantly, learning from them. My graduate course professors were ever helpful in my study of relevant fields and I am thankful to the high quality of teaching at the Ottawa-Carleton Institute of Electrical and Computer Engineering (OCIECE). A special thanks must go out to my colleague Mohammed Fouad, who is pursuing his Ph.D in Prof. Dansereau's research group at the time of writing of this document. Fouad was extremely cooperative in helping me implement some of the super-resolution techniques used in this thesis.

I would not be here without the unwavering support of my parents who have given up a lot, to see me stand where I am today. I would like to thank my brother, Adnan, for being my moral, social and financial conscience.

Last but not the least, I would like to thank all my friends at the Systems and Computer Engineering Department for making this educational journey an immensely enjoyable experience.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

> "There will come a time when you believe everything is finished. That will be the beginning."
>
> Louis L'Amour.

The goal of super-resolution (SR) is to increase the resolution of an image or a sequence of images beyond the resolving power of the imaging system. Image processing literature has traditionally used the term 'resolution' to refer to the image size in terms of the number of pixels. However, the conventional unit of pixel count is not an appropriate measure of resolution because increasing the number of pixels in the image may not contribute towards enhancing fine details in the image content. Correspondingly, the hardware-based solution of increasing resolution by reducing the pixel size and increasing the number of pixels per unit area does not always lead to satisfactory results. In digital cameras, a pixel corresponds to the detector sensor which is usually a CCD (charge-coupled device) or CMOS (complementary metal-oxide semiconductor) sensor. These photo-sensitive detectors work by integrating the available light impinging on them for a specified period called the aperture time. When the user presses the camera shutter release, each of these pixels has a 'photosite' which is uncovered to collect and store photons in a cavity. Once the exposure finishes, the photosites are closed and an assessment of how many photons fell into each cavity is made. The relative quantity of photons in each cavity are then sorted into various intensity levels, whose precision is determined by the bit depth[1]. A reduction in sensor size leads to the collection of photons over a smaller area and hence, an increased sensitivity to shot noise. Thus, resolution enhancement by sensor manufacturing techniques is limited by the minimum sensor area beyond which shot noise overwhelms the image. It has been reported that the imaging industry has already reached this pixel size limit of 40 $\mu m^2$ for a 0.35 $\mu m$ CMOS process [2]. Moreover, hardware-based solutions are often too expensive to employ in the ubiquitous applications of digital images. A software-based method is more attractive because it can be applied after the image sequence has already been captured and can thus leverage higher computational power utilizing more complex algorithms in an off-line scenario.

Super-resolution falls under the umbrella of image restoration techniques which have a wide variety of documented applications. From the simple need to enhance the resolution of digital photographs to more sophisticated computer-vision algorithms, the demand for higher resolution (HR) images is pervasive in various areas. The more popular applications alluded to in recent literature include surveillance applications like facial and license plate recognition; medical imaging applications fusing images from different modalities; resolution enhancement of astronomical and satellite imagery; and converting low resolution video to high resolution video amongst others.

As opposed to single-image interpolation techniques, super-resolution algorithms utilize multiple images to recover resolution that is lost in the imaging process. Each image in the set of mutually shifted images contributes unique information to aid in the recovery of a higher resolution view of the scene. Super-resolution is a particularly effective technique for low resolution (LR) images suffering from aliasing. Although natural scenes contain an infinite amount of detail, typical imaging devices are only capable of capturing the scene at a limited

---

[1] This simplistic description ignores the wavelength (color) dependence of the sensor

resolution. In other words, natural scenes are not band-limited signals and any attempt to digitally acquire an image of such a scene will inevitably result in aliasing. However, the camera optics typically serve as anti-aliasing filters because of their low-pass behavior. While this can mitigate the effects of aliasing, the resolving power of the sensor array in digital cameras is weaker than that of the camera optics. As a result, a small amount of aliasing is invariably present even when optical pre-filtering is carried out. In its most naive form, super-resolution by simply registering each low resolution image onto a higher-resolution sampling lattice can achieve an increase in resolution up to the diffraction limit [3]. Further increase in performance beyond the diffraction limit can be achieved by more elaborate super-resolution algorithms which require additional knowledge concerning the various blur degradations that the desired image undergoes as well as prior knowledge about the nature of the scene to be reconstructed.

A typical super-resolution algorithm involves three sub-tasks: registration, fusion and de-blurring. The set of low resolution images are first mutually aligned on a common high resolution lattice. This is generally referred to as image registration. The aligned pixel values (usually resulting in a non-uniform sampling) are then interpolated over the reference high resolution lattice to obtain a fused high resolution image. A subsequent de-blurring of the fused image results in a higher resolution image provided that a sufficient number of low resolution images are available and that the image alignnment is carried out to sub-pixel accuracy. While these sub-tasks have been separately identified for conceptual clarity, they are often performed in a joint fashion.

The observed low resolution images in a given set may differ from each other in various ways. Recent work on super-resolution using alternative cues like blur has been reported [4] (information from differently blurred images). However, this thesis primarily deals with the more common scenario of extracting useful information from images that differ from each other via either camera motion or object motion in the scene or both. The use of motion as a cue in enhancing vision is biologically inspired [3]. The human eye undergoes micro-tremblings called saccades to prevent image fade-out in the retina. Due to these random movements, the human eye is also capable of sensing an alias-free version of the scene. While the scene is not actually reconstructed on a finer grid, a finer localization of edges is made possible leading to higher perceived resolution [3]. It is interesting to draw a parallel from biological vision systems and see how they apply to digital cameras. The human retina has a limited density of photo-receptors which limits visual acuity (ability to resolve fine detail) to 1′ of visual angle. However, the human eye was found to be able to resolve certain stimuli at a much higher resolution of the order of 5″ of visual angle. This enhanced capacity is termed hyperacuity [5]. Quantitative tests were carried out in [6] to validate the hypothesis that hyperacuity results from micro-movements of the eye. These results motivate the use of motion as a cue to improve the resolution of images captured from digital cameras.

We define the problem under consideration in this thesis in Section 1.1. A short introduction is given in Section 1.2 regarding the specific imaging sensor used. The organization of the thesis is detailed in Section 1.3. Finally, a highlight of the contributions made in the thesis is given in Section 1.4.

## 1.1 Problem definition

In this thesis, we are mainly concerned with the resolution enhancement of a single image using spatial diversity from multiple views. In particular, we examine the resolution enhancement that can be obtained when an imaging sensor like the ViewPlus ProFUSION25 camera array is used to image a particular scene. Typical super-resolution algorithms trade off temporal resolution in a video sequence for higher spatial resolution [7]. For example, every five frames in the video could be replaced by one SR frame. The ProFUSION25 camera array provides us with multiple views of the same scene. In essence, there is no temporal variation because each camera in the array is synchronized to capture the scene at the same time. Such a scenario in which spatial resolution is enhanced by the use of multiple views presents unique challenges that have not been directly addressed previously to the best of our knowledge. In a typical imaging scenario, loss of resolution occurs mainly due to the following:

1. Aliasing: In this scenario, images are obtained with negligible optical blur, but the required fine details are unavailable because of the low spatial sampling density of the camera.
2. Blur: In this scenario, the sensor sampling density is adequate to capture the required details, but the images are degraded by some blurring function.

In aliasing dominated cases, an increase in pixel count can lead to more detail. However, in blur dominated cases, an increase in pixel count may not necessarily lead to more detail. In such a case, details are revealed by removing the blur degradation through a de-convolution process. Since we do not know a priori which of these effects dominate in a given image, it is difficult to give a precise definition of what is meant by an improvement in resolution. In this thesis, we use a subjective albeit perceptually motivated definition of resolution enhancement like the one used in [8]:

> The super-resolved image should demonstrate an improvement in the perceived detail content compared with that of the low resolution images. This typically involves restoration of the high frequency content, which may require an increase in pixel density.


(a) high resolution license plate image (b) Resolution loss due to motion blurring (c) Aliasing effects due to insufficient sensor density

Figure 1.1: Blurring and aliasing effects in the loss of resolution

Examples of both aliasing and blur dominated low resolution images are shown in Figure 1.1. The original high resolution is that of a license plate[1]. Both the blurred and aliased versions lack sufficient resolution. The aliasing effects in Figure 1.1(c) can be characteristically identified by the jagged edges that appear in the image.

With the definition stated above and the particular scenario of multi-view images taken from the ProFUSION25 camera array, a description of the problem can be stated as:

> Given a set of low resolution multi-view images $\{g_i, i = 0, 1, ..., k\}$, of size $M \times N$ pixels taken with the ProFUSION25 camera array, and a magnification factor $m$, reconstruct a higher resolution image of size $mM \times mN$ pixels that satisfies the definition of resolution improvement stated above.

## 1.2 The ProFUSION25 camera array

The ProFUSION25 system is a camera array of 25 cameras arranged on a 5x5 grid. A picture of the camera array is shown in Figure 1.2[2]. We chose to use the ProFUSION25 camera for a variety of reasons. Each camera in the array is capable

---

[1] The image was retrieved from `http://www.coolpl8z.com`

[2] The image was retrieved from `http://www.viewplus.co.jp`

Figure 1.2: The ProFUSION25 camera array

of delivering an image sequence at 25 frames per second (FPS). However, for the purposes of this thesis, the camera has only been used in one-shot mode to restrict the possibility of temporal motion of objects in the scene. As a result, any displacement occurring between the images is entirely due to a change in viewing position. The ProFUSION25 camera array outputs raw 8-bit gray-scale images of pixel resolution $640 \times 480$. The results discussed in this thesis are thus applied only to gray-scale images. Multi-view images like the ones captured by the ProFUSION25 are ideal for super-resolution applications. The small baseline between each camera in the array allows the multiple views to adequately sample the high resolution image. This is shown to be a condition for obtaining nearly optimal super-resolution performance in [9]. While many papers merely state that super-resolution from multi-view images is possible, very few carry out experiments on a representative data set. Furthermore, an array like the ProFUSION25 essentially uses 25 different cameras each of which may have a different point spread function (PSF) and other imaging defects. These are issues that raise interesting questions regarding the performance of super-resolution algorithms.

The critical reader may wonder what motivates us to use an array like the ProFUSION25 rather than a single higher resolution camera? The principal motivation of this work is to be able to render a higher-resolution three-dimensional representation of a scene using the diversity presented by the multiple views. For convenience, we imagine a scene consisting of a sculpture in a museum. Is it possible to use the disparity between multiple views of this sculpture to generate a high resolution three-dimensional representation of this sculpture? Can we render a super-resolved image from a novel view-point, i.e., from a view that was not one of the original viewing positions? This scenario is the original motivation for work with this particular camera sensor array. Before we can address the questions raised above, we need to examine whether any resolution enhancement can be attained from viewpoints that already exist. We pose this as the following questions to be answered during the course of this thesis:

1. Can we achieve an enhancement in resolution for simple planar scenes from current viewpoints?
2. If so, can we achieve similar resolution enhancement for scenes with depth discontinuities (more complex scenes with 3D objects)?
3. What do we need to know about the ProFUSION25 camera array to achieve resolution enhancement?
4. How can we combine all the components required for super-resolution to create a complete system?

Some of the questions raised above have been addressed in isolation in the current literature. However, our specific scenario brings up issues that have not been addressed together as it relates to super-resolution. A literature survey of relevant papers is provided in Chapter 2.

## 1.3 Organization of the thesis

Following this introductory chapter, a brief review of existing work in super-resolution is presented in Chapter 2 as they relate to the objectives of the thesis. Some relevant background material is also presented. Chapter 3 deals with regularization methods in super-resolution. We present super-resolution as an inverse problem which is typically solved using regularization. A comparison of both algebraic and stochastic approaches is made and the use of a total-variation based regularizer is also advocated, which is solved using the method of level sets. We postulate in Chapter 4 that accurate knowledge of the imaging pipeline is essential for the success of any super-resolution algorithm. Consequently, the components of the ProFUSION25 camera array are characterized. We identify the camera blur and demonstrate the effect of camera specific degradations like vignetting. Super-resolution methods are known to suffer from artifacts due to inevitable errors in displacement estimation. In Chapter 5 an adaptive weighting approach is proposed to alleviate the effect of incorrect displacement estimates. We also demonstrate occlusion-aware super-resolution by using an occlusion detection scheme together with dense displacement maps when reconstructing images of scenes with depth discontinuities. Finally, we make concluding remarks in Chapter 6, summarize the contributions of the thesis and present avenues for future work.

## 1.4 Thesis highlights

An exhaustive list of contributions is provided in Chapter 6, however, a brief mention is made here to prime the reader for the contributions made in subsequent chapters. In Chapter 2, a locally affine but globally smooth optical flow estimation technique [10] is described which has already been used in medical image registration. However, the integration of this algorithm within our super-resolution algorithm is a novel contribution as such a technique has not been tested specifically in super-resolution applications using dense displacement estimates. In Chapter 3, no original contributions are made. However, the implementation of a TV regularizer using mean curvature evolution has not been previously used in super-resolution of multi-view images. In Chapter 4, a characterization of the ProFUSION25 camera is carried out via its point spread function (PSF) and its vignetting function. This is an original contribution of this thesis. Moreover, a novel PSF identification technique is proposed. Finally, in Chapter 5, an intensity based occlusion detection scheme is used to remove artifacts related to erroneous displacement estimates. This particular method is another novel contribution. At present, no study has been carried out on the feasibility of using a specific sensor for super-resolution. The trend in super-resolution literature is to use image sequences that are artificially shifted, blurred or sub-sampled. These are not representative of real world scenarios. Consequently, the fact that we have used a real sensor (the ProFUSION25) with real images is not only an original contribution but also a significant one.

# Chapter 2

# Background

> "Euclid taught me that without
> assumptions there is no proof.
> Therefore, in any argument,
> examine the assumptions."
>
> - Eric T. Bell

The well-studied field of super-resolution draws from related research in image restoration and motion estimation. These fields are immense research areas by themselves and a comprehensive review of related work in all pertinent areas is infeasible. In Section 2.1, a brief review of super-resolution is presented for issues that are directly related to those considered in this thesis. Super-resolution is formulated as an inverse problem in Chapter 3. Such a formulation first requires a forward (or observation) model that is specified in Section 2.2. The displacement estimation techniques used in this thesis have been briefly introduced in Section 2.3. Finally, chapter highlights are presented in Section 2.4.

## 2.1 Literature review

Super-resolution has been extensively studied and a vast body of research is available in the literature. We will not repeat a drawn out review of existing techniques but point the reader towards the excellent review articles [2] and [7]. We will however present a review of topics related to super-resolution as they apply to this thesis in the following sub-sections.

### 2.1.1 On super-resolution restoration methods

The restoration method forms the backbone or engine that drives the super-resolution process. A targeted review of super-resolution is carried out to choose the best reconstruction to use in this thesis. The earliest attempt at super-resolution is that by Tsai and Huang [11] which is carried out in the frequency domain. Their approach is based on formulating a system of equations that relates the discrete Fourier transform (DFT) coefficients of the low resolution images to samples of the continuous Fourier transform of the original scene. The recovered samples, obtained after solving the system of equations, are used to obtain a high resolution image using the inverse DFT. However, their observation model only considered ideal down-sampling while neglecting the effects of blurring and noise. A related class of methods are those based on the multi-channel sampling theorem. A recent paper using the multi-channel sampling theorem is by Vandewalle, Susstrunk and Vetterli [12] which computes sub-pixel shifts between the low resolution images and registers them on a common high resolution grid. The registered pixel values are non-uniformly spaced and hence, a non-uniform interpolation is carried out to obtain a high resolution image on a uniformly spaced grid. The resulting image is subsequently de-blurred. This method is commonly used in the literature because it is conceptually intuitive. However, it ignores the role of the observation model. Such methods are also referred to as interpolation-restoration methods because they consist of a non-uniform interpolation stage followed by a subsequent de-blurring stage. Irani and Peleg [13] propose the iterative back-projection (IBP) approach that was borrowed from methods in computer-aided tomography. The observation model is used to generate a set of simulated low resolution images from an initial high resolution image. The high resolution image is then updated until the error between the simulated low resolution images and the observed low resolution images is minimized. The updates are carried out by passing the error through a back-projection operator. The selection of the back-projection operator is not clear and the approach forces the high resolution image to match the observed data. This might not necessarily be the best strategy in the presence of noisy images. These shortcomings led to the class of regularized super-resolution algorithms which are categorized as either stochastic or algebraic regularization methods. Regularization methods are used to include useful a priori information that help to disambiguate between the possible solutions of the ill-posed super-resolution problem. Schultz and Stevenson [14] utilize a Bayesian framework to formulate the super-resolution problem and solve it using MAP estimation. A stochastic image prior is used to regularize the problem. The specific prior used is the Huber-Markov random field (HMRF) which is able to preserve discontinuities and hence serves to better represent edges in an image. A block-based displacement estimation scheme is used to register the low resolution images in their work. As opposed to the stochastic case, a deterministic algebraic regularizer called the total-variation regularizer is used by Ng et al. [15]. The total-variation regularizer has been receiving burgeoning interest from the image processing community because of its demonstrated effectiveness in preserving edges. Ng et al. jointly estimate a dense displacement field together with the high resolution image. The experimental results compare the different types of total-variation regularizers including those using the $L_1$-norm and bi-lateral total variation (BTV). The seminal paper by Baker and Kanade [16] presents a sequence of results which show that as the magnification factor increases, the observation model provides far less useful constraints. The regularizer dominates at high magnification factors leading to overly smooth high resolution images. As a result, a recognition-based prior is proposed that is learnt from a database of high resolution images of specific images like faces or text. A whole class of recognition-based super-resolution algorithms has evolved as a result but we resist from taking such an approach because the image prior is only applicable to a specific class of images that it has been trained on.

Based on the literature surveyed on reconstruction methods, we believe that regularization techniques are the most promising approach to use as the engine that drives our super-resolution algorithm. The frequency domain methods restrict the displacement model to global models which are not conducive to the type of scenes being considered in this thesis. Regularization methods allow us to leverage a priori information about the observation model as well as impose desired characteristics on the high resolution image which other methods cannot fully incorporate. A further study is required to choose a suitable regularization method which is the topic of Chapter 3.

### 2.1.2 On super-resolution using multi-view images

The super-resolution community has traditionally only considered temporal sequences of images in super-resolution experiments because they are easily available in the form of a video sequence. However, there has been an increasing interest in super-resolution from multi-view images. This is motivated by the inherent trade-off between capturing a larger area of the scene and preserving detail in the image. Given a certain pixel resolution, a larger view of the scene results in fewer pixels being used to represent individual features in the scene. A way to compensate for this is to use multiple overlapping images of a scene to create a larger composite image. An interesting question that arises then is whether the multiple views of the scene can be used to increase the perceived detail in a scene using super-resolution principles. This is the primary motivation of this thesis. Capel and Zisserman [17] consider computer vision applications of super-resolution using multi-view images. However, they only consider global displacement models which relate different views of a scene through a homography. A photometric registration stage is also considered which compensates for intensity changes between images. The photometric registration considered only accounts for a brightness and contrast difference between the images. A more complicated photometric distortion cannot be accounted for using such a procedure. The high resolution image is recovered using a maximum a posteriori (MAP) estimation approach. A number of different priors including the Huber-Markov random field are utilized to obtain the high resolution image.

We feel that there is a paucity of super-resolution papers that directly examine multi-view images. While super-resolution techniques for temporal image sequences can be directly applied to multi-view imagery, a study of this specific scenario (not without its peculiarities) is useful. This is the theme that we maintain throughout the thesis.

### 2.1.3 On blur identification and super-resolution

The problem of de-blurring a single image has been extensively studied under the title of *blind* de-convolution. An excellent review of single image blind deconvolution techniques can be found in [18]. However, the problem of simultaneously estimating an unknown blur from the observed low resolution images and performing super-resolution is yet to be satisfactorily solved [19]. Existing techniques either treat the blur identification separately from the super-resolution part or jointly estimate the high resolution image and the unknown blur. Nguyen, Milanfar and Golub [20] take the latter approach and perform multi-image deconvolution which utilizes information from multiple images to obtain a de-blurred high resolution image. However, they parameterize the blur to depend on only one parameter. Such a parameter could represent the width of a gaussian blur or the radius of an out-of-focus blur. Moreover, the PSF support is assumed to be known or is set to be equal to the magnification factor in each dimension. A similar approach where a one-parameter blur is jointly estimated with the high resolution image is carried out in [21] and [22]. The restriction placed by a one-parameter blur is considered to be very limiting for most real applications.

Lertrattanapanich and Bose [23] also perform super-resolution using multiple images but separately identify the blur which is used to de-blur a non-uniform interpolation of registered low resolution images. A Delaunay triangulation is used to create a set of triangular patches for the non-uniformly spaced sampling locations that result from registering the low resolution images. A surface approximation using bi-variate polynomials is constructed for each triangular patch that is used to convert the non-uniform sampling to a uniformly spaced high resolution grid. The image that is obtained after the non-uniform interpolation is noise-filtered and de-blurred using a point spread function that is estimated using multiple low resolution images. A modified Richardson-Lucy algorithm [24] is used in the multi-image blur identification.

The most recent published work on simultaneous super-resolution and blur identification is that carried out by Sroubek, Cristobal and Flusser [25]. The super-resolution problem is formulated as a regularization problem involving both the image domain and the blur domain. This regularization problem is solved using an alternating minimization approach. In the alternating minimization approach using steepest descent, the iterations alternate between a descent in the image sub-space followed by a descent in the blur sub-space. This is a viable approach because each term in the energy function is quadratic independently but not as a whole. A further improvement suggested in [25] was to use the centroid of the blur estimated at the high resolution scale to infer shifts between the images (reportedly to a sub-pixel) accuracy. The work that is most similar to that contained in this thesis is that of Chen, Luo and Hu [26] who also use an alternating minimization between the image sub-space and the blur sub-space. They utilize an anisotropic diffusion framework to solve the restoration problem. However, the observation model used considers blurring to take place before warping which is different from the observation model used in this thesis. The warping model is also constrained to consist only of translational shifts.

Based on these papers, we feel that it would be worthwhile to investigate the effects of correctly identifying the blur in each ProFUSION25 camera. We expect an increase in super-resolution performance over assumed blurs which is validated in Chapter 4.

### 2.1.4 On super-resolution using dense displacement maps

A survey of super-resolution techniques brings up the surprising finding that few attempts have been made in applying dense displacement maps (also used synonymously with optical flow) to super-resolution. The first attempt of super-resolution using optical flow is reported by Baker and Kanade [27] using images taken from a multi-baseline stereo camera. This is one of the few papers using multi-view images rather than a temporal sequence of images. Their study is motivated by the use of super-resolution in face recognition. They use a hierarchical Lucas-Kanade [28] optical flow algorithm to register the low resolution images on the high resolution grid. The registered images are then fused by an averaging operation across the registered images. Motivated by the same application in face recognition, Lin et al. [29] also examine super-resolution using optical flow. They state that face images are non-planar, non-rigid, non-Lambertian and prone to self occlusions. These characteristics of faces cause global parametric displacement models to fail in providing sufficient resolution enhancement for face recognition tasks. Their technique utilizes an interpolation-restoration framework. However, they obtain the optical flow directly at the high resolution scale by estimating the flow between low resolution images that are interpolated using bi-cubic splines.

Sub-pixel shifts are not estimated and the registered images are fused using a median operation. This is also a simplistic fusion approach which results in a blurred high resolution image.

Zhao and Sawhney [30] address the important issue of the feasibility of super-resolution using optical flow. They allege that most super-resolution algorithms implicitly or explicitly assume that images in a sequence are related through global parametric transformations. A simple iterative back-projection (IBP) reconstruction method is used to reconstruct the high resolution image. While definitive conclusions are not made, an important remark made is that the optical flow estimates should be consistent in the forward and backward direction for the best performance. The regions where the forward and backward flow are inconsistent tend to be occluded regions or areas with little texture where reliable flow estimates cannot be made.

An investigation of basic optical flow methods on planar and omni-directional images is carried out by Nagy and Vamossy [31]. They use the non-uniform interpolation framework to obtain the high resolution image. The Lucas-Kanade [28] and Horn-Schunck [32] optical flow algorithms along with block-based displacement estimation techniques are used to estimate a dense displacement field at the low resolution scale. This displacement field is used to obtain displacement-compensated images on the sampling grid of a chosen reference image. These images are finally averaged to obtain a fused high resolution image. We feel that averaging in such a manner is an overly simplistic way of image fusion which ignores the observation model. A post-processing step is suggested to sharpen the image.

The recent work by Krylov and Nasonov [33] uses optical flow together with a regularization framework to solve the super-resolution problem. They use a feature-based image registration algorithm together with Lucas-Kanade optical flow to obtain an accurate dense displacement field. This is used in a Tikhonov regularization approach to restore the high resolution image. Fransens, Strecha and Van Gool [34] present one of the first reported approaches to combine occlusion detection and optical flow estimation within the super-resolution process. An expectation-maximization (EM) algorithm is used to solve the problem which is formulated as a maximum a posteriori estimation problem. The EM algorithm alternates between i) detecting the occlusions based on photometric mis-match and ii) optimizing the dense displacement fields and the high resolution image. The image prior used in the MAP framework is similar to a Gauss-Markov random field which results in blurring across edges. The occlusion detection used in this work depends on intensity mis-match which generates a visibility map. This visibility map is not a binary map of visible/occluded pixels but rather behaves like a weighting image. The intensity of each pixel in the visibility map is a weight that reflects the possibility of that pixel being occluded. These visibility maps are viewed as hidden variables and are updated iteratively with the high resolution image. The demonstrated results are only compared to a poor nearest neighbor interpolation of the low resolution image. While the super-resolution result is superior to the nearest neighbor interpolation, we feel that the test sets used do not contain significant occlusion areas to comment on the effectiveness of the visibility maps used. Shen et al. [35] propose segmentation of the image, estimation of the displacements and finally recovery of the high resolution image, all done in a joint fashion. A cyclic coordinate descent scheme is used to estimate the displacement fields, segmentation fields and high resolution image alternatingly given the other two. The segmentation field is partitioned into objects which are separately labeled and tracked through the low resolution image sequence. The displacement of each object is estimated using an affine homography. While such a scheme does not give dense correspondences for each pixel, it results in displacement fields that are more accurate than those arising from global models. Ng et al. [15] allow for non-visible (occluded) pixels within their observation model but details on occlusion detection are not given in their paper. The experimental setup mistakenly associates occluded pixels with 'missing' pixels. A synthetic scenario is constructed where regions of the image are blacked out. Occlusion may create missing displacement vectors but the images themselves do not contain missing pixel values. In any case, conclusive experiments demonstrating occlusions are lacking.

Based on the literature surveyed in this sub-section, we feel that the super-resolution of multi-view images would benefit from the use of a dense displacement estimation technique. Furthermore, a conclusive paper on methods to account for occlusion in super-resolution is absent. This motivates the study in Chapter 5.

## 2.2 Observation model

The observation model is sometimes referred to as the forward model in this thesis to emphasize the fact that super-resolution is an inverse problem. An accurate description of the observation model is vital for the success of the super-resolution algorithm. Such a description involves characterizing the imaging sensor as completely as possible and making appropriate assumptions about the type of scene being imaged. Such assumptions impact the choice of displacement model used and the prior knowledge incorporated into the super-resolution algorithm.

A skeleton diagram of a general observation model is shown in Figure 2.1. The fundamental components comprise the warp operator, the blur operator and the down-sampling operator. The warp operator describes the displacement that takes places between two images in a sequence. The displacement between images could arise from camera motion, object motion in the scene or a combination of both. The most general description of such an operator would take the form of densely populated displacement estimates providing horizontal and vertical displacement vectors for each pixel in the image. However, simplified parametric models are often used to varying degrees of success. Such models include translational models, affine models and more general homographies. Piece-wise parametric models have also been successfully used either over blocks or arbitrarily shaped regions in the image. The blurring operator describes the cumulative blurring effects from sensor averaging, motion blur and an out-of-focus blur etc. These cumulative effects can be combined into a single blurring operator represented by a single point spread function. Commonly used assumptions include a rectangular averaging blur to represent sensor averaging and a circular disk to represent de-focus. An important modeling ambiguity is pointed out in [36] regarding the correct order of the warp and blur operators. The physical process of generating each LR image that agrees with imaging physics is that of warping followed by blurring. Many super-resolution algorithms implicitly follow this warp-blur paradigm. However, certain algorithms like the interpolation-restoration algorithms advocate the blur-warp model which assumes that blurring precedes the warping operation. This paradigm also models physical reality with the difference being that atmospheric blur is considered the major contributor to the blurring process. On the other hand, the warp-blur model considers the blur induced by the camera to be dominant. Following the notation used in [37], consider a set of $k+1$ low resolution images $g_i, i \in [0, 1, ..., k]$. Each of these images associate a two-dimensional coordinate $\mathbf{x}$ on a low resolution lattice $\Gamma$ with an intensity value $g_i(\mathbf{x})$. The goal is to estimate a high resolution image $f_0$ from the same point of view as the reference low resolution view $g_0$ but on a higher resolution lattice $\Lambda$.

Considering the case of arbitrary displacements, a position $\mathbf{x}$ in $f_0$ is mapped onto a position $\mathcal{T}_i(\mathbf{x})$ in $g_i$ by a free-form transformation $\mathcal{T}_i$. This transformation can be decomposed into an warping operator $\mathcal{F}_i$ and a down-sampling operator $\mathcal{D}$ such that $\mathcal{T}_i = \mathcal{D} \circ \mathcal{F}_i$. The warping operator $\mathcal{F}_i$ maps positions $[x, y]$ onto $[x + u_i(x, y), y + v_i(x, y)]$ where $u_i(x, y)$ and $v_i(x, y)$ are the horizontal and vertical displacement field components. The down-sampling operator $\mathcal{D}$ maps $[x, y]$ onto $[x/m, y/m]$ where $m$ is the magnification factor.

The resulting transformation is

$$\mathcal{T}_i : \Lambda \rightarrow \Gamma : \begin{bmatrix} x \\ y \end{bmatrix} \rightarrow \begin{bmatrix} (x + u_i(x, y))/m \\ (y + v_i(x, y))/m \end{bmatrix},$$

where $\Lambda$ is the high resolution lattice and $\Gamma$ is the low resolution lattice. The inverse warping operator $\mathcal{T}_i^{-1}$ associates a two-dimensional pixel co-ordinate $\mathbf{x}$ in $g_i$ (on the low resolution lattice $\Gamma$) to a corresponding position $\mathcal{T}_i^{-1}(\mathbf{x})$ (on the high resolution lattice $\Lambda$) in $f_0$. Descriptively, given a horizontal ($u_i(x, y)$) and vertical ($v_i(x, y)$) displacement field, the two-dimensional coordinates $\mathcal{T}_i^{-1}(\mathbf{x})$ describe the original locations of the pixels on the high resolution grid of the reference image $f_0$.

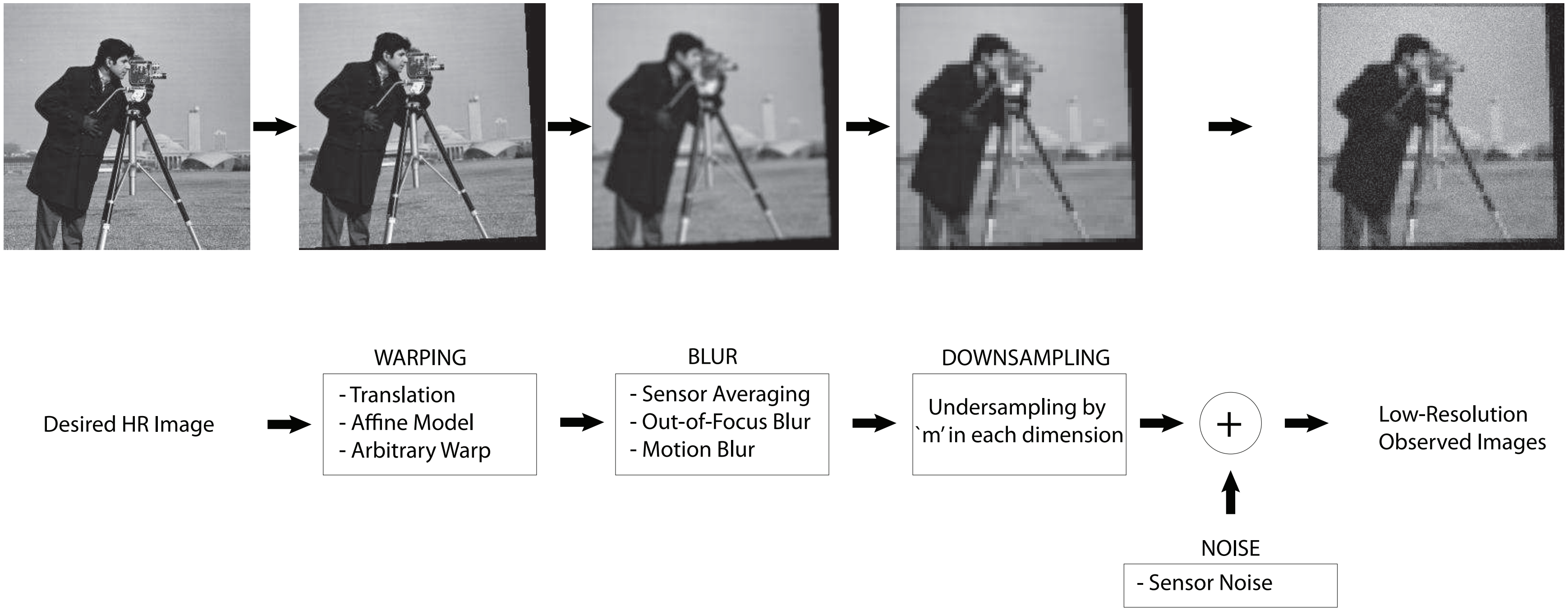


Figure 2.1: A general observation model

The low resolution images $g_i$ are assumed to be generated by the high resolution reference image $f_0$. First, the relative displacement between $g_i$ and $f_0$ is realized by applying the warping operator $\mathcal{F}_i$ to $f_0$. The warped image $f_i$ is then convolved with the point spread function $\mathcal{B}_i$. The PSF kernel $\mathcal{B}_i$ can represent the cumulative effects of sensor averaging, de-focus and other effects like motion blur. Finally, a down-sampling operator $\mathcal{D}$ is applied which performs ideal impulse sampling. As in any measurement system, the low resolution images are subject to additive noise $n_i$.

The imaging model can then be mathematically formulated as

$$g_i(\mathbf{x}) = \mathcal{D} \downarrow (\mathcal{B}_i * f_0(\mathcal{F}_i(\mathbf{x}))) + n_i(\mathbf{x}), \tag{2.1}$$

where $\mathcal{D} \downarrow$ refers to the down-sampling operation. We have abused notation slightly by using the same warping operator $\mathcal{F}_i$ and down-sampling operator $\mathcal{D}$ in a transformation of the domain from $\Lambda \rightarrow \Gamma$ as well as a transformation in the signal space from $f_i \rightarrow g_i$. Nonetheless, it is more convenient to express each of the operators as equivalent matrix multiplications to develop the discussion in subsequent chapters. This requires a lexicographic ordering of the images involved. Each of the $M \times N$ low resolution images $g_i$ can be ordered into a vector $\mathbf{g}_i$ of size $MN \times 1$. Similarly, the $mM \times mN$ high resolution reference image $f_0$ can be ordered into a column vector $\mathbf{f}_0$ of size $m^2MN \times 1$. The warping operator $\mathcal{F}_i$ can be represented by a square matrix $F_i$ of size $m^2MN \times m^2MN$. Convolution with the PSF kernel $\mathcal{B}_i$ can be represented by an equivalent convolution matrix $B_i$ of size $m^2MN \times m^2MN$. The structure of this matrix $B_i$ depends on the boundary conditions used. Finally, the down-sampling operator $\mathcal{D}$ can be represented as a matrix $D$ of size $MN \times m^2MN$. The additive noise is a column vector of size $MN \times 1$. The imaging model can now be expressed as

$$\mathbf{g}_i = DB_iF_i\mathbf{f}_0 + \mathbf{n}_i. \tag{2.2}$$

This observation model is essential for the formulation in Chapter 3. It is also instructive to relate the topics discussed in the thesis to where they fit into the observation model.

## 2.3 Displacement estimation

Displacement estimation is a vast field which has been extensively studied in literature with varied applications in video compression and other computer vision tasks. It is often used synonymously with the term ‘motion estimation’, ‘disparity estimation’, ‘warp estimation’, ‘image alignment’ or ‘image registration’ depending on the application domain being studied. A good review of displacement estimation algorithms is given in [38]. Since this thesis deals primarily with multiview images of static scenes, there is no real motion taking place between the images. As a result, we prefer to use the term ’displacement’ to refer to changes in spatial location. The use of the term ‘motion’ indicates a temporal aspect which is not dealt with in this thesis. However, in general discussions, we will continue to use the term ‘motion estimation’. Displacement estimation forms an integral part of super-resolution because:

1. knowledge about the displacements between images provides essential constraints to aid in the solution of the ill-posed super-resolution problem.
2. inaccurate estimates of the displacement leads to objectionable artifacts in the high resolution image.
3. the degree of resolution enhancement that can be achieved depends on the

sub-pixel accuracy of the displacement information [39].

In certain applications like satellite imaging, the relative displacement between low resolution images can be calculated based upon the orbital path and velocity of the satellite. However, in most cases, the displacement has to be estimated as a pre-processing step for super-resolution. The motion of objects in the real world can be described by a three-dimensional (3D) velocity field. Its projection on the image plane can be described by the corresponding two-dimensional (2D) motion field. When an estimate of the projected motion is required, we are faced with the motion estimation or displacement estimation problem. The estimated motion is typically described using instantaneous velocity (flow) or displacement vector fields [39] which are equivalent descriptions under the assumption of constant velocity motion between images. However, in an imaging system, only the spatio-temporal variation of the scene radiance is available as information. This variation in scene radiance arises from complex interactions between scene illumination, the motion of objects and the parameters of the camera (orientation, focal length, etc.) [39]. The projected 2D motion field has to be inferred using this intensity variation. However, not all changes in image intensity contribute to the 2D motion field. In Chapter 4, an intensity distortion is examined which causes intensity changes between images that is not motion related. Nevertheless, even in the presence of such ambiguities, it is possible to recover an approximation to the 2D displacement field which is popularly referred to as optical flow.

Consider a point $\mathbf{X} = [x, y, z]^T$ on an object in three dimensional space. A camera at a particular viewing position indexed by $i$ projects the 3D point $\mathbf{X}$ on the image plane at $\mathbf{x}_i = [x_i, y_i]^T$. Given two such viewing positions indexed by $i, j$ and corresponding image intensities $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ (or alternatively, $f_i(\mathbf{x})$ and $f_j(\mathbf{x})$), the position of the projection of the 3D point $\mathbf{X}$ on the image plane from viewing position $i$ may be related to that from the viewing position $j$ in two ways:

1. The forward displacement or disparity $x_j - x_i$ describes the displacement of the projected 3D point when there is a change in viewing position from position $i$ to position $j$. When the forward displacement is used to predict image $f_j(\mathbf{x})$ from the intensity values of the pixels of image $f_i(\mathbf{x})$, we term this process *forward registration*.

2. The backward displacement or disparity $x_i - x_j$ describes the displacement of the projected 3D point when there is a change in viewing position from position $j$ to position $i$. When the backward displacement is used to predict image $f_i(\mathbf{x})$ from the intensity values of the pixels of image $f_j(\mathbf{x})$, we term this process *backward registration*.

The estimation of displacement fields is most easily categorized under non-parametric and parametric displacement fields. A non-parametric displacement field is a representation of the 2D motion field on a finite set of points in the image plane. It is common to choose the set of points to correspond with the uniformly spaced, discrete image sampling lattice $\Gamma$. Such a method is the most general approach to displacement estimation and is advantageous because it can recover arbitrary motion fields. The displacement between sampling points can be obtained through a suitable interpolation. However, the large number of parameters to be estimated makes the problem under-determined and difficult to solve without the inclusion of additional constraints. These non-parametric methods are also typically called dense because they can recover correspondences between regions that may be as small as a pixel. A parametric displacement field, on the other hand, represents the 2D motion field as a continuous function of the spatial location $\mathbf{x}$ that is determined by some parameters. Common parametric models range from a two parameter shift to a twelve parameter homography. However, it is not possible to represent arbitrary displacement using global parametric methods. With this short introduction to displacement estimation, we present the displacement

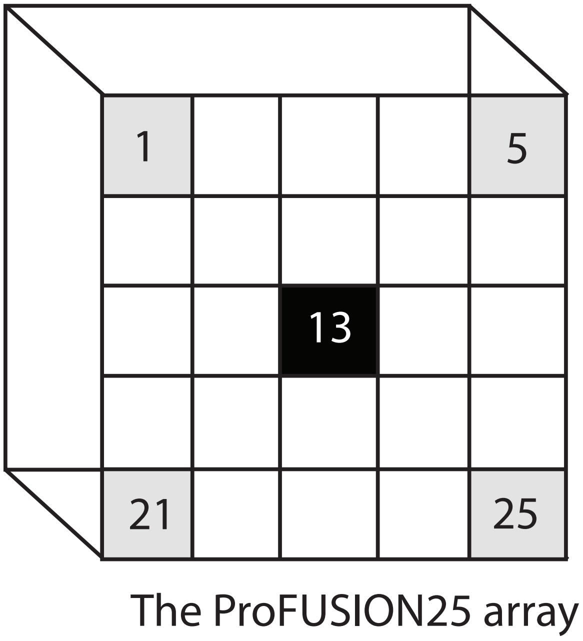


Figure 2.2: The $5 \times 5$ grid on which the ProFUSION25 sensors are arranged.

models and the estimation techniques used in this thesis.

### 2.3.1 Sub-pixel shift estimation

The ProFUSION25 camera array consists of individual cameras that are arranged on a regular $5 \times 5$ grid as shown in Figure 2.2. Each observed low resolution image can be considered to have been obtained by translating a camera to one of the 25 viewing positions on the $5 \times 5$ grid. In light of such a scenario, the most intuitive displacement model is that of a two parameter translational shift. It is expected that the pixels in the image undergo a translation that is proportional to the translation in the camera position. A spatial-domain technique proposed by Keren et al. [40] has been used to estimate the two-parameter shift in this thesis. The algorithm proposed in [40] is a gradient-based technique and was originally intended to estimate shifts as well as a rotation around the origin. However, in this thesis, only the shift parameters have been estimated using this algorithm. Consider a shifted image $g(x, y)$ which is related to a reference image $f(x, y)$ through the relation $g(x,y) = f(x+p_1, y+p_2)$. The horizontal and vertical shift parameters can be represented by the parameter vector $\mathbf{p} = [p_1 \;\; p_2]^T$. To cast the problem in an optimization framework, an error function is defined by

$$E(\mathbf{p}) = \sum_{x,y} \left[f(x+p_1, y+p_2) - g(x,y)\right]^2 . \tag{2.3}$$

However, the function in Equation 2.3 is non-linear in the parameters $\mathbf{p}$. Using the first order Taylor series expansion of the reference image intensity, the shifted image can be expressed in terms of the reference image by

$$g(x,y) \approx f(x,y) + p_1 \frac{\partial f(x,y)}{\partial x} + p_2 \frac{\partial f(x,y)}{\partial y}, \tag{2.4}$$

where $p_1$ and $p_2$ are the horizontal and vertical shift parameters to sub-pixel accuracy. The error function can then be expressed as a linear function of the shift parameters,

$$E(\mathbf{p}) = \sum \left[ f(x,y) + p_1 \frac{\partial f(x,y)}{\partial x} + p_2 \frac{\partial f(x,y)}{\partial y} - g(x,y) \right]^2 . \tag{2.5}$$

Differentiating $E(\mathbf{p})$ with respect to $p_1$ and $p_2$ and equating the respective partial derivatives to zero, provides the necessary equations to solve for the shift parameters. The shift parameters $p_1$ and $p_2$ can be obtained by solving,

$$\begin{aligned} \mathbf{p} &= M^{-1}\mathbf{k}, \\ \text{where} \;\; M &= \begin{bmatrix} \sum f_x(x,y)^2 & \sum f_x(x,y) f_y(x,y) \\ \sum f_x(x,y) f_y(x,y) & \sum f_y(x,y)^2 \end{bmatrix} \\ \text{and} \;\; \mathbf{k} &= \begin{bmatrix} \sum g(x,y) - f_x(x,y) f(x,y) \\ \sum g(x,y) - f_y(x,y) f(x,y) \end{bmatrix} . \end{aligned} \tag{2.6}$$

In Equation 2.6, $f_x(x,y)$ and $f_y(x,y)$ refer to the horizontal and vertical spatial derivatives. A multi-scale approach is used whereby at each level, the shift parameters are propagated to the next finer level by appropriate scaling of the shift parameters. Such an approach serves to reduce the computational burden in the iterative process.

### 2.3.2 Affine displacement estimation

The translational motion model proves to be inadequate even when imaging sufficiently planar scenes. This is due to the fact that individual cameras in the $5 \times 5$ ProFUSION25 camera array are not perfectly aligned. This results in images that may differ in scale and/or by small rotations. Hence, a six parameter affine displacement model was chosen to provide better image registration. The affine model can account for shear, rotation, scale and translation. Consider again the warped image $g(x,y)$ which is related to the reference image $f(x,y)$ through the affine motion model $g(x,y) = f(p_1x + p_2y + p_5, p_3x + p_4y + p_6)$. The parameters $p_1, p_2, p_3$ and $p_4$ form the affine matrix and the parameters $p_5, p_6$ form the translation vector,

$$A = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix} \quad \text{and} \quad \mathbf{t} = \begin{bmatrix} p_5 \\ p_6 \end{bmatrix}.$$

The goal of affine parameter estimation is to estimate the parameter vector $\mathbf{p} = [p_1, p_2, p_3, p_4, p_5, p_6]^T$ which is done through minimization of the error function,

$$E(\mathbf{p}) = \sum_{x,y} \left[ f(p_1x + p_2y + p_5, p_3x + p_4y + p_6) - g(x,y) \right]^2. \tag{2.7}$$

The error function is expressed as a linear function of the parameters $\vec{\mathbf{p}}$ using a first-order Taylor series expansion given by

$$E(\mathbf{p}) = \sum_{x,y} \left[ f(x,y) + (p_1 x + p_2 y + p_5 - x) f_x(x,y) + (p_3 x + p_4 y + p_6 - y) f_y(x,y) - g(x,y) \right]^2 . \tag{2.8}$$

Equation 2.8 can be written more compactly in the form,

$$E(\mathbf{p}) = \sum_{x,y} \left( \mathbf{c}^T \mathbf{p} - k \right)^2 , \tag{2.9}$$

where the coefficient vector $\mathbf{c}$ and the scalar $k$ are defined as,

$$\mathbf{c} = \left[ x f_x(x,y) \;\; y f_x(x,y) \;\; x f_y(x,y) \;\; y f_y(x,y) \;\; f_x(x,y) \;\; f_y(x,y) \right]^T$$

and

$$k = \left[ f(x,y) - g(x,y) - x f_x(x,y) - y f_y(x,y) \right] .$$

By differentiating the error function and equating it to zero, the parameter vector $\mathbf{p}$ can be obtained by solving,

$$\mathbf{p} = \sum_{x,y} \left( \mathbf{c}\mathbf{c}^T \right)^{-1} \left( \mathbf{c} k \right) . \tag{2.10}$$

The implementation details of this affine estimation technique are provided in [10]. The authors also provide a MATLAB implementation[1] of this technique as a subroutine in their dense displacement field estimation algorithm which is discussed next.

---

[1] `http://www.cs.dartmouth.edu/farid/research/registration.html`

### 2.3.3 Dense displacement estimation technique

A recent optical flow estimation technique proposed in [10] is used in this thesis to provide dense displacement estimates. An implementation is provided online by the authors which is used in this thesis. This method models the warping between images as locally affine but globally smooth. The affine model in the previous sub-section is applied to a small region (typically a rectangular block) in the images being registered. A smoothness assumption is then applied which constrains the model parameters to vary smoothly across space. A new energy function is defined which incorporates a term like the one in Equation 2.7 and a new smoothness term. This energy function is minimized in an iterative fashion. Similar to previous approaches, a coarse-to-fine scheme is used in order to contend with larger motions. The original application of this displacement estimation technique was for medical images where it performs satisfactorily. However, the registration of two $480 \times 480$ images takes approximately 90 minutes. The algorithm was tested on a laptop with a Intel Core2 Duo T9600 processor running at a clock speed of 2.8GHz. Despite its computational complexity, we use this dense displacement estimation algorithm because of the availability of an implementation.

## 2.4 Summary

This chapter presents certain background material that is essential to the work carried out in subsequent chapters. A literature review was presented in Section 2.1 which presented some of the shortcomings in the literature that this thesis hopes to address. The observation model was developed in Section 2.2. The assumptions made in the observation model have a profound effect on the success of the super-resolution process which is repeatedly pointed out in this thesis. Finally, the displacement estimation techniques used in this thesis are presented in Section 2.3.

# Chapter 3

# Regularization methods in super-resolution

> "Far better an approximate
> answer to the right question,
> which is often vague, than an
> exact answer to the wrong
> question, which can always be
> made precise."
>
> - John W. Tukey

In this chapter, super-resolution is introduced as an ill-posed inverse problem. A brief introduction to inverse problems is given in Section 3.1. This formulation as an inverse problem is pertinent because it justifies the use of *regularization* methods to make the problem well-posed. These concepts are further expounded in Section 3.2. In Sections 3.3-3.5, the most widely used regularization methods in image processing and super-resolution literature are discussed at a level that is sufficient to demonstrate the key features of each method. Finally, chapter highlights are listed in Section 3.6.

## 3.1 Inverse problems and super-resolution

In systems theory, given knowledge about the parameters and/or the physical processes involved, the outcome of an associated measurement can be predicted. The process of predicting the result of such measurements involving known characteristics of a system is called the forward problem. The corresponding inverse problem consists of using the actual result of such a measurement to infer the values of the parameters that characterize the system [41]. The forward and inverse problems form a chicken-and-egg scenario in which the formulation of one involves all or part of the solution of the other. In certain cases, the forward model is easier to solve while in others, the inverse problem is more tractable. For example, consider the forward problem of finding the roots of a polynomial and the inverse problem of constructing the polynomial from its known roots. In this case, the forward model is the harder one. Super-resolution has traditionally been cast as a inverse problem of inferring the formative high resolution image from the observed low resolution images. The associated forward model is the imaging process that generates the low resolution observations given the high resolution image. Not only is the inverse super-resolution problem harder to solve, it is also an ill-posed problem with numerical stability issues. The definition of ill-posedness was first described by Hadamard in the field of partial differential equations [42]. A wide variety of applications involving inverse problems has shifted the focus from ill-posed problems being studied exclusively as mathematical curiosities to ones of great practical interest (early vision problems, computer tomography, etc.) [43]. A well-posed problem is one for which (a) a solution exists, (b) the solution is unique and (c) the solution depends continuously on the initial data. In other words, the solution must be stable to perturbations in its arguments. The third condition does not imply robustness. For the solution to be robust, the problem must also be numerically well-conditioned. A problem that violates any of these conditions is said to be ill-posed. A hand-waving argument for the ill-posedness of super-resolution can be made by examining each of the Hadamard conditions [39].

1. Non-existence of the solution: The presence of noise in the imaging process (given its characteristics) may result in a sequence of observed low resolution images that is inconsistent with any high resolution image. As a result, the forward model is non-invertible and a high resolution image cannot be

estimated.

2. Non-uniqueness of the solution: Any portion of the high resolution image that lies outside the band-limit of the imaging sensor (often modeled as a low-pass filter) belongs to the null space of the system operator. Since such out-of-band data could be explained by more than one high resolution image, the solution can be argued to be non-unique. Furthermore, because the number of constraints in super-resolution is typically less than the number of parameters to estimate, the system is under-determined which leads to non-unique solutions.

3. Discontinuous dependence of the solution on the data: Consider a blur operator in the forward model that asymptotically goes to zero as the spatial frequency increases. While theoretically invertible, in practice, even a small noise component at a large enough frequency will cause a large spurious signal in the reconstructed high resolution image. Such a scenario leads to stability issues.

The ill-posedness of the inverse problem is inextricably linked to the information loss that occurs as a result of the realization of the forward model because the low resolution images contain less information than the high resolution image. Regularization refers to the methods which attempt to compensate for this information loss by leveraging a priori information about the desired high resolution image. This information is termed a priori because it cannot be typically estimated from the low resolution observations. The a priori information serves to constrain the set of feasible solutions by imposing characteristics like positivity, smoothness and minimum energy on the desired high resolution image.

From an optimization point of view, regularization methods are used to 'scalarize' multi-criterion optimization problems through the use of regularization parameters. Multi-criterion problems involve objective functions that are vector-valued functions. Each component of such a vector-valued function comes from one of the multiple and often conflicting criteria used in the objective function. The regularization parameters can be thought of as weights that signify the relative importance of these different criteria. Typically, the terms in the objective function represent a trade off between conflicting resources. The subsequent minimization of a regularized problem results in a Pareto-optimal point. The choice of regularization parameter selects different Pareto-optimal points along the optimal trade-off surface. While each solution along the optimal trade-off curve has the same value for the objective function, they may physically represent drastically different solutions.

## 3.2 Regularization in an optimization framework

In [44], the super-resolution problem is shown to fall under the class of image restoration problems. Recalling the matrix-based observation model introduced in Section 2.2, the mathematical formulation of super-resolution can then be cast in the familiar form of an image restoration problem,

$$\begin{bmatrix} \mathbf{g}_0 \\ \vdots \\ \mathbf{g}_k \end{bmatrix} = \begin{bmatrix} D_0 B_0 F_0 \\ \vdots \\ D_k B_k F_k \end{bmatrix} \mathbf{f} + \begin{bmatrix} \mathbf{n}_0 \\ \vdots \\ \mathbf{n}_k \end{bmatrix} = \begin{bmatrix} H_0 \\ \vdots \\ H_k \end{bmatrix} \mathbf{f} + \mathbf{n}$$

$$\mathbf{g} = H\mathbf{f} + \mathbf{n}, \quad \text{where} \tag{3.1}$$

$$H = \begin{bmatrix} H_0 \\ \vdots \\ H_k \end{bmatrix} \quad \text{and} \quad \mathbf{n} = \begin{bmatrix} \mathbf{n}_0 \\ \vdots \\ \mathbf{n}_k \end{bmatrix}.$$

To further elucidate the notation in Equation 3.1, $\mathbf{g}$ denotes a stacked version of the lexicographic ordering for each low resolution observed image $\mathbf{g}_i$. The corresponding column vector $\mathbf{g}$ is of size $(k+1)MN \times 1$. Similarly, the matrix $H$ represents a stacking of the generative matrices $H_i$ for each observed image. The resulting matrix $H$ is of size $(k+1)MN \times m^2MN$. The column vector $\mathbf{n}$ is of similar construction. The high resolution image to be estimated is represented by $\mathbf{f}$ (the zero sub-script is dropped for convenience). This mathematical notation is introduced to conform to the conventional notation used in super-resolution and image restoration literature.

We can now attempt to solve the super-resolution problem by a straight-forward objective function that is constructed in the classical least-squares sense. This objective function represents how closely the estimated high resolution image matches the observation model. Consequently, it is also referred to as a data fidelity term because it measures the fidelity of the high resolution estimate to the observed low resolution images.

$$\begin{aligned} \mathbf{f}^{LS} &= \arg\min_{\mathbf{f}} \mathcal{J}_d\left(\mathbf{f}, \mathbf{g}\right), \\ \mathcal{J}_d\left(\mathbf{f}, \mathbf{g}\right) &= \left\|H\mathbf{f} - \mathbf{g}\right\|^2. \end{aligned} \tag{3.2}$$

In Equation 3.2, the norm $\|\cdot\|$ refers to the $L_2$ norm. When the system matrix $H$ has full column rank (the columns are linearly independent of each other), the solution to the least-squares formulation is obtained by solving the set of normal equations given by

$$\left(H^T H\right)\mathbf{f} = H^T\mathbf{g}. \tag{3.3}$$

A full column rank implies that the null-space of $H$ is empty and guarantees a unique solution to the system of equations. However, this is seldom the case and hence the problem of a non-unique solution crops up. The null-space of the system matrix $\mathcal{N}(H)$ is the set of input images which result in a zero image when passed through the imaging system (observation model) $H$. For a predominantly low-pass filtering system $H$, an example of an image belonging to the null-space $\mathcal{N}(H)$ would be an image containing frequencies in the stop band of the system $H$. In such a case, there will be an infinite number of high resolution images that produce the same set of low resolution observations. If $\mathbf{f}_p$ is a particular solution to Equation 3.3, then the optimization problem defined solely in terms of the data fidelity criterion will generate an infinite number of solutions given by

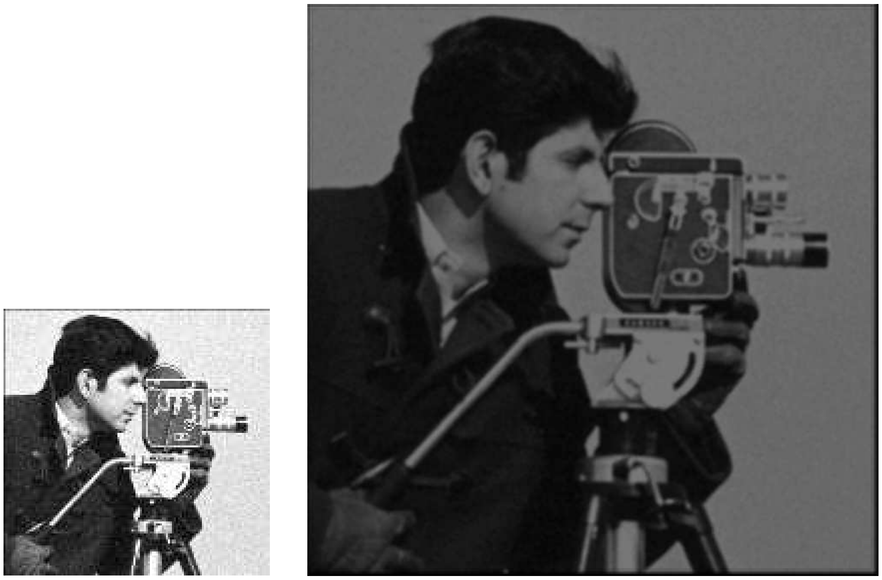
$$\mathbf{f}^{LS} = \mathbf{f}_p + \mathbf{f}_{\mathcal{N}}, \quad \text{where} \quad \mathbf{f}_{\mathcal{N}} \in \mathcal{N}(H). \tag{3.4}$$

It is often infeasible to construct the system matrix $H$ and the computational cost involved in its inversion can be prohibitive. The subsequent operations are carried out more easily through point-wise shifting for warping, convolution for blurring and interpolation/decimation for re-sampling operations. These can be combined with iterative procedures to solve the normal equation 3.3 using gradient descent as follows,

$$\begin{aligned}
\mathbf{f}_{n+1}^{\mathrm{LS}} &= \mathbf{f}_{n}^{\mathrm{LS}} - \epsilon \nabla \mathbf{f}_{n}^{\mathrm{LS}} \\
\nabla \mathbf{f}_{n}^{\mathrm{LS}} &= H^T H \mathbf{f}_{n}^{\mathrm{LS}} - H^T \mathbf{g} \\
\mathbf{f}_{n+1}^{\mathrm{LS}} &= \mathbf{f}_{n}^{\mathrm{LS}} - \epsilon \left( H^T H \mathbf{f}_{n}^{\mathrm{LS}} - H^T \mathbf{g} \right).
\end{aligned} \tag{3.5}$$

The action of the operator $H$ in the imaging process $H\mathbf{f} = DBF\mathbf{f}$ is the aggregate of first warping the high resolution estimate $\mathbf{f}$ with the operator $F$, followed by blurring with the operator $B$ and finally down-sampling with $D$. The adjoint operator of the imaging system $H^T = F^T B^T D^T$ is interpreted as first up-sampling by zero-insertion through $D^T$, blurring by the filter represented by $B^T$ and finally undoing the warping operation (backward motion) through the operator $F^T$. The blurring operation represented by $B^T$ is carried out by convolution with the space-reversed kernel $b(-\mathbf{x})$. If the original PSF $b(\mathbf{x})$ is symmetric around the origin, then $B^T$ is the same as the original convolution matrix $B$. We demonstrate in Figure 3.1 that depending on the initial estimate chosen, the iterative minimization of Equation 3.2 leads to different solutions among the solution space represented by Equation 3.4. By using two different initial estimates, a zero image and a uniformly random image, the least-squares optimization results in different high resolution estimates. In all the experiments performed in this chapter, synthetic scenarios have been used to demonstrate each concept. The experiments are carried out by first shifting a given image using only translational shifts. The original high resolution image that is used is the famous *cameraman* image. The particular version of the *cameraman* image used in this thesis was obtained by scanning (at a 600 dpi resolution) a photographic print of the original cameraman image that was provided by Dr. William F. Schreiber, Professor Emeritus, Massachusetts Institute of Technology. The shifted images are then blurred using a rectangular averaging filter of kernel size $3 \times 3$ and finally down-sampled by ideal impulse sampling to generate a set of low resolution images. A total of four low resolution images are generated through this process, each with a different sub-pixel shift on the low resolution grid. The resulting four low resolution images are then used in each of the schemes described in subsequent sections.

To deal with the non-uniqueness of the least-squares solution that is demonstrated in Figure 3.1, regularization terms, which are often referred to as smoothing terms, are imposed as constraints. Perhaps the simplest regularization term is that seeking a solution of minimum energy. It is referred to as the generalized inverse and can be interpreted as the solution $\mathbf{f}^\dagger$ having minimum energy $\|\mathbf{f}\|$ in the solution space represented by Equation 3.4. The optimization problem has the



(a) low resolution image

(b) Least-squares super-resolution result with a zero image as initial estimate (×4 magnification)

(c) Least-squares super-resolution result with a random (uniform) image as initial estimate (×4 magnification)

Figure 3.1: Non-uniqueness of the least-squares solution

form,

$$\begin{aligned} \mathbf{f}^{\dagger} &= \arg\min_{\mathbf{f}} \{\mathcal{J}_s(\mathbf{f}) \;\; \text{such that} \;\; \mathcal{J}_d(\mathbf{f},\mathbf{g}) = 0\} \\ \text{where} \;\; \mathcal{J}_d(\mathbf{f},\mathbf{g}) &= \|H\mathbf{f}-\mathbf{g}\|^2 \;\; \text{and} \;\; \mathcal{J}_s(\mathbf{f}) = \|\mathbf{f}\|^2. \end{aligned} \tag{3.6}$$

The generalized inverse solution gives equal importance to all frequency bands. However, natural images have an exponentially decaying spectrum and thus a regularizer that is able to weight different frequency bands is more useful. As we shall demonstrate in Sections 3.3, 3.4 and 3.5, the physical plausibility of the solution rather than its uniqueness is a more important concern in regularization analysis. This physical plausibility translates into an appropriately chosen smoothing term $\mathcal{J}_s(\mathbf{f})$.

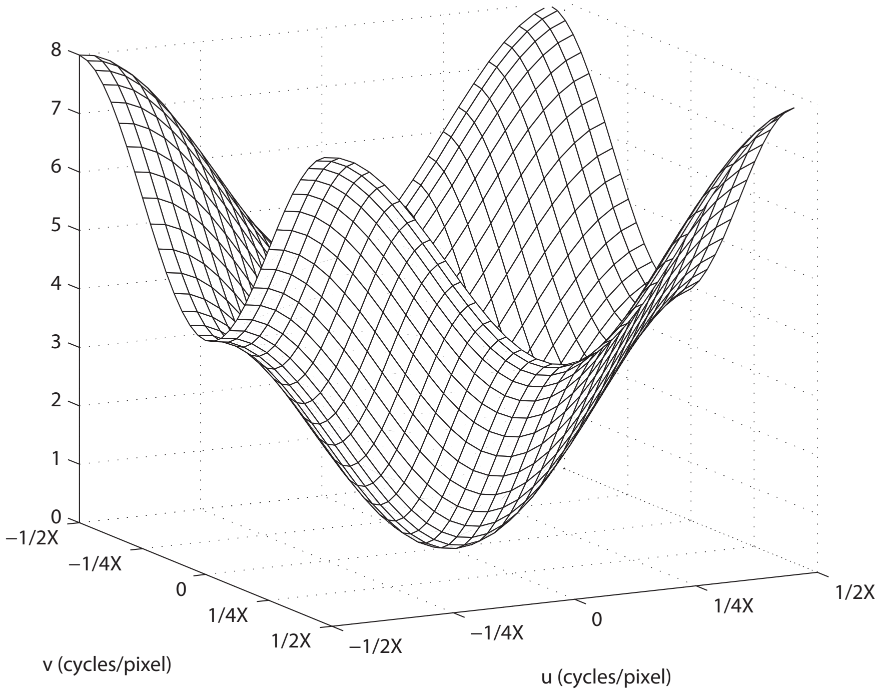
## 3.3 Tikhonov regularization

The most widely used class of regularizers are the Tikhonov regularizers having a smoothing term $\mathcal{J}_s(\mathbf{f}) = \|C\mathbf{f}\|^2$. A high-pass filter is a common choice for the operator $C$ that is used in the regularization term. In this thesis, the Laplacian filter, which serves as an approximation to the second-order derivative of the image $f(\mathbf{x})$, is used. The frequency response of the Laplacian filter and its corresponding blur kernel are shown in Figure 3.2.

Based on the high-pass operator $C$ used, Tikhonov regularization employs a specific class of so-called 'stabilizing functionals' to restrict admissible solutions to spaces of smooth functions. This reflects our expectation that natural images have minimum energy in higher frequency bands. Since noise patterns typically reside in higher frequencies, Tikhonov regularizers also serve to suppress noise and spurious high frequency artifacts. However, perceptually important edges also appear as high frequency components. Since the Tikhonov regularizer cannot differentiate

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

(a) Laplacian blur kernel



(b) Frequency response of the Laplacian filter

Figure 3.2: The Laplacian filter in the spatial and frequency domains

between noise and edges, the resulting high resolution image undergoes undesirable smoothing. An example of super-resolution using Tikhonov regularization is shown in Figure 3.3(b). The original high resolution image used is also shown in Figure 3.3(a). A reference low resolution image is shown in Figure 3.1(a). The optimization problem using Tikhonov regularization is formulated as,

$$\begin{gathered} \mathbf{f}^{\mathrm{Tik}} = \arg\min_{\mathbf{f}} \mathcal{J}_d(\mathbf{f},\mathbf{g}) + \lambda \mathcal{J}_s(\mathbf{f}), \quad \text{where} \\ \mathcal{J}_d(\mathbf{f},\mathbf{g}) = \|H\mathbf{f}-\mathbf{g}\|^2 \text{ and } \mathcal{J}_s(\mathbf{f}) = \|C\mathbf{f}\|^2. \end{gathered} \tag{3.7}$$

The minimizer of the optimization problem in Equation 3.7 is given by the normal equations,

$$H^T\mathbf{g} - (H^TH + \lambda C^TC)\mathbf{f}^{\mathrm{Tik}} = 0. \tag{3.8}$$

The parameter $\lambda$ is called the regularization parameter which allows us to control the trade-off between the stability of the solution in the presence of noise and the nearness of the regularized solution to the generalized solution in the ideal case of no noise. A large value of the regularization parameter leads to smoother images (and better noise suppression). On the other hand, a smaller regularization parameter ensures a closer match to the observed images. There are many techniques to select the optimal value of the regularization parameter which have not been discussed in this thesis. A collection of methods to select optimal values of the regularization parameter is presented in [45]. A more recent approach geared towards choosing the regularization parameters for super-resolution problems can be found in [46]. We solve Equation 3.8 iteratively using a gradient descent ap-

(a) Original high resolution image

(b) Tikhonov regularized result (×4 magnification)

Figure 3.3: Tikhonov regularized super-resolution result

proach.

$$\begin{aligned}
\mathbf{f}_{n+1}^{\text{Tik}} &= \mathbf{f}_{n}^{\text{Tik}} - \epsilon \nabla \mathbf{f}_{n}^{\text{Tik}} \\
\nabla \mathbf{f}^{\text{Tik}} &= H^T H \mathbf{f}_{n}^{\text{Tik}} - H^T \mathbf{g} + \lambda C^T C \mathbf{f}_{n}^{\text{Tik}} \qquad (3.9) \\
\mathbf{f}_{n+1}^{\text{Tik}} &= \mathbf{f}_{n}^{\text{Tik}} - \epsilon \left( H^T (H \mathbf{f}_{n}^{\text{Tik}} - \mathbf{g}) + \lambda C^T C \mathbf{f}_{n}^{\text{Tik}} \right)
\end{aligned}$$

In Equation 3.9, the subscript $n$ is used to refer to the iteration number. The algorithm is initialized with a nearest-neighbor interpolation of the reference low resolution image that we are attempting to magnify. The step-size $\epsilon$ is chosen to be small enough to converge within a suitable number of iterations without large oscillations in the value of the objective function. However, after performing Tikhonov regularization we can observe the excessive smoothing in Figure 3.3(b). Tikhonov regularization encounters serious difficulties in application to real life visual problems because global smoothness assumptions (implied by the minimization of high-frequency content) do not hold indiscriminately across visual discontinuities (depth, intensity, motion). The difficulty is in part also due to the fact that the quadratic nature of the prescribed $L_2$ norm offers no spatial control over its smoothness properties [47]. We address this shortcoming of excessive smoothing in the following sections.

## 3.4 Statistical regularization

A statistical regularization approach views the measurement noise $\mathbf{n}$ and the desired high resolution image $\mathbf{f}$ as random fields. The most widely used random field model is the Markov random field (MRF) made popular in computer vision applications by Geman and Geman [48]. The underlying principle behind Markov random fields is the conditional dependence of a pixel's intensity on a finite neighborhood around it. The super-resolution problem can then be cast into a Bayesian framework and maximum *a posteriori* (MAP) estimation of the high resolution image can be carried out. Bayesian estimation distinguishes between solutions based on an image prior model (a probability density for the image). A common assumption is that of global smoothness for the image which arises from a Gaussian prior for the high resolution image. This prior when combined with the Markov property leads to a Gauss-Markov random field (GMRF). However, such an assumption makes the appearance of sharp edges statistically unlikely, effectively suppressing the high-frequency information like Tikhonov regularization. A more plausible prior is one assuming that the high resolution image is piecewise smooth which is often the case with natural images. Images of real-world scenes correspond to smooth regions separated by sharp discontinuities. The Huber-Markov random field model (HMRF) is a prior which represents piece-wise smooth data. It was first used in super-resolution by Schultz and Stevenson [14] and in single image interpolation [49] showcasing dramatic visual improvement in resolution. The MAP estimation procedure can be mathematically represented as,

$$\begin{aligned}\mathbf{f}^{\mathrm{MAP}} &= \arg\max_{\mathbf{f}} \Pr\{\mathbf{f}|\mathbf{g}\} \\ &= \arg\max_{\mathbf{f}} \frac{\Pr\{\mathbf{g}|\mathbf{f}\}\Pr\{\mathbf{f}\}}{\Pr\{\mathbf{g}\}} \quad \text{(using Bayes theorem)} \\ &= \arg\max_{\mathbf{f}} \{\ln\Pr\{\mathbf{g}|\mathbf{f}\} + \ln\Pr\{\mathbf{f}\} - \ln\Pr\{\mathbf{g}\}\}.\end{aligned} \tag{3.10}$$

The maximization in Equation 3.10 is converted to a minimization after neglecting the term $\ln\Pr\{\mathbf{g}\}$ because it is independent of the variable over which minimization is being carried out. This minimization can be written as

$$\mathbf{f}^{\mathrm{MAP}} = \arg\min_{\mathbf{f}} \{-\ln\Pr\{\mathbf{g}|\mathbf{f}\} - \ln\Pr\{\mathbf{f}\}\}. \tag{3.11}$$

In the following discussion, we will show that the constituent terms of Equation 3.11 can be exactly interpreted as the data-fidelity term $\mathcal{J}_d(\mathbf{f}, \mathbf{g})$ and a smoothing regularization term $\mathcal{J}_s(\mathbf{f})$. The conditional probability density $\Pr\{\mathbf{g}|\mathbf{f}\}$ is referred to as the likelihood of observing the low resolution images $\mathbf{g}$ given a certain high resolution image $\mathbf{f}$. With the assumption of the noise $\mathbf{n}$ being i.i.d Gaussian distributed with a variance of $\sigma^2$, the likelihood term is represented by

$$\Pr\{\mathbf{g}|\mathbf{f}\} = \frac{1}{(2\pi\sigma^2)^{MN/2}} \exp\left(\frac{-\|H\mathbf{f} - \mathbf{g}\|^2}{2\sigma^2}\right). \tag{3.12}$$

In Equation 3.12, $M$ and $N$ refer to the width and height (in number of pixels) of the low resolution image. The high resolution image prior is represented by the Gibbs random field that is given by

$$\Pr\{\mathbf{f}\} = \frac{1}{\mathcal{Z}} \exp\left(\frac{-1}{\lambda} \sum_{c \in \mathcal{C}} V_c(\mathbf{f})\right). \tag{3.13}$$

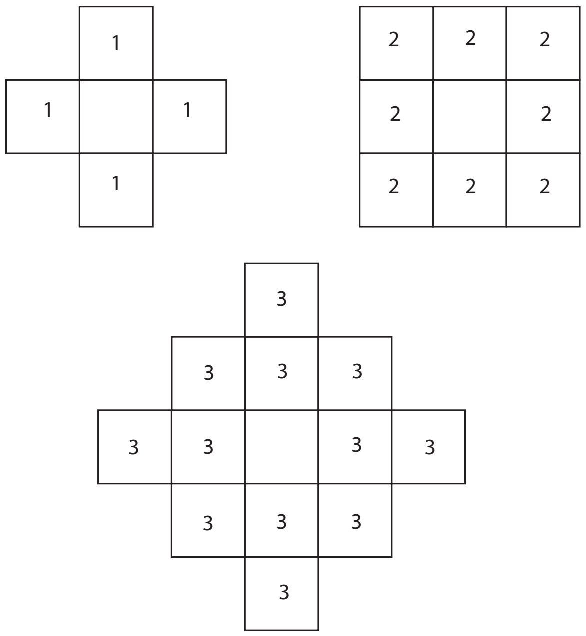


Figure 3.4: $1^{st}$, $2^{nd}$ and $3^{rd}$ order neighborhoods

The Hammersely-Clifford theorem [50] proves the equivalence of the Gibbs random field and a Markov random field. The Gibbs random field is therefore used as a practical generating distribution for the MRF because it is characterized by neighboring site interactions which lead to intuitive image models. The parameter $\mathcal{Z}$ represents a normalizing constant also called the partition function, $\lambda$ is the temperature parameter of the density (also plays the role of the regularization parameter) and $V_c(\mathbf{f})$ is the potential function of a clique $c$. A clique is defined as a single site or a set of sites that are all neighbors of each other [51] which are used to measure some local property of the image. The summation in Equation 3.13 is carried out over the set of all cliques $\mathcal{C}$ in the image. The Markovian property of decoupling future from the past is extended to two dimensions through the familiar notion of neighborhoods. A neighborhood is denoted as either first, second or higher orders based on the number of pixel neighbors that are considered. A pictorial depiction is shown in Figure 3.4. Substituting Equation 3.12 and 3.13 in the objective function and ignoring the terms independent of the high resolution image $\mathbf{f}$, we arrive at an equation that consists of the data fidelity term $\mathcal{J}_d(\mathbf{f}, \mathbf{g}) = \|H\mathbf{f} - \mathbf{g}\|^2$ and a smoothing term $\mathcal{J}_s(\mathbf{f}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{f})$.

$$\mathbf{f}^{\mathrm{HMRF}} = \arg\min_{\mathbf{f}} \frac{1}{2\sigma^2}\|H\mathbf{f} - \mathbf{g}\|^2 + \frac{1}{\lambda}\sum_{c \in \mathcal{C}} V_c(\mathbf{f}). \tag{3.14}$$

In the particular choice of the Huber-Markov random field as the image prior, the clique potentials are a function of second-order directional derivatives which provide a measure of smoothness (or lack of smoothness). The prior density is represented by,

$$\Pr\{\mathbf{f}\} = \frac{1}{\mathcal{Z}} \exp\left(\frac{-1}{\lambda}\sum_{c \in \mathcal{C}} \rho(\mathbf{d}_c^T \mathbf{f})\right) \tag{3.15}$$

The coefficient vector $\mathbf{d}_c$ represents a spatial activity measure for each clique. Depending on the type of clique (classified as horizonal, vertical, diagonal and anti-diagonal), the spatial activity measure for that clique is a discrete approximation to a second-order directional derivative. These directions are shown in Figure 3.6.

When using a Gauss-Markov random field, the function $\rho(\cdot)$ is quadratic in nature which leads to a convex function that has a unique global minimum. However, the quadratic nature results in blurring of the edges that is very similar in spirit to that occurring in Tikhonov regularization. The second-order derivatives are small in smooth image regions but have a large magnitude at edges. The quadratic 'penalty' function drives these large magnitudes even higher thus severely penalizing the occurrence of edges. It is then more suitable to select a function that does not penalize large derivatives by a large amount. The Huber function is given by

$$\rho(x) = \begin{cases} x^2 & |x| \leq T \\ T^2 + 2T(|x| - T) & |x| > T. \end{cases} \tag{3.16}$$

The Huber function is parameterized by the parameter $T$ which is the threshold at which the behavior of the function changes from quadratic to linear. For small arguments, the Huber function remains quadratic but at the threshold $T$, the Huber function becomes linear. A linear behavior does not penalize larger derivatives (edges) excessively and hence serves to preserve edges in the high resolution image. The Huber function and the quadratic penalty function are compared in Figure 3.5. The use of the Huber function has two major implications on the optimization process:

1. The Huber function is convex in nature and hence the corresponding objective function remains convex.
2. The decreased sensitivity to large derivatives (outliers) allows discontinuities to be preserved in the high resolution estimated image.

The clique potentials in the smoothing term $\mathcal{J}_s(\mathbf{f}) = \sum_{c \in \mathcal{C}} \rho(\mathbf{d}_c^T \mathbf{f})$ employ second-order derivatives along the horizontal, vertical, diagonal and anti-diagonal directions. Each of these derivatives require interactions between three pixels in the

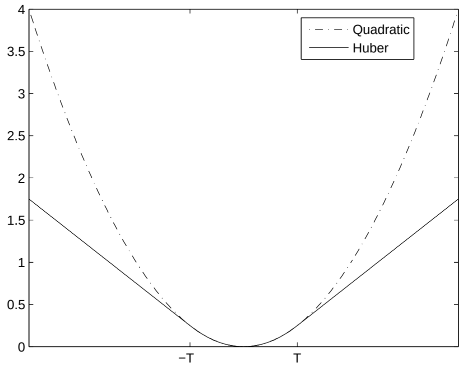


Figure 3.5: The Huber penalty function parameterized by the threshold $T$

chosen direction. The cliques involved and their directions are shown in Figure 3.6.

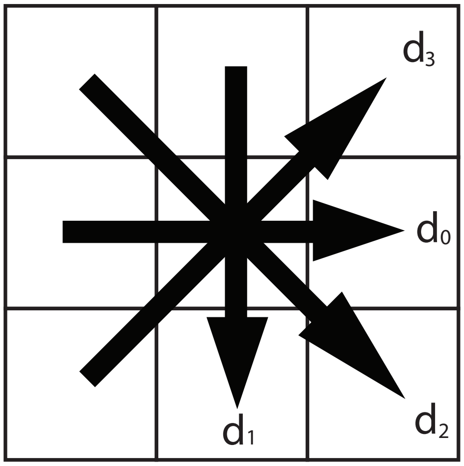


Figure 3.6: The four cliques and their associated directions

The second order derivatives are calculated using finite difference approximations (convolution with an equivalent $3 \times 3$ filter kernel). The minimization is carried out by the gradient descent method which results in an update equation given by

$$\mathbf{f}_{n+1}^{\mathrm{HMRF}} = \mathbf{f}_{n}^{\mathrm{HMRF}} + \epsilon \left( -\nabla \mathcal{J}_s(\mathbf{f}_{n}^{\mathrm{HMRF}}) - \frac{\lambda}{\sigma^2} H^T (H\mathbf{f}_{n}^{\mathrm{HMRF}} - \mathbf{g}) \right). \tag{3.17}$$

Following the same notation used before, $\epsilon$ refers to the step-size used in the update and the subscript $n$ denotes the iteration number. Notice that the super-script has been changed from $\mathbf{f}^{\text{MAP}}$ to $\mathbf{f}^{\text{HMRF}}$ to reflect the use of the Huber-Markov random field as a prior image. A super-resolution using the same synthetic scenario involving the *cameraman* image is shown in Figure 3.8. It can be observed that the HMRF is able to preserve edges to a greater degree than Tikhonov regularization. However, the HMRF-based regularization suffers from certain drawbacks identified below that make it less attractive [52].

1. The optimization process depends on three parameters viz., the regularization parameter $\lambda$, the noise variance $\sigma^2$ and the threshold of the Huber function $T$.

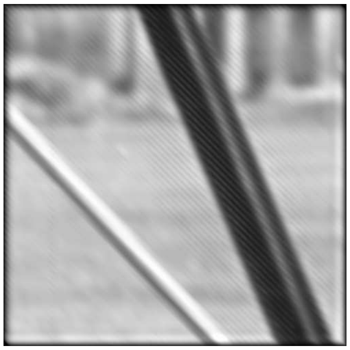
2. The HMRF prior is unable to preserve texture well because it reduces the efficacy of the Markov random field in modeling such regions.

3. The smoothing afforded by the four clique measures is not adequate to prevent smoothing across edges. This effect is demonstrated in Figure 3.7 which shows the independent effect of each clique measure (by setting the others to zero) on an arbitrarily oriented edge. We can observe that with the exception of the clique representing the anti-diagonal direction in Figure 3.7(d), all the other clique measures in Figures 3.7(a)-(c) perform smoothing across the edge. The result using all the cliques involves a weighted sum of contributions from each clique measure. Thus smoothing across edges inevitably occurs even when using the HMRF prior. These are the primary motivations leading to the use of a variational regularizer which is discussed in the next section.

## 3.5 Variational regularization

Variational regularization is derived from applications in the related fields of thermodynamics, diffusion and the heat equation. It is also involved in the theory of

(a) Clique 0 Contribution (b) Clique 1 Contribution (c) Clique 2 Contribution



(d) Clique 3 Contribution

Figure 3.7: Contributions of the different cliques in smoothing

Figure 3.8: Huber-Markov random field regularized super-resolution result (×4 magnification)

minimal surfaces in pure mathematics and in engineering applications like elasticity and fluid dynamics [53]. The use of variational regularizers in image processing was initiated by [54] for edge detection and in [55] for noise removal. Since those landmark papers, variational regularizers have been widely used in image restoration, segmentation and motion estimation. More recently, they have been used in [52] for regularized image up-sampling with a new observation model. The demonstrated success of the variational regularizer in such applications motivates its use in super-resolution. The various variational regularizers used in the literature have a unified form for the smoothing term $\mathcal{J}_s(f)$, which is given by

$$\mathcal{J}_s(f) = \int_{\Omega} \mathcal{L}(|\nabla f|) d\Omega. \tag{3.18}$$

Equation 3.18 assumes that the high resolution image is continuous in nature and is defined over the region $\Omega$, which is a rectangular bounding box in $\mathbb{R}^2$. The functional $\mathcal{L}(\cdot) \geq 0$ is an increasing function of $\nabla f$, where $\nabla$ represents the gradient operator. The derivative of the function $\mathcal{L}'(\cdot) > 0$ and $\mathcal{L}'(\nabla f) = 0$ when $\nabla f = 0$. A discussion of some of the more successful functions $\mathcal{L}(\cdot)$ used in literature can be found in [56]. We choose to use the function $\mathcal{L}(|\nabla f|) = |\nabla f|$ in this thesis which is in the spirit of *total variational* regularization. However, we continue the following discussion with a general function $\mathcal{L}(\cdot)$.

The energy functional $\mathcal{J}_s(f)$ can be minimized (similar to steepest descent) by moving in the negative direction of the gradient through the following partial differential equation, which uses an artificial time parameter $t$,

$$\frac{\partial f}{\partial t} = -\nabla \mathcal{J}_s(f) = \text{div}\left(\mathcal{L}'(|\nabla f|)\frac{\nabla f}{|\nabla f|}\right). \tag{3.19}$$

The gradient of the functional $\mathcal{J}_s(f)$ is shown to be equal to $-\text{div}\left(\mathcal{L}'(|\nabla f|)\frac{\nabla f}{|\nabla f|}\right)$ in [56]. The desired high resolution image can be recovered through the minimization of the energy functional given by

$$\mathcal{E}(f) = \int_{\Omega} \mathcal{L}(|\nabla f|) d\Omega + \frac{\lambda}{2}\|H\mathbf{f} - \mathbf{g}\|^2. \tag{3.20}$$

If $\mathcal{E}(f)$ has a unique minimizer, then it satisfies the Euler-Lagrange equation,

$$H^T\mathbf{g} - H^T H\mathbf{f} + \lambda \operatorname{div}\left(\frac{\mathcal{L}'\left(|\nabla f|\right)}{|\nabla f|}\nabla f\right) = 0. \tag{3.21}$$

Equation 3.21 can be expanded by developing the divergence term. This is done by decomposing the divergence term using local image structures [57]. In particular, the normal and tangential directional vectors to isophote curves in the image are used. Isophotes are constant intensity curves in the image. An example is shown in Figure 3.9. Only three levels have been used in Figure 3.9(b), yet a large part of the image structure is observable.

At each point $\mathbf{x} = (x, y) \in \mathbb{R}^2$, we define the vectors $N(\mathbf{x}) = \frac{\nabla f(\mathbf{x})}{|\nabla f(\mathbf{x})|}$ and $T(\mathbf{x}), \|T(\mathbf{x})\| = 1$ to be orthogonal to $N(\mathbf{x})$. Using the notation $f_x, f_{xx}$ to represent first and second derivatives of the image $f$ in the subscripted directions, the Euler-Lagrange equation 3.21 is expressed as

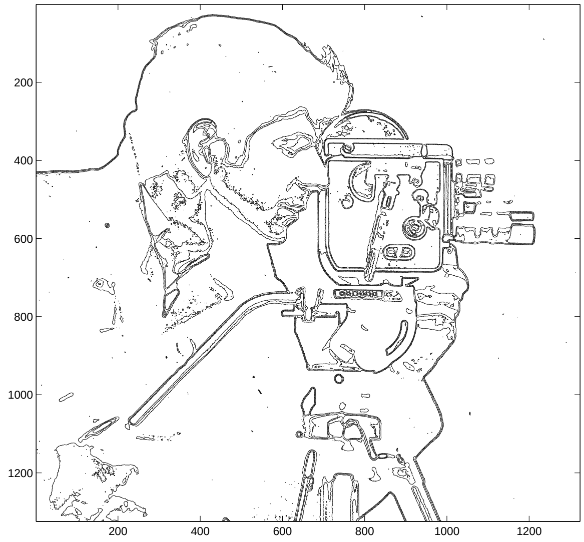
$$H^T\mathbf{g} - H^T H\mathbf{f} + \lambda \operatorname{div}\left(\frac{\mathcal{L}'\left(|\nabla f|\right)}{|\nabla f|} f_{TT} + \mathcal{L}''\left(|\nabla f|\right) f_{NN}\right) = 0. \tag{3.22}$$

The second order directional derivatives $f_{NN}$ and $f_{TT}$ are given by

$$\begin{aligned} f_{NN}(\mathbf{x}) &= \frac{1}{|\nabla f|^2}\left(f_x^2 f_{xx} + f_y^2 f_{yy} - 2 f_x f_y f_{xy}\right) \\ \text{and} \quad f_{TT}(\mathbf{x}) &= \frac{1}{|\nabla f|^2}\left(f_x^2 f_{yy} + f_y^2 f_{xx} + 2 f_x f_y f_{xy}\right). \end{aligned} \tag{3.23}$$

The decomposition in Equation 3.22 is useful because the action of the operators in the tangential and normal directions is more clear. The quantity $f_{TT}$ is the

(a) Original image



(b) 3 level iso-intensity plot

Figure 3.9: Constant intensity curves of an image

second order directional derivative that is tangential to the isophote lines. It can also be interpreted as being normal to the gradient $\nabla f$ at $\mathbf{x}$. Similarly, the quantity $f_{NN}$ is the second order directional derivative normal to the isophote line which can be viewed as being along the gradient $\nabla f$. The argument of the function $\mathcal{L}$ is the $L_1$-norm of the gradient of the high resolution image. A small gradient is characteristic of noise or textured region, while a large gradient indicates the presence of an edge. We are then faced with the following two scenarios.

1. When the variation in image intensity is weak (small gradients), smoothing should be encouraged equally in all directions. Such an isotropic smoothing condition is obtained by imposing the conditions that $\mathcal{L}'(0) = 0$, $\lim\limits_{s \to 0^+} \dfrac{\mathcal{L}'(s)}{s} = \lim\limits_{s \to 0^+} \mathcal{L}''(s) = \mathcal{L}''(0) > 0$. Thus, at those points where $|\nabla f|$ becomes small, Equation 3.22 becomes,

   $$H^T \mathbf{g} - H^T H \mathbf{f} + \lambda \, \mathcal{L}''(0) \, (f_{TT} + f_{NN}) = 0. \tag{3.24}$$

   However, since $f_{TT} + f_{NN} = f_{xx} + f_{yy} = \Delta f$,

   $$H^T \mathbf{g} - H^T H \mathbf{f} + \lambda \, \mathcal{L}''(0) \, (\Delta f) = 0. \tag{3.25}$$

   Equation 3.25 represents an equation having strong regularizing properties in all directions [57]. It behaves similar to the Tikhonov regularization detailed in Section 3.3.

2. The presence of an edge is indicated by a large change in image intensity and a correspondingly large image gradient. To preserve the edge, it is preferable to diffuse along the edge (in the tangential direction $T(\mathbf{x})$) rather than blurring it by diffusing across the edge. In order to accomplish this, it is sufficient to ensure that the coefficient of the derivative in the normal direction

$f_{_{NN}}$ vanishes in Equation 3.22 and that the coefficient of the derivative in the tangential direction $f_{_{TT}}$ remains finite.

The decomposition in Equation 3.22 demonstrates that the diffusion process behaves like an energy dissipation in two orthogonal directions. The diffusion of the intensity values of $f(\mathbf{x})$ across the image isophotes (along the spatial gradient) leads to blurred edges in the image which is an undesirable property. On the other hand, the diffusion along the image isophotes has an edge-preserving effect along with smoothing of unwanted texture and noise. Diffusion in an orthogonal direction to the gradient can be viewed as edge-directed filtering without any prior computation of edge-maps [52]. As a result, the intensity and location of the edges is preserved while smoothing along them and maintaining their crisp quality.

We choose the function $\mathcal{L}(|\nabla f|) = |\nabla f|$ put forward by Rudin, Osher and Fatemi in [55] which is often termed the ROF model in image processing literature. Such a choice leads to the *total variation* norm which represents the diffusion process given by

$$\frac{\partial f}{\partial t} = \operatorname{div}\left(\frac{\nabla f}{|\nabla f|}\right) = \kappa. \tag{3.26}$$

Equation 3.26 defines the mean curvature $\kappa$ of $f(\mathbf{x})$. The use of the total variation norm defines a space of high resolution images having bounded total variation in which discontinuities are permissible. As mentioned before, this is advantageous because natural images comprise of smooth regions separated by sharp edges. In fact, the space of functions of bounded variation has been described in [55] as the proper space for performing many image processing tasks.

To solve the optimization problem in Equation 3.20, we use mean curvature evolution which relies on the theory of level set methods. We refer the reader to [58] for a review of level set methods. Osher and Sethian [58] developed the level set method (LSM) to analyze the motion of a front in two or three dimensions. The goal of level set methods is to compute and analyze the motion of a front under a velocity field. This velocity can be dependent on a number of factors like position, time and the geometry of the front [59]. In our case, the front is represented by the constant intensity curves or *isophotes* of the image. The underlying principle of level set methods when applied to image processing is to propagate these fronts with different types of motion. Depending on the characteristics of the motion that the fronts are subject to, we can obtain effects such as [52];

1. straightening out oscillatory contours
2. shrinking isolated noise patterns
3. preserving the boundaries of small objects and so on.

The formulation of the problem using Equation 3.20 and choosing $\mathcal{L}(|\nabla\mathbf{f}|) = |\nabla\mathbf{f}|$ is given by

$$\mathbf{f}^{TV} = \arg\min_{\mathbf{f}} \int_{\Omega} |\nabla\mathbf{f}| d\Omega + \frac{\lambda}{2}\|H\mathbf{f} - \mathbf{g}\|^2. \tag{3.27}$$

This minimization problem is solved by the steady-state solution to the partial different equation arising from the Euler-Lagrange equation 3.21,

$$\frac{\partial f}{\partial t} = \kappa - \lambda H^T(H\mathbf{f} - \mathbf{g}), \tag{3.28}$$

where $\kappa$ is the mean curvature defined in Equation 3.26 and $\lambda$ is a regularization parameter. Equation 3.28 can then be converted into the following iterative process [52],

$$\mathbf{f}_{n+1}^{TV} = \mathbf{f}_n^{TV} + \epsilon\left(\kappa - \lambda H^T(H\mathbf{f}_n^{TV} - \mathbf{g})\right). \tag{3.29}$$

The step-size $\epsilon$ is also called an artificial time-step. The numerical implementation of the partial differential equations in the mean curvature term $\kappa$ is carried out using second-order central differences which is suggested in [60]. The mean curvature term is given by

$$\kappa = \frac{f_{xx}f_y^2 - 2f_xf_yf_{xy} + f_{yy}f_x^2}{(f_x^2 + f_y^2)^{3/2}}. \tag{3.30}$$

The second-order partial derivatives are approximated using central differences which are given by,

$$\begin{aligned}
f_x(x,y) &= \frac{f(x+1,y) - f(x-1,y)}{2}, \\
f_y(x,y) &= \frac{f(x,y+1) - f(x,y-1)}{2}, \\
f_{xx}(x,y) &= f(x+1,y) - 2f(x,y) + f(x-1,y), \\
f_{yy}(x,y) &= f(x,y+1) - 2f(x,y) + f(x,y-1), \\
f_{xy}(x,y) &= \frac{f(x-1,y-1) + f(x+1,y+1) - f(x-1,y+1) - f(x+1,y-1)}{4}.
\end{aligned} \tag{3.31}$$

The second-order derivatives in Equation 3.31 can be conveniently calculated by convolution with the corresponding kernels,

$$\begin{aligned}
h_x &= \begin{bmatrix} 0.5 & 0 & -0.5 \end{bmatrix}, \\
h_y &= \begin{bmatrix} 0.5 & 0 & -0.5 \end{bmatrix}^T, \\
h_{xx} &= \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}, \\
h_{yy} &= \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}^T, \\
h_{xy} &= \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}.
\end{aligned} \tag{3.32}$$

We have given a mathematical treatment of the various regularizers in this

Figure 3.10: Total-variation (TV) regularized result (×4 magnification)

chapter for the sake of completeness and to understand their properties and impact on super-resolution. The mechanism of the super-resolution process can be understood when it is viewed in a graphical (flow-chart) form as shown in Figure 3.13. This particular flow-chart represents a single iteration in the iterative total-variation based super-resolution algorithm. Using the total-variation regularizer, the same sequence of *cameraman* images has been super-resolved in Figure 3.10. As can be observed, the result is sharper than the Tikhonov and HMRF based regularization. Furthermore, note the preservation of texture in the cameraman's hair as opposed to the HMRF-regularized result. For ease of comparison, we have included the different regularizers used as well as single image bi-cubic and nearest neighbor interpolation in Figure 3.12 to demonstrate the visual increase in resolution. An objective image quality metric that is often used is the peak signal-to-noise ratio (PSNR). The PSNR requires the availability of an original high resolution image to compare with the super-resolution result. It is mathematically

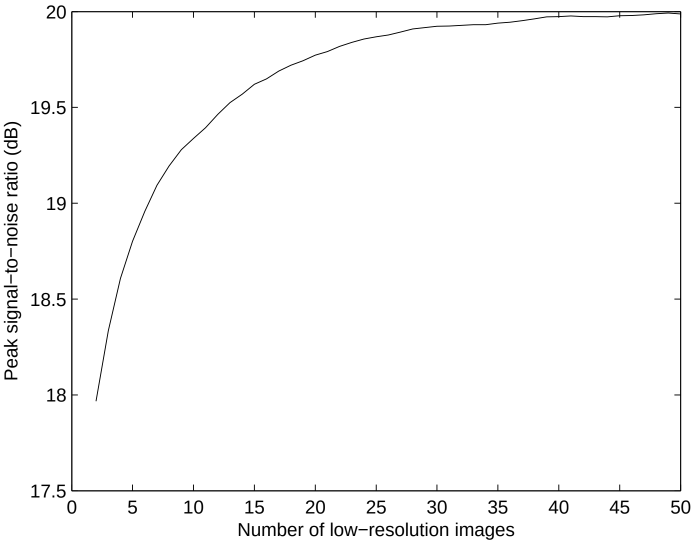


Figure 3.11: Super-resolution performance (PSNR) as a function of the number of low resolution (LR) images

expressed as

$$\begin{aligned} \text{PSNR} &= 10\log\left(\frac{1}{\text{MSE}}\right), \\ \text{where } \ \text{MSE} &= \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\|f(x,y)-g(x,y)\|^2. \end{aligned} \tag{3.33}$$

In Equation 3.33, $M$ and $N$ refer to the number of pixels in the horizontal $(x)$ and vertical $(y)$ image axes. The image intensity values have been assumed to be normalized between $0$ and $1$. Using the PSNR value, the effect of the number of low resolution images on super-resolution performance can be quantified numerically. A characteristic graph using the total variation regularizer is shown in Figure 3.11. This graph shows a steep increase in performance which begins to saturate after a certain number of low resolution images is used. This is due to the decrease in useful non-redundant information that each low resolution image contributes to the high resolution image.

(a) Bi-cubic Interpolation (b) Nearest Neighbor Interpolation

(c) Tikhonov Regularization (d) HMRF Regularization

(e) TV Regularization

Figure 3.12: Comparison of the different super-resolution results

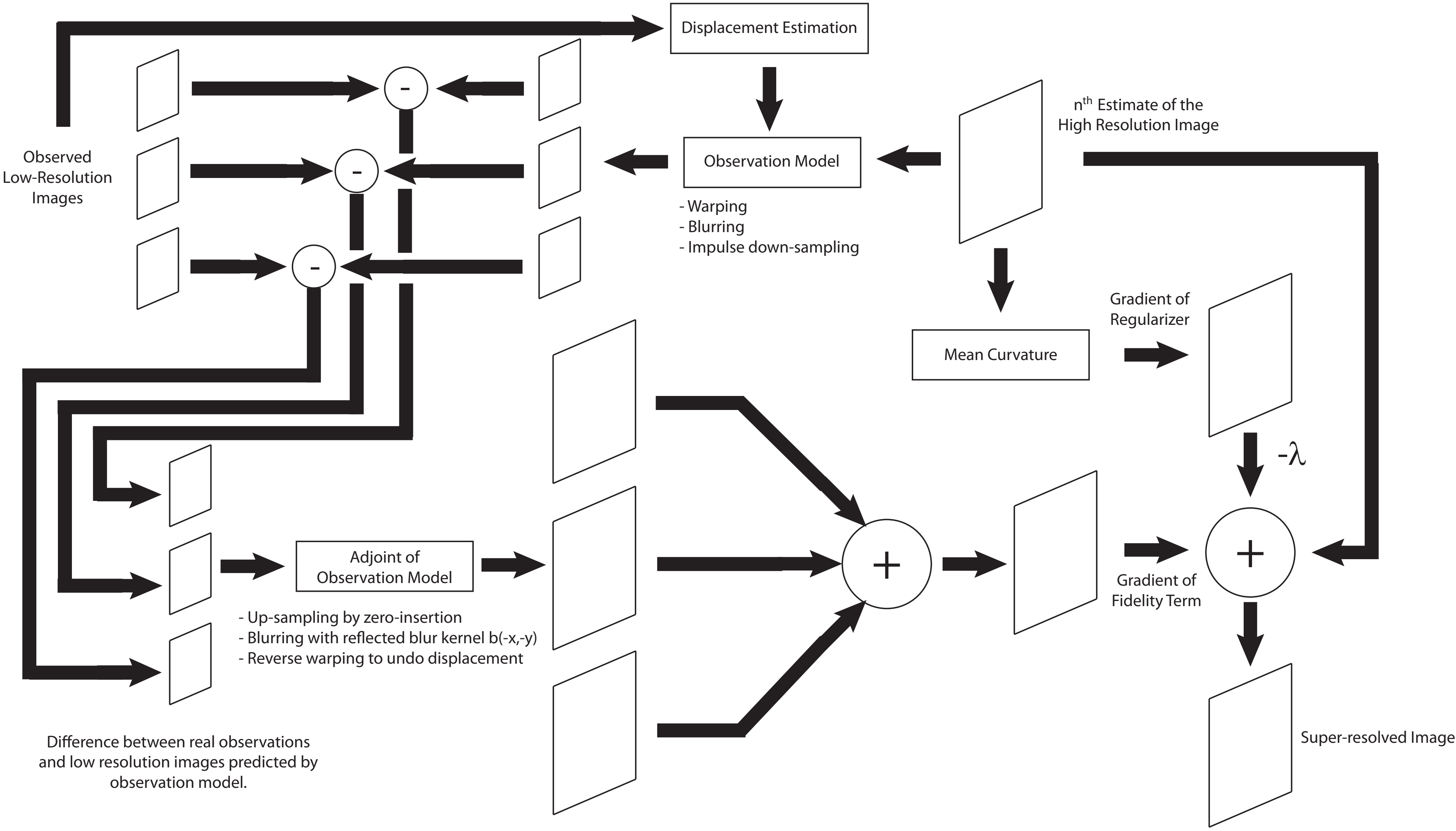


Figure 3.13: A single iteration in regularized super-resolution restoration

## 3.6 Summary

The discussion in this chapter is mainly aimed towards formulating super-resolution as an inverse problem and justifying the need for regularization. From an observation model point of view, the most physically plausible prior (or regularizer) is investigated. The impact of some popular regularizers on super-resolution are demonstrated. The results show that the total-variation regularizer results in the most noticeable resolution enhancement. In light of this observation, all subsequent results in this thesis use the total-variation regularizer unless otherwise stated.

# Chapter 4

# Characterizing the ProFUSION25 camera array and its implication on super-resolution

> "The best material model of a cat is another, or preferably the same, cat."
>
> - Norbert Weiner

Most of the recent advances in super-resolution performance can be attributed to the use of increasingly specific prior models for the high resolution image. The performance of three specific priors, viz., a Tikhonov regularizer, an HMRF prior and the total-variation regularizer, were demonstrated in Chapter 3. However, we feel that current literature has paid little attention towards characterizing the imaging sensor which can play a major role in the super-resolution process. Existing papers on super-resolution tend to focus on better regularization methods to achieve incremental gains in performance. This approach is myopic because the problem is not considered as a whole. The oft-cited paper by Baker and Kanade [16] demonstrates the use of a learnt prior as the regularizer which can 'break' the super-resolution limit at high magnifications. They argue that a general smoothness prior, like the ones used in this thesis, tends to overly smooth the image at higher magnifications. This has sparked an increase in the number of techniques utilizing learnt priors for different categories of images like faces in specialized applications. However, rather than taking such an approach, we ask the question whether knowing more about the imaging process itself would be useful in improving super-resolution performance. The techniques employed rarely account for the different operations in the imaging pipeline that implicitly occur during the capture of the scene. Such operations in the imaging pipeline can adversely affect various computer vision tasks if they are not accounted for properly. A review of the typical components in an imaging pipeline can be found in [61].

In the following discussion, a specific photometric distortion introduced by the ProFUSION25 camera called *vignetting* is first examined in Section 4.1. A photometric distortion (as opposed to a geometric distortion) results in pixel intensities that are not constantly proportional to the luminance of objects in the scene. The implications of vignetting on super-resolution performance are also examined in Section 4.2. After correcting for the vignetting, the point spread function (PSF) of the ProFUSION25 is identified in Section 4.3; this is shown to significantly impact super-resolution performance in Section 4.4. Finally, a summary of the chapter is presented in Section 4.5.

## 4.1 Vignetting effects in super-resolution

Vignetting refers to the radial fall-off of intensity from the center of an image. It can be observed as the gradual darkening of the image towards the corners. Vignetting effects can prove to be problematic for a variety of applications. It can affect graphics applications in which sequences of images are stitched together, such as image mosaicing, image-based rendering, etc. It also affects vision applications which use the brightness constancy assumption and intensity-based correlation methods to recover scene structure like optical flow [62]. Since super-resolution involves both estimating the displacement between images and fusing the image sequence into a higher-resolution image, vignetting effects prove to be twice implicated in the detrimental performance of super-resolution algorithms: once in the estimation of displacements and next in combining the low resolution images. This motivates us to examine its effect on super-resolution and develop methods to correct for it.

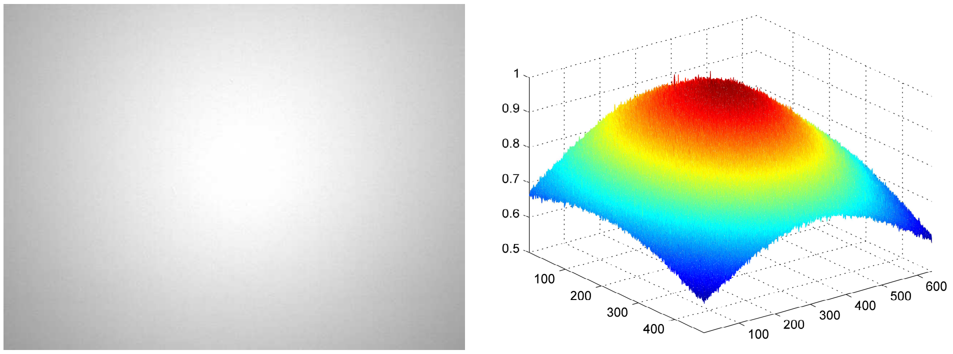
The four main sources of vignetting as categorized in [62] are:

1. **Natural Vignetting**: refers to radial intensity fall-off due to optical geometry. While the irradiance is proportional to the radiance, it decreases as the fourth power of the cosine of the angle that a ray makes with the optical axis. This angle changes as the focal distance is varied. It is usually assumed to be the dominant vignetting effect present.
2. **Pixel Vignetting**: refers to radial fall-off due to angle dependence of digital sensors. Light that enters the camera at a right angle to the camera plane produces a stronger signal than that entering at oblique angles. This type of vignetting is characteristic mainly of digital cameras which have photon wells of finite depth in digital sensors. This causes light striking at steeper angles to be partially occluded by the sides of the well [62].
3. **Optical Vignetting**: refers to radial fall-off due to light rays blocked in the lens body by the lens diaphragm. It is also referred to as physical vignetting or artificial vignetting and is a function of aperture width. As the aperture size is decreased, the effect of optical vignetting decreases because a smaller aperture limits light equally at the center and the edge of the image [62].
4. **Mechanical Vignetting**: is a less common effect of intensity fall-off due to other camera elements blocking light paths. Such elements can include filters or hoods attached to the front of the lens body.

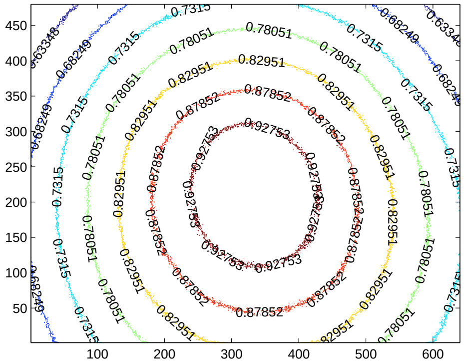
Vignetting can thus arise from any of the above mentioned sources. Several researchers have attempted to estimate models or vignetting functions that can explain the radial fall-off of intensity. However, the various sources of vignetting do not facilitate estimation of a simple vignetting function. As a result, we first perform a naive calibration to estimate a vignetting function that is expected to best explain the intensity fall-off. The estimated function (called the prototype vignetting function) is considered to be the ground truth for practical purposes.

The prototype vignetting function is obtained by capturing an image of a sufficiently *matte* paper card using the ProFUSION25 camera array. A matte surface refers to one that has little shine, reflectiveness or gloss. The matte card represents a homogenous white background that if imaged under ideal conditions should appear to be uniformly white. However, with the ProFUSION25 camera, it was observed that imaging such a background led to a non-uniform image with the characteristic radial intensity fall-off shown by vignetting effects. A real image taken with one of the 25 cameras in the ProFUSION25 sensor is shown in Figure 4.1(a). It can be immediately observed that the center of the image is brighter than the periphery. These calibration images obtained by imaging the matte card are first filtered by a 3x3 gaussian filter for de-noising purposes. This filtered version of the calibration image is stored as the prototype of the vignetting effect. A representative example amongst the prototype vignetting functions obtained is displayed as a surface plot in Figure 4.1(b) and as a contour plot in Figure 4.1(c). Every image that is captured by the ProFUSION25 array is assumed to have its intensity value scaled by the prototype vignetting function. The vignetting effect is removed by scaling a particular image with the inverse of the prototype using a pixel-by-pixel operation.

While such a simplistic calibration is able to eliminate the vignetting satisfactorily, it is useful to attempt to fit a model to the vignetting function. Natural vignetting is assumed to have the most dominant effect and a cosine-fourth model is used to fit the vignetting function. While this model may not be the most appropriate fit to the ground truth, it is still able to eliminate most of the vignetting and also gives a significant visual performance increase over the case where vignetting is neglected.



(a) Vignetting effect from the ProFUSION camera (b) Surface plot of the vignetting response



(c) Contour plot of the vignetting response

Figure 4.1: The prototype vignetting function

A number of vignetting models are used in the literature to parametrize the characteristic radial intensity fall-off. In particular, the cosine-fourth model and the Kang-Weiss model have been used in both single-image and multiple-image vignetting estimation [63, 64].

While working with the ProFUSION25 camera array, it was observed that the center of the vignetting function, i.e., the point at which the intensity is brightest, does not correspond to the geometric center of the image. Moreover, each of the 25 cameras in the array displayed vignetting functions with different centers. As a result, the center of the vignetting function $(x_v, y_v)$ is also included as a paramater to be estimated in addition to the parameters of the vignetting function $v(r)$. Since the vignetting function is assumed to be radially symmetric, the conventional Cartesian notation $v(x, y)$ is replaced by the radial distance $v(r)$, where $r = \sqrt{(x - x_v)^2 + (y - y_v)^2)}$. A simple cosine-fourth model [65] is sufficient to fit the vignetting response obtained by imaging the matte paper card. The cosine-fourth model is represented as,

$$v(r) = \alpha \cos^4 \left( \tan^{-1} \left( \frac{r}{F_c} \right) \right) + (1 - \alpha). \tag{4.1}$$

The vignetting function affects the image intensity values through,

$$f(x, y) = v(x, y) i(x, y). \tag{4.2}$$

In Equation 4.2, $i(x, y)$ refers to the undistorted image that is free from vignetting and $f(x, y)$ refers to the distorted image that is affected by vignetting. When $i(x, y)$ is a white paper card and the intensity values have been scaled in the range $[0, 1]$ ($0$ denotes black and $1$ denotes white), then the distorted imaged $f(x, y) = v(x, y)$ is the same as the vignetting function. To find the parameters of the model that best fits the prototype function, the problem is formulated as a least-squares minimization problem that is given by

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\alpha}} \|v_{\boldsymbol{\alpha}}(x,y) - f(x,y)\|^2. \tag{4.3}$$

The vignetting function has been parametrized by a four element parameter vector $\boldsymbol{\alpha} = \begin{bmatrix} \alpha & F_c & x_v & y_v \end{bmatrix}$. These elements represent the scaling factor $\alpha$, the focal length of the camera in pixels $F_c$ and the center of the vignetting function $(x_v, y_v)$ which may be different from the image center. The minimization of Equation 4.3 is carried out using the simplex search method [66] (with the MATLAB routine *fminsearch*). Once the parameter set $\boldsymbol{\alpha}$ is estimated, the vignetting function $v(x,y)$ is generated by the cosine-fourth model in Equation 4.1 and an image free from the vignetting effect is obtained by

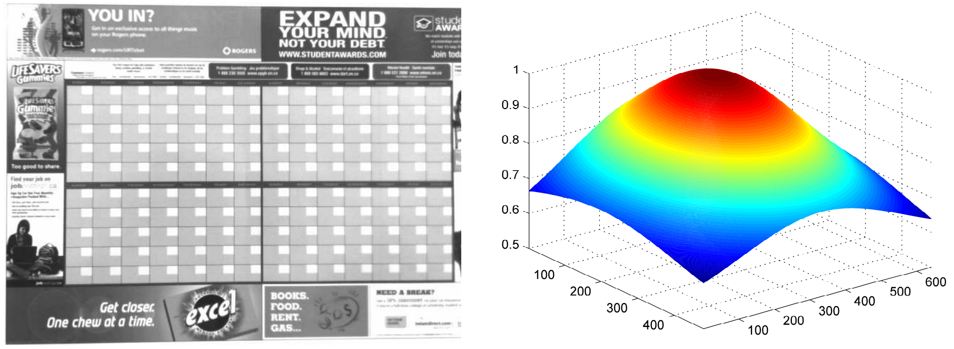
$$i(x,y) = \frac{f(x,y)}{v(x,y)}. \tag{4.4}$$

The operation in Equation 4.4 is carried out in a per-pixel fashion where each pixel of the image $f(x,y)$ is scaled by the corresponding value in $\frac{1}{v(x,y)}$. A real image (called the *Calendar* image) taken by the ProFUSION25 array is shown in Figure 4.2(a). This image suffers from vignetting as can be seen from the characteristic bright center and darker peripheries. The *de-vignetted* image using the calibration response is shown in Figure 4.2(b) while the *de-vignetted* image using the estimated cosine-fourth model is shown in Figure 4.2(c).

To demonstrate the fit to the calibration response, the estimated vignetting response is also plotted as a surface plot in Figure 4.2(d) and a contour plot in Figure 4.2(e). The amount of model match can be seen from how close the shape and characteristics of these plots appear to match those of the prototype in Figure 4.1(b) and Figure 4.1(c) respectively. Furthermore, the *de-vignetted* image using either the calibration or the estimated response appear to be the same, at least perceptually.

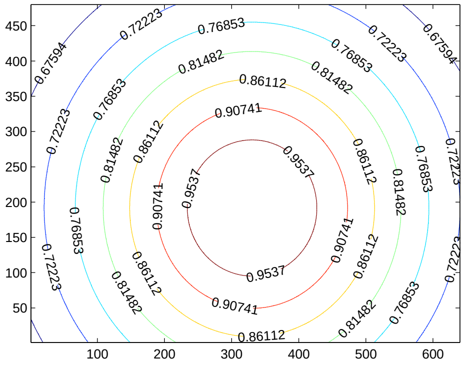
(a) Original *calendar* image suffering from vignetting

(b) *De-vignetted* image using the calibrated response



(c) *De-vignetted* image using the estimated response

(d) Surface plot of the vignetting response



(e) Contour plot of the vignetting response

Figure 4.2: Fitting a cosine-fourth model to the ProFUSION25 vignetting function

Any mismatch between the calibrated and estimated reponse is atttributed to errors in finding the center of the vignetting function. An incorrect center of the vignetting affects the value of all pixel intensities due to the radial dependence of the vignetting function. The vignetting effect is assumed to occur in the forward model before the blurring operator. A diagrammatic placement of the vignetting is shown in Figure 4.3.

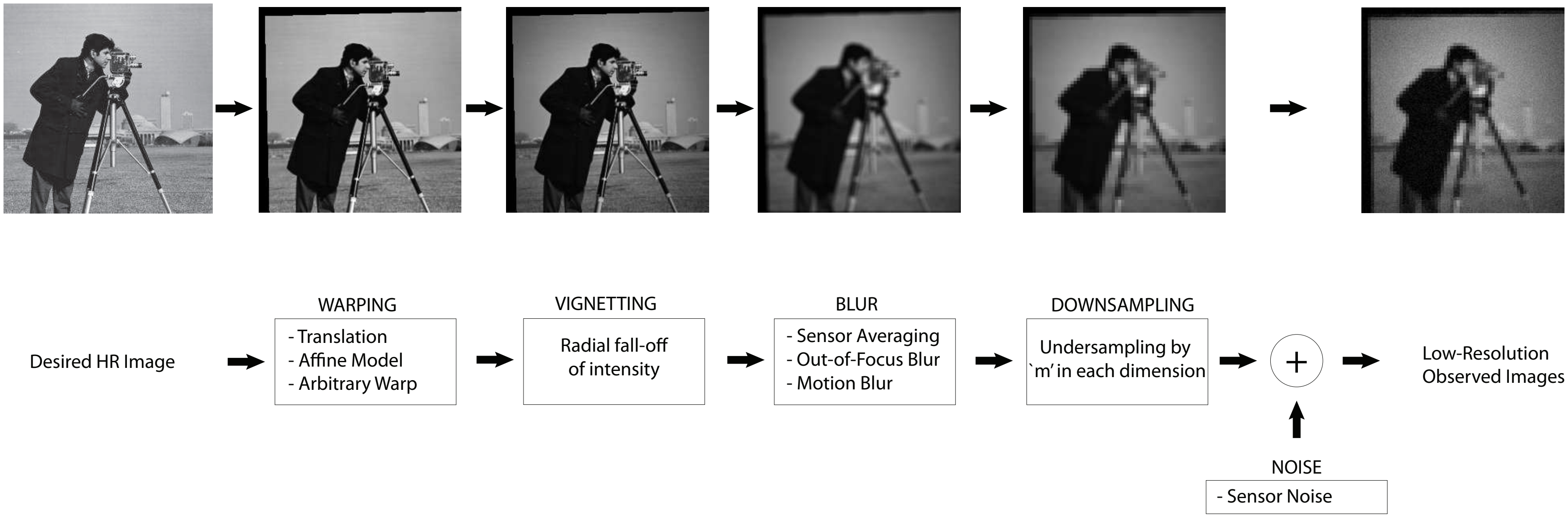


Figure 4.3: Vignetting in the forward model

## 4.2 Implications of vignetting in the super-resolution process

Super-resolution can be broken down into a displacement estimation and an image fusion part. Due to the displacement that takes place between each multi-view image, vignetting appears as a photometric mis-match (intensity distortion). The most commonly employed displacement estimation algorithms use a similarity measure that depends on comparing pixel intensities. Since vignetting causes the brightness constancy assumption to be violated, the accuracy of the displacement estimates is expected to be negatively affected. The assumption of constant brightness is used in displacement estimation to refer to the premise that an object is imaged with the same pixel intensity even when its position changes. If a translational shift is assumed to relate the low resolution images, the accuracy of the displacement estimation can be measured by the standard deviation of the error in the horizontal and vertical shift parameters. We demonstrate the implication of vignetting on displacement estimation by the following experimental procedure. A test image is first shifted and then distorted by vignetting. The shift (in horizontal and vertical directions) is then estimated using the shift estimation technique [40] described in Chapter 2. The estimated shift is compared to the known shift that was artificially introduced. This process is performed for 100 randomly selected shifts (ranging from $-50$ to $50$ pixels). The results of the experiment are tabulated in Table 4.1. The mean error $(\mu_x, \mu_y)$ in each case is very close to zero which implies that there is little bias in the estimation of the shifts. When no vignetting is introduced, the standard deviation of the displacement estimation error $(\sigma_x, \sigma_y)$ is significantly lower. More noticeably, when vignetting is introduced together with the blurring present in the forward model, the standard deviation of the error increases even further. This leads to two conclusions:

1. Vignetting does affect displacement estimation negatively by itself but the

problem is compounded with the introduction of blurring. It is prudent to point out that the error in the displacement estimates arising from vignetting alone does not constitute too much of a problem for super-resolution purposes. While super-resolution demands sub-pixel accuracy in displacement estimation, the vignetting-induced errors are of the order of a tenth of a pixel, and as such are not very significant.

2. The increase in the standard deviation of the error is far more significant (of the order of half a pixel) when vignetting is considered along with blurring. This increase is mostly attributed to the blurring process itself which negatively affects displacement estimation. In the next section, the identification of the correct blur kernel is examined, which is motivated from this experimental outcome.

Table 4.1: Error analysis of displacement estimates in images that are blurred and/or affected by vignetting

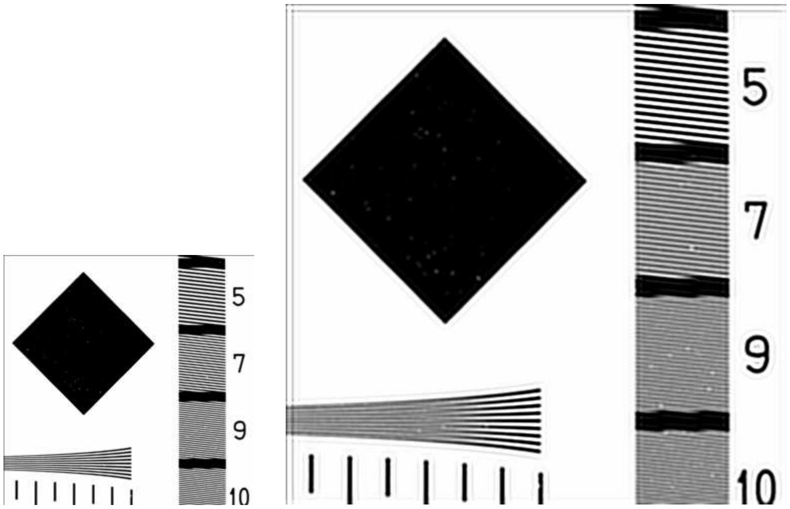
| | $\sigma_x$ | $\sigma_y$ | $\mu_x$ | $\mu_y$ |
|---|---|---|---|---|
| No vignetting | 0.0069 | 0.0048 | -0.0005 | -0.0017 |
| With vignetting only | 0.0942 | 0.0985 | 0.0015 | -0.0034 |
| With blurring only | 0.3322 | 0.3171 | 0.0011 | -0.0441 |
| **With vignetting and blurring** | **0.4042** | **0.3464** | -0.0891 | -0.0903 |

While we can conclude that vignetting alone does not significantly affect the accuracy of the displacement estimates, it does affect the perceived quality of the super-resolution result as shown in Figure 4.2(c). If the vignetting effect is not removed prior to super-resolution, the radial intensity fall-off remains in the super-resolved image, detracting from the perceptual quality of the image. The super-resolution result in Figure 4.2(c) was obtained by using four low resolution images that were corrupted by a vignetting function generated using the cosine-fourth model. In Chapter 1, the definition of resolution enhancement was stated to include an "improvement in perceived content" which can only be achieved when the vignetting effect is suitably removed. In the following discussion, the vignetting effect has been removed from all the ProFUSION25 images using the methods discussed in Section 4.1.

## 4.3 Identifying the blurring operator

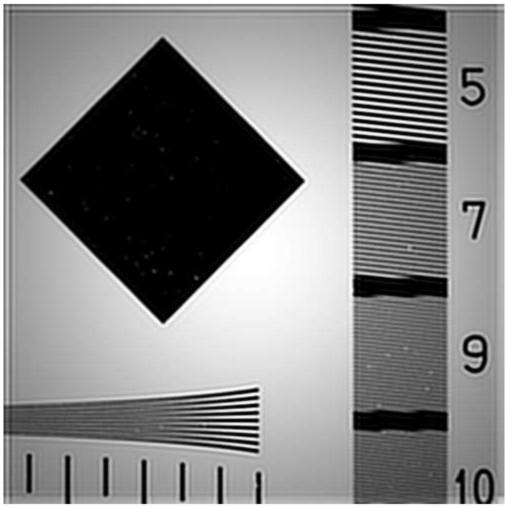
In Chapter 2, the observation model which includes the blurring operator was introduced. For a digital camera like the ProFUSION25 sensors, the most likely point spread function is either a rectangular averaging blur or a circularly symmetric out-of-focus blur or a combination of both effects that we refer to as a combination blur in this thesis. The rectangular averaging blur models the spatial averaging that takes place when incident light is integrated over the available sensor area. The out-of-focus blur models the blurring effects due to the optical lens used in the camera. The out-of-focus blur causes a point source to be imaged as a circular disk of radius $R$. The radius $R$ depends on the focal length and aperture number of the lens as well as the distance between the camera and the object being imaged [67]. The combination blur represents both these effects and can be mathmatically obtained by performing a two-dimensional convolution of the averaging blur and the out-of-focus blur. Representative examples of these blurs as both discrete-space kernels and their corresponding frequency response magnitudes are shown in Figure 4.5.

Most super-resolution algorithms assume that the blurring PSF is known. While this is a reasonable assumption in synthetic scenarios, it is often not applicable when dealing with a sensor like the ProFUSION25 camera array. When the dominant blurring effect is that of sensor averaging, the rectangular averaging kernel is assumed as the blurring kernel. In cases where the object being imaged is out-of-focus, the out-of-focus blur is assumed as the blurring kernel. However, in most



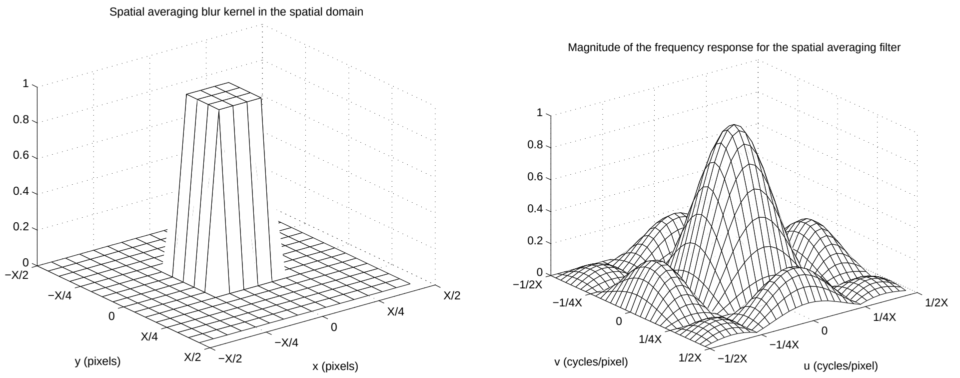
(a) low resolution image

(b) Super-resolution in the absence of vignetting (×4 magnification)



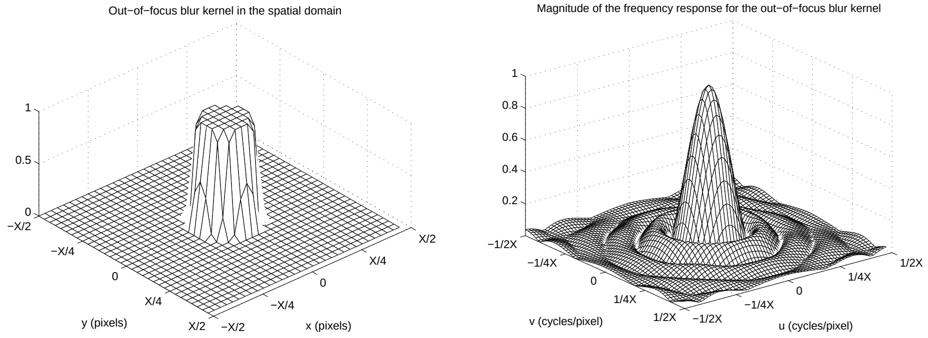
(c) Super-resolution in the presence of vignetting (×4 magnification)

Figure 4.4: Super-resolution in the presence of vignetting



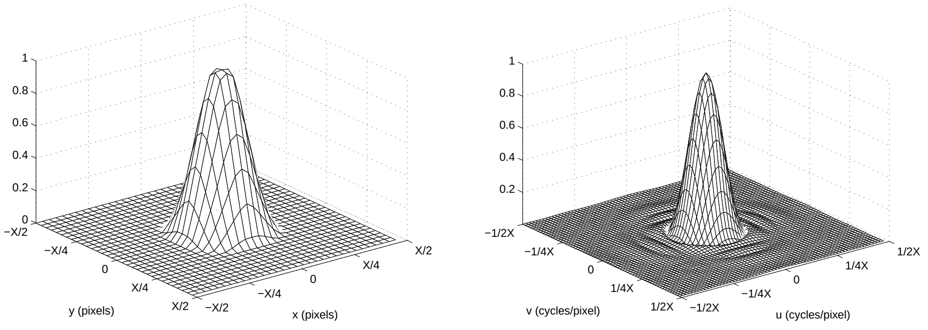
(a) Blur kernel representing spatial averaging (a 'rect' function)

(b) Frequency response of the rect function



(c) Blur kernel representing out-of-focus (a 'circ' function)

(d) Frequency response of the circ function



(e) Blur kernel representing combination blur

(f) Frequency response of the combination blur

Figure 4.5: Common blurring kernels and their frequency response

cases, the dominant blurring effect is not known and furthermore there may be other effects such as camera shake that introduce other blurring effects. We use the words blur kernel and point spread function interchangeably in this chapter and the remainder of this thesis since our implementations are carried out in discrete-space.

The simplest method of identifying the blur of the camera would be to take an image of an ideal point source of light. The corresponding image of the point source is exactly the point spread function of the camera. However, when using a digital camera, the effect of light integration over a finite sensor area may lead to an incorrect estimate of the PSF. Moreover, an ideal point source is difficult to construct. As a result, the image processing community has used the slanted edge [68] technique successfully in many applications. In this technique, an image is taken of a slanted step edge. The step edge is slightly slanted to increase the sampling rate of the edge. The corresponding image of the slanted step edge is called the edge spread function (ESF). By differentiating a one dimensional profile perpendicular to the edge, the line spread function (LSF) can be obtained. The LSF represents the cross-section of the two-dimensional PSF in a given direction. This method typically assumes that the PSF is circularly symmetric and that the two-dimensional PSF can be obtained from the one dimensional LSF. Other published work on blur identification revolved around the regular pattern of zero-crossings in the frequency response of the point-spread function [69, 70, 71]. The structure of the zero-patterns characterizes the type and degree of blur. However, such an approach limits the type of blur to the class of motion blurs or out-of-focus blurs. Additionally, such frequency-domain methods are ineffective in the presence of higher noise levels [72]. More recently, maximum likelihood (ML) estimation techniques have been used by assuming that the latent unblurred image can be described using the 2D auto-regressive model [67]. Such an approach includes a larger class of blur types. Nonetheless, ML estimation techniques suffer from the restriction of stationarity placed on the image by the auto-regressive model as well as the assumption of the noise being white and Gaussian in nature. In addition, the log-likelihood function involved is highly non-linear and has many local maxima which makes the optimization process computationally difficult.

The recovery of the blur kernel from a single blurred image is an inherently ill-posed problem [73]. The observed blurred image alone does not provide enough information to recover an arbitrary PSF. Many combinations of unblurred images and PSF can be convolved together to give the same blurred image. As a result, all the techniques mentioned above employ parametric constraints on the type of blur to disambiguate the solutions [73]. Another type of parametrization is possible by constraining the size of the blur kernel. For example, the blur kernel may be assumed to exist on a $5 \times 5$ support, which requires the estimation of 25 parameters. Such a parametrization does not restrict the type of blur. We use such an approach in this thesis. The blur kernel is assumed to be constrained to a finite support of size $N \times N$.

The blurring process, as commonly modeled in image restoration problems, is given by

$$g(\mathbf{x}) = b(\mathbf{x}) * f(\mathbf{x}) + n(\mathbf{x}). \tag{4.5}$$

In Equation 4.5, $g(\mathbf{x})$ refers to the blurred image, $f(\mathbf{x})$ refers to the sharp or unblurred image, $b(\mathbf{x})$ represents the blurring kernel and $n(\mathbf{x})$ is the observation noise. The $(*)$ operation represents a two-dimensional convolution operation given by,

$$f(\mathbf{x}) * b(\mathbf{x}) = \sum_{\tau \in \Omega_b} b(\boldsymbol{\tau}) f(\mathbf{x} - \boldsymbol{\tau}), \tag{4.6}$$

where $\Omega_b \subset \mathbb{R}^2$ is the support of the blur kernel $b(\mathbf{x})$. In non-blind image restoration, the PSF is typically assumed to be known and the unblurred image $f(\mathbf{x})$ is required to be estimated. The deconvolution process involved faces many difficulties due to the ill-conditioning of the blur operation. As a result, regularization techniques (discussed in Chapter 3) are often employed to solve the restoration problem. Using a Tikhonov regularization scheme, the following cost function $\mathcal{E}(f)$ is minimized to obtain the unblurred image:

$$\mathcal{E}(f) = \frac{1}{2}\sum_{\mathbf{x}\in\Omega_f}\left[g(\mathbf{x}) - b(\mathbf{x}) * f(\mathbf{x})\right]^2 + \frac{\lambda}{2}\sum_{\mathbf{x}\in\Omega_f}\left[c(\mathbf{x}) * f(\mathbf{x})\right]^2. \tag{4.7}$$

In Equation 4.7, $\lambda$ is a regularization parameter and $c(\mathbf{x})$ is a high-pass filter like the Laplacian filter used in Chapter 3. The summation is carried out over $\Omega_f \subset \mathbb{R}^2$, which is the support of the image. However, the task at hand is not to estimate the unblurred image $f(\mathbf{x})$. Instead, we are faced with a symmetric problem to that represented by Equation 4.7, in which the unknown PSF is to be identified, and the sharp image is assumed to be known. This is in contrast to the image deconvolution problem in Equation 4.7 where the blur kernel is known but the sharp image is to be estimated. The dilemma we are faced with when using the ProFUSION25 camera to capture a scene, is that the unblurred image is not available. A calibration experiment detailed below is devised to work around this problem.

A target calibration image is constructed having shapes with edges in different orientations. This image is shown in Figure 4.7(a). The calibration image is printed using a laserjet printer and then captured by the ProFUSION25 camera array. Consequently, both the sharp and blurred versions of the image are now available. The blurred image $g(\mathbf{x})$ is the image of the printed calibration image. The sharp image $f(\mathbf{x})$ is available as a vector graphic which was created using Adobe Illustrator. The symmetric problem of estimating the unknown PSF using a regularization approach is carried out in a fashion similar to Equation 4.7. The corresponding cost function is given by

$$\mathcal{E}(b) = \frac{1}{2} \sum_{\mathbf{x} \in \Omega_f} \left[ g(\mathbf{x}) - b(\mathbf{x}) * f(\mathbf{x}) \right]^2 + \frac{\lambda}{2} \sum_{\mathbf{x} \in \Omega_b} \left[ c(\mathbf{x}) * b(\mathbf{x}) \right]^2. \tag{4.8}$$

This cost function is minimized using steepest descent with the gradient of the cost function given by

$$\frac{\partial \mathcal{E}(b)}{b} = p(\mathbf{x}) = -\left[ g(\mathbf{x}) - b(\mathbf{x}) * f(\mathbf{x}) \right] * f(-\mathbf{x}) + \lambda \left[ c(\mathbf{x}) * b(\mathbf{x}) \right] * c(-\mathbf{x}). \tag{4.9}$$

We have included the regularization term in Equation 4.8 to show the similarity between image deconvolution and blur identification. The implementation used in this thesis does not include a regularization term when identifying the blur. The blur kernel is constrained to lie on a finite support of size $N \times N$ which corresponds to the estimation of $N^2$ parameters. The number of blur kernel coefficients $N^2$ are far fewer than the number of constraints available from a typical ProFUSION25 image of size $640 \times 480$ pixels. This results in an over-determined system for which a least-squares estimate, without the regularization term in Equation 4.8, is sufficient. However, we have used hard constraints in each iteration of the least-squares estimation. The blur kernel is normalized so that the sum of all the coefficients is one. Also, the kernel is forced to have quadrantal symmetry. A blur kernel $b(x, y)$ is said to possess quadrantal symmetry if,

$$b(x, y) = b(-x, y) = b(x, -y) = b(-x, -y).$$

In order to verify the validity of the least-squares estimation procedure, we used a synthetic scenario where the calibration image $f(\mathbf{x})$ is blurred using a $5 \times 5$ gaussian blur kernel ($\sigma = 3$) after adding Gaussian noise of zero mean and variance $\sigma^2 = 0.001$. The blur kernel $b(\mathbf{x})$ is shown in Figure 4.6(a). Using the approach outlined above, the recovered PSF $\hat{b}(\mathbf{x})$ is shown in Figure 4.6(b). The recovered PSF matches the actual PSF quite closely. However, as mentioned before, we do not have access to the sharp image when dealing with the ProFUSION25 camera array. The difficulty with the proposed calibration method lies in the fact that the captured image does not correspond exactly to the sharp calibration target. The calibration target has been geometrically and photometrically transformed in the image capture process. Typically, the captured image is geometrically related to the original sharp image through a scaling, rotation and translation. This observation allows us to fit an affine homography relating the sharp and the blurred image. The image is captured under laboratory conditions and care is taken in keeping the ProFUSION25 camera parallel to the printed paper so that perspective effects can be avoided. The affine homography is estimated using the gradient-based affine displacement estimation technique described in Chapter 2. To recover the blur kernel, the affine transformation must be accounted for by a registration procedure and only then can the least-squares estimation approach be applied to the blur identification task. However, the registration procedure invariably contains errors and the images cannot be aligned perfectly. More importantly, the vignetting effect also introduces a photometric mis-match between the pixel intensities which cannot be fully removed. This imperfect geometric and photometric registration causes the blur identification task to converge to a kernel that does not result in a satisfactory result after subsequent deblurring. An image that has been deblurred by using a kernel recovered by the described regularization process is shown in Figure 4.9. To demonstrate the geometric misalignment, the sharp image and the blurred image are shown as a pseudo-color image in different color channels. The misalignment can be clearly seen in the edge areas of Figure 4.8. The extent of the problem does not end at the geometric mis-alignment. The sharp image is an artificially constructed black and white image. However, the captured image has also been affected with the vignetting mentioned before as well as other unidentified effects which cause variations in intensity between the sharp and blurred images that cannot be explained by the blurring process. Since the blur identification task essentially relies on solving a set of equations relating pixels in the sharp image to those in the blurred image, these effects cause inaccuracies in the blur identification task. The deblurred results shown in this chapter have been obtained using the Lucy-Richardson method [74]. This was implemented in MATLAB using the *deconvlucy* routine. A dampening parameter is used to restrict iterative updates at pixels deviating from their original value by a certain threshold. This threshold was chosen experimentally to reduce ringing.

$$\begin{bmatrix} 0.0318 & 0.0375 & 0.0397 & 0.0375 & 0.0318 \\ 0.0375 & 0.0443 & 0.0469 & 0.0443 & 0.0375 \\ 0.0397 & 0.0469 & 0.0495 & 0.0469 & 0.0397 \\ 0.0375 & 0.0443 & 0.0469 & 0.0443 & 0.0375 \\ 0.0318 & 0.0375 & 0.0397 & 0.0375 & 0.0318 \end{bmatrix}$$

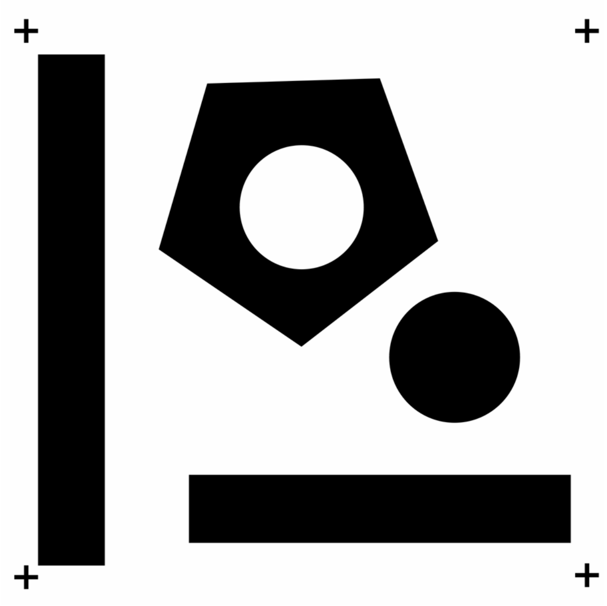
(a) The correct blur kernel

$$\begin{bmatrix} 0.0314 & 0.0384 & 0.0402 & 0.0384 & 0.0314 \\ 0.0377 & 0.0451 & 0.0430 & 0.0451 & 0.0377 \\ 0.0390 & 0.0475 & 0.0501 & 0.0475 & 0.0390 \\ 0.0377 & 0.0451 & 0.0430 & 0.0451 & 0.0377 \\ 0.0314 & 0.0384 & 0.0402 & 0.0384 & 0.0314 \end{bmatrix}$$

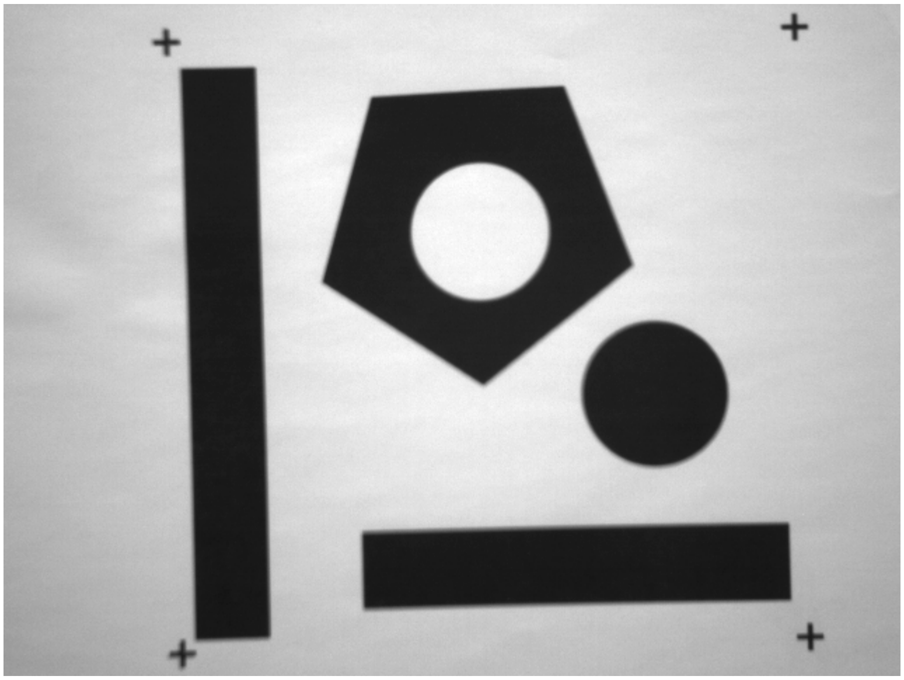
(b) The identified blur kernel

Figure 4.6: Regularized blur kernel identification

The factors mentioned above led us to choose an alternative blind deconvolution procedure that does not rely on knowing the sharp image before-hand to identify the blur. We chose to use a state-of-the-art algorithm by Shan et al. [1] to perform the blur identification task. The algorithm proposed in [1] uses a unified probabilistic model of both blur kernel identification and non-blind image restoration. It employs a maximum *a posteriori* (MAP) estimation approach to recover the



(a) The sharp target used to identify the blur



(b) The image of the target using the ProFUSION25 camera array

Figure 4.7: The sharp and blurred calibration image pair

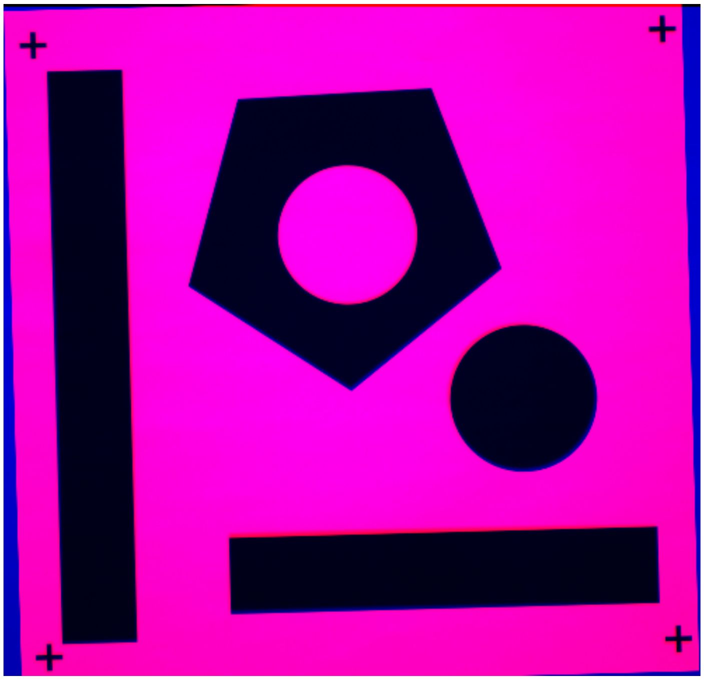


Figure 4.8: Super-imposition of the target and registered captured image to demonstrate mis-alignment (seen near edges)

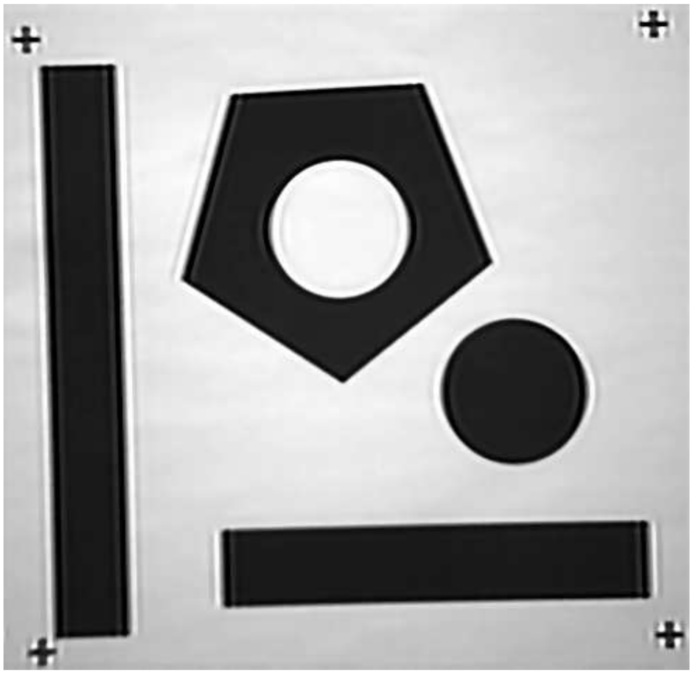


Figure 4.9: The deblurred result using the kernel identified by the regularized calibration approach

sharp, unblurred image. The deblurred version of the calibration target in Figure 4.7(a) is shown in Figure 4.10. This result shows some improvement over the result in Figure 4.9 using the least-squares identification method. The algorithm [1] is purported to perform better that other published techniques because of the use of a novel prior for the noise term which takes into account its spatial randomness. Additionally, the prior for the unblurred image also includes a term that reduces ringing artifacts. The algorithm [1] was originally developed for motion blur artifact removal occurring during camera shake. However, the results of the deblurring procedure on images taken with the ProFUSION25 camera show that it is extensible towards more general blur identification tasks. We refer the reader to [1] for further details on the algorithm. An implementation has been made available online by the authors which was used in this thesis[1]. The result of using the blur identification algorithm in [1] is demonstrated in Figures 4.11(b) and 4.12(b) which show significant sharpening. The identified blur kernels are stored and used in the subsequent super-resolution part. The results obtained by first identifying the blurs are significantly better than those obtained without any blur identification. We show super-resolution results with both synthetic and real scenarios in the next section. A comparison of the frequency response of the identified blurs using both the least-squares and the blind deconvolution approach is provided in Figure 4.13. The frequency response of the identified blurs is not similar to any of the likely blurs shown in Figure 4.5. However, a common element between the identified blurs in Figure 4.13(a) and 4.13(b) is the significant energy in the sidelobes. These hint that perhaps the real camera blurs are not spatially invariant. In retrospection, looking back at the image in Figure 4.7(b), each of the four '+' markers have been blurred in a different manner. This also indicates that the real camera blur is indeed spatially variant. Nonetheless, the assumption of a spatially-invariant

---

[1] `http://www.cse.cuhk.edu.hk/~leojia/projects/motion_deblurring/index.html`

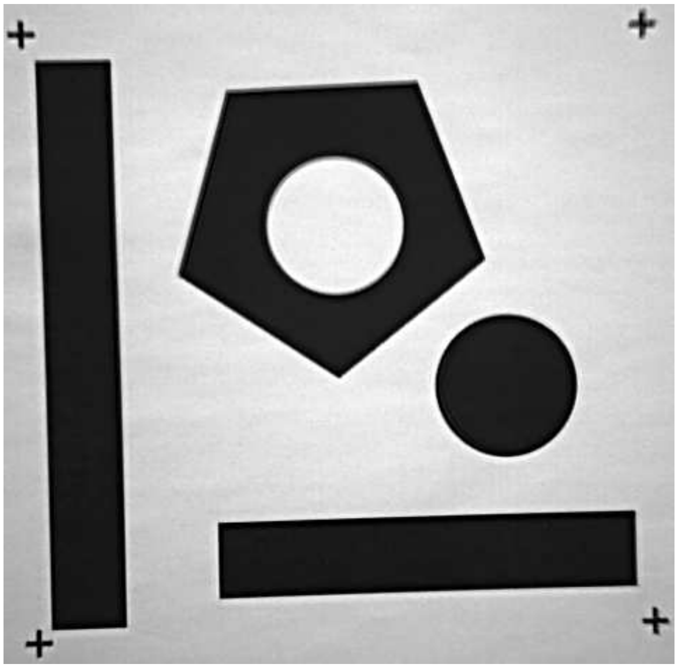


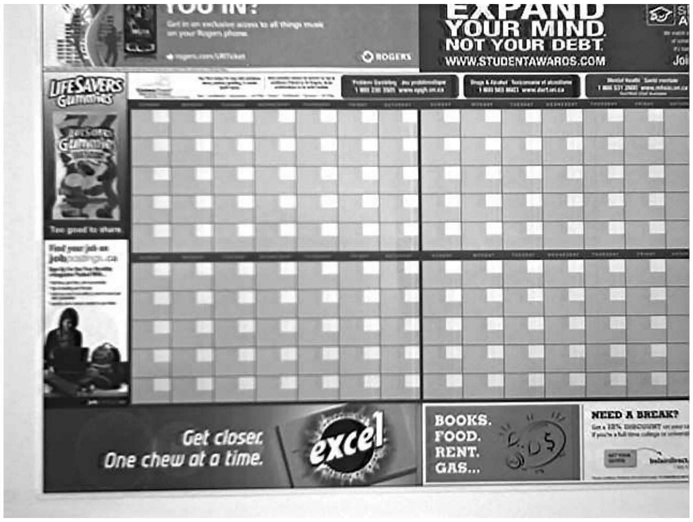
Figure 4.10: The deblurred result using the blind deconvolution approach [1]

blur makes the super-resolution analysis more tractable. We will, therefore, persist in using the identified blur in Figure 4.13(b) because it results in a suitably sharp image after de-blurring.

## 4.4 Implications of blur identification in the super-resolution process

To demonstrate the advantage of inserting a blur identification step in the super-resolution algorithm, we first construct a synthetic scenario. An image is passed through the observation model described in Chapter 2. This involves shifting the image, blurring with an appropriately chosen blur kernel, subsequent down-sampling and finally contamination with noise. In the first experiment, a rectangular averaging blur of size $5 \times 5$ is used as the blurring operator in the observation model. A sequence of $4$ low resolution images are used in the super-resolution process by assuming a particular blur type without any identification of the blur.

(a) A blurred image taken by the ProFUSION25 camera (*Calendar* image)



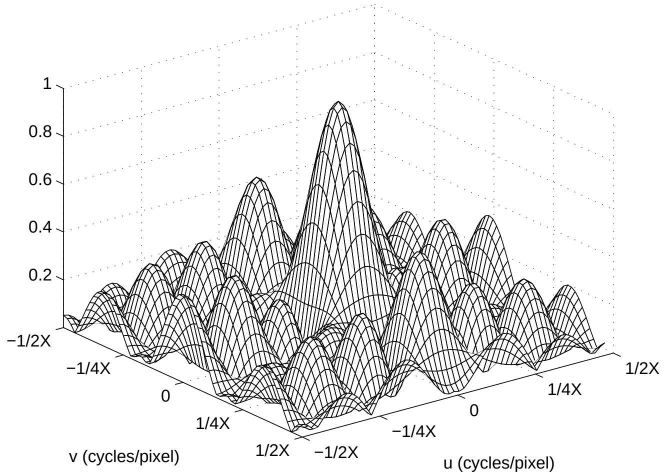
(b) The deblurred version of the *Calendar* image

Figure 4.11: Deblurring the *Calendar* image

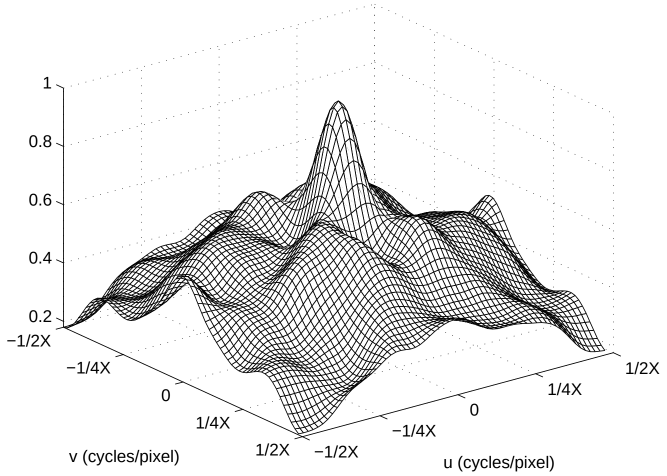
(a) A blurred image taken by the ProFUSION25 camera (*Dinosaur* image)

(b) The deblurred version of the *Dinosaur* image

Figure 4.12: Deblurring the *Dinosaur* image



(a) A blur kernel identified using the least-squares calibration approach



(b) A blur kernel identified using the blind deconvolution approach

Figure 4.13: A comparison of the identified ProFUSION25 camera blurs

We deal with three types of blur that are:

1. A rectangular averaging blur to mimic the spatial averaging of the sensor elements.
2. A circular disk blur to mimic the effects of out-of-focus blur.
3. A combination of spatial averaging and out-of-focus that is obtaining by a two-dimensional convolution of the two blurs.

The total-variation super-resolution results are shown in Table 4.2. Similar experiments are carried out by using the out-of-focus and combination blurs in the observation model and investigating the effect of assuming other likely blur types in the super-resolution process. These results are also tabulated in Table 4.3 and 4.4.

Table 4.2: Super-resolution performance (PSNR, in dB) of different likely blurs when the averaging blur is used in the forward model

| | **Average Blur** | Out-of-focus Blur | Combination Blur |
|---|---|---|---|
| *Satellite* Image | **25.9648** | 24.6273 | 24.0438 |
| *Boat* Image | **23.0336** | 22.0421 | 21.5339 |
| *Elaine* Image | **20.3095** | 16.8261 | 15.5790 |

Table 4.3: Super-resolution performance (PSNR, in dB) of different likely blurs when the out-of-focus blur is used in the forward model

| | Average Blur | **Out-of-focus Blur** | Combination Blur |
|---|---|---|---|
| *Satellite* Image | 24.1992 | **27.2351** | 26.6452 |
| *Boat* Image | 21.8441 | **24.5846** | 24.1240 |
| *Elaine* Image | 17.2004 | **20.0395** | 19.2879 |

The experiment is performed on three different test images which are shown in Figure 4.14. It is evident that correct identification of the blurring function used in the forward model significantly impacts the super-resolution result as is shown

Table 4.4: Super-resolution performance (PSNR, in dB) of different likely blurs when a combined blur is used in the forward model

| | Average Blur | Out-of-focus Blur | **Combination Blur** |
|---|---|---|---|
| *Satellite* Image | 22.2313 | 27.0462 | **28.5668** |
| *Boat* Image | 19.8885 | 24.2105 | **25.7723** |
| *Elaine* Image | 14.9690 | 20.1455 | **22.2956** |

(a) The *Satellite* image

(b) The *Boat* image

(c) The *Elaine* image

Figure 4.14: The three test images used in experiments for numerical results

by the peak signal-to-noise ratio increase whenever the correct blur is used in the super-resolution process. To corroborate the advantage of knowing the correct blurring function, experiments are also carried out using real ProFUSION25 images. The blur identification step is introduced into the super-resolution process as shown diagrammatically in Figure 4.15. The blur identification is carried out using the low resolution images as a separate step. The identified blurs are then used in the subsequent super-resolution algorithm. As opposed to other super-resolution algorithms, it is worthwhile to point out that each of the 25 images in the ProFUSION25 camera array are taken by different cameras. As a result, blur identification is separately performed for each low resolution image to output a possibly different blur for each camera. In conventional super-resolution algorithms utilizing temporal sequences, all the images are obtained from the same camera. Consequently, only one blur kernel needs to be identified for the camera.

Since each camera in the ProFUSION25 camera array blurs the high resolution image differently, we intuitively expect a better super-resolution result as opposed to the case in which all low resolution images are blurred with the same kernel. This is expected because diversity can be exploited not only from the sub-pixel displacement between views but also due to the different blurring that relates each low resolution pixel to a group of high resolution pixels. Further details about super-resolution from differently blurred images can be found in the recent work [4]. An experiment is conducted by using different blur kernels (arbitrarily choosing one of the three mentioned before) for each low resolution image in its forward model generation. The set of generated images is then input to the super-resolution algorithm. We examine the two scenarios in which:

1. all the low resolution images are assumed to be blurred with the same blur kernel.
2. the correct blur kernel corresponding to each low resolution image is used in

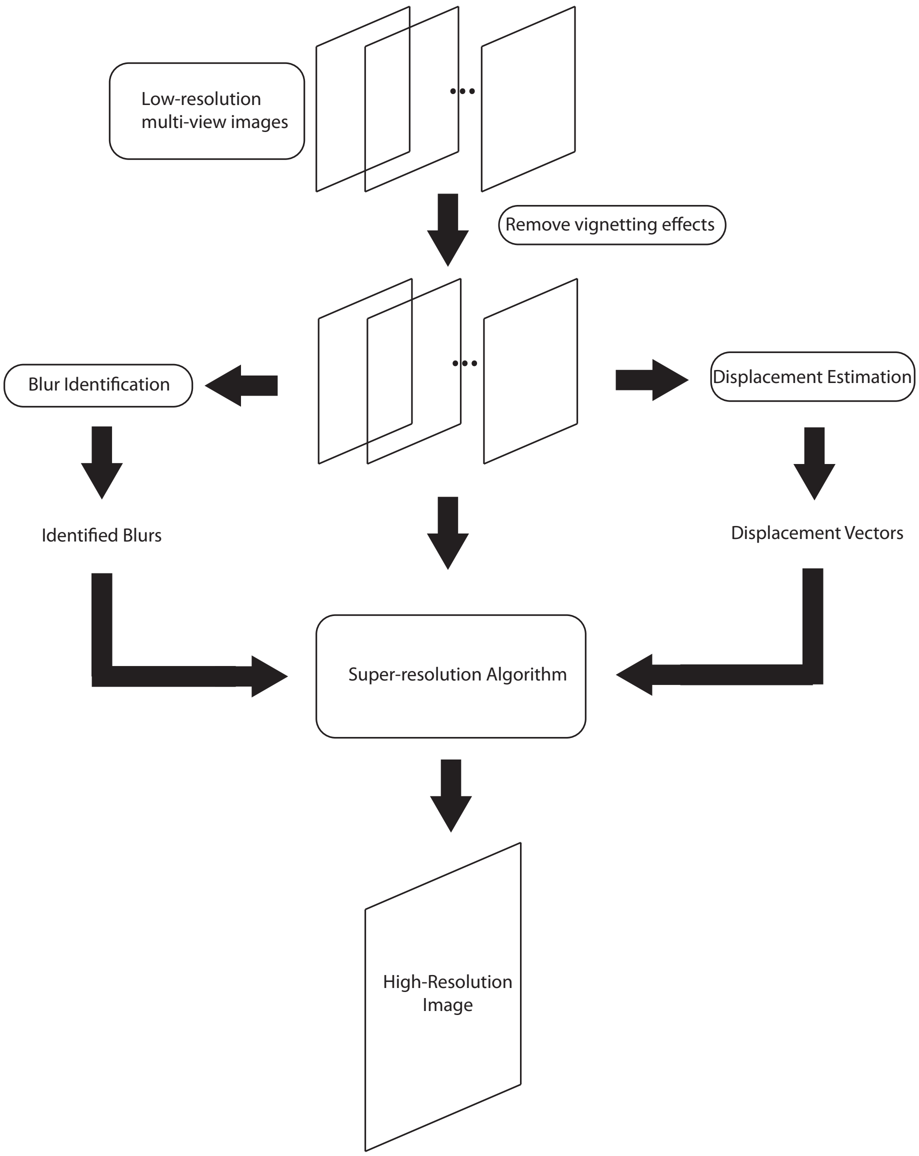


Figure 4.15: A block diagram of the super-resolution process involving blur identification

the super-resolution process.

The super-resolution results from the experiment are displayed in Table 4.5 for the three test images used throughout this chapter. There is an obvious improvement in the PSNR value when the correct blur in each 'channel' is used. Having demonstrated that accurate knowledge of the blur kernels does afford an objective performance increase, we now consider the case of real images taken with the ProFUSION25 camera array.

Table 4.5: Super-resolution performance (PSNR, in dB) when multiple blurs are used in the forward model

| | **Correct Blurs** | Average Blur (5x5) | Out-of-focus Blur (5x5) |
|---|---|---|---|
| *Satellite* Image | **30.3198** | 28.4922 | 28.5164 |
| *Boat* Image | **24.5803** | 23.3841 | 23.4369 |
| *Elaine* Image | **27.7707** | 26.4592 | 26.4824 |

When dealing with real images, super-resolution performance cannot be numerically evaluated because a high resolution version of the image to compare results with is not available. Consequently, we can only comment on the perceptual quality of the super-resolution results. Two different image sets are considered. The first image set is of the *Calendar* image which represents a sufficiently planar scene. The second set of images is that of the *Dinosaur* image which represents a more complex scene with depth discontinuities. In these cases, a noticeable improvement in sharpness is noticed with blur identification as oposed to the case of super-resolution with an assumed blur kernel. In particular, the text in the *Calendar* image in Figure 4.17 is more legible than the other results. Similarly, in Figure 4.19, the texture of the dinosaur as well as creases on its skin are sharper as compared to the scenario involving an assumed blur in Figure 4.18. The *Dinosaur* image, however, suffers from artifacts that arise from the choice of a simplistic translational model for the displacement occurring between the multi-view images. The removal of such displacement-related effects forms the subject matter of the next chapter.

## 4.5 Summary

The vignetting function and the PSF of the sensors in the ProFUSION25 camera array are identified in this chapter. We hypothesize that the vignetting function results in a photometric distortion that is expected to affect both displacement estimation and image fusion. The numerical results demonstrate that the accuracy of displacement estimation is not affected significantly by vignetting alone. However, removing the vignetting function is a cheap pre-processing step which greatly enhances the visual quality of the image. Consequently, all images taken with the ProFUSION25 and used in this thesis are de-vignetted before being used in any super-resolution algorithm. A second hypothesis made in this chapter is that assuming a parametric blur and performing super-resolution on real images is less effective than first identifying the PSF of the imaging sensor. This hypothesis is tested by different numerical experiments which show that knowledge of the correct PSF in each image channel results in an increase in super-resolution performance. Two methods are used to identify the blur kernel: a novel least-squares method which requires a latent unblurred image and a blind method which does not require the sharp image to be known. Both methods are tested and the latter is shown to result in better deconvolution. This performance increase is attributed to the photometric and geometric mis-match between the sharp and blurred images in the least-squares approach. The identified blurs are compared to the most probable blur types and are found to be inconsistent with them. This observation is not unexpected in light of the spatially variant blurring observed in the captured images. In the next chapter, all super-resolution results are presented after using a blur identification stage to identify the blur from each ProFUSION25 sensor.

Figure 4.16: Super-resolution of the *Calendar* image with an assumed $5 \times 5$ averaging blur ($\times 4$ magnification)

Figure 4.17: Super-resolution of the *Calendar* image with identified blurs ($\times 4$ magnification)

Figure 4.18: Super-resolution of the *Dinosaur* image with an assumed $5 \times 5$ averaging blur ($\times 4$ magnification)

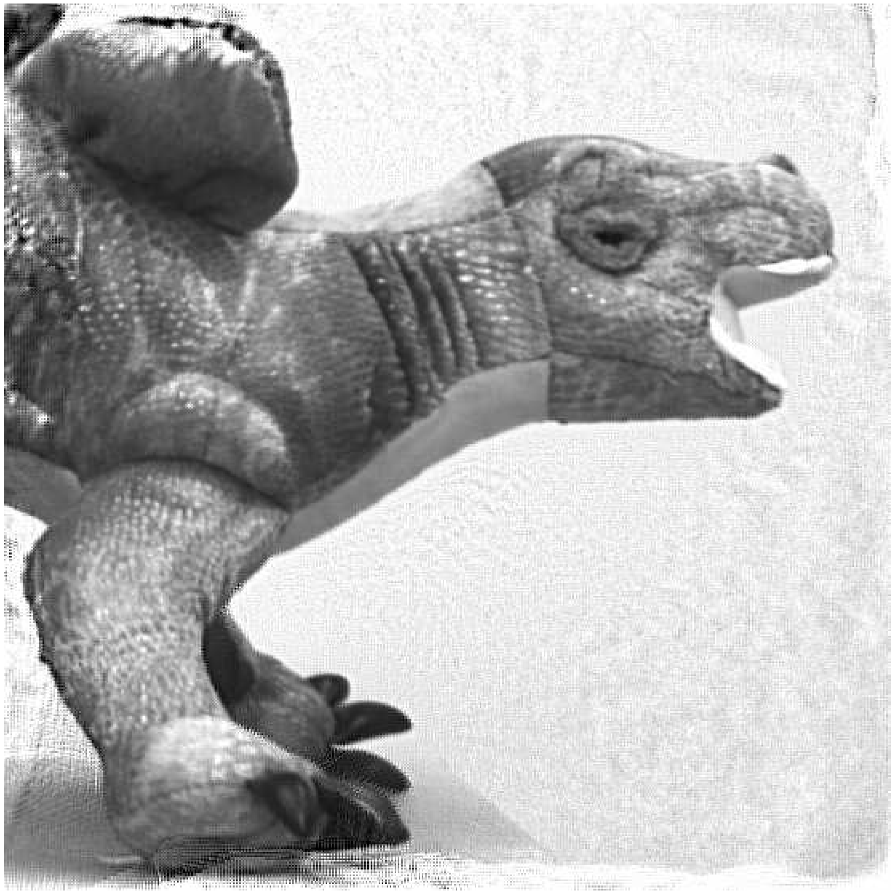


Figure 4.19: Super-resolution of the *Dinosaur* image with identified blurs (×4 magnification)

# Chapter 5

# Occlusion-aware super-resolution using dense displacement fields

> "Nothing hurts a new truth
> more than an old error."
>
> \- Johann Wolfgang von Goethe

In Chapter 3, the issue of a plausible image prior was examined, which relates to properties of the high-resolution image itself. The blurring part of the observation model was investigated in Chapter 4. However, little has been mentioned so far about the warping part of the observation model. In this chapter, the warping part of the observation model is examined from a multi-view perspective. This is the final results chapter in this thesis which presents results that build on those using the total-variation regularizer from Chapter 3 and the identified blurs using the algorithm in [1] from Chapter 4.

The multi-view nature of the images taken from the ProFUSION25 camera array presents a scenario that is different from the traditional treatment of displacement (or rather motion) in temporal sequences of images. However, a common need for accurate sub-pixel displacement estimates is identified in the super-resolution literature for both multi-view and temporal image sequences. This need for accurate estimation of correspondences between pixels in each of the multi-view images is identified as the limiting factor in most super-resolution algorithms. The experiments in this chapter attempt to answer the question, "How can the accuracy of displacement estimates be ensured?". Some researchers have resigned themselves to the inevitability of displacement estimation errors and have preferred to employ adaptive techniques to suppress the artifacts arising from such errors. On the other hand, some research groups have used more complicated models and displacement estimation techniques to ensure the availability of accurate displacement estimates. The utility of both approaches is examined in Section 5.1 and 5.2. However, the use of more complicated models comes with its own share of caveats which is treated in Section 5.3. Each of the Sections 5.1-5.3 respectively deal with effects related to displacement estimation in super-resolution arising from the following:

1. The artifacts arising from the propagation of erroneous displacement estimates in iterative super-resolution algorithms.
2. The use of a displacement model that is too simplistic to capture the displacement taking place between the multi-view images. For example, a two parameter translational shift model is not suited to parameterize scenarios when the images are rotated with respect to each other.
3. The artifacts arising from regions in the images that are occluded. This scenario is especially relevant to multi-view images because they often involve self-occlusions of 3D objects as well as occlusions from disparate objects.

## 5.1 Adaptive suppression of displacement-error artifacts

We begin our discussion by first addressing the issue of erroneous displacement estimates. At this point, we do not concern ourselves with the source of the displacement errors. To put matters into perspective, the reader is asked to refer to Figure 3.13 where a typical super-resolution flow chart is shown. The super-resolution algorithms in this thesis do not include displacement estimation within the restoration loop. In other words, the high resolution image and the displacement vectors are estimated independently from each other and not in a joint fashion. As a result, any error in the displacement estimation is carried forward in every iteration of the high resolution image restoration. We acknowledge the importance of jointly estimating the displacement vectors, but reserve this topic for consideration as an issue in future endeavors. The reasons for taking such a stance are discussed in Section 6.2.

When using a two parameter translation shift model or even a six parameter affine model, small errors in any of the parameters can lead to significant artifacts in the super-resolution result. To demonstrate such an effect, the *cameraman* image is artificially warped and the displacement parameters are intentionally perturbed. When the super-resolution algorithm is applied using perturbed displacement parameters, the result suffers from ghosting artifacts which can clearly be seen in both Figures 5.1(a) and 5.1(b). The image in Figure 5.1(a) was obtained by perturbing one of the displacement estimates by an error of 11.5 pixels (in both directions). The image in Figure 5.1(b) was obtained by perturbing the rotation angle of one of the affine parameter estimates by $5$ degrees. While such large displacement errors are highly exaggerated, it serves to show the resulting artifacts more clearly.

To mitigate the effects of such displacement estimation errors, a weighting term is introduced into the super-resolution algorithm that reflects the degree of confidence in the accuracy of the registration. The contribution of each low resolution observed image to the estimated high resolution image is weighted by considering the registration error caused by inaccurate displacement estimates. The proposed approach is similar in spirit to that proposed in [75] and [76]. The weighting term $\alpha_i$ is designed to satisfy certain properties such that it is:

1. inversely proportional to the magnitude of the data-fidelity term $\mathcal{J}_d(\mathbf{f}, \mathbf{g}_i)$.

2. greater than zero.

(a) SR image with inaccurate shift parameter estimates (×4 magnification)

(b) SR image with inaccurate affine parameter estimates (×4 magnification)

Figure 5.1: Inaccurate displacement estimates cause ghosting effects in the super-resolved image

The subscript $i = 0, 1, ..., k$ indicates individual low resolution images. The first property dictates that when the estimated displacements cause a large error in data fidelity, the corresponding image should be weighted less in the update process. In light of these, a weighting function of the following form is proposed [76],

$$\alpha_i = \mathcal{L}\left(T_i(\mathbf{f}) \cdot \frac{1}{\mathcal{J}_d(\mathbf{f}, \mathbf{g}_i)}\right), \tag{5.1}$$

where $\mathcal{L}(\cdot)$ is a monotonically increasing function and $T_i(\mathbf{f})$ is a scaling factor that also takes into account the mis-registration taking place in the other low resolution images. Using the previously defined data-fidelity term in Equation 3.2, two choices of the weighting function are examined,

$$\alpha_i = \sqrt{T_i(\mathbf{f}) \cdot \frac{1}{\|H_i\mathbf{f} - \mathbf{g}_i\|^2 + \delta_i}} \quad \text{(Type 1)} \tag{5.2}$$

and

$$\alpha_i = \log\left(T_i(\mathbf{f}) \cdot \frac{1}{\|H_i\mathbf{f} - \mathbf{g}_i\|^2 + \delta_i} + 1\right) \quad \text{(Type 2)}. \tag{5.3}$$

The scaling factor is given by,

$$T_i(\mathbf{f}) = \frac{\sum_{j=0}^{k} \|H_j\mathbf{f} - \mathbf{g}_j\|^2}{\|H_i\mathbf{f} - \mathbf{g}_i\|^2 + \delta_i}. \tag{5.4}$$

The term $\delta_i$ in Equations 5.2, 5.3 and 5.4 prevents the denominator from going to zero. The data fidelity or residual term $\|H_i\mathbf{f} - \mathbf{g}_i\|^2$ decreases with the number of iterations because there is a better fit between the observation model and the high resolution image. However, the residual remains significant in images that are associated with displacement errors. This causes the contribution of the mis-registered image to be weighted less in accordance with the weighting function used. The effectiveness of this technique can be seen in Figure 5.2 which is able to satisfactorily remove the artifacts in Figure 5.1(a). An experimental setup was also constructed to numerically evaluate the effectiveness of this approach. In this setup, a sequence of $4$ low resolution images is synthetically created by passing it through the forward model described in Chapter 2. The warping model used is either a translational shift or an affine warp. The warped images are then blurred with a $5 \times 5$ rectangular averaging blur kernel and finally ideally down-sampled by a factor of two. The displacement estimates in both cases are known and are input to the super-resolution algorithm after perturbing one of them to emulate outlier displacement errors. We define an outlier displacement estimate as a large error in the displacement actually occurring. These perturbations are randomly chosen to lie between $1.5$-$7.5$ pixels (uniformly distributed) at the low resolution scale. The super-resolution performance for different test images is quantified using the PSNR measure in Table 5.1. We would like to point out that such outlier displacement estimates are not totally impractical and can arise in certain situations. For example, consider two planar objects, object $\mathcal{A}$ and object $\mathcal{B}$ which are placed at different depths in a scene. Since the apparent displacement between images is depth dependent, the two objects appear to be displaced by different amounts when the scene is imaged from different positions. A global shift estimation technique cannot account for both these shifts simultaneously and hence, a compromise shift is estimated, which can be considered to be an outlier displacement estimate. This is precisely the scenario faced by us in the scene represented by Figure 5.9 that is described in a subsequent section.

The improvement in super-resolution performance by using the proposed adaptive suppression technique is apparent in Table 5.1 over the case in which the erroneous displacement estimates are used without any check in place. However, the surprising observation is that the super-resolution performance using the weighting technique does not perform better than single image bi-cubic interpolation.

Table 5.1: Performance comparison of adaptive suppression technique (PSNR, in dB)

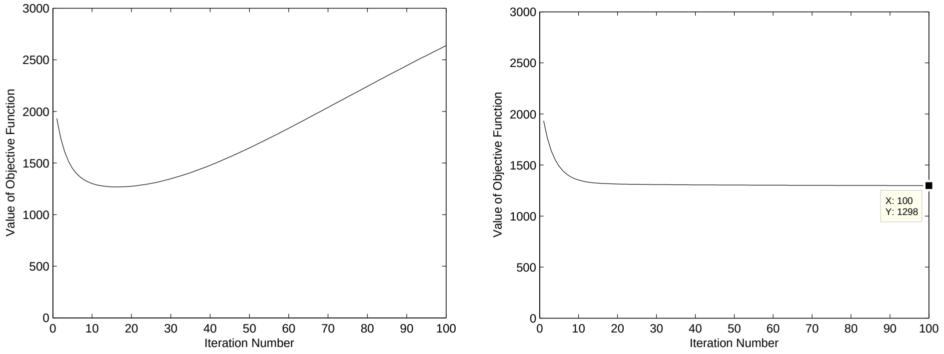
| | *Satellite* Image | *Boat* Image | *Elaine* Image |
|---|---|---|---|
| SR with outlier displacements | 15.7739 | 14.7051 | 15.5108 |
| SR with adaptive suppression (type 1) | 21.3882 | 24.1110 | 27.0281 |
| SR with adaptive suppression (type 2) | **21.8124** | **24.8648** | **27.8965** |
| Bi-cubic Interpolation | 22.1720 | 25.1729 | 27.9460 |
| SR with no outlier displacements | 23.2891 | 26.6197 | 28.7806 |

Figure 5.2: Using the adaptive weighting approach to reduce displacement-related artifacts (×4 magnification)

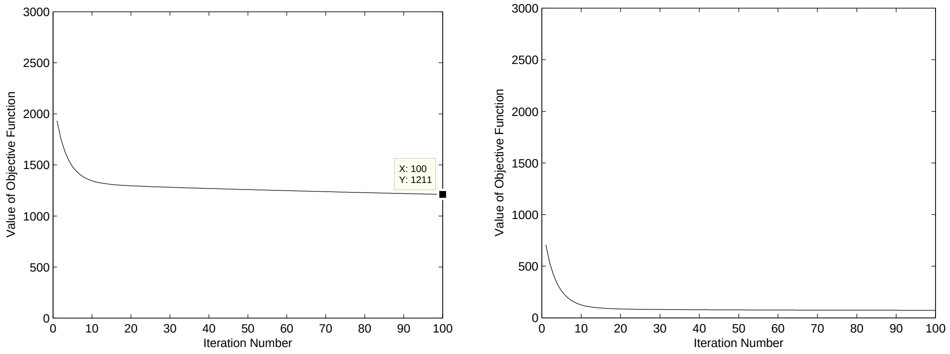
The suppression of artifacts is achieved by heavily dampening the contribution of the mis-registered images which reduces the amount of information that is available in the super-resolution restoration. When there is no displacement error, the super-resolution result has a much higher PSNR than single image bi-cubic interpolation. On comparing both the functions in Equation 5.2 and Equation 5.3, we notice that the $\log$ function consistently performs better than the function in Equation 5.2. Consequently, it is used in all experiments henceforth because of the incremental PSNR performance gain it affords.

The effect of the erroneous or outlier displacement estimates on the optimization process can be gauged by examining the behavior of the objective function. From Figure 5.3(a), it can be observed that the outlier displacement estimates cause the objective function to increase after a certain number of iterations. We conjecture that it takes a few iterations for the information from the low resolution images to be fused into the high resolution image during which the objective function decreases. After that point, the contribution of the mis-registered image dominates the objective function and causes it to increase. This is in contrast to Figures 5.3(b), 5.3(c) and 5.3(d) which show no increase in the value of the objective function. The steepest decrease in the objective function is seen in 5.3(d) which represents the case when no outlier displacements are present. The value of the objective function does not decrease to zero in Figures 5.3(b) and 5.3(c) because the weighting term does not allow the images with outlier displacement estimates to contribute to the high resolution image.

We now demonstrate results on real images taken with the ProFUSION25 array. Only 5 images from the set of 25 available images are used in this example. The viewing positions of the images used are shown in Figure 5.4(e). The images are registered using a dense displacement estimation scheme (optical flow) which is discussed in more detail in Sections 2.3 and 5.2. The registered images which suffer from significant errors are shown in Figure 5.4. We can observe from comparing the images in Figure 5.4 that the displacement estimation technique has failed to recover the correct displacement estimates in Figure 5.4(a) which is labeled as '1' in the ProFUSION25 grid. This is evident in the manner in which the dinosaur looks distorted in comparison to the other images. Nevertheless, this set of displacement estimates is used to recover a high-resolution image in an admittedly exaggerated demonstration of the resulting artifacts. The corresponding TV super-resolution result is shown in Figure 5.6(a). The high resolution image



(a) Super-resolution with no suppression of mis-registered image (b) Super-resolution with adaptive suppression (type 1)



(c) Super-resolution with adaptive suppression (type 2) (d) Super-resolution with no displacement errors

Figure 5.3: Trends in the value of the objective function

is expectedly plagued by annoying artifacts. On using the adaptive suppression technique, the artifacts are suitably removed as shown in Figure 5.6(b). Although the artifacts have been removed, the result does not demonstrate significant resolution enhancement over the bi-cubic interpolation in Figure 5.5 because of the suppression of the mis-registered images. However, this is only the case when a small number of low resolution images are used to recover the high-resolution image. In this example, five low resolution images were used. When a larger number of images are used, the effect of suppressing a lone or few mis-registered images does not detract from the enhancement of resolution as is expected from the general trend in Figure 3.11.

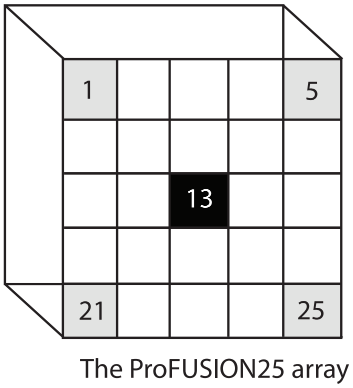
While this method works well to remove the artifacts caused by erroneous displacement estimates, it does so by suppressing the information from individual low resolution images that are judged to have been registered incorrectly. This is disadvantageous because every pixel in the image is penalized similarly by the weighting term even if certain regions have been correctly aligned (one weight for the entire image). Such a scenario commonly occurs when there are independently moving objects in the scene. From a multi-view perspective of a static scene, such a scenario can occur when two planar objects are at different depths in the scene. When a two parameter translational shift model is used, only one of the objects can be correctly registered. In such cases, a adaptive weighting of the entire image is not an optimal solution to eliminate displacement related effects. However, when the situation demands the use of simple displacement models that are of low computational complexity, such a method can be very useful in getting rid of displacement artifacts. In Section 5.3, a general treatment is described that is more selective in the weighting that is performed.

(a) The reference image 13 registered to image 1

(b) The reference image 13 registered to image 5

(c) The reference image 13 registered to image 21

(d) The reference image 13 registered to image 25



(e) Viewing positions of the images under consideration

Figure 5.4: Registration of the reference image to each of the observed low resolution images

Figure 5.5: Single image bi-cubic interpolation of the *Dinosaur* image (×2 magnification)

## 5.2 On using a more general displacement model

Many super-resolution algorithms assume simple parametric models like a translational shift for the displacement that occurs between the observed low resolution images. Such algorithms then proceed to demonstrate a dramatic increase in the performance of their algorithm over traditional single-image interpolation. While the theoretical analysis of such algorithms may be rigorous, they are of limited practical use simply because displacements under real world imaging conditions rarely correspond to simple shifts. The dramatic performance increase is mainly because the assumed translational shift model exactly matches the manner in which the synthetic images are created. In [9], the authors prove that assuming the displacement to be known leads to an almost 10-25% increase in performance (in terms of mean square error) as opposed to estimating the pixel correspondences from the observed low resolution images. This oft neglected result explains the su-

(a) SR image with no suppression of mis-registered images (×2 magnification)

(b) SR image with suppression of mis-registered images (×2 magnification)

Figure 5.6: Adaptive suppression of mis-registered images using real ProFUSION25 images (and dense displacement fields)

perior results reported by algorithms using synthetically generated low resolution image sets. The need to produce accurate estimates of the displacement vectors and the high-resolution image jointly makes the problem drastically harder. To tie this into the bigger theme of the thesis, the assumed displacement model in the observation model greatly impacts the super-resolution process. In the discussion so far, we have assumed a simple translational shift model for the displacement between multi-view images. While such an assumption is appropriate for planar scenes like the *Calendar* image and for small cropped regions from more complex scenes, it fails when applied to non-planar scenes. An example is shown in Figure 5.7(a) which demonstrates a TV super-resolution result assuming the translational shift model. The *Dinosaur* image represents a scene consisting of a 3D object with depth discontinuities. Since, the displacement of points in a scene is depth dependent, the global two parameter translational model cannot account for the different displacements of regions occurring at different depths. The artifacts arising from such a scenario where the displacement model itself is inadequate is most apparent in the areas near the legs of the dinosaur. As a comparison, the results using a more complicated six parameter global affine model is also shown in Figure 5.7(b). While the super-resolution image using an affine displacement model is perceptually sharper, the artifacts near the legs of the dinosaur are still apparent. These results demonstrate the inadequacy of global displacement models for complex scenes.

While the shift and affine models are not suitable as global representations of the scene displacement, they are viable models over smaller regions of the image which can be amalgamated to provide dense displacement vectors for arbitrary displacements. These smaller regions can be as small as a single pixel or even as large as an arbitrarily shaped object in the scene. This is the premise of piece-wise parametric models, block-based models and optical flow estimation techniques

[77]. As an alternative to global displacement models, a locally affine but globally smooth optical flow technique [10] that is described in Chapter 2 is chosen to estimate the displacement vectors for the scenes being considered in this thesis. Using such a displacement model allows us to obtain a dense displacement field that can account for more complex scenes. We show the result of using this optical flow technique on super-resolution performance in Figure 5.8. It can be seen that the high-resolution image does not suffer from the artifacts near the dinosaur's legs that were visible in Figures 5.7(a) and 5.7(b). However, it does suffer from objectionable artifacts mainly in the background regions containing little texture where reliable displacement estimates cannot be found. Additionally, the boundaries of the image, which can be treated as occlusion areas, display undesirable artifacts because corresponding regions cannot be found in other low resolution observations. The removal of such occlusion-related artifacts is treated in the next section.

## 5.3 Accounting for occlusion in super-resolution

While optical flow can be reliably estimated between image regions visible in both images being considered, it cannot be estimated in areas that are occluded. Occlusion refers to the phenomenon where an image region that is visible from a particular viewing position is covered up and can no longer be seen when the scene is viewed from a different viewing position. This occlusion can be caused either by the scene structure itself or due to the motion of objects within the scene. An example of a real scenario using the ProFUSION25 array is shown in Figure 5.9. The example shown in Figure 5.9 consists of a scene containing two books at different depths. Due to a change in the viewing position, the book that is closer occludes a certain region of the book that is placed further away. A consequence of occlusion is that it is not possible to find correspondences for presently visible points that become covered when the scene is viewed from a different viewing position.

(a) SR assuming a translational model (×2 magnification)

(b) SR assuming a affine model (×2 magnification)

Figure 5.7: Inadequacy of parametric displacement models

Figure 5.8: Super-resolution result using optical flow estimation (×2 magnification)

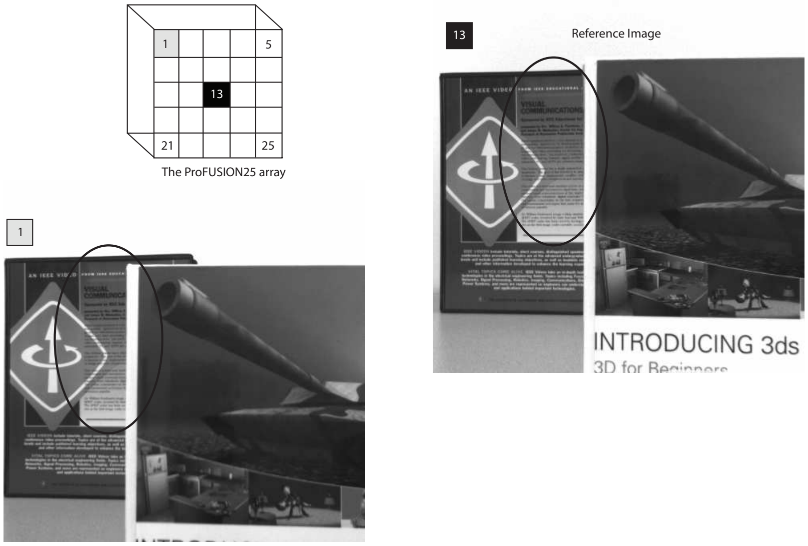


Figure 5.9: A scene showing occlusion between two objects at different depths

Similarly, regions that become uncovered have no displacement vectors pointing into them. To account for occlusion-related artifacts in the super-resolution result, we propose using only the visible pixels in the super-resolution algorithm. This process calls for identifying the occlusion areas and disabling their contribution in the super-resolution algorithm. A three step method is proposed for handling occlusion areas in the super-resolution process. First, optical flow is estimated at all pixel locations in the reference image. The estimated optical flow will result in unreliable displacement vectors in occlusion areas. These unreliable vectors are used to identify occlusion regions. Once the occlusion regions are identified, they are disabled from taking part in the super-resolution algorithm by using an occlusion mask (which acts as a binary weighting).

The occlusion detection stage itself is performed on the basis of intensity mismatch. The estimated displacement fields are used to predict each low resolution image from the pixel values of the reference image. The predicted images are then subtracted from the observed low resolution images. The difference image is squared and thresholded to obtain a binary image which indicates the occluded areas. The occlusion detection procedure is diagrammatically represented in Figure 5.10. A consequence of the intensity mis-match approach described above is that regions with wrong displacement estimates are also masked in addition to the occlusion areas. The displacement-related artifacts are a combination of effects from wrong displacement estimates and occluded regions. Once the occlusion maps have been obtained, the question arises as to how the occlusion maps can be used within the super-resolution framework. The following two scenarios are possible. In Figure 5.11, a typical super-resolution iteration using the total-variation regularizer is shown. The occlusion maps can be used at two stages marked as *Stage 1* and *Stage 2*. At *Stage 1*, the current estimate of the high resolution image has been warped in accordance with the estimated *forward* displacement vectors (or flow fields) as it passes through the observation model. Consequently, the occlusion maps can be obtained by the intensity mis-match between the *forward* registration of the reference image (to the sampling lattice of the corresponding low resolution observation) and that particular low resolution image itself. At *Stage 2*, the difference images arising from the data fidelity term are passed through the adjoint of the system operator represented by the forward observation model. The adjoint operator warps each difference image with the corresponding *backward* displacement estimates. Additionally, this is performed at the magnified spatial resolution because the adjoint operator includes a zero-insertion up-sampling component. The occlusion maps to be used at *Stage 2* are then obtained from an interpolated version of the intensity mis-match between the *backward* registration of each low resolution image (to positions on the sampling lattice of the reference image) and the reference image itself. Using the same scene of two books at different depths in Figure 5.9, an occlusion map is produced and shown in Figure 5.12(a). The occlusion map has been obtained with the intent of using it at *Stage 2*. A corresponding occlusion map with the intent of usage at *Stage 1* is shown in Figure 5.12(b). These are different because of the reasons mentioned in the preceding paragraph. The dominant areas that show up in these occlusion maps are the regions that are outside the field of view of the image being registered. The occluded area between the books indicated by ellipses in Figure 5.9 also shows up. The occlusion detection stage depends on the threshold parameter used, which is chosen heuristically. A very low threshold parameter leads to a noisy occlusion map whereas a high theshold parameter leads to an occlusion map that is dominated by the pixels that are beyond the viewing field of the reference camera. The advantage of using occlusion maps over the full-image adaptive suppression approach in Section 5.1 is that only those regions in the image which are occluded or not aligned properly are selectively suppressed from contributing to the high resolution image. This is in contrast to the earlier approach where the entire image was weighted in the same manner. A TV super-resolution result using the occlusion maps is shown in Figure 5.14. The corresponding single image bi-cubic interpolation result is also shown in Figure 5.13. A second example using the *Dinosaur* image is shown with the corresponding single image interpolation in Figures 5.16 and 5.15 respectively. These images do not suffer from annoying artifacts and are also sharper than the corresponding single image interpolated versions.

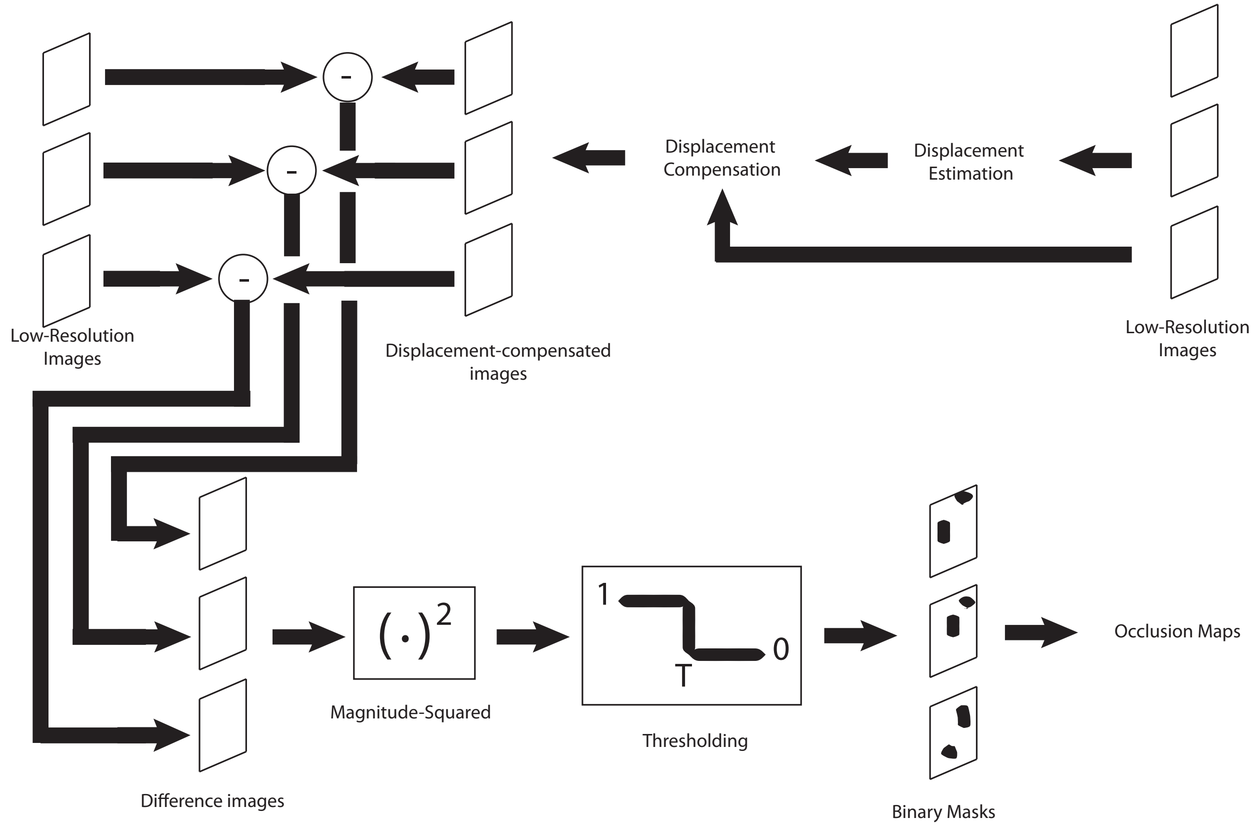


Figure 5.10: The proposed occlusion detection algorithm

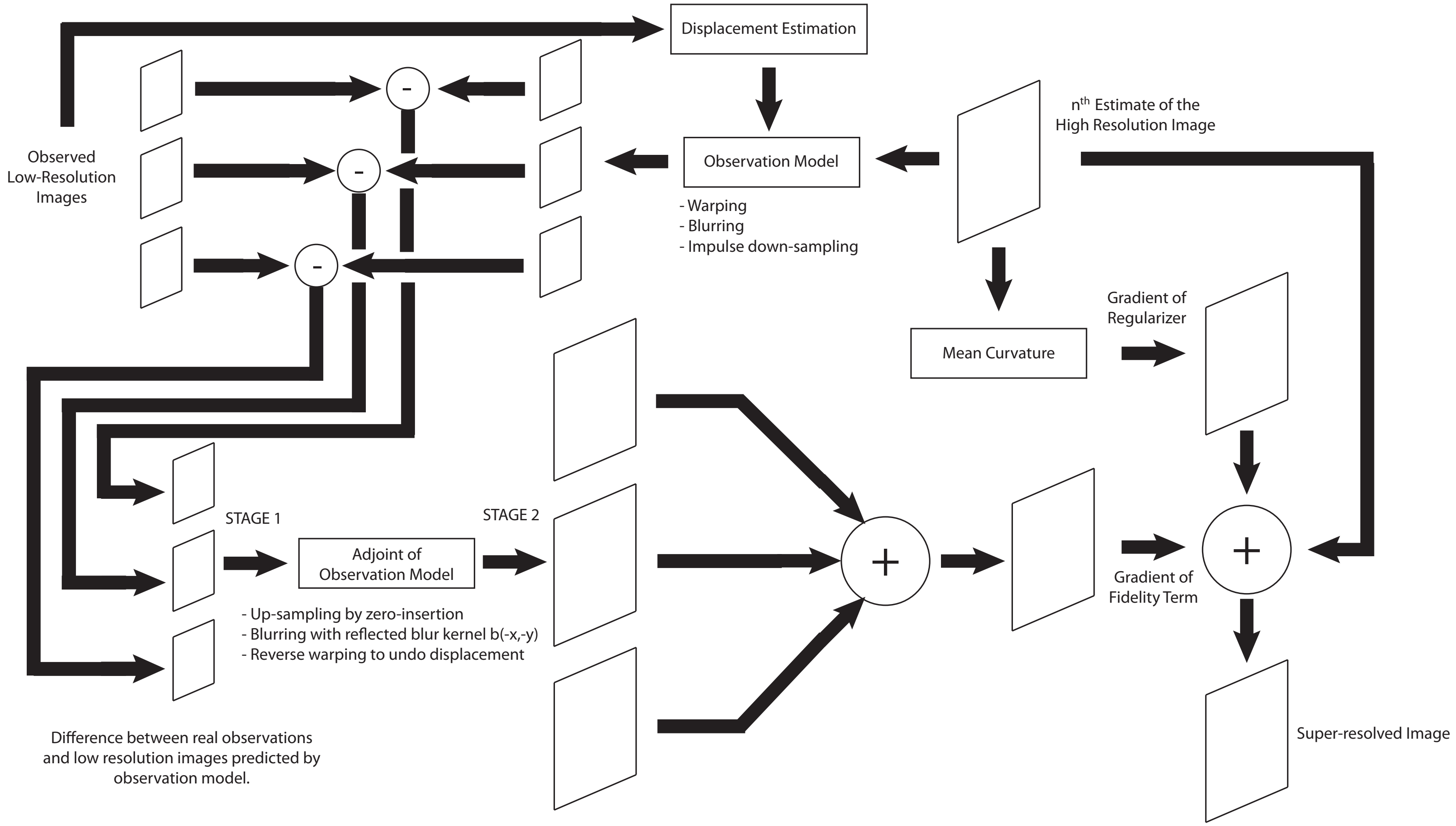
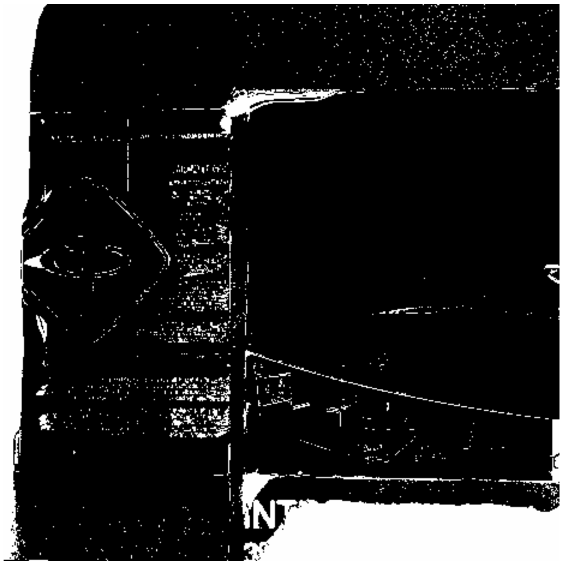


Figure 5.11: A single iteration within the super-resolution algorithm

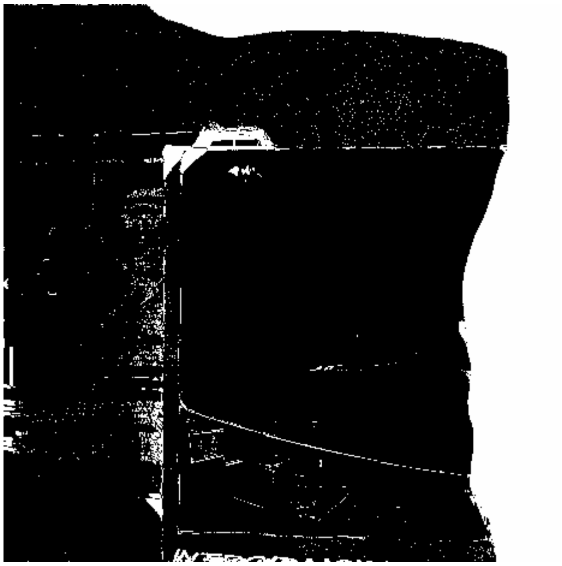
To demonstrate that using occlusion maps is more advantageous than the weighting term used in Section 5.1, an experiment is setup as follows. We use real ProFUSION25 images and pass each image through an observation model. The observation model used consists of blurring with a rectangular averaging filter and ideal down-sampling by 2. The images have already been warped due to their capture from different viewing positions. The set of 25 images is then input to the super-resolution algorithm and magnified by a factor of two using different displacement models together with the suppression and occlusion masking techniques. We perform this experiment so that a reference high resolution image is available to objectively compare the super-resolution result. This allows us to characterize super-resolution performance using a numerical measure like the PSNR. The results are shown in Table 5.2. Several independent trends can be noticed from this table:

1. When adaptive suppression is used with any displacement model, the super-resolution result demonstrates an increase in PSNR over the case when no adaptive suppression is performed. This is an expected result when the full complement of 25 images is used in the super-resolution process. When a lower number of images is used, the PSNR is not higher because of excessive dampening of the few images that are being considered. As a result, the information in those images is only weakly incorporated in the high resolution image.

2. When occlusion detection is included with the optical flow displacement model, the PSNR of the super-resolution result is the highest amongst all cases considered. This is also an expected result because the occlusion regions as well as mis-aligned regions are excluded from the super-resolution process. While a region may be occluded from a particular viewing position, it is a visible region from a different viewing position. As a result, even



(a) Occlusion map after registering Image 1 to spatial grid of Image 13



(b) Occlusion map after registering Image 13 to spatial grid of Image 1

Figure 5.12: Occlusion maps produced using the proposed approach

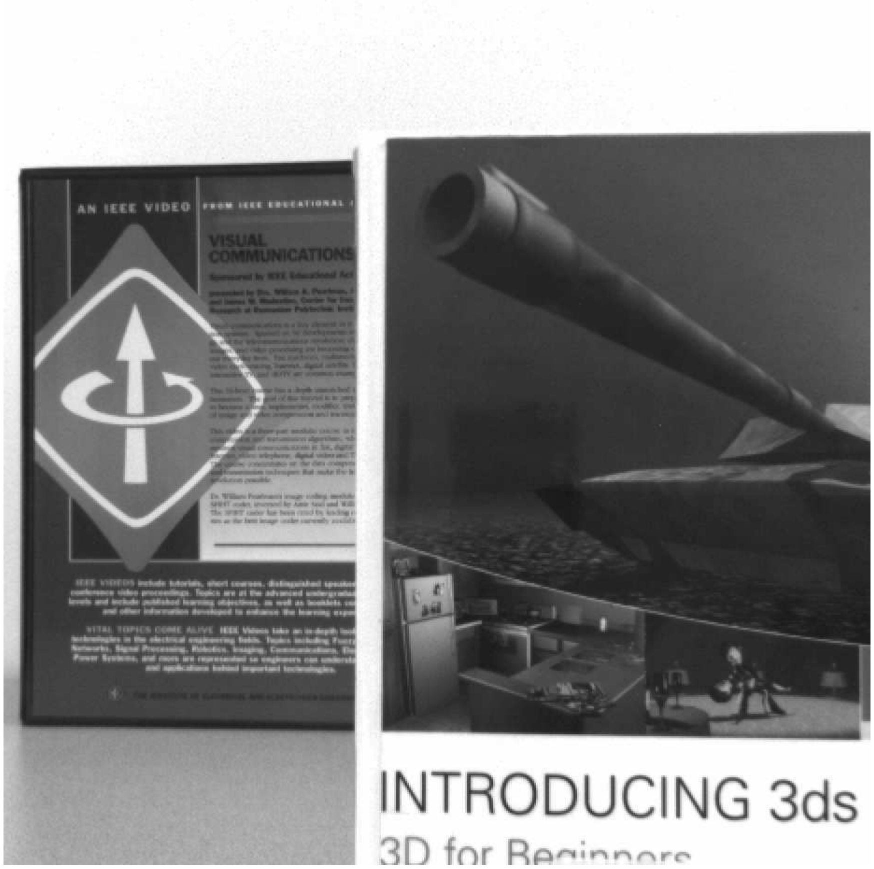


Figure 5.13: Single image bi-cubic interpolation of the *Book* image (×2 magnification)

Figure 5.14: Super-resolution of the *Book* image using optical flow and occlusion detection (×2 magnification)

Figure 5.15: Single image bi-cubic interpolation of the *Dinosaur* image (oblique view, ×2 magnification)

Figure 5.16: Super-resolution of the *Dinosaur* image (oblique view) using optical flow and occlusion detection (×2 magnification)

if certain regions are excluded from a particular low resolution image, the spatial diversity due to different viewing positions allows a net increase in PSNR. The only exception is the *Dinosaur* image (side view) in which super-resolution with occlusion detection results in only a small PSNR gain over the bi-cubic interpolation case. We believe this is due to the fact that the dinosaur has been imaged at a greater distance than the other two images. To capture as much of the dinosaur as possible in the camera's field of view, we had to place the dinosaur at a greater distance than the other two images considered. As a result, changes in viewing position do not result in significant self-occlusions and there are no new regions convered/uncovered at the boundaries like the *Book* image. The occlusion detection stage has little impact besides mitigating some of the effects in the background.

3. When using occlusion masks with optical flow displacement estimation, the PSNR is higher than the scenario in which the adaptive weighting technique is used. This is also an expected result because only occluded or mis-aligned regions are excluded from contributing to the high-resolution image as opposed to weighting of the entire image.

4. When just a shift or affine displacement model is considered, the PSNR is higher than that of super-resolution results using optical flow. This is not an intuitive result. We expect optical flow to perform better because it can produce more accurate displacement vectors for the type of displacements taking place in scenes with 3D objects. However, the optical flow estimation technique is not perfect and it fails to produce accurate displacement vectors in regions with little texture like the background wall and in covered/uncovered regions. Inaccurate displacement estimates in these areas forces the PSNR to drop even though there is significant resolution enhance-

ment in well aligned regions as evidenced in the comparison between Figures 5.7 and 5.8.

Table 5.2: Performance comparison of techniques to alleviate displacement-related effects (PSNR, in dB)

| | *Dinosaur* Image (Oblique Profile) | *Books* Image *Books* Image | *Dinosaur* Image (Side Profile) |
|---|---|---|---|
| SR with shift model | 25.2852 | 20.8836 | 25.0365 |
| SR with shift model and adaptive suppression | 26.0286 | 21.7521 | 25.8573 |
| SR with affine model | 24.6357 | 18.5842 | 22.3095 |
| SR with affine model and adaptive suppression | 25.8917 | 20.6895 | 25.8994 |
| SR with optical flow | 22.7832 | 22.8830 | 23.1087 |
| SR with optical flow and adaptive suppression | 25.8811 | 26.6022 | 26.0463 |
| SR with optical flow and occlusion detection | **29.2154** | **28.3106** | **28.6029** |
| Bi-cubic Interpolation | 28.0299 | 27.2429 | 28.4378 |

## 5.4 Summary

The aim of this chapter was to make the reader aware of the need for accurate displacement estimates in obtaining good super-resolution performance. An adaptive weighting approach was proposed to mitigate the effect of erroneous displacement estimates that is mostly effective when global parametric models are imposed on the displacement taking place between images. While this method was shown to be effective in removing displacement related artifacts, it does so by sacrificing resolution enhancement when a small number of low resolution images are used. When a larger number of images are used, it is successful in both resolution enhancement as well as reducing displacement-related artifacts. A locally affine but globally smooth dense displacement estimation technique is adopted for use with scenes containing 3D objects. Such a model allows the estimation of arbitrary displacements that is demonstrated to result in better super-resolution performance rather than global parametric models like translation shift or affine models. Finally, a novel intensity based method for detecting occlusions, which is a more general form of the weighting approach, is proposed and the resulting super-resolution result not only has reduced displacement-related artifacts but also demonstrates a resolution enhancement over single image bi-cubic interpolation.

The results in this chapter represent a cumulative increase in performance by using the components developed in Chapters 3-5. The use of the total-variation regularizer was shown to preserve edges to a better degree than the other regularizers used. The process of identifying the PSF was then also demonstrated to result in a subjective as well as an objective performance increase over the scenario where a reasonable blur is assumed for the camera PSF. Putting all these results together with the use of a dense displacement estimation technique results in the performance increases seen in this chapter. Furthermore, the use of an occlusion detection stage helps remove the artifacts associated with inaccurate displacement estimates. These different components applied together are essential to produce a clean, pleasing, visually superior high resolution image. It is important to mention that all results in this chapter with real ProFUSION25 images were super-resolved by a magnification factor of $2$. While the improvement over bi-cubic interpolation at this magnification is certainly noticeable, we believe that the improvement will be more significant for higher magnifications. However, due to the computational costs of the displacement estimation stage, we were unable to provide results at higher magnifications.

# Chapter 6

# Conclusions and future research

> "I hope that posterity will judge me kindly, not only as to the things which I have explained, but also to those which I have intentionally omitted so as to leave to others the pleasure of discovery."
>
> - Rene Descartes

In this chapter, we summarize the study carried out during the course of this thesis, suggest avenues for future work and highlight the main contributions. In Chapter 1, this work was introduced as a smaller part of a larger project. This umbrella project envisions a system that allows a user to experience a realistic, high-resolution three-dimensional view of a certain environment. For example, imagine a user navigating a virtual environment created from panoramic images of a museum. The user should be able to walk up to a certain exhibit, for instance, a statue or other 3D objects and be able to view the exhibit in high resolution. An added benefit would be to be able to view the exhibit as a 3D representation. This representation could take the form of a stereo pair. A 3D representation is possible only when the structure of the exhibit can be inferred. This is largely carried out by using multiple views of the same object. There is a fundamental tradeoff between the size of the scene captured and the resolution at which it is captured. As the captured scene area becomes larger, the resolution of an object in the scene decreases. In the overall quest for a panoramic viewing experience, resolution is often sacrificed. In the light of such an ambitious project, the objective of this thesis was to investigate super-resolution performance from multi-view images taken by the ProFUSION25 camera array which can then be incorporated within the panoramic environment. This was motivated by two main observations from the literature surveyed:

1. Existing super-resolution algorithms are geared towards super-resolution from a temporal sequence of images and not from a multi-view perspective. As a result, the performance of super-resolution algorithms are mainly available for super-resolution from temporal sequences of images.

2. A majority of the algorithms perform super-resolution without identifying the characteristics of the imaging sensor and assuming simplistic displacement models.

In Section 6.1, major conclusions based on the entire study are given. We make general comments on super-resolution itself as well as comments specific to multi-view images and the ProFUSION25 camera array. This study is by no means complete and there are several issues for future consideration that are outlined in Section 6.2. Finally, the main contributions of the thesis are identified in Section 6.3.

## 6.1 Conclusions

The over-arching question that this thesis attempts to answer is, ”Is there a possibility for resolution enhancement of a captured scene using multi-view images from the ProFUSION25 array?”. If so, how can we achieve this? If not, what are the factors that make this difficult? In our quest towards answering these questions, we first examined the most appropriate super-resolution reconstruction method that should be used as the core of our algorithm. The surveyed literature pointed towards best results using regularized super-resolution. As a result, an in-depth study of regularization-based approaches was conducted in Chapter 3. Super-resolution was formulated as an inverse problem and a justification for using a regularization-based approach was given. The three most popular regularizers used, viz., Tikhonov, HMRF and TV regularizers were compared within a super-resolution scenario. The Tikhonov regularizer was shown to cause excessive smoothing and resulted in a blurry high resolution image. The HMRF-based statistical regularization performed better than Tikhonov regularization. However, its dependence on the clique activity measure in specified directions caused smoothing across edges which is not desirable. Finally, the TV regularizer was demonstrated to preserve edges as well as produce the most visually appealing high resolution image. This was attributed to the fact that the TV regularizer does not smooth the image across edges. Additionally, the effect of the number of images on super-resolution performance was also demonstrated. We observed that as the number of low resolution images used was increased, the PSNR of the high resolution image also increased until a saturation level was reached. Beyond this point, an increase in the number of low resolution images led to very small changes in PSNR. This is due to the redundant information in the large number of low resolution images. We found that when using real images captured from the ProFUSION25 camera array, the increase in super-resolution performance (visually and using PSNR) was not as marked as for synthetic scenarios. This is primarily due to the fact that the displacement taking place between images is not known nor can it be estimated to a high degree of accuracy. Furthermore, the imaging process itself affects the low resolution image by introducing blurring and other photometric distortions. These effects are not known a priori and formed the subject of investigation in Chapter 4.

The multi-view images taken by the ProFUSION25 camera are affected by a photometric distortion which has a characteristic radial fall-off called vignetting.

This photometric distortion is expected to affect the accuracy of intensity-based displacement estimation techniques. We found that while vignetting does affect the accuracy of displacement estimation technique, it does not do so to a degree such that super-resolution becomes infeasible. The standard deviation in the displacement estimation error was found to be of the order of a tenth of a pixel which only becomes significant when the magnification factor is more than ten. However, we concluded that an intensity distortion detracts from the overall visual quality of the high resolution image. As a result, vignetting was corrected by a suitable calibration technique. This gave satisfactory results in removing the vignetting effect. An exercise in finding a vignetting model that best explained the ProFUSION25 camera vignetting effect was also carried out. We found that the cosine-fourth model can be used to fit the vignetting effect generated by the ProFUSION25 camera. The fit was carried out using a least-squares technique. The advantage of knowing the correct PSF in the super-resolution process is also investigated. We found that when the correct PSF is known, the PSNR of the high resolution image is at least 2 dB higher than when a likely blur is assumed. The ProFUSION25 camera array consists of 25 different sensors with the possibility of different PSFs for each sensor. We showed experimentally that knowing the correct blur in each channel or sensor also increases super-resolution performance significantly. The blur identification was carried out in two ways. First, a calibration experiment using a known target was used to obtain a least-squares estimate of the blur. This did not produce satisfactory results because of inaccurate geometric and photometric registration. As an alternative, a blind de-convolution technique was used to identify the camera PSF. On subsequent de-blurring with the identified blur, the ProFUSION25 images were shown to be much sharper and clearer. These identified blurs were also used in the super-resolution process with significant visual increase in resolution. The conclusion we can draw from these experiments is that the identification of the PSF plays an important role in increasing super-resolution performance. Moreover, the use of assumed parametric blurs is not sufficient because they do not correspond to the actual ProFUSION25 camera blurs. In fact, we believe that the real camera blur is not spatially-invariant as explained in Chapter 4.

In the final Chapter 5, occlusion-aware super-resolution with dense displacement estimates was examined. Initially, a global parametric displacement model was used in the observation model which was shown to be prone to outlier displacement estimates due to its inability to account for complex displacements. An adaptive weighting mechanism was proposed to weight the contribution of misregistered images which led to a largely artifact-free high resolution image. However, the weighting term did not allow the low resolution images to contribute much towards overall enhancement of resolution. As a result, we proposed a more general displacement model that was based on a locally affine but globally smooth dense displacement estimation technique. This technique was shown to work well except where reliable displacement estimates could not be found (in regions with little texture) and in occluded regions. Subsequently, an intensity-based occlusion detection scheme was proposed which was able to remove artifacts due to inaccurate displacement estimates as well as provide resolution enhancement in the high resolution image. This was demonstrated using both visual results and the numerical PSNR measure.

Going back to the question raised at the beginning of this section, we were able to demonstrate a visual as well as quantitative increase in super-resolution performance using multi-view images. The increase in visual quality over bi-cubic interpolation is significant but not remarkable. In Chapter 1, aliasing and blurring were identified to be the major causes of loss in resolution. However, the ProFUSION25 camera does not produce images with significant blurring. More importantly, the ProFUSION25 images do not suffer from extensive aliasing. In most super-resolution literature, aliasing is forced into the images by synthetically down-sampling the image sequence by a large factor. This accounts for the remarkable resolution gains that can be obtained. Finally, the computational costs involved in the estimation of displacements as well as the eventual reconstruction may be too costly for the increase in resolution that super-resolution from the ProFUSION25 array affords. Having said that, using a simple displacement estimation technique in conjunction with the adaptive weighting approach does not have a very steep computational cost. In such a scenario, super-resolution could be feasible and cheap to implement.

## 6.2 Future work

Super-resolution using multi-view images offers a variety of avenues for future consideration. With the bigger picture of the fore-mentioned project in mind, the next step would be super-resolution of a novel viewpoint. While the ProFUSION25 camera array provides 25 different viewpoints, a user in an immersive environment would prefer a continuous rendering of high resolution views of a scene. This would typically involve super-resolution from viewpoints that are not part of the original set of viewing positions. Such a scenario presents difficulties that need to be addressed. For example, the issue of computing displacement estimates between the novel viewing position and a reference view is an interesting issue. Furthermore, since the intensity values at that position are not known, occlusion regions cannot be identified with certainty. A common approach that is undertaken in the literature is to perform a calibration between the cameras to obtain their extrinsic and intrinsic parameters. These parameters relate the projection of a point in 3D space to the image plane. Using this information, a cloud of 3D points can be generated from the multiple views to build a three-dimensional representation. A novel viewpoint can be obtained if information about extrinsic and intrinsic parameters of a camera placed at that viewing position can be obtained. These parameters are then used together with the 3D cloud of points to generate the novel view.

The issue of super-resolution of color images is another important research front. Mono-chrome processing by applying super-resolution to each color channel independently is not optimal because it does not take into account the spectral correlation between the channels [19]. If the channels can be de-correlated using a transform like the Karhunen-Loeve transform (KLT) [78] or in a suitable color space, then the super-resolution algorithm can be applied to each de-correlated channel separately and transformed back to the original domain or color space. However, the ProFUSION25 camera array only outputs gray-scale images. We envision a scenario where each camera sensor in the array can be overlayed with an optical color filter. Each camera in the array can thus output different color channels which can later be combined using available techniques into a color image.

The specific work performed in this thesis can also be built upon. With regard to regularization, we have not explored the possibility of recognition-based priors which was mentioned in the literature review in Chapter 2. For example, if we were to use the ProFUSION25 camera to obtain 3D portraits of a face, a database of high resolution face images can be used to learn a more relevant prior (or regularizer). Similarly, if we constrain our application to museum artifacts, a corresponding database of images of artifact like objects can be used to learn an application specific prior. These would be particularly useful at higher magnifications where the observation model constraints cannot provide sufficient information.

We comment in Chapter 4 that the real camera blur is probably spatially-variant as is observed in the captured images. This opens up a whole field of research using space-variant blurs. A straight-forward approach to space-variant blur identification would be to approximate a space-variant blur as space-invariant blurs in different sections of the image. This suggests local image processing operations on sectioned regions of the image which need to be handled with properly chosen boundary conditions. The advantage of using such an approach over the space-invariant assumption has not been widely studied in super-resolution literature. The encouraging results in this thesis of knowing the correct blur hint that such an approach may be promising.

In Chapter 5, we have chosen to use a relatively simple dense displacement estimation technique which enforces smoothness constraints by ensuring that the affine parameters between local regions change smoothly. A more recent approach [79] that includes stronger constraints could also be used to obtain more accurate displacement estimates. This is especially relevant if the ProFUSION25 camera array is not used in one-shot mode. If a video stream of multi-view images is obtained, the problem is made harder due to the possibility of temporal motion. In such a case, a smoothness constraint in the temporal dimension also has to enforced by using an additional term in the energy function being minimized.

We also remark in Chapter 5 that the joint estimation of displacement fields and the high resolution image could prove to be beneficial. We have not included displacement estimation within the restoration loop because of the added computational cost involved. However, a computationally cheap dense displacement estimation technique used within the restoration loop could lead to potentially better displacement estimates in each iteration. This is especially relevant for high magnification factors because accurate displacement estimates are harder to obtain from extremely small aliased images.

As a final note, the occlusion detection scheme used in this thesis was based on intensity mis-match. While this gives satisfactory performance in our case, it may not prove to be useful in situations that are affected by large intensity distortions. For example, in a typical scene, the lighting conditions may vary from point to point which may lead to significant intensity differences. These could cause the occlusion detection to go awry. A geometry-based occlusion detection scheme [80] could prove to be useful in such situations. Geometry-based occlusion detection is based on the fact that uncovered regions (occluded regions in the reference frame) do not have displacement vectors pointing to them. This can be used as a distinguishing feature to detect occluded regions. While there are other issues for consideration, we feel that these topics are especially relevant to the case of super-resolution from multi-view images. In the next section, a brief highlight of the thesis contributions is given.

## 6.3 Thesis contributions

At the risk of repeating ourselves, we feel that it is necessary to reiterate the contributions of this thesis and the work carried out during the course of this thesis.

1. A comprehensive survey of super-resolution literature was carried out to begin with. The literature survey identified a need for consideration of the case of multi-view images which has received little attention. An original contribution of this thesis is a novel super-resolution application of multi-view images taken using the ProFUSION25 camera array.

2. After choosing the regularization-based approach as the core of our super-resolution algorithm, an in-depth study of regularization-based approaches was performed to determine the most suitable regularizer. The experiments performed at this stage helped us realize the importance of blur identification and the need for dense displacement estimates. Each of the regularized approaches mentioned were implemented using MATLAB. The Tikhonov and HMRF approaches were implemented ourselves. The TV-based regularizer

was implemented using level set methods. This involved computation of the mean curvature which was obtained using an implementation provided by Dr. Eric Dubois (part of a thesis [52] which was supervised by Dr. Dubois).

3. The ProFUSION25 camera was found to suffer from a photometric defect called vignetting. We corrected for this effect using a calibration method. A cosine-fourth model was also used to fit the observed vignetting effect. All ProFUSION25 images were subsequently de-vignetted before applying any super-resolution algorithm. Next, a calibration experiment was setup to identify the camera blurs using a known sharp and blurred image pair. This is an original contribution of this thesis although we have not performed an extensive literature review on this topic. Consequently, a state-of-the-art blind de-convolution procedure [1] is adopted to identify the camera blurs. An implementation released by the original authors is used. The use of this particular blind de-convolution approach has not been previously reported in super-resolution literature. To sum it up, the characterization of the ProFUSION25 camera via its PSF and its vignetting effect is an original contribution.

4. The identified camera blurs are used in all subsequent super-resolution stages. A shift or affine displacement estimation technique were initially used to obtain the required displacement estimates. The shift estimation technique was implemented ourselves while an implementation was available for the affine estimation technique. These global models turned out to be inadequate to represent the displacements occurring in the scenes considered. Consequently, an optical flow estimation technique was proposed to obtain dense displacement estimates. The optical flow algorithm implementation was also available. However, its use in a super-resolution setting is a novel contribu-

tion. In addition, a weighting method to suppress the artifacts from erroneous displacement estimates is implemented. Due to the shortcomings of this weighting technique, a more general intensity mis-match based occlusion detection scheme is implemented. The application of such an occlusion detection scheme is novel in a super-resolution scenario.

5. Since this thesis forms a smaller part of a larger intended project, a super-resolution toolbox was created for use in future endeavors. This toolbox contains a collection of functions for implementing each component in the super-resolution process as discussed in this thesis. It is largely intended for multi-view images, although it can easily be used for temporal sequences. This toolbox is an original contribution of this thesis. We intend to make this toolbox available online in the interest of reproducible research.

# Bibliography

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

[2] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[3] P. Q. Tuan, "Spatio-tonal adaptivity in super-resolution of under-sampled image sequences," Ph.D. dissertation, Delft University of Technology, 2006.

[4] S. Chaudhuri and J. Manjunath, *Motion-Free Super-Resolution*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.

[5] M. Trimeche, "Super-resolution image reconstruction using non-linear filtering techniques," Ph.D. dissertation, Tampere University of Technology, 2006.

[6] M. Hennig and F. Worgotter, "Eye micro-movements improve stimulus detection beyond the Nyquist limit in the peripheral retina," *Advances in Neural Information Processing Systems*, vol. 16, 2004.

[7] S. Borman and R. Stevenson, "Super-resolution from image sequences - A review," in *Proceedings of the 1998 Midwest Symposium on Circuits and Systems*. Notre Dame, IN, USA: IEEE, August 1998, pp. 374–378.

[8] D. Capel, *Image Mosaicing and Super-Resolution (Cphc/Bcs Distinguished Dissertations.)*. Springer-Verlag, 2004.

[9] D. Robinson and P. Milanfar, "Statistical performance analysis of superresolution image reconstruction," in *Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004.*, vol. 1, November 2004, pp. 144–149.

[10] S. Periaswamy and H. Farid, “Elastic registration in the presence of intensity variations,” *IEEE Transactions on Medical Imaging*, vol. 22, no. 7, pp. 865–874, July 2003.

[11] R. Y. Tsai and T. S. Huang, “Multiframe image restoration and registration,” in *Advances in Computer Vision and Image Processing*, vol. 1. JAI Press Inc., 1984, pp. 317–339.

[12] P. Vandewalle, S. Süsstrunk, and M. Vetterli, “A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution,” *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. Article ID 71 459, 14 pages, 2006.

[13] M. Irani and S. Peleg, “Improving resolution by image registration,” *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.

[14] R. R. Schultz and R. L. Stevenson, “Extraction of high-resolution frames from video sequences,” *IEEE Transactions on Image Processing*, vol. 5, pp. 996–1011, 1996.

[15] E. Y. L. Michael K. Ng, Huanfeng Shen and L. Zhang, “A total variation regularization based super-resolution reconstruction algorithm for digital video,” *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. Article ID 74 585, 16 pages, 2007.

[16] S. Baker and T. Kanade, “Limits on super-resolution and how to break them,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1167–1183, 2002.

[17] D. Capel and A. Zisserman, “Computer vision applied to super resolution,” *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 75–86, May 2003.

[18] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, May 1996.

[19] M. Ng and N. Bose, "Mathematical analysis of super-resolution methodology," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 62–74, May 2003.

[20] P. M. Nhat Nguyen and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1299–1308, September 2001.

[21] N. Woods, N. Galatsanos, and A. Katsaggelos, "EM-based simultaneous registration, restoration, and interpolation of super-resolved images," in *International Conference on Image Processing*, vol. 2, Sept. 2003, pp. II–303–II–306.

[22] D. Rajan and S. Chaudhuri, "Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1102–1117, 2003.

[23] S. Lertrattanapanich and N. Bose, "High resolution image formation from low resolution frames using delaunay triangulation," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1427–1441, Dec 2002.

[24] D. S. C. Biggs and M. Andrews, "Asymmetric iterative blind deconvolution of multi-frame images," in *Proceedings of the SPIE Conf. on Advanced Signal Processing Algorithms, Architectures, and Implementations VIII*, vol. 3461, Notre Dame, IN, USA, 1998, pp. 328–338.

[25] F. Sroubek, G. Cristobal, and J. Flusser, "Simultaneous super-resolution and blind deconvolution," *Journal of Physics: Conference Series*, vol. 124, pp. 1–8, 2008.

[26] Y. Chen, Y. Luo, and D. Hu, “A general approach to blind image super-resolution using a pde framework,” in *Visual Communications and Image Processing*, vol. 5960, no. 1. SPIE, 2005, pp. 1–12.

[27] S. Baker and T. Kanade, “Super resolution optical flow,” Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-99-36, October 1999.

[28] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision (darpa),” in *Proceedings of the 1981 DARPA Image Understanding Workshop*, April 1981, pp. 121–130.

[29] F. C. Lin, C. B. Fookes, V. Chandran, and S. Sridharan, “Investigation into optical flow super-resolution for surveillance applications,” in *APRS Workshop on Digital Image Computing: Pattern Recognition and Imaging for Medical Applications*, 2005, pp. 1–7.

[30] W. Zhao and H. S. Sawhney, “Is super-resolution with optical flow feasible?” in *Proceedings of the 7th European Conference on Computer Vision-Part I*. London, UK: Springer-Verlag, 2002, pp. 599–613.

[31] A. Nagy and Z. Vamossy, “Super-resolution for traditional and omni-directional image sequences,” *Acta Polytechnica Hungarica*, vol. 6, no. 1, pp. 117 – 130, 2009.

[32] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[33] A. Krylov and A. Nasonov, “Fast super-resolution from video data using optical flow estimation,” in *9th International Conference on Signal Processing*, Oct. 2008, pp. 853–856.

[34] R. Fransens, C. Strecha, and L. V. Gool, "Optical flow based super-resolution: A probabilistic approach," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 106 – 115, 2007.

[35] H. Shen, L. Zhang, B. Huang, and P. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 479–490, February 2007.

[36] Z. Wang and F. Qi, "On ambiguities in super-resolution modeling," *IEEE Signal Processing Letters*, vol. 11, pp. 678–681, August 2004.

[37] R. Fransens, C. Strecha, and L. V. Gool, "A probabilistic approach to optical flow based super-resolution," in *Proceedings of the Workshop on Generative Model Based Vision*, 2004, pp. 6–12.

[38] B. Zitova, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, October 2003.

[39] S. Borman, "Topics in multiframe super-resolution restoration," Ph.D. dissertation, University of Notre Dame, 2004.

[40] S. P. D. Keren and R. Brada, "Image sequence enhancement using sub-pixel displacements," in *Computer Vision and Pattern Recognition*, 1988, pp. 742–746.

[41] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics (SIAM), 2005.

[42] J. Hadamard, *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. Yale University Press, 1923.

[43] T. Poggio, V. Torre, and K. C., "Computational vision and regularization theory," *Nature*, vol. 317, no. 26, pp. 314–319, 1985.

[44] M. Elad and A. Feuer, “Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images,” *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646–1658, December 1997.

[45] W. C. Karl, “Regularization in image restoration and reconstruction,” in *Handbook of Image and Video Processing*, A. Bovik, Ed. San Diego, CA: Academic Press, 2000, ch. 3.6, pp. 141–160.

[46] M. V. Zibetti, F. S. Bazán, and J. Mayer, “Determining the regularization parameters for super-resolution problems,” *Signal Processing*, vol. 88, no. 12, pp. 2890 – 2901, 2008.

[47] D. Terzopoulos, “Regularization of inverse visual problems involving discontinuities,” *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 4, pp. 413–425, 1986.

[48] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[49] R. R. Schulz and R. L. Stevenson, “A Bayesian approach to image expansion for improved definition,” *IEEE Transactions on Image Processing*, vol. 3, pp. 233–242, May 1994.

[50] J. Besag, “Spatial interaction and the statistical analysis of lattice systems,” *Journal of Royal Statistical Society B*, vol. 36, no. 12, pp. 192 – 225, 1974.

[51] P. F. J. Zhang and D. Wang, “Random field models,” in *Handbook of Image and Video Processing*, A. Bovik, Ed. San Diego, CA: Academic Press, 2000, ch. 7.2, pp. 645–654.

[52] H. Aly, "Regularized image up-sampling," Ph.D. dissertation, University of Ottawa, School of Information Technology and Engineering, 2004.

[53] T. Chan and J. Shen, *Image Processing And Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, 2005.

[54] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629–639, July 1990.

[55] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," in *Proc. Int. Conf. of the Center for Nonlinear Studies on Experimental Mathematics : Computational issues in nonlinear science*. Elsevier North-Holland, Inc., 1992, pp. 259–268.

[56] Y. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE Transactions on Image Processing*, vol. 5, pp. 1539–1552, November 1996.

[57] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, ser. Applied Mathematical Sciences. Springer-Verlag, 2006, vol. 147.

[58] S. Osher and J. A. Sethian, "Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.

[59] S. Osher and R. P. Fedkiw, "Level set methods," in *Imaging, Vision and Graphics*. Springer, 2000.

[60] A. Marquina and S. Osher, “Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal,” *SIAM Journal on Scientific Computing*, vol. 22, no. 2, pp. 387–405, 2000.

[61] R. Ramanath, W. Snyder, Y. Yoo, and M. Drew, “Color image processing pipeline,” *IEEE Signal Processing Magazine*, vol. 22, no. 1, pp. 34–43, Jan. 2005.

[62] D. Goldman and J.-H. Chen, “Vignette and exposure calibration and compensation,” *Tenth IEEE International Conference on Computer Vision*, vol. 1, pp. 899–906, Oct. 2005.

[63] Y. Zheng, S. Lin, and S. B. Kang, “Single-image vignetting correction,” in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 461–468.

[64] Y. Zheng, J. Yu, S. B. Kang, S. Lin, and C. Kambhamettu, “Single-image vignetting correction using radial gradient symmetry,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[65] Y. Altunbasak, R. M. Mersereau, and A. J. Patti, “A fast parametric motion estimation algorithm with illumination and lens distortion correction,” *IEEE Transactions on Image Processing*, vol. 12, no. 4, pp. 395–408, April 2003.

[66] M. H. W. J. C. Lagarias, J. A. Reeds and P. E. Wright, “Convergence properties of the nelder-mead simplex method in low dimensions,” *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.

[67] R. L. Lagendijk and J. Biemond, “Basic methods for image restoration and blur identification,” in *Handbook of Image and Video Processing*, A. Bovik, Ed. San Diego, CA: Academic Press, 2000, ch. 3.5, pp. 167–181.

[68] S. E. Reichenbach, S. K. Park, and R. Narayanswamy, “Characterizing digital image acquisition devices,” *Optical Engineering*, vol. 30, no. 2, pp. 170–177, 1991.

[69] D. B. Gennery, “Determination of optical transfer function by inspection of frequency-domain plot,” *Journal of Optical Society of America*, vol. 63, pp. 1571–1577, December 1973.

[70] T. C. T. Stockham and R. Ingebresten, “Blind deconvolution through digital signal processing,” in *Proceedings of the IEEE*, vol. 63, April 1975, pp. 678–692.

[71] A. T. R. Lagendijk and J. Biemond, “Maximum likelihood image and blur identification: a unifying approach,” *Optical Engineering*, vol. 29, no. 5, pp. 422–435, May 1990.

[72] Y. You and M. Kaveh, “A regularization approach to joint blur identification and image restoration,” *IEEE Transactions on Image Processing*, vol. 5, no. 3, pp. 416–428, March 1996.

[73] R. S. N. Joshi and D. J. Kriegman, “PSF estimation using sharp edge prediction,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[74] W. H. Richardson, “Bayesian-based iterative method of image restoration,” *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55–59, 1972.

[75] W. B. Lim, M. K. Park, and M. G. Kang, “Spatially adaptive regularized iterative high-resolution image reconstruction algorithm,” in *Visual Communications and Image Processing*, vol. 4310. SPIE, 2000, pp. 10–20.

[76] E. S. Lee and M. G. Kang, "Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration," *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp. 826–837, July 2003.

[77] J. Konrad, "Motion detection and estimation," in *Handbook of Image and Video Processing*, A. Bovik, Ed. San Diego, CA: Academic Press, 2000, ch. 3.10, pp. 207–224.

[78] B. Hunt and O. Kubler, "Karhunen-Loeve multispectral image restoration, part I: Theory," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 3, pp. 592–600, Jun 1984.

[79] M. J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, January 1996.

[80] S. Ince and J. Konrad, "Occlusion-aware optical flow estimation," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1443–1451, 1972.