TOPICS IN MULTIFRAME SUPERRESOLUTION RESTORATION

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Sean Borman, B.Sc. (Eng.), M.S.E.E.

Robert L. Stevenson, Director

Graduate Program in Electrical Engineering
Notre Dame, Indiana
April 2004

TOPICS IN MULTIFRAME SUPERRESOLUTION RESTORATION

Abstract

by

Sean Borman

Multiframe superresolution restoration refers to methods for still image and video restoration from multiple low-resolution, degraded observed images derived from an underlying scene. These methods extend classical single frame image restoration methods by simultaneously utilizing information from *multiple* observed images to achieve restoration at resolutions higher than that of the original data. Since multiframe methods are capable of restoring spectral information beyond that which is available in any of the observed images, they are referred to as *superresolution* methods.

Multiframe superresolution restoration methods register the observed images to a common reference frame in order to formulate multiple observed data constraints for a given region of the restoration. The image registration process thus requires knowledge of the visual motion occurring in the observed image sequence. Since this typically is not known, the motion information must be estimated from the observed image sequence in order to effect the restoration process.

An exhaustive survey of the multiframe superresolution restoration literature is presented along with some of the necessary background material. A detailed introduction to the problem of motion estimation is also provided. The performance characteristics of a common class of block-matching subpixel motion estimators is

examined in an effort to understand the limitations that these motion estimators impose on the performance of multiframe superresolution restoration frameworks. A new Bayesian multiframe superresolution restoration procedure, which computes multiple restored frames simultaneously, is introduced. The proposed method enables the incorporation of temporal as well as spatial constraints on the restored image sequence. A novel approach for incorporating general interframe motion and linear, but spatially-varying, degradations in an observation matrix relating the observed low-resolution image sequence to the unknown restoration is presented.

Stemming from the work conducted on multiframe superresolution restoration, the equivalence between two mathematical models describing a pinhole camera used in the computer graphics and computer vision communities is proved.

To my parents for their endless optimism and support throughout a difficult and protracted journey.

To Sami for all the reasons that make my future worth looking forward to.

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

We humans are social creatures. Many other dissertations deal with this proposition in great detail with erudite perspicacity. For me, however, it suffices to infer that the fruits of an individual's labors are also fruits borne of those who surround, nurture and support him.

There have been a great many who have played a role in the process which culminates in this modest achievement and it is an honor for me to acknowledge them here. In truth, their contributions befit much more than parenthetical mention among pages of equations.

During my High School years, Andrew Farlam, Katherine Elks, Suzy Brokensha and Doria Joubert were critical in illuminating possible futures and prospects. I recognize now how their influences, which at the time seemed relatively small, shaped my outlook as a maturing individual.

During my undergraduate years at the University of Cape Town I am grateful for the guidance and friendship of John Greene, Stephen Schrire, Gerhard de Jager and Lynn Gillis. I must, however, reserve my greatest thanks for Norman Morrison who inspired me when I was all but ready to change academic paths. Dr. Morrison's influence came at a critical time and set me on a path which led directly to this dissertation. His contribution cannot be overstated.

Through the Fulbright Scholarship Program I was able to obtain a Master's degree in Electrical Engineering at the University of Notre Dame. This degree enabled me to contribute back to South Africa by returning to lecture at the University of

Cape Town. I am extremely grateful for these opportunities. I also wish to acknowledge my academic adviser Ken Sauer for his support during my Master's program.

To the Millers — Connie, Chris, Lorcan and Maeve, thank you so much for your incredible hospitality and support during my last major South Bend sojourn. As any acquaintance of Connie and Chris will asseverate, they are fabulous friends and truly gracious people.

Likewise to Tracy and Kia, who welcomed me to the Bay Area and into their home. Through many stimulating and sometimes impassioned discussions ranging from technology to politics, you have been fabulous hosts and selflessly generous in your support. I am in your debt.

Now at the conclusion of my doctorate, I have the opportunity to express my gratitude to my academic adviser, Bob Stevenson, who has granted me enormous latitude, both geographic and academic, and given me fortitude in the face of a protracted epic which has finally reached its denouement. I have enjoyed our working relationship and I am immensely grateful to have had the opportunity to work with him.

My thanks also to the members of my dissertation defense committee, Peter Bauer, Pat Flynn, Tom Fuja and Anré Venter.

Finally, my greatest thanks is reserved for my family and Sami who together form the social foundations in which my life, values and outlook are grounded. With their unwavering support, I know my life will be meaningful and fruitful. We are, after all social creatures. This work is dedicated to them. My only regret is that it is far too modest to do them justice.

CHAPTER 1

INTRODUCTION

This dissertation addresses the problem of multiframe superresolution restoration of images and video — that is, the problem of restoring high-quality images and video from multiple observed frames. The term superresolution is used to describe the ability of these multiframe methods to achieve far higher fidelity restoration than is possible using traditional single frame restoration methods. As shall be shown in this work, superresolution methods are capable of restoring information lost in the imaging process.

The multiframe superresolution restoration problem is multifaceted and draws extensively from the areas of motion estimation and image restoration, both of which have an extensive literature. It is therefore impossible to understand and approach the topic without a strong background in these areas. Therefore, this work endeavors to provide thorough overviews of these areas before exploring new ground.

The novel results presented in this dissertation pertain to motion estimation as well as restoration algorithms, consistent with the multipronged approach taken toward understanding and developing the theory and practice of multiframe superresolution restoration.

1.1 Overview of the Dissertation

Following this introductory text, Chapter (2) presents some of the background theory necessary for a formal definition of the multiframe superresolution restoration problem. The chapter also discusses the relationship between multiframe superresolution restoration methods and other image restoration techniques. An effort is made to classify broadly superresolution restoration methods into a logical structure which differentiates between spatial, temporal and spatiotemporal restoration methods. Throughout the dissertation, the concept of an inverse problem and especially that of ill-posed inverse problem, will recur. For this reason, a brief introduction to this interesting subject is also provided.

In Chapter (3) an extensive survey of the literature on multiframe superresolution image and video restoration is presented. This chapter is an excerpt from a more extensive study of multiframe superresolution restoration methods [17]. The key findings of this study were presented later in [18]. This survey divides multiframe superresolution restoration algorithms into two main classes — frequency domain methods and spatial domain methods. Within each class, research is classified according to the formulations used. The result is an almost exhaustive review of the restoration approaches which have been applied. This chapter concludes with a summary and a comparison of the many techniques discussed.

One of the obvious conclusions of Chapter (3) is the critical importance of interframe motion to the multiframe superresolution restoration problem. The problem of motion estimation is introduced and discussed in considerable detail in Chapter (4) which is also an excerpt from a larger body of work on image sequence processing which appeared as a chapter in the Dekker Encyclopedia of Optical Engineering [187]. This chapter provides the depth of knowledge on the motion estimation problem that will be required in later chapters.

In Chapter (5) the limitations of the commonly-used block-matching motion estimation method are explored in an empirical study of the performance of these methods in an idealized best-case scenario. The motivation is to determine the suitability of applying these methods to the multiframe superresolution restoration problem. This is done by studying the performance that can be anticipated from such subpixel motion estimators. This is important since the performance of the motion estimator can be the limiting factor on the performance of the multiframe superresolution restoration method. The work presented in this chapter appears in [16].

Chapter (6) may at first appear somewhat tangential to the main thrust of this work as it straddles the areas of computer vision and computer graphics and is not directly related to the problem of multiframe superresolution restoration. The main results of this chapter emerged from work presented in Chapter (8) in which techniques from the computer vision and computer graphics literature were used in connection with the development of new results in the area of multiframe superresolution restoration. Chapter (6) presents results which connect camera models commonly used in computer graphics raytracing with the camera matrix representation used in the computer vision field.

In Chapter (7) the state of the art in multiframe superresolution restoration is extended with the introduction of a novel generalization of a superresolution restoration framework which simultaneously restores multiple superresolution frames given multiple input frames. In the proposed restoration method, multiple superresolution frames are computed simultaneously rather than sequentially and independently as is the case with existing methods. The proposed formulation is unique in that it enables the incorporation of temporal constraints in addition to the usual spatial constraints on the restored video sequence. The proposed estimation framework also includes a method for mitigating the effects of motion estimation errors on the restoration. This work was presented in [19].

One of the major components in any restoration problem, superresolution or otherwise, is the observation model which relates the known observations to the unknowns to be estimated and which takes into account the imaging process relating the two. Restoration is essentially the process of inverting the observation process to obtain an estimate of the unknowns giving rise to the observations. As a result, inaccuracies in the observation model tend to limit the performance and accuracy of the restoration process. In Chapter (8) it is shown how the linear observation model, which is widely used in multiframe restoration, can be extended to include spatially-varying degradations as well as nonaffine image registration accounting for the motion occurring between frames. This work is based on ideas used the computer graphics community, but heretofore never applied to the development of observation models for multiframe superresolution restoration. The work discussed in this chapter appears in [20, 21].

In Chapter (9), the content of the dissertation is summarized briefly, conclusions are drawn and directions for future research are suggested.

1.2 Other Work

Apart from the main body of work presented in this dissertation, the author was involved in other projects which are not presented here.

One of these is concerned with the generation of High Dynamic Range (HDR) images. In real-world scenes, the dynamic range of the illumination (the ratio of the intensities of the brightest highlight to the darkest shadow) far exceeds the range which can be captured using solid state image sensing devices. As a result, digitized images captured in natural or uncontrolled lighting situations often lose information in the highlights or shadows (or both). In this work a method was introduced which computes HDR images which represent the full dynamic range of the scene, derived from multiple differently exposed images of the scene. The resulting floating-point representation contains the full range of the illumination information present in the scene. The idea for this work originated with the author and was executed in collaboration with Robertson and Stevenson [156, 157].

Also not included in this dissertation is work conducted in collaboration with Andrew Lumsdaine (formerly of the Department of Computer Science and Engineering at the University of Notre Dame) and Robert Stevenson on the development of a rapid prototyping graphical user environment for image and video processing. Called the Video Enhancement Workshop (ViEW), the software package was developed in collaboration with a number of students in the Computer Science Department. The author served as the leader of the development team and was responsible for the overall architectural design of the project. Though this project was successfully completed, this work remains unpublished.

CHAPTER 2

BACKGROUND

2.1 Introduction

In this chapter some of the background knowledge necessary for understanding and exploring the multiframe superresolution restoration problem is presented. An attempt is made to provide a perspective on how modern superresolution restoration techniques relate to earlier work, as well as to emerging research areas.

In Section (2.2), a high level overview of the image restoration problem is presented and the meaning of the term *superresolution* is defined. Single-input and multi-input restoration frameworks are discussed and a hierarchy of spatial and temporal restoration categories is presented. A problem statement which describes a very general case of spatiotemporal superresolution restoration is provided. In Section (2.3), the theory of inverse problems, which will play a important role in the chapters ahead, is introduced. Finally, in Section (2.4), some of the applications of superresolution image restoration techniques are discussed.

2.2 Image Restoration

A very brief overview of the enormous field of image restoration is provided here. Image restoration remains an important research topic and one of the major applications driving the theory and practice of image processing since digital computers made processing large amounts of data possible. This section is not meant to provide a review of the extensive literature on image restoration, (the reader is referred to the texts [7, 102, 119, 97] for that) but rather to provide some perspective on how modern multiframe superresolution restoration algorithms grew out of the existing body of research and how these new multiframe methods define directions for future work.

Image restoration techniques may be very broadly categorized into two classes based on the number of observed frames. Specifically, the categorization is into the classes of single-input and multi-input restoration methods.

2.2.1 Single Frame Restoration

The classical image restoration problem is concerned with restoration of a single output image from a single degraded observed image, also known as Single-Input Single-Output (SISO). The literature on the restoration of a single input frame is extensive and spans several decades. A wide variety of techniques exist to tackle this problem and for a concise but representative review, the reader is referred to [97]. More extensive treatments of specific classes of single frame restoration algorithms may be found in [7, 102, 119].

More recently, with the growing interest in digital processing of image sequences, researchers have begun to address the problem of multiframe image restoration. The techniques developed for single frame restoration have often provided the theoretical basis for extending to the multiframe case. Indeed, much of the work in superresolution spatial restoration which is reviewed in Chapter (3) borrows from techniques developed for the single frame restoration problem.

2.2.2 Multiframe Restoration

While the field of single frame image restoration appears to have matured, digital video has provided many new restoration problems for image processing researchers [191]. Since video typically consists of a sequence of similar, though not identical frames, it becomes possible to utilize the inter-frame motion information in processing the video data. This led to the development of image sequence processing techniques such as motion compensated filtering, image sequence interpolation and standards conversion. Image restoration researchers also recognized the potential for image restoration at increased spatial resolution using the totality of information contained in an image *sequence* as compared with that available from a single image. This led naturally to algorithms which apply motion compensation and image restoration techniques to produce high-quality still images from image sequences. These Multi-Input Single-Output (MISO) methods in turn spurred the development of more general Multi-Input Multi-Output (MIMO) approaches. In the review of the multiframe restoration literature presented in Chapter (3) many of these algorithms are discussed, with special attention given to those which have the potential for superresolution.

2.2.3 Superresolution

Though the terminology varies considerably from field to field, superresolution is broadly understood to mean bandwidth extrapolation beyond the diffraction limit of the optical system. Though the use of the diffraction limit as the threshold defining superresolution is common, it is often inappropriate in electronic imaging applications. In Chapter (5) it is shown that in typical electronic imaging systems, resolution is limited not by the diffraction limit of the optical system, but rather by spatial averaging at the Focal Plane Array (FPA) detector elements. Therefore, for such applications it is more meaningful to consider an algorithm as superresolving if it is capable of restoring information beyond the bandlimit of the imaging system, rather than the diffraction limit of the optical system. Of course, should the algorithm be capable of bandwidth extrapolating beyond the bandlimits of both the imaging and optical systems, all the better.

Superresolution restoration thus requires not only the amelioration of the degradations caused in the imaging process (the classical restoration problem), but also the extrapolation of frequency content beyond that which is present in the observed data. Superresolution restoration therefore requires the restoration of lost information. At first this may seem an impossibility. Indeed, superresolution was described as a "myth" by Andrews and Hunt in their influential monograph on image restoration [7]. In Section (2.3), however, a framework is presented which explains how and why superresolution restoration is in fact possible. Hunt has since recanted his earlier assertion [85] and has gone on to publish numerous papers in the superresolution field [83, 84, 174, 173, 175, 132, 85, 103, 65, 181]. Jain provides a brief introduction to superresolution for the single frame restoration problem [97].

Superresolution techniques may be applied to SISO problems as well as to the more general MISO or MIMO cases. Superresolution techniques were investigated in the single image restoration problem long before modern multiframe superresolution methods became prominent.

Though SISO superresolution restoration algorithms provide performance improvements over standard restoration techniques, their performance is fundamentally limited by the quantity of the available observed data. With multiframe input algorithms, the totality of information present in an image *sequence* may be utilized to provide additional observation constraints, typically resulting in performance improvements in the restoration process as compared with the single frame scenario.

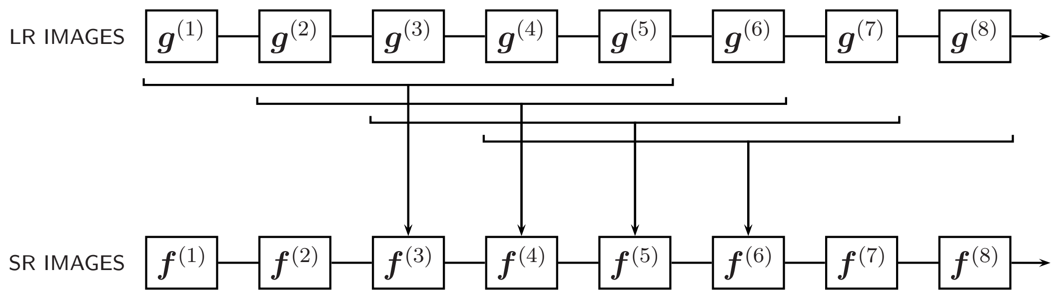


Figure 2.1. Superresolution video restoration consisting of multiple superresolved video stills.

2.2.4 Video Restoration from Superresolution Images

The vast majority of the superresolution restoration algorithms reviewed in Chapter (3) use a short sequence of low-resolution input frames to produce a *single* superresolved high-resolution output frame (the MISO case). All of these techniques may, however, be applied to *video* restoration by using a shifting window of processed frames as illustrated in Figure (2.1). For a given superresolution frame, a "sliding window" determines the subset of low-resolution frames to be processed to produce a given superresolution output frame. The window is moved forward in time to produce successive superresolved frames in the output sequence. Various approaches may be taken to determine the subset of low-resolution frames used to compute the superresolution frames corresponding to the start and end of the observed low-resolution image sequence. An exception to this is the simultaneous multiframe restoration technique developed in Chapter (7).

2.2.5 A Hierarchy of Spatial and Temporal Restoration Techniques

Figure (2.2) illustrates the relationship between five major research thrusts in image and video restoration.

At the lowest level are the single frame image restoration techniques. These include the entire spectrum of single frame restoration, interpolation and superresolution methods referred to previously.

MISO restoration algorithms are shown at the second level. This category contains the bulk of research which is presently applied to superresolution spatial restoration. These MISO techniques are often applied to the problem of spatial resolution enhancement for video (at level three) using the method described in Section (2.2.4).

MIMO methods are found at the third level of the hierarchy. By virtue of having generalized to the multi-output case, it is possible to consider temporal as well spatial restoration using this class of restoration methods. Note that in contrast only spatial restoration is possible at levels one and two. Temporal resampling, also called image sequence interpolation, is one application of these MIMO methods.

In the category of spatial restoration, a wide range of superresolution algorithms are discussed in Chapter (3). The problem of temporal restoration, which includes image sequence interpolation as a special case, is not addressed in this work.

At the highest level of the hierarchy are algorithms which provide spatial as well as temporal resolution restoration. This most general category includes problems such as video standards conversion where video data is converted from one spatiotemporal sampling lattice to another. There is significant scope for future work in the area of superresolution spatiotemporal restoration.

2.2.6 Superresolution Spatiotemporal Restoration Problem Statement

In a fairly general way, the multiframe superresolution spatiotemporal restoration problem may be described as follows:

Let $f(x_1, x_2, t),\ x_1, x_2, t \in \mathbb{R}$ denote the time-varying virtual image of the scene in the image plane coordinate system. Given a sequence of P low-resolution, typ-

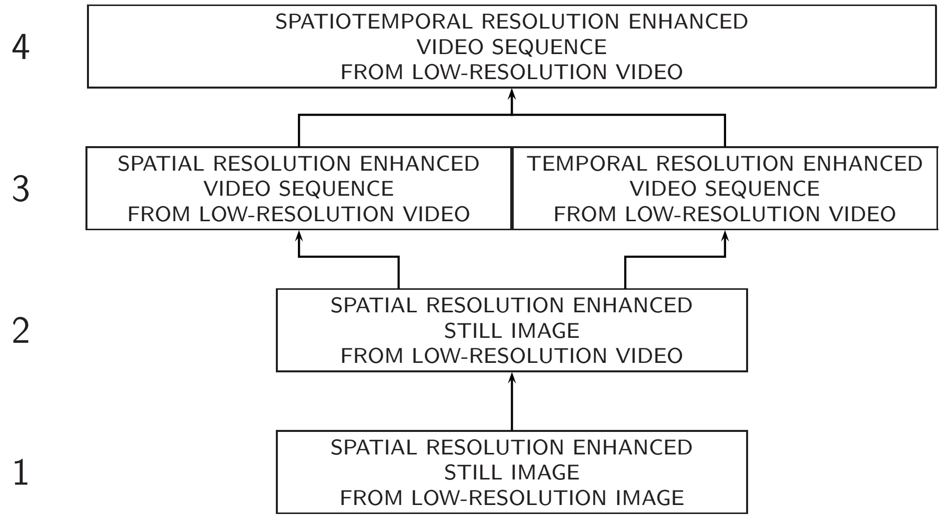


Figure 2.2. Hierarchy describing spatial and temporal superresolution restoration techniques.

ically noisy and undersampled images $g[m_1, m_2, p]$ with $m_1 \in \{1, 2, \ldots, M_{1_p}\}$, $m_2 \in \{1, 2, \ldots, M_{2_p}\}$ and $p \in \{1, 2, \ldots, P\}$ acquired by imaging of the scene $f(x_1, x_2, t)$ at times $t_1 \leq t_2 \leq \cdots \leq t_p \leq \cdots \leq t_P$, the objective is to form S estimates $\hat{f}[n_1, n_2, s]$, $s \in \{1, 2, \ldots, S\}$ of $f(x_1, x_2, \tau_s)$ on the discrete sampling grid indexed by $[n_1, n_2]$ with $n_1 \in \{1, 2, \ldots, N_{1_s}\}$, $n_2 \in \{1, 2, \ldots, N_{2_s}\}$ at the arbitrary time instants $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_s \leq \cdots \leq \tau_S$. Typically $N_{1_s} > M_{1_p}$, $N_{2_s} > M_{2_p}, \forall\ p, s$ and / or $S > P$. Superresolution refers to the restoration of a sequence of images $\hat{f}[n_1, n_2, s]$ that has information content beyond the spatial and/or temporal bandlimit of the imaging system (bandwidth extrapolation).

The research presented in this dissertation is primarily concerned with the problem of superresolution spatial restoration.

2.3 Inverse Problems

In this section a brief overview of some basic concepts in mathematical inverse problems are provided. The discussion here is nontechnical so that the reader may obtain an intuitive grasp of the fundamental ideas prior to the presentation of greater detail in subsequent chapters.

2.3.1 Definition of an Inverse Problem

Keller [104] provides the following description of an inverse problem:

> We call two problems *inverses* of one another if the formulation of each involves all or part of the solution of the other. Often, for historical reasons, one of the two problems has been studied extensively for some time, while the other has never been studied and is not so well understood. In such cases, the former is called the *direct problem*, while the latter is the *inverse problem*.

This description hints at the arbitrariness of the definition as to which problem is considered direct and which is the inverse. Considering some classical examples will clarify the terminology.

Perhaps the historically most celebrated example concerns the solution of the heat equation. Fourier, in his 1816 treatise, "Théorie Analytique de la Chaleur," pioneered the use of the trigonometric series which now bears his name to solve the *direct* heat equation: given an initial temperature distribution at time $t = t_0$, determine the evolution of the temperature profile for times $t > t_0$.

Consider, however, the following: assume that a temperature profile at time $t = t_f > t_0$ is provided. The challenge is to determine the original temperature profile at the earlier time $t = t_0$. This is the *inverse problem* of the direct heat equation. It turns out however that while the direct problem is easily solved, the inverse problem is not.

Another familiar example of a direct/inverse problem pair concerns polynomials. Given a polynomial of order p the direct problem is finding the p roots of the polynomial. The corresponding inverse problem problem is finding the polynomial given the p roots. In the case of polynomials it is clear that the direct problem is more difficult.

The above examples illustrate another typical characteristic of direct/inverse problem pairs. The data for the direct problem is the desired solution of inverse problem and vice versa.

2.3.2 Well-Posed and Ill-Posed Problems

Hadamard, in his work on differential equations [67], classified a problem as *well-posed* if the solution to the problem:

1. exists for any data,
2. is unique, and
3. depends continuously on the data.

Consider again the problem of solving the heat equation. It can be shown that this problem meets all three of Hadamard's requirements, so it is well-posed. Given the initial temperature distribution at time $t = t_0$ and the boundary conditions, it is possible to uniquely determine the temperature distribution for any time $t > t_0$, thus satisfying the first two requirements. The third requirement warrants some further explanation. Intuitively the third condition requires that bounded changes in the initial temperature distribution at time $t = t_0$ lead to bounded changes in the final distribution at time $t > t_0$. This is reassuring as it would be undesirable if arbitrarily small changes in the initial temperature distribution at $t = t_0$ led to unbounded variations in the temperature profile at some later time $t > t_0$. The third condition is, in essence, a stability requirement on the solution.

In contrast, a problem which fails to satisfy any of the Hadamard conditions is said to be *ill-posed*.

Though the problem of solving the direct heat equation is well-posed, its corresponding inverse problem, that of determining the initial temperature distribution given the final temperature distribution, turns out to be be highly problematic. The difficulties are intimately connected to the irrecoverable loss of information.

First, consider the solution of the direct heat equation via the classical Fourier method where a trigonometric series is used to represent the initial temperature distribution at time $t = t_0$ as an infinite summation of sinusoids. Since the Partial Differential Equation (PDE) governing heat transfer is linear, it is straightforward to determine the evolution of the complete solution by considering the superposition of the solutions for the individual sinusoidal components in the Fourier series representing the initial temperature distribution. A characteristic of the heat equation solution is that higher frequency sinusoidal components are rapidly attenuated with increasing time.

Next, consider the effect of measurement accuracy. Any attempt to determine the initial temperature profile will involve some degree of uncertainty due to measurement error, corrupting noise, finite precision representation and so on. The consequence is that there exists some uncertainty in the knowledge of the initial temperature distribution. In the Fourier series representation of the temperature distribution, small Fourier coefficients may not therefore represent meaningful information. Since the direct heat equation rapidly attenuates Fourier coefficients (with more rapid attenuation with increasing frequency) as time evolves, more and more coefficients fall below the threshold dividing meaningful information from noise. The result is irrecoverable information loss. The information content of the solution is lower than that of the initial state. Technically, this is a situation of increasing entropy.

This irrecoverable loss of information does not present significant difficulties for the direct problem. If, however, the objective is to determine the initial temperature distribution from a distribution at some later time, that is, to solve the inverse problem, this information loss has serious consequences. It is not possible to reverse the process exactly and return to the initial condition due to the loss of information. In particular the loss of information implies that there exists a multitude of initial temperature distributions which could have given rise to the observed temperature distribution at time $t > t_0$. In this case, since the inverse problem fails to have a unique solution (Hadamard's second requirement), it is an ill-posed problem.

There are a great many problems of considerable practical importance which violate one or more of the Hadamard requirements. For additional examples, the reader is referred to [114, 50, 10, 194].

2.3.3 Regularization and the Solution to Ill-Posed Problems

Regularization is a term which refers to methods which utilize additional information to compensate for the information loss which characterizes ill-posed problems. This additional information is typically referred to as *a-priori* or *prior* information as it cannot be derived from the observations or the observation process and must be known "before the fact". Typically the prior information is chosen to represent desired characteristics of the solution, for example total energy, smoothness, positivity and so on. The role of the *a-priori* information is to constrain or reduce the space of solutions which are compatible with the observed data. A deterministic theory of regularized solutions to ill-posed problems was pioneered by A. N. Tikhonov [194].

In the Tikhonov approach, a family of approximate solutions to the inverse problem is constructed, with the family of solutions controlled by a nonnegative real-valued *regularization parameter*. The solution family is constructed so that in the noise-free case, in the limit as the regularization parameter tends to zero, the exact solution to the original problem is yielded, while for noisy data, a positive value of the regularization parameter yields an optimal approximate solution. Since Tikhonov's seminal work, stochastic and set-theoretic regularization theories have also been proposed. Stochastic regularization methods are applied in later chapters.

In all these regularization frameworks, the inclusion of *a-priori* information results, by construction, in a new well-posed problem which is closely related to the original ill-posed one. This well-posed problem, the solution of which satisfies the Hadamard requirements, is formulated so that its unique solution is meaningful with respect to the original ill-posed one. It is therefore very important to ensure that the *a-priori* constraints used accurately reflect the required characteristics of the solution. Often physical principles are a useful guide. For instance, in an imaging application, light intensities are always nonnegative — a powerful *a-priori* constraint in bandlimited reconstruction scenarios [41].

2.3.4 Discretization and Ill-Conditioning

In practice, the solution to inverse problems are determined using computational methods. An immediate consequence is that the observed data, the solution, as well as the models used, must be discretized. It has been noted that the discretization of ill-posed problems leads to ill-conditioned numerical problems [10, 69, 50]. In particular, the finer the discretization, the larger the condition number of the corresponding system matrix, leading ultimately to numerical instability in the computational solution of the inverse problem. Regularization methods are applied to these ill-conditioned discretized problems in order to obtain acceptable solutions.

2.3.5 Superresolution Restoration as an Ill-Posed Inverse Problem

One of the recurring issues in this work is that multiframe superresolution restoration is usually an ill-posed inverse problem. There are two dimensions to this assertion:

1. Multiframe superresolution restoration is an inverse problem:

   Referring to Section (2.2.6), a low-resolution image sequence is acquired by an imaging system recording a scene. The imaging process that results in the observed image sequence represents the direct problem. The corresponding inverse problem is that of determining estimate(s) of the scene given the observed image sequence and the characterization of the imaging process. Given the characteristics of the imaging process and system, the forward problem is one of *simulation*, while the inverse problem is one of *restoration*.

2. Multiframe superresolution restoration is an ill-posed problem:

   Recall that ill-posedness implies failure of one or more of the Hadamard conditions. The multiframe superresolution restoration problem may fail to satisfy one or more of these conditions. The failure may result from either the characteristics of the imaging system, or the observed data.

   (a) Nonexistence of the solution

       The presence of noise in the observation process may result in an observed image sequence which, given the imaging system characterization, is inconsistent with any scene. The result is that the system is noninvertible and the scene cannot be estimated from the observations.

   (b) Nonuniqueness of the solution

       When the operator which characterizes the imaging process is many-to-one, there exists a nontrivial space of solutions consistent with any given observed image sequence, that is, the solution to the inverse problem is nonunique. For example, in bandlimited imaging systems all out-of-band scene data represent the nullspace of the imaging process operator.

Even if the imaging operator is nonsingular, a simple lack of data, which represent constraints on the solution space, is sufficient to result in the nonuniqueness of the solution. For example, consider a discretized imaging scenario with P observed low-resolution images each consisting of N pixels. These observed data provide a maximum of PN independent constraints. Assume a single superresolution image containing $M > PN$ pixels is to be estimated from the data. Since the number of unknowns exceeds the number of constraints, it is clear that insufficient constraints exist to ensure the existence of a unique solution to the inverse problem.

Furthermore, since superresolution, by definition, requires the restoration of information that is lost in the imaging process it should be expected that the solution to the superresolution restoration problem is likely to be nonunique.

(c) Discontinuous dependence of the solution on the data

Depending on the characteristics of the imaging system, the inverse problem may be highly sensitive to perturbations of the data. For example, consider an imaging system with a spectral response which decreases asymptotically toward zero with increasing frequency. While such a system is invertible in theory, in practice the inverse is unstable — an arbitrarily small noise component at a sufficiently high frequency leads to an arbitrarily large spurious signal in the computed restoration. In practice such restorations are typically overwhelmed by the amplification of the noise.

While in rare circumstances it happens that the Hadamard conditions are satisfied, in general, practical applications involving multiframe superresolution restoration are invariably ill-posed.

Despite the difficulties caused by the ill-posedness, regularized solution methods enable high quality superresolution restorations as is shown in later chapters. The inclusion of *a-priori* information is crucial to achieving this.

2.4 Applications Utilizing Superresolution Methods

Applications for the techniques of superresolution restoration from image sequences appear to be growing rapidly as the theory gains exposure. Continued research and the availability of fast computational machinery have made these methods increasingly attractive in applications requiring the highest restoration performance. Superresolution restoration techniques have already been applied to problems in:

- Satellite imaging, [203].
- Astronomical imaging [84, 182].
- Video enhancement and restoration, [168].
- Video standards conversion, [147].
- Confocal Microscopy.
- Digital mosaicing, [124].
- Aperture displacement cameras, [121, 153].
- Medical computed tomographic imaging.
- Diffraction tomography.
- Video freeze frame and hard copy.

- Restoration of MPEG-coded video streams, [35, 4, 144, 53, 5] and more recently, [127, 128, 172, 6, 171, 66].

This list is far from exhaustive as new and imaginative applications are constantly being developed.

CHAPTER 3

SURVEY OF MULTIFRAME SUPERRESOLUTION RESTORATION TECHNIQUES

3.1 Introduction

In this chapter a comprehensive review of techniques for multiframe superresolution spatial restoration is presented. The techniques discussed are divided into two broad categories: frequency-domain methods and spatial-domain methods, depending on the domain in which the restoration method is effected.

Although many of the frequency-domain approaches discussed are intuitively simple and computationally attractive, they are poorly suited for accommodating general scene observation models and cannot easily utilize general *a-priori* constraints or noise models.

Spatial-domain formulations, however, provide considerable flexibility in the range of degradations and observation models which may be represented and are well suited to enable the inclusion of *a-priori* knowledge. These benefits come typically at the cost of increased computational complexity.

The literature on frequency and spatial-domain formulations of the multiframe superresolution restoration problem is discussed in Section (3.2) and Section (3.3) respectively.

Recall from Chapter (2) that superresolution restoration is an *ill-posed* inverse problem. Furthermore, it was noted that the problem of ill-posedness could be mitigated through *regularization* with the inclusion of *a-priori* knowledge of characteristics of the solution. In multiframe superresolution restoration, in which one or more superresolved frames are estimated from the observed low-resolution, degraded image sequence, *a-priori* information provides valuable constraints in the regularized solution of the inverse problem. More significantly, however, in the multiframe case improved restoration is possible by simultaneously utilizing information from multiple frames comprising the observed image sequence. In particular, with the availability of multiple observed frames comes the possibility of taking advantage of the relative motion occurring between frames to align or register the observed images so as to provide additional constraints for the restoration of the superresolved frames.

In many of the multiframe superresolution restoration techniques examined in this chapter, constraints resulting from image registration or *motion compensation* as well as *a-priori* constraints will be evident. Also apparent is the use of nonlinear *a-priori* constraints which provide the potential for bandwidth extrapolation beyond the band limit of the imaging system.

Among the spatial-domain restoration techniques, two procedures for computing regularized solutions to the ill-posed superresolution inverse problem have gained prominence: these are the estimation-theoretic, Maximum A-posteriori Probability (MAP) estimation method discussed in Section (3.3.4) and the set theoretic method of Projection Onto Convex Sets (POCS) presented in Section (3.3.5).

In the MAP estimation approach the superresolution inverse problem is viewed as a statistical inference problem and explicit use is made of *a-priori* information in the form of a prior probability density on the solution. This statistical method rests on a rigorous theoretical framework.

The POCS method defines the feasible solution space as the region of intersection of a collection of convex constraint sets. A consequence of the restriction to convex constraint sets is the existence of an iterative procedure to determine solutions satisfying the intersection constraint. An attractive characteristic of the POCS approach is that it provides an intuitive and convenient framework for the inclusion of *a-priori* information.

The MAP estimation and POCS approaches are somewhat complimentary. Hybrid methods which attempt to combine the most desirable characteristics of each technique are discussed in Section (3.3.6).

In Section (3.4) the merits and demerits of the frequency-domain and spatial-domain methods are summarized and within the category of spatial-domain restoration methods, a comparison of the MAP and POCS restoration methods is also presented.

In the discussion that follows, three important factors affecting the performance of multiframe superresolution restoration methods are identified:

1. Accurate and precise subpixel-resolution motion information is required.

2. The observation model must accurately describe the imaging system and its degradations. It must also be capable of describing a wide range of scene motions.

3. The restoration method must provide the maximum potential for inclusion of *a-priori* information.

3.2 Frequency-Domain Superresolution Restoration Methods

In this section restoration methods which formulate and solve the superresolution restoration problem in the frequency-domain are considered. The techniques discussed utilize the shifting property of the Fourier transform [26] to model global translational scene motion and take advantage of results from sampling theory to effect superresolution image restoration from the data available in the observed image sequence.

The early work on superresolution restoration is based on a frequency-domain formulation of the restoration problem. Although somewhat limited in terms of the generality of the approach, this initial work served as a catalyst for the development of more sophisticated frequency-domain formulations as well as the spatial-domain formulations that are discussed in Section (3.3).

Frequency-domain superresolution restoration methods typically rely on familiar Fourier transform properties, specifically the shifting and sampling theorems. Since these properties are generally very well known and understood, the frequency-domain approaches are easy to grasp and are intuitively appealing. Many of the frequency-domain approaches are based on assumptions which enable the use of efficient procedures for computing the restoration, the most important of which is the Fast Fourier Transform (FFT). Unfortunately the very relationship to Fourier transform properties which makes frequency-domain superresolution methods appealing is also responsible for their greatest drawback — they preclude the use of general observation and motion models.

The frequency-domain restoration algorithms discussed here are considered superresolving since they are able to recover frequency information beyond the Nyquist limit. This is achieved by restoring aliasing terms caused by undersampling. This is possible due to the availability of an observed image *sequence* in contrast to the classical single-frame restoration case where this is not possible.

The review of frequency-domain methods begins with the seminal work of Tsai and Huang [203] published in 1984.

3.2.1 Restoration via Alias Removal

The earliest formulation of and proposed solution to the multiframe superresolution restoration restoration problem was undertaken by Tsai and Huang [203] in 1984, motivated by the need to improve the resolution of images acquired by the Landsat 4 satellite. Landsat 4 acquires images of the same area of the earth in the course of its orbits, thus producing a sequence of similar, but not identical, images of a given region on the earth's surface. The observed images are modeled as undersampled images of a static, unknown scene with continuous spatial variables, which is subject to global translational motion. Ideal impulse sampling is assumed, but with the a sampling rate *below* the Nyquist rate [97]. The effects of blurring due to satellite motion during image acquisition and observation noise are ignored.

The authors propose a frequency-domain formulation based on the shift and aliasing properties [26, 155] of the continuous and discrete Fourier transforms for the restoration of a single bandlimited image from the set of undersampled and therefore aliased observed images. The shift and aliasing properties are used to formulate a system of equations which relate the aliased Discrete Fourier Transform (DFT) coefficients of the observed images to samples of the Continuous Fourier Transform (CFT) of the unknown scene from which the observations derive. This system of equations is solved for the frequency-domain coefficients of the original scene, which is then recovered using the inverse DFT. The formulation of the system of equations requires knowledge of the translational motion occurring between the observed frames to subpixel accuracy and precision. The solution of the resulting system of equations requires that the observations contribute *linearly independent* constraints which places restrictions on the range of inter-frame motion that contributes useful data.

Denoting the continuous scene as $f(x_1, x_2)$, global translational motion yields R

shifted images,

$$f_r(x_1, x_2) = f(x_1 + \Delta_{x_{1_r}}, x_2 + \Delta_{x_{2_r}}) \tag{3.1}$$

with $r \in \{1, 2, \ldots, R\}$. The CFT of the scene is given by $\mathcal{F}(u_1, u_2)$ and that of the translations by $\mathcal{F}_r(u_1, u_2)$. The shifted images are impulse sampled to yield R observed images,

$$y_r[m_1, m_2] = f(m_1 T_{x_1} + \Delta_{x_{1_r}}, m_2 T_{x_2} + \Delta_{x_{2_r}}) \tag{3.2}$$

with $m_1 \in \{0, 1, \ldots, M_1 - 1\}$ and $m_2 \in \{0, 1, \ldots, M_2 - 1\}$ where M_1 and M_2 are the dimensions of the sampled images. The sampling periods in the x_1 and x_2 dimensions are given by T_{x_1} and T_{x_2} respectively. The R corresponding 2D DFTs are denoted as $\mathcal{Y}_r[k_1, k_2]$. The CFT of the scene and the DFTs of the shifted and sampled images are related via aliasing [155] by,

$$\mathcal{Y}_r[k_1, k_2] = \frac{1}{T_{x_1} T_{x_2}} \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} \mathcal{F}_r \left(\frac{k_1}{M_1 T_{x_1}} + p_1 f_{s_1}, \frac{k_2}{M_2 T_{x_2}} + p_2 f_{s_2} \right), \tag{3.3}$$

where $f_{s_1} = 1/T_{x_1}$ and $f_{s_2} = 1/T_{x_2}$ are the sampling rates in the x_1 and x_2 dimensions respectively. The aliasing expression of Equation (3.3) establishes the relationship between the DFT of the observed images and the CFT of the translations of the unknown scene.

The shifting property [26] of the continuous Fourier transform,

$$\mathcal{F}_r(u_1, u_2) = e^{i2\pi(\Delta_{x_{1_r}} u_1 + \Delta_{x_{2_r}} u_2)} \mathcal{F}(u_1, u_2), \tag{3.4}$$

describes the frequency-domain consequence of the spatial-domain translation by $\Delta_{x_{1_r}}$ and $\Delta_{x_{2_r}}$. If $f(x_1, x_2)$ is bandlimited, there exist $L_i \in \mathbb{N}$ with $i \in \{1, 2\}$ such that $\mathcal{F}(u_1, u_2) \to 0$ for $|u_i| \geq L_i f_{s_i}$. Assuming $f(x_1, x_2)$ is bandlimited, the infinite summations in Equation (3.3) may be reduced to finite sums. Combining the shifting property of Equation (3.4) it is possible to rewrite the alias relationship of Equation (3.3) in matrix form as,

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{F}. \tag{3.5}$$

In this expression, $\mathbf{Y}$ is a $R \times 1$ partitioned vector, the r^{th} element of which contains the DFT coefficients $\mathcal{Y}_r[k_1, k_2]$ of the observed image $y_r[m_1, m_2]$. $\mathbf{\Phi}$ is a matrix which relates the DFTs of the observed image data to the samples of the unknown CFT of $f(x_1, x_2)$ contained in the $4L_1L_2 \times 1$ vector $\mathbf{F}$.

Superresolution restoration is therefore reduced to finding the DFTs of the R observed images, determining $\mathbf{\Phi}$, solving the system of equations of Equation (3.5) for $\mathbf{F}$ and then using the inverse DFT to obtain the restored image. Tsai and Huang demonstrate that $\mathbf{\Phi}$ may be factored so as to reduce the computational cost in solving for the CFT coefficients in $\mathbf{F}$.

Since the construction of the system matrix $\mathbf{\Phi}$ requires knowledge of the translation parameters Δ_{x_1} and Δ_{x_2} which are not typically known *a-priori*, these parameters must be estimated before restoration is possible. Superresolution restoration is thus effected using a two step process: motion estimation to determine the translation parameters, followed by restoration of the improved-resolution image.

The authors address the problem of registration using a novel approach which appears not to have gained the attention that it probably warrants. Since the observed images $y_r[m_1, m_2]$ are undersampled, there is some question as to the accuracy of standard techniques for motion estimation which typically utilize two (though sometimes more) frames when computing motion estimates. It is widely recognized [168] that the accuracy of the motion estimates is arguably *the* limiting factor in superresolution restoration performance, so any fruitful consideration of this problem promises significant returns. A simultaneous multiframe image registration algorithm is proposed and is shown to deliver reliable registration parameters even under the conditions of severe undersampling, provided that a sufficiently large number of observed frames are available. This approach does not appear to have been sufficiently explored in the superresolution community thus far, and may hold promise as a means for improving the performance of multiframe superresolution restoration methods. It should be noted, however, that the proposed method addresses only the relatively simple case of estimating global translational motion.

The Tsai-Huang frequency-domain restoration approach, though computationally attractive and intuitively simple, has several disadvantages. The assumption of ideal sampling is unrealistic since real world imaging systems are characterized by sampling sensors which perform spatial as well as temporal integration (during the *aperture time*).

Additionally, no consideration is given to the effects of the optical system on the recorded image data. The issue of noise degradation is also ignored. This is a major shortcoming given that the presence of noise has a significant deleterious effect on the feasibility of solving Equation (3.5).

The global translational motion model considered by the authors, while arguably appropriate for the problem domain addressed, is unacceptable in more general contexts. It is essential to recognize that the global translational motion model is intimately bound with the frequency-domain approach. Indeed, as discussed above, global translational motion in the spatial domain has an exact dual in the Fourier-domain, manifested as phase shifts. It is exactly this fortuitous relationship which makes the frequency-domain formulation possible. However, it is not difficult to see that this relationship is exceptional. More general motion models, especially those which are spatially-varying, do not have Fourier-domain analogs and are thus not amenable to a frequency-domain restoration approach. This inherent limitation represents probably the most serious disadvantage of any frequency-domain approach to multiframe superresolution restoration. This is unfortunate since many of the most interesting applications of multiframe superresolution restoration involve general motion fields.

Two of the limitations of the Tsai-Huang method are addressed by Tekalp, Ozkan and Sezan [193]. The authors propose a frequency-domain approach which extends [203] by including the effects of a Linear, Shift-Invariant (LSI) Point Spread Function (PSF) $h(x_1, x_2)$ as well as observation noise. Periodic sampling is still assumed but impulse sampling is not since the PSF of the sensor can be incorporated into the system PSF. As in [203], a global translational motion model is used which enables the formulation of a Fourier-domain system model which relates the noisy, degraded and undersampled observed images to the original 2-D scene. By utilizing L^2 observed images, each L times undersampled in the x_1 and x_2 dimensions, a system of L^2 equations in L^2 unknowns is formulated for each sampled frequency-domain point. These equations are solved in the least squares sense due to the presence of observation noise.

The authors assume the spatial-domain observation model,

$$y_r[m_1, m_2] = \sum_{p_1=0}^{M_1-1} \sum_{p_2=0}^{M_2-1} \Big\{ \big[f(x_1 - \Delta_{x_{1_r}}, x_2 - \Delta_{x_{2_r}}) * h(x_1, x_2) \big] \cdot \delta(x_1 - p_1 T_{x_1}, x_2 - p_2 T_{x_2}) + n_r(p_1 T_{x_1}, p_2 T_{x_2}) \Big\}, \tag{3.6}$$

where $*$ denotes convolution. Similar to [203], the equivalent frequency-domain equations are derived for Equation (3.6) which relate the DFT coefficients of the undersampled observed images of the shifted and blurred original scene to samples of the CFT of the unknown original image. After computing the least squares solution to the frequency-domain equations, an estimate of the original scene is obtained via the inverse DFT of the sampled CFT data.

In addition to the extensions to [203], the authors propose a spatial-domain POCS restoration method as well as a spatial-domain interpolation-restoration approach. These are discussed in Section (3.3.5) and Section (3.3.1) respectively.

In the text, "Digital Video Processing" [192], Tekalp dedicates a chapter to a review of selected techniques in multiframe superresolution restoration. Tekalp provides an introduction to the superresolution problem and discusses various observation models which are capable of including the effects of scene motion and sensor and optical system PSFs. The frequency-domain model of Tsai and Huang [203], as well as the extensions proposed in [193] are presented in a tutorial exposition. The POCS method proposed in [193] (see Section (3.3.5)) is also reviewed and examples are presented.

Kaltenbacher and Hardie [100] utilize the frequency-domain formulation proposed by Tsai and Huang [203] for the restoration of a single unaliased frame from a sequence of undersampled low-resolution frames but propose an alternate method for estimating the inter-frame global translational motion parameters required for restoration. Their subpixel motion estimation procedure, based on a single frame Taylor series expansion, is cheaper computationally than that of [203], though the theoretical foundations for the proposed estimation of motion from the gradient information in a single frame is open to question.

3.2.2 Recursive Least Squares Methods

This section discusses computationally efficient Recursive Least Squares (RLS) solution to Equation (3.5).

Kim, Bose and Valenzuela [108] utilize the frequency-domain restoration framework as well as the global translational motion model proposed by Tsai and Huang [203], however they extend the formulation to consider observation noise as well as the effects of spatial blurring. An excellent review of the frequency-domain restoration method precedes the authors' primary contribution — a RLS and a weighted RLS solution method for solving the linear system of equations of Equation (3.5) in the presence of observation noise. Sufficient conditions to ensure that the system matrix $\mathbf{\Phi}$ in Equation (3.5) is nonsingular are derived in terms of the global translational motion parameters relating the undersampled observations of the scene. The proposed recursive solution approach is computationally efficient and the least squares formulation provides the advantage of improved solutions in the case of noise or an underdetermined or overdetermined system in Equation (3.5).

Although the authors address some of the limitations of the Tsai-Huang [203] method by considering LSI blur which removes the restriction of impulse sampling, Linear, Shift-Varying (LSV) blur and observation noise, the restriction to a global translational motion observation model remains. Additionally, the problem of estimating the motion parameters is not addressed. It is assumed that these parameters are known *a-priori*. The authors observe that the presence of zeros in the spectral representation of the LSI blur results in the restoration problem being ill-posed. This is due to the nonuniqueness of the solution caused by the loss of information at the response zeros.

Kim and Su [109, 110] extend the work of Kim, Bose and Valenzuela [108] by addressing the issue of the ill-posedness of the restoration inverse problem caused by the presence of zeros in the spectral representation of the blur PSF. The formulation of the superresolution restoration problem follows the Tsai-Huang [203] approach, but includes a LSI PSF in the frequency-domain observation model. Since zeros in the spectrum of the blur PSF result in an ill-posed inverse problem [119], the authors propose replacing their earlier RLS solution for Equation (3.5) with a method based on Tikhonov regularization [194].

In the presence of observation noise, the system of equations $\mathbf{Y} = \mathbf{\Phi F}$ will typically be inconsistent as $\mathbf{F}$ will lie outside the range space of $\mathbf{\Phi}$. Additionally,

$\mathbf{\Phi}$ is usually ill-conditioned or singular due to zeros in the spectral representation of the blur PSF. A regularized solution [194] may be obtained by finding $\mathbf{F}$ which minimizes the expression,

$$\|\mathbf{\Phi F} - \mathbf{Y}\|^2 + \gamma(\mathbf{F}), \tag{3.7}$$

where $\gamma(\cdot)$ is a *regularizing* or *stabilizing functional*. In [109, 110],

$$\gamma(\mathbf{F}) = \lambda\|\mathbf{F} - \mathbf{c}\|^2,$$

where $\mathbf{c}$ is an approximation to the (as yet unknown) solution. Algebraic manipulation yields the minimum of the expression in Equation (3.7) as,

$$\hat{\mathbf{F}} = \left(\mathbf{\Phi}^\top\mathbf{\Phi} + \lambda\mathbf{I}\right)^{-1}\left(\mathbf{\Phi}^\top\mathbf{Y} + \lambda\mathbf{c}\right). \tag{3.8}$$

Further manipulations show that it is possible to solve Equation (3.8) using an iterative method, thus obviating the matrix inversion step. Since the solution is unknown, $\mathbf{c}$ is initially set to zero and thereafter set to the result of the previous iteration.

Although this paper addresses the ill-posedness of the superresolution restoration inverse problem, the proposed approach does have shortcomings. The stabilizing functional, based on the squared error, is unrealistic for images as it leads to overly smoothed solutions. Furthermore, the use of an estimate of the unknown solution $\mathbf{c}$ leaves unanswered questions as to the stability and convergence properties of the proposed recursive solution method. This important point is not addressed.

Su and Kim [190] partially address another limitation of the Tsai-Huang [203] formulation — the inability to accommodate nonglobal motion models. The basic Tsai-Huang frequency-domain restoration framework is, however, still used. The proposed method allows a limited degree of nonglobal translational motion by utilizing the Tsai-Huang approach on overlapping motion compensated subblocks of the observed image sequence.

In this method, observed images are decomposed into overlapping blocks and inter-frame translational motion is estimated for these blocks using the procedure described in [111]. The Tsai-Huang frequency-domain restoration procedure is then performed independently for each block. Interpolation is used to correct for coverings and uncoverings which may be present in the observed image sequence. The resulting restored blocks are assembled to produce the final image.

The extension of the Tsai-Huang [203] method to accommodate nonglobal translational motion is a useful addition to the frequency-domain superresolution framework, but the limitation to *translational motion* remains. This is due to the fact that the restoration procedure remains the frequency-domain approach based on the Fourier shifting property which models only translational motion. There remains the nontrivial problem of choosing the best block-based decomposition from the image sequence, as well as the performance of the algorithm in cases where there is rapid motion or multiple independent motions within blocks. Although effective in the test cases, the algorithm presented is ad-hoc in nature especially with regard to the merging of the restored blocks.

3.2.3 Recursive Total Least Squares Methods

An extension to the RLS approach discussed in Section (3.2.2) is that of Recursive Total Least Squares (RTLS) which is known to provide a degree of robustness to errors in the matrix $\mathbf{\Phi}$ as well as to errors in the observation vector $\mathbf{Y}$ in the solution of linear systems of the form $\mathbf{Y} = \mathbf{\Phi}\mathbf{F}$. In the case of multiframe superresolution restoration, errors in $\mathbf{\Phi}$ are likely and result from errors in motion estimation and uncertainty in the observation model. RTLS is a computationally efficient implementation of the Total Least Squares (TLS) method, the theory of which is well developed. The reader is referred to [61, 62] for additional information.

Apart from the use of TLS to obtain improved solutions to the superresolution restoration problem, methods considered in this section differ little from those already discussed. This is especially true with respect to the observation models used.

Bose, Kim and Valenzuela [23, 22] extend the ideas developed in their earlier work [108, 110] to include a degree of robustness to errors in both the matrix $\mathbf{\Phi}$ and the observation vector $\mathbf{Y}$ in Equation (3.5) by using the TLS [61, 62] approach. Since motion estimation errors (manifested as errors in the system matrix $\mathbf{\Phi}$) tend to have serious adverse effects on the restoration quality, the appeal of TLS methods, which provide some robustness to such errors, is clear.

The observation model used is of the form,

$$\mathbf{Y} = [\mathbf{\Phi} + \mathbf{E}]\,\mathbf{F} + \mathbf{N}, \tag{3.9}$$

where $\mathbf{E}$ is the error in $\mathbf{\Phi}$ resulting from errors in the motion estimates. The TLS problem in [23, 22] is formulated as a minimization problem,

$$\begin{aligned} &\text{minimize} && \|\, [\ \mathbf{N} \vdots \mathbf{E}\ ]\, \| \\ &\text{subject to} && \mathbf{Y} - \mathbf{N} = [\mathbf{\Phi} + \mathbf{E}]\mathbf{F}. \end{aligned} \tag{3.10}$$

The Frobenius norm (the matrix norm defined as the square root of the sum of the squares of the absolute value of the matrix elements) is used. A RTLS algorithm based on [39, 40] is used to solve Equation (3.10) efficiently. Since the formulation of the restoration problem is in the frequency domain, utilizing complex DFT coefficients, Equation (3.10) is transformed to an equivalent real-valued minimization. As with their earlier work [108], the proposed algorithm is well suited to massively parallel implementation since the DFT values of $\mathbf{F}$ may be computed in a decoupled manner.

Though attention is directed to the problem of uncertainties in the global translational motion parameter estimates, the proposed method does not address more fundamental issues such as the inherent limitations of the underlying frequency-domain approach which cannot incorporate general scene or camera motion models. Though not explicitly demonstrated in the paper, the proposed method can incorporate a LSI blur as discussed in [108, 109, 110]. Motion blur due to nonzero aperture time is not considered.

Bose, Kim and Zhou [24], and Kim [106] provide a theoretical analysis of the performance of their RTLS restoration algorithm based on existing work in TLS theory. They show that under certain assumptions, the TLS formulation reaches the Cramér-Rao bound.

3.2.4 Multichannel Sampling Theorem Methods

Ur and Gross [204] propose a superresolution restoration method based on the generalized sampling theorem of Papoulis [142] and a variant thereof, the multichannel sampling theorem of Brown [28]. Although the implementation of the restoration is achieved in the spatial domain, the technique is fundamentally a frequency-domain technique as it relies on the shift property of the Fourier transform to model the translation of the source image.

The authors consider a σ-bandlimited (bandlimited to $-\sigma<\omega<\sigma$) function $f(x)$ which is passed through R "mutually independent" linear channels, the outputs of which are sampled at the rate $2\sigma/R$ (undersampled at $1/R$ times the Nyquist rate) to produce R discrete signals $y_r(mT)$ with $T = R/2\sigma,\ m \in \mathbb{Z}$ and $r \in \{1, 2, \ldots, R\}$. By the multichannel sampling theorem [28], $f(x)$ may be perfectly reconstructed from $y_r(mT)$ by passing the subsampled signals through R linear filters, summing the outputs and performing bandlimited interpolation. Specifically,

$$\hat{f}(x) = \sum_{r=1}^{R} \hat{f}_r(x) = \sum_{r=1}^{R} \sum_{m=-\infty}^{\infty} y_r(mT) \cdot h_r(x - mT), \tag{3.11}$$

where $\hat{f}(x)$ is a sampled version of $f(x)$ meeting the Nyquist criterion and which may be interpolated to recover $f(x)$ exactly. The R filters with impulse response $y_r(m)$ may be found according to equations given in [142] and [28]. A frequency-domain derivation is used to obtain this result. In particular, in Brown's formulation [28] of the multichannel sampling theorem, matrices containing samples from the frequency-domain representation of the R sampled signals are used to determine the set of filters $h_r(\cdot)$ which are applied in the time domain to the R degraded and subsampled signals to yield a sampled version of $f(t)$. This is similar to the frequency-domain formulation of Tsai and Huang [203], in which systems of equations in the frequency domain are solved to yield the frequency-domain representation of $f(t)$. The multichannel sampling theorem yields a time-domain method (multichannel linear filtering) which computes the equivalent restoration as the Tsai-Huang method which operates by solving a linear system of frequency-domain equations.

Ur and Gross consider the linear degradation channels to include the effects of a blur PSF as well as global translational motion which is modeled as a delay. Observing that the operations of blurring and translation commute and assuming a single blur common to all the channels, it is shown that the superresolution restoration problem may be separated into two distinct processes: "merging" the undersampled signals into a single σ-bandlimited function, followed by deblurring of the merged signal. Since the deblurring operation is independent of the merging process, it is possible to derive a closed form solution for computing the merged signal from the R degraded and undersampled channel outputs. Any standard image restoration algorithm may then be applied to the merged signal to correct for the blur degradation.

The Ur-Gross method requires that the global translational motion parameters be known *a-priori*. The motion estimation problem is not addressed.

The proposed restoration approach is, in effect, a spatial-domain analog of the Tsai-Huang frequency-domain formulation, the only significant difference being the inclusion of a single PSF common to all the observations. Observation noise and motion blur is not considered. Since the Ur-Gross proposal is effectively a spatial-domain implementation equivalent to the frequency-domain methods discussed in Section (3.2.1), it suffers from the same limitations in regard to motion models. Like the frequency-domain formulations discussed earlier, this work is limited to global translational motion models.

3.2.5 Summary of Frequency-Domain Superresolution Restoration Methods

Frequency-domain superresolution restoration methods offer the following advantages:

- Simplicity

  The theoretical principles which underlie frequency-domain approaches to superresolution restoration are readily understandable in terms of elementary Fourier theory. Though additional complexities exists in the implementation, the fundamental ideas are clear.

- Computational complexity

  Many of the techniques discussed in this section are computationally attractive. In the Tsai-Huang formulation and its variants, the system of frequency-domain equations are decoupled, that is, a single frequency sample point of the superresolution image may be computed independently of any other. This fact makes this formulation highly amenable to massively parallel implementation.

- Intuitive superresolution mechanism

The techniques discussed are all based on the dealiasing techniques of [203]. This formulation makes explicit the underlying mechanism of resolution improvement — restoration of frequency components beyond the Nyquist limit of the individual observed images. In many of the methods discussed later in this work the mechanism whereby superresolution is achieved is less transparent.

Unfortunately there are significant drawbacks to frequency-domain superresolution restoration methods which have to be addressed if more general video restoration problems are to be tackled:

- Limitation to a global translational motion model

  It is important to note that all the frequency-domain methods utilize an essentially global translational motion model. Even [190], which describes a method with local motion by a block decomposition of the low-resolution observed images, falls into this category since only a global translation model for the decomposed blocks is possible. This limitation is no coincidence — it is a fundamental limitation of the frequency-domain approach.

- Inflexibility regarding motion models

  Frequency-domain methods require the existence of a Fourier-domain transformation which is the dual of the spatial-domain motion model. This requirement imposes severe limitations on the range of feasible motion models due to the inability to formulate Fourier-domain transformations which are equivalent to spatially varying motion models typically required in sophisticated multiframe restoration applications.

- Degradation models

For reasons similar to the previous point, it is also difficult to include spatially varying degradation models in the frequency-domain restoration formulation.

- Inclusion of spatial-domain *a-priori* knowledge for regularization

  Since superresolution restoration is typically an ill-posed inverse problem, regularized solution methods are often required to achieve acceptable results. Though some of the methods discussed include regularizing functionals, these functionals are, by virtue of their formulation in the Fourier domain, spatially invariant. Often the most useful *a-priori* constraints are expressed in the spatial domain and may be spatially-varying. Frequency-domain restoration methods are poorly suited to the inclusion of such constraints.

Where the global translational motion model is appropriate, there are significant benefits to be gained by utilizing frequency-domain approaches to multiframe superresolution restoration. Many emerging superresolution restoration application are, however, characterized by the need to include more general motion models, as well as degradations. In these cases the frequency-domain approach is not suitable.

In the sections that follow, the literature of spatial-domain multiframe superresolution restoration methods is explored. These methods address all the disadvantages noted for frequency-domain approaches, but typically at the cost of a significant increase in computational demands.

3.3 Spatial-Domain Superresolution Restoration Methods

In the previous section, the literature on frequency-domain approaches to the multiframe superresolution problem was reviewed in some detail. The second major category of multiframe superresolution restoration comprises methods which are based on a *spatial-domain* formulation and solution of the restoration problem. The major advantage of these spatial-domain methods is that they provide:

- Generality regarding observation models.

  Formulating the observation model directly in the spatial domain provides enormous flexibility in modeling realistic image scenarios and degradations. In particular it is possible to accommodate:

  – arbitrary motion models (global or nonglobal)

  – motion blurring

  – optical system degradations (spatially varying or invariant)

  – nonideal sampling (spatially varying or invariant)

  – complex degradations (such as compression blocking artifacts)

- Powerful methods for the inclusion of *a-priori* constraints.

  In order to address the ill-posedness of the multiframe superresolution restoration problem, it is preferable to utilize solution methods which facilitate the inclusion of *a-priori* constraints. Formulating and solving the restoration problem in the spatial domain allows the use of powerful spatial-domain image models. Examples include:

  – Markov Random Field (MRF) models

  – set-based constraints (POCS formulation)

  – nonlinear models capable of bandwidth extrapolation

  – nonstationary models

This discussion of spatial-domain multiframe superresolution restoration techniques will span a broad range and begins with the intuitive, but oversimplistic, interpolation-based methods in Section (3.3.1). This is followed in Section (3.3.2), by a technique borrowed from the computed tomography community — algebraic filtered back-projection. This method is based on a discretized linear observation model. Similar linear models are used in almost all modern spatial-domain superresolution restoration formulations. The computed tomographic approach is extended in Section (3.3.3) where iterative back-projection methods are examined. This is followed by a detailed review of stochastic methods in Section (3.3.4) and the competing set theoretic methods presented in Section (3.3.5). The bulk of this presentation is dedicated to the examination of these two important approaches. Hybrid restoration methods, which combine elements of the stochastic and set theoretic approaches to restoration, are discussed in Section (3.3.6). Methods derived from optimal and adaptive filter theory are presented in Section (3.3.7) and those based on Tikhonov regularization in Section (3.3.8). Section (3.3.9) includes a summary of all the techniques discussed.

3.3.1 Interpolation of Nonuniformly-Spaced Samples

In this section a simple approach to constructing superresolved images from an image sequence based on spatial-domain interpolation is described. All the frames in the low-resolution observed image sequence are geometrically registered to a fixed reference frame, resulting in a composite image of nonuniformly-spaced samples. These nonuniformly-spaced sample points are then interpolated and resampled on a regularly-spaced high-resolution sampling lattice. Though this approach may initially appear attractive due to its highly intuitive nature, it is overly simplistic as it does not take into consideration the fact that the samples in the observed low-resolution images are nonideal. In reality, the image samples are a function of the characteristics of the imaging system and may include effects such as spatial and temporal blurring caused by the lens and sensor. The drawback of this simplifying assumption is that the reconstructed image does not contain the full range of frequency content that could be restored given a more realistic imaging model.

Keren, Peleg and Brada [105] describe a spatial-domain approach to image registration using a global translation and rotation model and a two stage approach to superresolution restoration.

The image registration procedure assumes R observed frames from which a gradient-based procedure estimates translation parameters and rotation parameters for each of $R-1$ images relative to a chosen reference image. With the image registration parameters estimated, the first stage of the restoration procedure registers the observed images and a high-resolution restoration grid is imposed on the "stack" of observed images. Each pixel in the high-resolution image is then chosen as the outlier-trimmed mean of the values in the set of observed image pixels the centers of which fall within the area of the high-resolution pixel under consideration.

The resulting "merged" image, which is a composite of the R observed low-resolution images may be subjected to further processing such as deblurring or may be used as an initial estimate in an iterative restoration algorithm as is the case in the second phase of the restoration procedure described in [105]. This iterative procedure is discussed later in Section (3.3.3).

Though [105] does contribute early ideas concerning image registration, the proposed image interpolation procedure is overly simplistic in its approach to the superresolution problem. No attempt is made to dealias the observed data — a feature of all the frequency-domain methods in Section (3.2) and though perhaps not as explicitly, many of the spatial-domain methods discussed later. In effect, only a fraction of the available information is utilized in the restoration. The proposed method does not provide a systematic framework for achieving superresolution restoration.

There is little indication as to the nature of the image improvement and no analysis of the method is presented.

Another interpolation-based approach is proposed by Aizawa, Komatsu and Saito [2]. The authors examine the problem of acquiring high-resolution images from stereo cameras. Though ostensibly distinct from the problem of superresolution restoration from an image sequence, it is in fact an analog in that an image pair captured using a single camera with relative camera/scene motion produces, like a stereo pair, similar but not identical views of the scene. Of course the range of possible motion is more general in the case of temporally distinct images, however under the observation model assumptions used, this is not a significant complication.

By considering the spatial frequency response of a regularly-spaced two dimensional array of sampling apertures (the FPA pixels) the authors show that there is a limit, governed by Nyquist's theorem, to the extent of frequency components which may be reconstructed from the sampled image. By considering the possibility of sampling at spatial positions *between* the array pixels, it is demonstrated that the effective frequency response of the combined (double image) system is increased.

In practice this is implemented by utilizing two (and in their later papers, more) cameras which are positioned so as to provide images which are shifted by subpixel distances relative to the first (reference) image. The image from the second camera is registered relative to the reference image sampling lattice to subpixel accuracy using gradient-based or block-matching techniques. Due to effects such as perspective projection, differing camera alignment, etc., the samples from the second camera do not, in general, fall regularly between the reference sampling grid, but are spaced nonuniformly. The result is a set of regularly-spaced samples (from the reference image) among which are interspersed irregularly-spaced samples from the second image. In order to reconstruct a high-resolution frame, the authors are thus faced with the problem of interpolation and resampling to obtain the desired high-resolution image. This is effected using a method for restoration from nonuniformly-spaced samples [37]. The idea is demonstrated in the one dimensional case. Consider the function $f(x)$ sampled at the nonuniformly-spaced locations x_m for $m \in \mathbb{Z}$. If a one-to-one mapping $\gamma(\cdot)$ and positive constant T exist such that $mT = \gamma(x_m)$ and if $f(\gamma^{-1}(x))$ is bandlimited to $\omega_0 = \pi/T$ then it can be shown that $f(x)$ may be reconstructed as

$$f(x) = \sum_{m=-\infty}^{\infty} f(x_m) \frac{\sin[\omega_0(\gamma(x) - mT)]}{\omega_0(\gamma(x) - mT)}. \tag{3.12}$$

This is immediately recognized as bandlimited sinc interpolation with a nonlinear mapping used to correct for the nonuniform spacing of the samples. The case for two dimensions entails minor additional complications regarding the choice of $\gamma(\cdot,\cdot)$, but the principle remains the same.

In later papers, Aizawa *et. al.* extend these ideas to the case of multiple cameras (closer to the multiframe image restoration problem), and an iterative simulate-and-correct approach based on the Landweber method [120] is used to estimate the uniformly-spaced samples of the superresolution restoration. These papers are discussed later in this section.

Tekalp, Ozkan and Sezan [193] propose, among other techniques discussed in Section (3.2.1) and Section (3.3.5), a two step procedure where the upsampling of the low-resolution images is followed by restoration to correct for the sensor PSF. The low-resolution frames are registered and combined to form an intermediate high-resolution image consisting of nonuniformly-spaced samples. This image is then interpolated and resampled on a uniform grid to produce a high-resolution frame with uniformly-spaced samples. The authors suggest the use of the thin-plate spline method of Franke [58], the iterative Bayesian method of Sauer and Allebach [159] or the POCS-based method of Yeh and Stark [208] to achieve the interpolation and resampling.

Since the interpolated image still suffers from sensor blur as well as errors resulting from the interpolation process itself, a final restoration step is applied. The authors suggest the use of any of the commonly known deconvolution methods that take into account the presence of noise.

Aizawa, Komatsu, Saito and Igarashi extend [2] by introducing an iterative superresolution restoration procedure based on their earlier merging and interpolation work. In [3, 116, 115], the problem of acquiring high-resolution imagery from two or more cameras is considered and is extended in [117, 158] to include cameras which differ in the dimensions of the FPA pixels. The objective is to produce images at a resolution which exceeds the physical resolution limits of the individual sensor arrays. As discussed earlier, this problem is closely related to that of superresolution image restoration from a single camera image sequence.

As in [2], the low-resolution observed images are merged using motion compensation to form a nonuniformly-sampled high-resolution image $\tilde{\mathbf{z}}$. Desired, however, is the uniformly-sampled image $\mathbf{z}$. Various interpolation-based techniques are considered [37, 107, 57, 30], however a procedure based on the Landweber method [120] is used to iteratively estimate the uniformly-sampled high-resolution image from the nonuniformly-sampled image produced by the registration process. The desired uniformly-sampled image $\mathbf{z}$ and the merged image $\tilde{\mathbf{z}}$ are related by

$$\tilde{\mathbf{z}} = \mathbf{A}\mathbf{z}, \tag{3.13}$$

where $\mathbf{A}$ represents the nonuniform sampling process. Inversion of Equation (3.13) is infeasible due to the dimensionality and likely singularity of $\mathbf{A}$. Instead, an iterative procedure which corrects for the nonuniform sampling process is used. An initial guess $\mathbf{z}^{(0)}$ for the high-resolution image (with uniformly-spaced samples) is made and is then iteratively updated using the recurrence formula,

$$\mathbf{z}^{(j+1)} = \mathbf{z}^{(j)} + \alpha \cdot \mathbf{A}^* \left(\tilde{\mathbf{z}} - \mathbf{A}\mathbf{z}^{(j)} \right), \tag{3.14}$$

where $\mathbf{A}$ is a linear approximation to the nonuniform sampling process, $\mathbf{A}^*$ is the adjoint operator of $\mathbf{A}$, α is a control parameter and $\tilde{\mathbf{z}}$ is the merged image with nonuniformly-spaced samples. The parameter α is chosen so as to ensure that the iteration described by Equation (3.14) is convergent.

In Section (3.3.3) an identical Landweber-type iteration is encountered, however it is important to note that the operation represented by $\mathbf{A}$ in that case models the *imaging process*, not the mapping between a pair of superresolution images as is the case here.

The authors later propose a single camera temporal integration imaging procedure which utilizes a sequence of images from a single camera to compute a single superresolved image [136, 135]. Unlike their earlier multiple camera work, these papers address the familiar multiframe superresolution restoration problem. The proposed method utilizes motion estimation and image registration using deformable quadrilateral patches, followed by a restoration phase which interpolates from nonuniformly-spaced samples. An innovation is that the restoration method includes spatial smoothness constraints and therefore may be considered to fall into the class of regularized solution methods. An iterative procedure is used to solve the resulting variational problem. The proposed method does not, however, consider the effects of noise or image blurring.

The techniques discussed in this section have the advantage of simplicity but do not address several important issues. The observation models used are generally unrealistic as they do not properly account for the effects of optical blurring, motion blurring or noise. Though [193] considers an observation PSF, there is some question as to the optimality of *separate* merging and restoration steps. Furthermore, few of the methods discussed include *a-priori* information in an attempt to address the ill-posedness that is typical of the multiframe superresolution restoration problem.

3.3.2 Algebraic Filtered Back-Projection Methods

Some work has been done in applying methods commonly used in the computed tomography community to the problem of multiframe superresolution restoration. This section addresses superresolution restoration by means of the method known as algebraic Filtered Back-Projection (FBP). In this method, the restoration problem (more commonly referred to as a reconstruction problem in the tomography literature) is formulated as a discrete, linear inverse problem which is then solved in a single step. A "back-projection" operator, which is often the adjoint operator of the linear operator approximating the observation model, is used in the solution, thus lending the name "back-projection". It is also common to include an inverse filtering step to improve the quality of the restoration thus explaining the term "filtered".

An early algebraic tomographic FBP approach to superresolution restoration is that of Frieden and Aumann [59]. The authors do not consider the problem of superresolution image restoration from an image sequence, but the related problem of superresolution image restoration from multiple 1-D scans of a stationary scene by a linear FPA. Noting that the PSF in the 1-D scan system represents a line integral while the PSF of the multiframe superresolution restoration problem represents a spatial integral, it is clear that the problems differ only in the form of the imaging system PSF. The salient point is that similar restoration techniques are applicable in both cases. In [59], the dimensions of the detector elements in the linear sensor array are assumed to be greater than the limiting resolution of the optical system. The same is typically true for area-scan imaging systems. In [59] the imaging geometry provides overlapping scans of a given scene area, enabling restoration at a resolution higher than the limiting spatial sampling rate of the sensor array.

The imaging process is described by the discrete linear model

$$\mathbf{y} = \mathbf{H}\mathbf{z}, \tag{3.15}$$

where $\mathbf{y}$ is a vector of the data observed during successive image scans, $\mathbf{H}$ is the system matrix which models the imaging process and $\mathbf{z}$ is a vector containing the unknown image to be estimated. Observation noise and/or a lack of data, as well as the considerable size of the matrix $\mathbf{H}$ preclude the use of direct inversion to estimate $\mathbf{z}$. Instead, the authors propose that an estimate be formed using the back-projection of the observed data as,

$$\hat{\mathbf{z}} = \mathbf{H}^{\top}\mathbf{y}. \tag{3.16}$$

It is important to note that the back-projection matrix $\mathbf{H}^{\top}$ is not, in general, equivalent to the inverse $\mathbf{H}^{-1}$. Using Equation (3.15) gives,

$$\hat{\mathbf{z}} = \mathbf{H}^{\top}\mathbf{H}\mathbf{z}. \tag{3.17}$$

If and only if $\mathbf{H}^{\top}\mathbf{H} = \mathbf{I}$ will the back-projection be equivalent to the inverse. In the given application, the matrix $\mathbf{H}^{\top}\mathbf{H}$ is Toeplitz and is thus associated with a well-defined point spread function, allowing fast Fourier-domain inverse filtering to restore the back-projected observed data. In the computed tomography literature, this technique is known as Filtered Back-Projection. In the application considered by Frieden and Aumann, however, the point spread function differs from that typically found in the case of computed tomographic reconstruction.

It is instructive to note that the application of an inverse filter for $\mathbf{H}^{\top}\mathbf{H}$ yields an estimate,

$$\hat{\mathbf{z}} = \left(\mathbf{H}^{\top}\mathbf{H}\right)^{-1}\mathbf{H}^{\top}\mathbf{y}, \tag{3.18}$$

which should be immediately recognized as the Linear Minimum Mean Square Error (LMMSE) estimate for $\mathbf{z}$.

Frieden and Aumann make no allowances for observation noise. This has serious consequences since inverse filtering is known to be highly noise sensitive due to the typically increasing amplitude response of the inverse filter with increasing frequency. The proposed restoration method does not fall into the class of regularized methods and as such cannot be expected to yield acceptable performance in the presence of noise or missing data. Despite these drawbacks, this paper has the distinction of being an early attempt to apply tomographic image reconstruction techniques to the superresolution restoration problem. Later papers extend these basic ideas to include iterative and regularized solutions to the linear inverse problem formulation of the superresolution restoration problem.

3.3.3 Iterative Back-Projection Methods

A large class of iterative multiframe superresolution restoration methods have in common a simple but powerful simulate-and-correct approach to restoration. Since a very similar approach finds widespread use in computed tomographic reconstruction where it is known as Iterative Back-Projection (IBP), this name will be adopted to describe the approach when applied to the problem of multiframe superresolution restoration. The general procedure is introduced in an informal way first, then specific approaches are discussed.

A set of low-resolution, possibly degraded images which derive from an unknown high-resolution scene are observed. The objective is to determine the unknown high-resolution scene. Although the scene remains unknown, it is assumed that some *estimate* of the scene is available. Given knowledge of the imaging process relating the scene to the observed image sequence, it becomes possible to *simulate* the output of the imaging system given the estimate of the original scene. The simulated images may then be compared with the observed data and a residual error can be formulated. A small residual error suggests that the scene estimate is an accurate one, whereas a significant residual error indicates that the estimate is poor. The error information can, however, be used to improve the scene estimate. By a process called "back-projection," the error residual is used to form an updated estimate of the scene which is a better approximation of the original. The approximation is better in the sense that the result of the simulated imaging process using the updated estimate has reduced residual error as compared with the earlier estimate. This process may be iterated in the hope of reducing the simulation residual error to a minimum. The iterative process thus comprises two steps: simulation of the observed images, and back-projection of the error to correct the estimate of the original scene.

In this section several approaches which fall into the category of iterative back-projection methods are examined while in later sections more sophisticated methods which extend these basic ideas are developed. The discussion will show that it is possible to cast these restoration procedures into a common theoretical framework, with the differences in approach evident in the choice of the projection and back-projection operators as well as the error functional to be minimized.

Consider the linear observation model relating the observed low-resolution images (described by the stacked vector $\mathbf{Y}$ of lexicographically ordered pixels of each image) to the unknown discretized scene $\mathbf{z}$ via the operator $\mathbf{H}$ which typically models motion compensation and subsampling. The situation is described algebraically as,

$$\mathbf{Y} = \mathbf{H}\mathbf{z}. \tag{3.19}$$

Given an estimate $\hat{\mathbf{z}}^{(j)}$ of the superresolution restoration of the scene $\mathbf{z}$, it is possible to *simulate* the low-resolution observed images by applying Equation (3.19) to the superresolution estimate $\hat{\mathbf{z}}^{(j)}$ as,

$$\hat{\mathbf{Y}}^{(j)} = \mathbf{H}\hat{\mathbf{z}}^{(j)}. \tag{3.20}$$

IBP procedures update the estimate of the superresolution restoration by "back-projecting" the error between the simulated low-resolution images $\hat{\mathbf{Y}}^{(j)}$ and the observed low-resolution images $\mathbf{Y}$ via the back-projection operator $\mathbf{H}^{BP}$ which models the apportioning of "blame" for errors in the projection to pixels in the superresolution estimate $\hat{\mathbf{z}}^{(j)}$. Typically $\mathbf{H}^{BP}$ is designed to approximate the inverse of the operator $\mathbf{H}$. This iterative procedure is formally described by,

$$\begin{aligned} \hat{\mathbf{z}}^{(n+1)} &= \hat{\mathbf{z}}^{(j)} + \mathbf{H}^{BP}\left(\mathbf{Y} - \hat{\mathbf{Y}}^{(j)}\right) \\ &= \hat{\mathbf{z}}^{(j)} + \mathbf{H}^{BP}\left(\mathbf{Y} - \mathbf{H}\hat{\mathbf{z}}^{(j)}\right) \end{aligned}. \tag{3.21}$$

It is interesting to note that this simulate-and-correct approach bears close similarity to many well known solution methods to seemingly unrelated problems. In computed tomographic reconstruction for instance, the iteration of Equation (3.21) is known as Iterative Back-Projection (IBP); the operator $\mathbf{H}$ is a projection matrix which describes the interaction of the penetrating radiation beam with the object and $\mathbf{H}^{BP}$ is a back-projection operator. In the solution of integral equations however, Equation (3.21) is of the form of the Landweber iteration [120]. The choice of the projection and back-projection operators in Equation (3.21) varies widely.

Peleg, Keren and Schweitzer [152] and Keren, Peleg and Brada [105] present the initial ideas which, in their later papers, are developed into the IBP algorithms for superresolution restoration. In [152], superresolution restoration from a set of globally translated images of an unchanging 2-D scene is considered, while in [105], a more general global translation and rotation model is used. The method for estimating the motion parameters for this model is briefly described in Section (3.3.1). In [152, 105], the effects of a spatially invariant PSF are considered.

The authors propose an iterative procedure for estimating the unknown scene. Given an estimate of the superresolution image, a simulated imaging process yields a set of low-resolution images which are compared with the observed low-resolution images. The superresolution estimate is then modified so as to reduce the error between the observed and simulated images. The process is terminated when the error has reached some predetermined level, or after a given number of iterations.

Images $y_r[m_1, m_2]$ with $r \in \{1, 2, \ldots, R\}$ are observed. A superresolution image, coincident with one of the R observed images, is to be reconstructed from the low-resolution sequence. The j^{th} estimate of the superresolution image is given by $\hat{z}^{(j)}[n_1, n_2]$. Assuming the imaging process which resulted in the observed images is known, it is possible to simulated the imaging process with the estimate $\hat{z}^{(j)}[n_1, n_2]$ as an approximation to the original scene, yielding R low-resolution simulated images $\hat{y}_r^{(j)}[m_1, m_2]$. The error between the j^{th} iteration simulated images and the original observed images is defined as,

$$\epsilon^j = \sum_{r=1}^{R} \sum_{m_1, m_2} \left| \hat{y}_r^{(j)}[m_1, m_2] - y_r[m_1, m_2] \right|. \tag{3.22}$$

Since it is desired that the simulated images be identical to the observed images, the error ϵ^j must be minimized. Peleg *et. al.* propose the following modification of the current estimate $\hat{z}^{(j)}[n_1, n_2]$. For some pixel $[n_1, n_2]$ if $\hat{z}^{(j)}[n_1, n_2] = l$, then consider modifications $\hat{z}^{(j)}[n_1, n_2] = l - 1$ and $\hat{z}^{(j)}[n_1, n_2] = l + 1$. Compute for these three pixel values $\{l-1, l, l+1\}$ the corresponding simulated low-resolution images and the corresponding errors. Choose the pixel value for which the error in Equation (3.22) is a minimum. Repeat the process for each pixel in the image over several iterations.

This simple optimization attempts to monotonically decrease the error on a pixel by pixel basis. From their experimental results, the authors observe that their iteration tends to terminate in local minima. This problem is addressed by utilizing a simulated annealing [113] approach for optimization. Though it has been shown in [60] that simulated annealing is provably convergent to the global minimum given a suitable *annealing schedule*, the required schedule is too slow to be of practical utility. Selecting an appropriate annealing schedule to deliver useful performance in acceptable time is often difficult and since the process is run for a finite time, there remains a finite probability of attaining a local minimum even with this algorithm. Convergence of the proposed optimization is thus slow.

The issue of the uniqueness of the solution is not addressed. Since the dimensionality of the unknown scene is greater than that of the observed data, it is likely that there exist multiple images which, under the given imaging process, yield sets of observed images which are identical. The proposed framework provides no manner in which to favor certain solutions over others. Furthermore, the particular solution reached in the space of possible solutions will be dependent on the initial estimate for the superresolution image, as well as the order in which pixels are updated. The effects of noise on the stability and quality of the solution is inadequately addressed.

The method proposed for back-projecting the error is highly simplistic as compared with common tomographic reconstruction methods for example. The proposed correction mechanism makes very limited use of the available knowledge concerning the characteristics of the imaging system.

Irani and Peleg [87, 88] extend the earlier work [152, 105] by improving the method of back-projecting the error between the simulated low-resolution images and the observed data. They modify the error functional in Equation (3.22) to the form,

$$\epsilon^j = \sqrt{\sum_{r=1}^{R} \sum_{m_1,m_2} \left(y_r[m_1,m_2] - \hat{y}_r^{(j)}[m_1,m_2] \right)^2}. \tag{3.23}$$

They also consider a simulated imaging process which includes a point spread function to model degradation in the imaging system. The most important modification to [152, 105] is, however, the use of a back-projection kernel in the update of the estimate $\hat{z}^{(j)}[n_1, n_2]$. In the proposed update procedure, the value of each high-resolution pixel $\hat{z}^{(j)}[n_1, n_2]$ is updated according to *all* low-resolution pixels which are dependent on its value (via the simulated imaging process). The contribution of a low-resolution pixel $\hat{y}_r^{(j)}[m_1, m_2]$ is the error $y_r[m_1, m_2] - \hat{y}_r^{(j)}[m_1, m_2]$ multiplied by a factor which measures the relative contribution of the high-resolution pixel $\hat{z}^{(j)}[n_1, n_2]$ to the low-resolution pixel $y_r[m_1, m_2]$. The result is that errors in low-resolution pixels which are strongly influenced by a particular high-resolution pixel contribute more strongly to the corrective term applied to the estimate $\hat{z}^{(j)}[n_1, n_2]$. The correction may be written in the form,

$$\begin{aligned} \hat{z}^{(j+1)}[n_1, n_2] &= \hat{z}^{(j)}[n_1, n_2] + \\ &\sum_{r=1}^{R} \sum_{\zeta} \left(\hat{y}_r[m_1, m_2] - \hat{y}_r^{(j)}[m_1, m_2] \right) \cdot h^{BP}[n_1 - m_1, n_2 - m_2], \end{aligned} \quad (3.24)$$

where ζ is the set of low-resolution pixels $[m_1, m_2]$ in the j^{th} low-resolution image dependent on the high-resolution pixel $\hat{z}^{(j)}[n_1, n_2]$. $h^{BP}[\cdot, \cdot]$ is a *back-projection kernel* which scales the correction to the high-resolution estimate depending on the influence of a given high-resolution pixel on the low-resolution pixel for which the error is computed. The authors observe that h^{BP} may be chosen so as to influence the solution to which the iteration converges. The back-projection kernel may therefore be used to incorporate additional constraints on the solution such as smoothness. It is important to note, however, that these constraints are linear in the data. Nonlinear constraints will be encountered in later papers. A theorem providing a sufficient condition on h^{BP} for convergence in the special case of deblurring is proved.

Irani and Peleg continue to develop their early ideas [87, 88] in [89] where they propose a very general procedure for superresolution restoration of scenes which contain arbitrary independent motion. Their primary contribution is not in the superresolution restoration algorithm per sé, which is unchanged from [87, 88], but is the incorporation of an algorithm which tracks multiple independent motions. This development allows superresolution restoration for partially occluded objects, transparent objects or a given object of interest.

The motion detection and estimation algorithms are discussed in in detail in [90] so the principles are only summarized here. The proposed approach uses an iterative procedure to identify and track multiple independent motions occurring in the scene. A single motion is first computed and an object corresponding to this motion is identified and tracked. This is called the "dominant" motion, and object, respectively. Once the dominant object is identified and tracked, it is excluded from the region of analysis and the process is repeated to identify other objects and their motions. The procedure is terminated when no further objects are identified. Object motion may be modeled using any 2-D parametric motion model. The procedure is also extended to include a "temporal integration" approach which builds an internal representation for each moving object over the course of several frames.

Shah and Zakhor [178, 179] utilize a Landweber [120] restoration method similar to that of Irani and Peleg. Their primary contribution, however, is not in the area of restoration techniques, but is a novel approach to resolving problems caused by unreliable motion estimates in the superresolution restoration. A per-pixel block-based motion estimation method in which a *set* of possible motion vectors are stored for each pixel in each low-resolution frame is suggested.

The authors illustrate how standard block matching techniques may select incorrect motion estimates even though they maximize some correlation measure. This fact is well recognized — motion estimation, like superresolution restoration, is an ill-posed problem (see Chapter (4)). Often multiple candidate motions are equally "optimal" with respect to the block correlation measure. Since the effectiveness of the superresolution restoration scheme is critically dependent on the accuracy of the estimated motion vectors [168], erroneous motion estimates are highly undesirable as they are manifested in the form of objectionable artifacts in the restored image. Given the likelihood of errors in motion estimation, the situation often arises where the true motion is rejected in favor of an incorrect estimate which nevertheless has a higher block correlation measure. By choosing to maintain a *set* of "near optimal" matches, however, recourse to other feasible motion estimates is possible if the initial motion estimate for a particular pixel is found to be inconsistent with those for surrounding pixels.

IBP methods provide a useful framework for solving the superresolution problem by providing a mechanism for constraining the superresolution restoration to conform to the observed data. Though this is intuitively appealing, there are several complications with this simple approach which should be considered.

- Accuracy of the imaging model

  The simulate-and-correct procedure is only as good as the imaging model. If the imaging model is inaccurate then it is impossible to recover the unknown scene from the observed data. It may only be possible to approximate the original scene within some error dictated by the accuracy of the observation model. This remains true for any restoration method which utilizes a model of the imaging system.

- Criteria for determining the error metric

  There are many choices regarding the way in which an error between the simulated and observed images is computed. The issue of optimality is immediate.

- The method for correcting the scene estimate given the simulation error

  Even when an error metric is determined there remains the question as to how to correct the estimate of the unknown scene given the error between the observed and simulated images. The back-projection of the error to correct the estimate requires an approximation of the inverse of the imaging system model which may be difficult to find or may exhibit instabilities.

- Solution existence

  Few of the papers discussed addressed the question of the existence of the solution. There are several scenarios which can lead to nonexistence of solution. For example, observation noise leads to conflicting constraints and an inconsistent system of observation equations. In this case it is impossible to find a solution to the restoration inverse problem.

- Solution uniqueness

  Often it is assumed that minimizing the simulation error implies a reasonable solution (if a solution even exists). In fact, minimizing the error does not necessarily imply a reasonable solution. In superresolution image restoration for example, the number of unknowns is often far greater than the number of measured datum. This implies that the imaging system represents a many to one mapping — there are many unknown images which are consistent with the observed data. Thus merely minimizing (or even nulling) the error between the observed and simulated images is insufficient to differentiate among the multitude of superresolution images which satisfy the observed data constraints.

  As a corollary, a convergent iteration need not converge to a unique solution. Differences in the initial guess for bootstrapping the iteration, or even the order of pixel updates can lead to convergence to quite different solutions, all of which satisfy the observed data constraints. Needless to say, it is necessary to determine conditions under which the iteration converges and whether or not that solution is unique.

- Stability of the solution with respect to the data

  In much of the work discussed, it was assumed that there was no observation noise and that the addition of small amounts of noise would not drastically affect the solution. Unfortunately these assumptions are often false. Even physically realizable systems are capable of exhibiting unbounded variation in the solution for a bounded variation of the observed data.

The reader should immediately recognize the three preceding points as the Hadamard requirements for a well-posed problem [67]. A powerful method for addressing the problems of nonexistence, nonuniqueness and continuity failings of ill-posed problems is through the use of *a-priori* knowledge which constrains the solution space. This is not easily achieved within the simulate-and-correct frameworks discussed thus far. Though Irani and Peleg [87, 88] recognize that the choice of the back-projection kernel may impose smoothness constraints on the restoration, their approach is neither flexible or general.

Tools to elegantly and effectively tackle ill-posed problems do, however, exist. In the following sections consideration is given to methods which utilize *a-priori* constraints to obtain *regularized* solutions to ill-posed inverse problems. Stochastic methods are discussed in Section (3.3.4) and the set-theoretic POCS formulation in Section (3.3.5). Approaches which utilize Tikhonov regularization are also briefly mentioned in Section (3.3.8).

It is interesting to note that though the stochastic and POCS methods are not derived from the Landweber or simulate-and-correct iterations discussed earlier, these methods typically include the Landweber iteration as a special case since they incorporate constraints which ensure that the superresolution restoration is consistent with the observed data. The methods differ, however, in that the restoration need *also* match the *a-priori* constraints. This inclusion of *a-priori* constraints makes high-quality solutions to even ill-posed inverse problems possible.

3.3.4 Stochastic Methods

In Chapter (2) it was shown that superresolution restoration is typically an ill-posed inverse problem. It was also noted that the inclusion of *a-priori* information in the solution process provides a mechanism for obtaining physically meaningful solutions to this class of problem. Recently, researchers have recognized the utility of stochastic methods for solving ill-posed problems. Unlike their deterministic predecessors, these methods use a stochastic model for the observation process and apply well known results from statistical estimation theory to obtain regularized solutions to the inverse problem. The so-called "Bayesian" methods are particularly well suited to the solution of ill-posed inverse problems as they provide a natural framework for the inclusion of *a-priori* constraints in the form of a prior probability density function on the unknown. These Bayesian techniques have proved highly successful and are now central to the solution of ill-posed inverse problems in a wide variety of applications. The Bayesian approach is almost synonymous with MAP estimation which is the starting point of this discussion of stochastic techniques. Other related methods are also briefly addressed.

Earlier, a linear, spatial-domain observation model was introduced to describe the observed image acquisition process from the unknown scene. The task of estimating the unknown scene thus becomes a problem of solving a linear inverse problem. The essential difference between the work presented thus far and the stochastic formulation that is discussed here is that instead of treating the observations and the unknown as deterministic, these quantities will be considered as stochastic — that is, it is assumed that the observations, unknowns and noise are realizations of stochastic processes (with associated probability distributions) and are related via stochastic equations. The presentation begins with a summary of the Bayesian MAP formulation as applied to the linear inverse problem.

MAP Restoration Methods

Assume that the linear stochastic observation model relating an unknown $\mathbf{f}$ to the observation $\mathbf{Y}$ which includes additive noise $\mathbf{N}$ is described by,

$$\mathbf{Y} = \mathbf{H}\mathbf{f} + \mathbf{N}. \tag{3.25}$$

The MAP approach to estimating the unknown vector $\mathbf{f}$, attempts to find the estimate $\hat{\mathbf{f}}_{\mathrm{MAP}}$ for which the *a-posteriori* probability, $\mathcal{P}\left\{\mathbf{f}|\mathbf{Y}\right\}$ is a maximum. More formally, the objective is to find $\hat{\mathbf{f}}_{\mathrm{MAP}}$ as,

$$\hat{\mathbf{f}}_{\mathrm{MAP}} = \arg\max_{\mathbf{f}} \left[\mathcal{P}\left\{\mathbf{f}|\mathbf{Y}\right\}\right]. \tag{3.26}$$

Using Bayes' rule [143] to expand Equation (3.26) yields,

$$\hat{\mathbf{f}}_{\mathrm{MAP}} = \arg\max_{\mathbf{f}} \left[\frac{\mathcal{P}\left\{\mathbf{Y}|\mathbf{f}\right\}\mathcal{P}\left\{\mathbf{f}\right\}}{\mathcal{P}\left\{\mathbf{Y}\right\}}\right] \tag{3.27}$$

The maximum of this expression for $\hat{\mathbf{f}}_{\mathrm{MAP}}$ is independent of $\mathbf{Y}$ so,

$$\hat{\mathbf{f}}_{\mathrm{MAP}} = \arg\max_{\mathbf{f}} \left[\mathcal{P}\left\{\mathbf{Y}|\mathbf{f}\right\}\mathcal{P}\left\{\mathbf{f}\right\}\right]. \tag{3.28}$$

Since the logarithm is a monotonic increasing function, this is equivalent to finding,

$$\hat{\mathbf{f}}_{\mathrm{MAP}} = \arg\max_{\mathbf{f}} \Big[\log\mathcal{P}\left\{\mathbf{Y}|\mathbf{f}\right\} + \log\mathcal{P}\left\{\mathbf{f}\right\}\Big], \tag{3.29}$$

where $\log\mathcal{P}\left\{\mathbf{Y}|\mathbf{f}\right\}$ is called the *log-likelihood function* and $\log\mathcal{P}\left\{\mathbf{f}\right\}$ is the logarithm of the *a-priori* density or *prior* for $\mathbf{f}$. Maximizing the *a-posteriori* probability thus requires maximizing a sum of two terms. The first term, $\log\mathcal{P}\left\{\mathbf{Y}|\mathbf{f}\right\}$, describes in a stochastic fashion, constraints relating the observations to the unknown to be estimated. The second term, $\log \mathcal{P}\{\mathbf{f}\}$, describes, again in a stochastic sense, knowledge concerning the form of the unknown. It is this term which makes the Bayesian formulation powerful, as it enables the inclusion of constrains which favor physically meaningful solutions within the space of all possible solutions.

Continuing with the derivation, since $\mathbf{Y} = \mathbf{Hf} + \mathbf{N}$, it is easy to see that the likelihood function is determined by the probability density of the noise $f_{\mathbf{N}}(\cdot)$ as,

$$\mathcal{P}\{\mathbf{Y}|\mathbf{f}\} = f_{\mathbf{N}}(\mathbf{Y} - \mathbf{Hf}). \tag{3.30}$$

Often the noise is assumed to be Gaussian and in this case the use of the natural logarithm in the above derivation cancels the exponentiation in the density $f_{\mathbf{N}}(\cdot)$. Regarding the choice of the prior density on $\mathbf{f}$, there is much flexibility. In recent years the MRF model, which is well suited to modeling image data, has gained prominence. MRFs possess a Gibbs probability density of the form,

$$\mathcal{P}\{\mathbf{f}\} = \frac{1}{\mathcal{Z}} \exp\left\{-\frac{1}{\beta}\mathcal{U}(\mathbf{f})\right\}, \tag{3.31}$$

where $\mathcal{Z}$ is a normalizing constant, β is the so called "temperature" parameter and $\mathcal{U}(\mathbf{f})$ is the "energy" of the configuration $\mathbf{f}$. Given the exponential form of the Gibbs density, the use of the natural logarithm in the formulation for the MAP solution greatly simplifies algebraic and computational manipulation.

If the noise is assumed to be Gaussian and the prior is chosen to be a convex function of $\mathbf{f}$, then it can be shown [166] that the optimization of Equation (3.26) is convex, so that the solution $\hat{\mathbf{f}}_{\mathrm{MAP}}$ is assured to exist and is unique. This is very significant since it implies that even an ill-posed problem can, with sufficient constraints on the unknown, have a solution that exists and is unique.

The survey of the MAP superresolution restoration literature begins with the work of Schultz and Stevenson. The relatively detailed development will provide a clear theoretical framework which is representative of many of the existing MAP restoration methods, allowing other work to be mentioned in terms of its departures from this work.

Schultz and Stevenson extend their earlier work [163, 164] on Bayesian image interpolation using a Huber Markov Random Field (HMRF) prior model to the problem of superresolution image restoration in [166, 165, 167, 168]. They propose a linear observation model which is characterized by a motion compensating subsampling matrix which accounts for both scene and camera motion. Blurring due to motion during a nonzero aperture time is not considered in their initial work. The observation model used by Schultz and Stevenson, which remains essentially unchanged in their later papers, is summarized below.

Assume that an odd number p, low-resolution frames $y[m_1, m_2, k]$ with $k \in \{c - \frac{p-1}{2}, \ldots, c, \ldots, c + \frac{p-1}{2}\}$ and $m_1 \in \{1, 2, \ldots, M_1\}$, $m_2 \in \{1, 2, \ldots, M_2\}$ are observed. The objective is to restore a superresolution image $f[n_1, n_2, c]$ with $n_1 \in \{1, 2, \ldots, qM_1\}$, $n_2 \in \{1, 2, \ldots, qM_2\}$ and $q \in \mathbb{N}$, coincident with $y[m_1, m_2, c]$, the center frame in the observed image sequence. A subsampling model which describes the spatial integration of sensors in the FPA is proposed for the c^{th} observed frame as,

$$y[m_1, m_2, c] = \sum_{n_1 = qm_1 - q + 1}^{qm_1} \sum_{n_2 = qm_2 - q + 1}^{qm_2} f[n_1, n_2, c]. \tag{3.32}$$

This relationship for the c^{th} (center) frame may be more succinctly written using lexicographic ordering of the low-resolution and superresolution images as,

$$\mathbf{y}_c = \mathbf{H}_c \mathbf{f}_c, \tag{3.33}$$

where $\mathbf{H}_c \in \mathbb{R}^{M_1 M_2 \times q^2 M_1 M_2}$ is the subsampling matrix relating the superresolution image $\mathbf{f}_c$ with the observed frame $\mathbf{y}_c$. The remaining observed images $\mathbf{y}_k$ are related to $\mathbf{f}_c$ via *motion-compensated* subsampling matrices which compensate for the effects of motion occurring between frames as,

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{f}_c + \mathbf{u}_k, \quad \text{for} \quad k \in \left\{ c - \frac{p-1}{2}, \ldots, c-1, c+1, \ldots, c + \frac{p-1}{2} \right\}. \tag{3.34}$$

$\mathbf{H}_k \in \mathbb{R}^{M_1 M_2 \times q^2 M_1 M_2}$ is the motion-compensated subsampling matrix which relates the k^{th} low-resolution observation to the superresolution image $\mathbf{f}_c$ which is temporally coincident with the center frame in the low-resolution sequence. The vector $\mathbf{u}_k$ contains pixels which are unobservable from $\mathbf{f}_c$, but present in $\mathbf{f}_k$. The elements of $\mathbf{u}_k$ are not known since $\mathbf{f}_c$ is unknown. Notice that rows of $\mathbf{H}_k$ which contain useful information are those for which elements of $\mathbf{y}_k$ are observed *entirely* from motion compensated elements of $\mathbf{f}_c$. To improve robustness, rows for which this is not true are removed, yielding a reduced set of equations,

$$\mathbf{y}'_k = \mathbf{H}'_k \mathbf{f}_c. \tag{3.35}$$

In practice, $\mathbf{H}'_k$ is unknown and must be estimated from the observed low-resolution frames $\mathbf{y}_k$ and $\mathbf{y}_c$. Since estimating $\mathbf{H}'_k$ requires that the motion be estimated from the observed low-resolution image sequence, it is all but certain that estimation errors will be introduced. To account for this, an additive noise term is introduced and an *estimated* observation matrix $\hat{\mathbf{H}}'_k$ extends the observation model to,

$$\mathbf{y}'_k = \hat{\mathbf{H}}'_k \mathbf{f}_c + \mathbf{n}_k, \tag{3.36}$$

where $\mathbf{n}_k$ contain the errors resulting from the use of the estimate $\hat{\mathbf{H}}'_k$. Schultz and Stevenson assume that the elements of $\mathbf{n}_k$ are Independent, Identically Distributed (IID) Gaussian random variables.

With p observed frames, this results in the system of equations,

$$\begin{array}{lllll}
\mathbf{y}'_{c-\left(\frac{p-1}{2}\right)} & = \hat{\mathbf{H}}'_{c-\left(\frac{p-1}{2}\right)} & \mathbf{f}_c & + & \mathbf{n}_{c-\left(\frac{p-1}{2}\right)} \\
\vdots & \vdots & \vdots & & \vdots \\
\mathbf{y}'_{c-1} & = \hat{\mathbf{H}}'_{c-1} & \mathbf{f}_c & + & \mathbf{n}_{c-1} \\
\mathbf{y}_c & = \mathbf{H}_c & \mathbf{f}_c & + & \mathbf{0} \\
\mathbf{y}'_{c+1} & = \hat{\mathbf{H}}'_{c+1} & \mathbf{f}_c & + & \mathbf{n}_{c+1} \\
\vdots & \vdots & \vdots & & \vdots \\
\mathbf{y}'_{c+\left(\frac{p-1}{2}\right)} & = \hat{\mathbf{H}}'_{c+\left(\frac{p-1}{2}\right)} & \mathbf{f}_c & + & \mathbf{n}_{c+\left(\frac{p-1}{2}\right)}.
\end{array} \tag{3.37}$$

The above system of equations may be written using stacked vectors and partitioned matrices in the simplified form,

$$\mathbf{Y} = \mathbf{H}\mathbf{f}_c + \mathbf{N}. \tag{3.38}$$

The unknown image $\mathbf{f}_c$ is estimated using the MAP criterion from the set of observed images $\{\mathbf{y}_k\}$ as,

$$\hat{\mathbf{f}}_c = \arg\max_{\mathbf{f}_c} \Big[\log \mathcal{P}\left\{\mathbf{f}_c | \{\mathbf{y}_k\}\right\} \Big], \tag{3.39}$$

which after applying Bayes' rule may be written as,

$$\hat{\mathbf{f}}_c = \arg\max_{\mathbf{f}_c} \Big[\log \mathcal{P}\left\{\mathbf{f}_c\right\} + \log \mathcal{P}\left\{\{\mathbf{y}_k\} | \mathbf{f}_c\right\} \Big]. \tag{3.40}$$

For the prior term $\log \mathcal{P}\left\{\mathbf{f}_c\right\}$, Schultz and Stevenson use the HMRF model which is a discontinuity-preserving image model which imposes smoothness constraints on the restoration while still allowing the restoration of visually important edge features (discontinuities).

It is assumed that the motion estimation errors are independent so the likelihood term may be written in the form $\mathcal{P}\left\{\{\mathbf{y}_k\}|\mathbf{f}_c\right\} = \prod_{\mathrm{k}} \mathcal{P}\left\{\mathbf{y}_k|\mathbf{f}_c\right\}$. Taking into account that $\mathbf{H}_c$ is known exactly, finding $\hat{\mathbf{f}}_c$ requires the solution of the constrained optimization,

$$\hat{\mathbf{f}}_c = \arg\max_{\mathbf{f}_c \in \mathcal{F}} \Big[\log \mathcal{P}\left\{\mathbf{f}_c\right\} + \log \mathcal{P}\left\{\{\mathbf{y}_k\}_{k \neq c} | \mathbf{f}_c\right\} \Big] \tag{3.41}$$

where the constraint set $\mathcal{F}$ is given by,

$$\mathcal{F} = \Big\{ \mathbf{f} : \mathbf{H}_c \mathbf{f} = \mathbf{y}_c \Big\}. \tag{3.42}$$

This constrained optimization is solved using a gradient projection technique [9]. The implementation details may be found in [168].

The observation model used by Schultz and Stevenson assumes local translational motion. Scene motion is estimated using an hierarchical subpixel-resolution block-matching estimator for both computational efficiency as well as to ensure consistency in the motion estimates.

Schultz and Stevenson have also successfully applied their superresolution restoration technique to the problem of scan conversion where interlaced-scan images are combined to produce motion compensated, deinterlaced, superresolution images [169].

Schultz, Stevenson and Meng, noting the importance of accurate motion information for superresolution restoration have examined techniques for subpixel motion estimation. In [161, 162] the MAP superresolution restoration procedure of [168] is reviewed and various motion estimation methods are studied. In particular, a new block motion estimation procedure which utilizes an eight parameter projective model (based on [125]), as well as the Horn-Schunck [81] optical flow method are discussed and utilized in various superresolution restoration experiments. Methods for detecting and eliminating inaccurate motion vector estimates are also presented. The result is improved superresolution restoration since objectionable artifacts resulting from incorrect motion vectors are minimized.

Noting that motion estimation is itself an ill-posed problem, Schultz and Stevenson [170] propose the use of a regularized motion estimation procedure which favors smoothness of the motion fields while still allowing discontinuities. This is achieved using Bayesian motion estimation techniques and an HMRF *a-priori* smoothness constraint with Horn-Schunck optical flow estimation. Superresolution image restorations from a low-resolution image sequence with motion estimated using the Bayesian motion estimator are compared with those computed using the block motion estimation technique of [168] and are found to produce visually superior results.

It is interesting to note that in [161, 170, 162] motion estimates are computed from only image *pairs*. It is reasonable to anticipate improved reliability in the motion estimates computed using multiple images. Also of note is the fact that the motion estimation techniques are not object-based. No attempt is made, as was the case with [89] to construct internal representations for multiple independent motions within the scene. These points appear to represent open research topics.

Hardie, Barnard and Armstrong [72] present a MAP superresolution restoration procedure which is essentially identical in form to that of Schultz and Stevenson [168]. They consider the cases of global as well as nonglobal motion estimation. For the former, they use the translation and rotation model of [88] and for the latter, they draw on optical flow techniques [191] with foreground/background segmentation. The superresolution formulation is equivalent to that of Schultz and Stevenson under the assumption of Gaussian noise and a Gaussian MRF prior model. This work is extended in [70] to consider the problem of simultaneous estimation of both the superresolution image $\hat{\mathbf{f}}$ and the motion parameters $\hat{\mathbf{s}}$. A MAP formulation similar to that of [72] is used, but included in the optimization are the unknown motion parameters as,

$$\hat{\mathbf{f}}, \hat{\mathbf{s}} = \arg\max_{\mathbf{f},\mathbf{s}} \Big[\mathcal{P} \left\{ \mathbf{f}, \mathbf{s} | \mathbf{Y} \right\} \Big]. \tag{3.43}$$

Since estimation of $\hat{\mathbf{f}}$ requires the motion estimates $\hat{\mathbf{s}}$, a procedure is suggested in which $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$ are estimated alternately, that is, $\hat{\mathbf{s}}$ is fixed and $\hat{\mathbf{f}}$ estimated, and then $\hat{\mathbf{f}}$ fixed and $\hat{\mathbf{s}}$ estimated. Though this procedure can be slow to converge, it does have advantages. Motion is no longer estimated from the low-resolution images directly as is the case in most superresolution restoration algorithms. Instead, the motion parameters are chosen as those which minimize the error between simulated low-resolution observation images (as computed using the observation model and the current superresolution image estimate) and the observed low-resolution frames. Gaussian priors are used for both the superresolution image and the motion. The optimization of Equation (3.43) is achieved using Iterative Coordinate Descent (ICD). The authors observe that it is not possible to use the faster conjugate gradient method, since the cost functional is nonconstant from iteration to iteration due to the joint estimation of both the motion parameters and the superresolution restoration. The alternating estimation of the motion parameters $\hat{\mathbf{s}}$ and the restoration $\hat{\mathbf{f}}$ is reminiscent of the Expectation Maximization (EM) algorithm [42] which provides a general framework for statistical estimation under the conditions of incomplete data (See Appendix A).

A more detailed exposition of an observation model which includes global rotation, translation and a sensor and optical system PSF may be found in [71] where image registration is performed using the method of Irani and Peleg [88] and Lucas and Kanade [123]. In this paper, which estimates motion in a separate step from the superresolution restoration phase, gradient descent and conjugate gradient descent methods are used to maximize the posterior probability.

Lorette, Shekarforoush and Zerubia [122] also present a MAP formulation of the superresolution restoration problem. Elsewhere [180] they consider the problem of 3-D superresolution restoration, but [122] studies three nonlinear penalty functions for their MRF prior model which allow edge restoration while ensuring smoothness.

This is similar in function to the edge preserving HMRF image model used by Schultz and Stevenson.

Cheeseman, Kanefsky, Kraft and Stutz [34, 32, 33] have applied MAP superresolution restoration techniques to images received from remote vehicles exploring the solar system, in particular from the Viking, Voyager and more recently Mars Pathfinder missions. Their formulation assumes Gaussian noise and utilizes prior terms which lead to a linear system of equations which are solved using Gauss-Jacobi methods.

Hunt, Sementilli, Mariappan, Marcellin, Nadar, Yuen and Sheppard have pursued work on a Poisson MAP superresolution formulation and have provided a theoretical analysis of the performance of this superresolution technique in [83, 84, 174, 175, 210, 85]. This work concentrates on the restoration of astronomical images and makes assumptions concerning bounded support of the reconstructed objects which are not valid in the more general problem of superresolution restoration of video. The assumption of Poisson photon statistics in the observation model, although essential for effective restoration of astronomical images, is less relevant to video restoration applications where low photon counts are not typically a problem. A substantial portion of this work concentrates on superresolution restoration techniques which are well known within the astronomical community but are of limited applicability to restoration of general video sequences. This body of research is noted here primarily for completeness, as well as a source for theoretical ideas.

Maximum Likelihood Methods

Tom and Katsaggelos [201, 197] examine the superresolution restoration problem as composed of three separate steps — registration of the low-resolution images, restoration of these images and an interpolation step which yields the superresolution restoration. In [201], a simultaneous registration and restoration approach is proposed in an effort to improve the restoration. The work is based on ideas on multichannel image identification and restoration developed in [196]. In [198] these ideas are extended to perform combined registration, restoration and interpolation.

These papers are essentially based on a Maximum Likelihood (ML) formulation of the registration, restoration and interpolation phases of superresolution restoration. A fundamental problem with this approach is the fact that ML estimation is poorly suited to the solution of ill-posed inverse problems since it is not possible to include *a-priori* information. ML estimation is in fact a special case of MAP estimation under the assumption of a uniform prior. Throughout this section, the importance of regularized solutions using *a-priori* constraints has been emphasized as a necessary condition for yielding high-quality restorations. These papers do not sufficiently address this issue despite the useful ideas contributed in terms of simultaneous restoration and registration.

Elad and Feuer [47, 48] have also considered (along with other techniques discussed later) a ML estimation formulation, but the same criticism may be applied to this work. In fact they modify the ML estimate to include a regularization functional in [47] which, in effect, is equivalent to a MAP formulation with a Gaussian prior or a Tikhonov [194] type regularization. It is also suggested that the simulate-and-correct methods yield the ML estimate of the superresolution restoration.

Summary of Stochastic Restoration Methods

The stochastic restoration techniques, the MAP formulation utilizing MRF prior models in particular, have become prominent in the superresolution restoration literature. There are several reasons for this:

- The MAP estimation framework enables direct incorporation of *a-priori* constraints on the solution — essential for finding high quality solutions to the ill-posed inverse problem of superresolution restoration.

- MRF priors are extensively used as they provide a convenient, intuitive and realistic method for modeling typical image characteristics using only a local neighbor interaction model. With a judicious choice of the penalty function and the spatial activity measure in the MRF model, priors may be defined which provide the smoothness constraints necessary to ensure acceptable restoration under conditions of limited or poor data, but also allow effective restoration of discontinuities (edges).

- The optimization inherent in the MAP estimation framework leads, under suitable conditions, to a globally convex optimization problem. For instance, under the usual assumption of Gaussian observation noise and convex prior model, the MAP estimation formulation leads to a globally convex optimization. This ensures not only the existence and uniqueness of the solution, but also enables the use of well understood and efficient gradient descent type optimization algorithms.

- Simultaneous motion estimation and restoration is possible within the MAP framework.

The MAP estimation formulation represents one of the most promising and flexible approaches to superresolution image restoration. The only competitors, in terms of convenience and flexibility, are the set theoretic restoration methods which are discussed next.

3.3.5 Set Theoretic Methods

This section addresses methods which solve the restoration problem by defining sets of constraints which must be satisfied by candidate solutions. At a minimum these include constraints to ensure that candidate solutions are consistent with the observed data, however it is common (especially with ill-posed problems) to include constraints which represent *a-priori* knowledge regarding the feasible solution space. In order to find solutions which simultaneously satisfy the totality of constraint sets, iterative procedures are often used. Where the constraint sets are known to be convex, this leads to the method of Projection Onto Convex Sets (POCS).

Projection Onto Convex Sets

Earlier, the prominence of Bayesian estimation techniques for superresolution image restoration was noted. In the superresolution restoration literature, the only major competitor to the Bayesian techniques are those based on the method of POCS. In the POCS formulation, constraint sets are used to define the feasible solution space for the superresolution restoration. Constraints are defined as convex sets in the vector space $\mathbb{R}^{N_1 \times N_2}$ which represents the space containing all superresolution images. Sets that represent desirable characteristics of the solution are defined on this space. These sets encapsulate constraints such as fidelity to the observed data, positivity, bounded energy, smoothness and so on. The solution space of the superresolution restoration problem is, by design, the intersection of the convex constraint sets. POCS is an iterative procedure which, given any point in the vector space, locates a point which simultaneously satisfies all the constraint sets.

Given k convex constraint sets in $\mathbb{R}^{N_1 \times N_2}$ such that the intersection of the sets is nonempty, POCS projects a point in the vector space onto each constraint set in turn, repeating the process until a point is reached which lies in the intersection of the k sets. This iteration converges provided that the constraint sets are convex. A detailed theoretical discussion of the POCS method can be found in [209]. Discussions on the use of POCS techniques in image restoration are presented in [141, 177].

POCS has attracted much attention in recent years in a multitude of image restoration and reconstruction applications. Three reasons for this stand out:

- Simplicity

  POCS is very intuitive and generally simple to implement. The only potential source of difficulty is in determining the projection operators.

- Flexible spatial-domain observation model

  Because the POCS method is typically formulated in the spatial domain, very general motion and observation models may be used. The complexity of the motion and observation model has little impact on the POCS solution procedure.

- Powerful inclusion of *a-priori* information

  Perhaps the most useful aspect of the POCS formulation is the ease with which *a-priori* information may be included. It is generally simple to define convex constraint sets which incorporate desired solution characteristics. These sets may impose restrictions such as positivity or bounded energy which are difficult to represent in terms of cost functionals.

The examination of POCS-based techniques begins with a detailed account of an early POCS superresolution restoration paper which demonstrates many of the fundamental ideas. This presentation will enable a more concise explanation of later work.

Stark and Oskoui [186] propose an early POCS approach to superresolution image restoration problems. In particular, the paper addresses the scanning linear array problem originally discussed by Frieden and Aumann (see Section (3.3.2)) as well as the problem of restoring a superresolution image from multiple planar array images. The discussion begins with a brief outline of the proposed imaging model. The mathematics for the case of a time invariant image is developed. It should be clear how the method is extended to the case of time varying imagery with multiple image acquisitions.

The optical system projects an image $f(x, y)$ onto the image sensor array which is assumed to consist of a regular array of M_1 by M_2 sensor elements. The output y_i of the i^{th} detector in the sensor array with spatial response characteristic $\sigma_i(x, y)$ is given by,

$$y_i = \int\int_{-\infty}^{\infty} f(x, y)\sigma_i(x, y) \quad , \quad 1 \le i \le M_1 M_2. \tag{3.44}$$

This integration over the continuous spatial variables (x, y) may be discretized on the superresolution restoration grid $[n_1, n_2]$ yielding,

$$y_i = \sum_{n_1}\sum_{n_2} f[n_1, n_2]\sigma_i[n_1, n_2] \quad , \quad 1 \le n_1 \le N_1 \ , \ 1 \le n_2 \le N_2, \tag{3.45}$$

where $N_1 > M_1$, $N_2 > M_2$ are the dimensions of the superresolution restoration array. The i^{th} detector spatial response characteristic $\sigma_i(x, y)$ is discretized to yield $\sigma_i[n_1, n_2]$ which is the fractional area of the superresolution pixel $[n_1, n_2]$ contained within the response region of the i^{th} low-resolution detector. This assumes that the response of the detector is uniform over its sensing region. In particular,

$$\sigma_i[n_1, n_2] = \begin{cases} 0 & \text{if SR pixel } [n_1, n_2] \text{ is completely outside of} \\ & \quad \text{the } i^{\text{th}} \text{ detector response region} \\ 1 & \text{if SR pixel } [n_1, n_2] \text{ is completely within} \\ & \quad \text{the } i^{\text{th}} \text{ detector response region} \\ r_i & (0 < r_i < 1) \text{ if SR pixel } [n_1, n_2] \text{ is partially within} \\ & \quad \text{the } i^{\text{th}} \text{ detector response region} \end{cases} \tag{3.46}$$

Lexicographic ordering of $f[n_1, n_2]$ and $\sigma_i[n_1, n_2]$ yields column vectors $\mathbf{f}$ and $\boldsymbol{\sigma}_i^\top$ making it possible to rewrite Equation (3.45) using the compact inner product notation as,

$$y_i = \boldsymbol{\sigma}_i^\top \mathbf{f}. \tag{3.47}$$

Next, consider the sets C_i defined as,

$$C_i = \left\{ \mathbf{f} : \boldsymbol{\sigma}_i^\top \mathbf{f} = y_i \right\} \quad \text{for} \quad i \in \{1, 2, \ldots, M_1 M_2\}. \tag{3.48}$$

For fixed i, C_i represents the set of all discretized superresolution images for which the response of the i^{th} sensor in the detector array is y_i, the observed value. Since the sensor array is composed of $M_1 M_2$ detectors, for each observed image there corresponds sets $\{C_i\}$ with $i \in \{1, 2, \ldots, M_1 M_2\}$. These sets constrain the possible values which may be assumed by an estimate for the solution $\mathbf{f}$. In particular, the constraint described by Equation (3.48) ensures that the superresolution image $\mathbf{f}$ is consistent with the observed datum y_i. Noticing that the constraints on $\mathbf{f}$ are linear, it is in principle possible to obtain a sufficient number of equations so that a solution for $\mathbf{f}$ may be found by matrix inversion.

For restoration of a superresolution image with $N_1 N_2$ pixels by matrix inversion, $N_1 N_2$ *independent* observation equations are required. Notice immediately that since $N_1 > M_1$, $N_2 > M_2$, data from more than one observed image are required to obtain a sufficient number of equations to enable a solution by matrix inversion. The requirement to obtain $N_1N_2 > M_1M_2$ independent equations implies that novel information must be obtained — information gained from additional observed images. This novel information is derived from differences between the observed images, typically caused by scene or camera motion. In imaging environments where the scene or camera motion is uncontrolled (the typical case), obtaining exactly the right motion to ensure a full-rank observation matrix may not be possible, or may require a large number of observed images. Even when the observation matrix is full-rank, the presence of observation noise can still preclude a solution via matrix inversion since the observed data may result in an inconsistent linear system (the data lie outside the column-space of the observation matrix).

In reality, additional image observations contribute further sets of the form of Equation (3.48). These additional constraints augment the underdetermined system of equations derived for a single image so that the system approaches the full-rank case.

The collection of sets $\{C_i\}$ is often referred to as data consistency constraints and in this case they can be shown to be closed and convex. This allows the definition of a projection operator $\mathcal{P}_i$ as,

$$\mathcal{P}_i\mathbf{f} = \begin{cases} \mathbf{f} & \text{if } \boldsymbol{\sigma}_i^\top \mathbf{f} = y_i \\ \mathbf{f} + \dfrac{y_i - \boldsymbol{\sigma}_i^\top \mathbf{f}}{\boldsymbol{\sigma}_i^\top \boldsymbol{\sigma}_i}\boldsymbol{\sigma}_i & \text{otherwise.} \end{cases} \tag{3.49}$$

$\mathcal{P}_i\mathbf{f}$ is the projection of the point $\mathbf{f}$ onto the set C_i. In the POCS formulation, an initial guess $\mathbf{f}^{(0)}$ for the superresolution image is projected onto each of the constraint sets C_i to ensure consistency with each measured datum y_i. With K representing the total number of observed pixel constraint sets, the n^{th} estimate $\mathbf{f}^{(n)}$ is updated using a sequence of projections,

$$\mathbf{f}^{(n+1)} = \mathcal{P}_1\mathcal{P}_2\mathcal{P}_3 \cdots \mathcal{P}_K\mathbf{f}^{(n)}. \tag{3.50}$$

This sequence of projections is applied repeatedly to yield an updated estimate of the superresolution image. Closedness and convexity of the constraint sets ensure convergence of the iteration to some point satisfying all the constraint sets [209]. It is imperative to note that this point is, in general, nonunique. It lies on the surface of the convex polytope defined by the intersection of the constraint sets. The solution is, in general, also dependent on the initial guess.

It is interesting to note that application of the POCS method using only the data consistency constraints in Equation (3.48) is identical to the Algebraic Reconstruction Technique (ART) in the computed tomography literature [77].

In addition to the data consistency constraints, additional constraints which represent *a-priori* knowledge concerning the form of the solution may be included. For example, the range of values in the solution image may be constrained to the set C_A,

$$C_A = \{\mathbf{f} : \alpha \leq f_i \leq \beta \ , \ \alpha < \beta\} \, . \tag{3.51}$$

Another possibility is to require that the solution have bounded energy using the constraint set,

$$C_E = \left\{\mathbf{f} : \|\mathbf{f}\|_2^2 \leq E\right\} . \tag{3.52}$$

If the solution is known to be similar to some reference image $\mathbf{f}_R$ then the solution may be constrained to lie within some distance ϵ_R of $\mathbf{f}_R$ using the sets,

$$C_R = \{\mathbf{f} : \|\mathbf{f} - \mathbf{f}_R\| \leq \epsilon_R\} \, . \tag{3.53}$$

Bounded support may also be imposed by defining a set of points A for which the solution is required to be zero, such as,

$$C_S = \{\mathbf{f} : f_i = 0 \ \text{ for } \ i \in A\} \, . \tag{3.54}$$

Additional constraints on the solution may be defined in a similar manner. The iteration of Equation (3.50) is augmented to include the additional constraint sets.

The inclusion of prior knowledge in this fashion constrains the solution space thus enabling robust performance in the presence of noise, inconsistent data or missing data.

The work of Stark and Oskoui represents the first application of the POCS method to superresolution image restoration. There are, however, several drawbacks to the approach taken in their paper, many of which are addressed in later work. For now, it is noted that the proposed observation model does not include noise. The observation model for the k^{th} observed image is of the form,

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{f}, \tag{3.55}$$

where $\mathbf{H}_k = \begin{bmatrix} \boldsymbol{\sigma}_{1_k}^\top & \boldsymbol{\sigma}_{2_k}^\top & \cdots & \boldsymbol{\sigma}_{M_{1_k} M_{2_k}}^\top \end{bmatrix}^\top$. In the case of p image observations,

$$\mathbf{Y} = \mathbf{H}\mathbf{f}, \tag{3.56}$$

where $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^\top & \mathbf{y}_2^\top & \cdots & \mathbf{y}_k^\top & \cdots & \mathbf{y}_p^\top \end{bmatrix}^\top$ and $\mathbf{H} = \begin{bmatrix} \mathbf{H}_1^\top & \mathbf{H}_2^\top & \cdots & \mathbf{H}_k^\top & \cdots & \mathbf{H}_p^\top \end{bmatrix}^\top$.

Tekalp, Ozkan and Sezan [193] propose three different superresolution restoration approaches. Only the POCS-based approach which extends the earlier work by Stark and Oskoui [186] is examined here. The other approaches, which are a frequency-domain formulation and an interpolation/restoration method, are discussed in Section (3.2.1) and Section (3.3.1) respectively. The primary contribution of this paper is an observation model which includes noise. In particular, the authors propose a single image observation model of the form,

$$\mathbf{y} = \mathbf{H}\mathbf{f} + \mathbf{n}, \tag{3.57}$$

where the system matrix $\mathbf{H}$ includes the effects of a sensor PSF and global translational motion. In order to account for observation noise, the data consistency constraint for the i^{th} pixel observation, which Stark and Oskoui model as an equality constraint in Equation (3.48), is modified to allow for some error δ_0 as,

$$C_i = \{\mathbf{f} : |r_i| < \delta_0\}, \tag{3.58}$$

where r_i is the i^{th} element of the residual $\mathbf{r} = \mathbf{y} - \mathbf{H}\mathbf{f}$. Noticing that $\mathbf{y} - \mathbf{H}\mathbf{f} = \mathbf{n}$, it is clear that the residual $\mathbf{r}$ is nothing more than the observation noise $\mathbf{n}$ and so δ_0 represent the confidence in the observation and may be set according to the noise statistics. The authors define the required projection operator for the set in Equation (3.58) and show the results of experiments in which a superresolved image is restored from several globally translated, undersampled frames. Though the paper addresses the problem of observation noise, the motion model used assumes only global translation between images and does not consider the effects of motion blur. This is not, however, a result of any inherent limitation of the POCS method which should become clear from the discussion of [146] below.

Patti, Sezan and Tekalp [146] address the shortcomings of [193]. Their paper utilizes the same general observation model $\mathbf{y} = \mathbf{H}\mathbf{f} + \mathbf{n}$ and a POCS formulation for the solution of these equations. Their novel contribution is the motion model included in the system matrix $\mathbf{H}$ which differs from earlier work. Most significantly, the matrix $\mathbf{H}$ incorporates a Spatially-Varying Point Spread Function (SVPSF) which models the degradations caused by sampling with a low-resolution sensor, as well as the effects of relative motion between the sensor and the scene (which includes blurring due to a nonzero aperture time). The SVPSF results from three modeling stages: first the blur caused by the relative motion between the scene and the sensor during the aperture time; next the effects of the physical dimensions and response characteristics of the low-resolution sensors; and finally the effects of sampling by the sensor array. The combined model defines the system matrix $\mathbf{H}$.

The motion model proposed, which appears to be inspired by Sawchuk [160] allows for the inclusion of a very general class of scene motion. In particular, the authors show how it is in theory possible to include general scene motion, provided occlusions are not present. A theoretical derivation illustrates how the SVPSF may be found for this general case. This model is illustrated using the example of temporally piecewise constant motion velocities, with the case of global translational motion fully developed and used in experiments. Restoration of the superresolution image is effected using a POCS framework which follows their earlier work [193].

In summary, [146] provides a POCS-based framework for superresolution restoration which accounts for aliasing, sensor blur, motion blur and additive noise. It is interesting to note that spatial blurring due to motion occurring during the aperture time had not been addressed prior to this paper. The work also relaxes the restriction to LSI blurs and spatially-invariant noise.

In addition to their work addressing the superresolution restoration problem, Patti, Sezan and Tekalp [147] have also published in the related field of video standards conversion. The objective is the conversion of video material between differing spatiotemporal sampling lattices. Some examples of common standards among which material is converted include Phase Alternating Line (PAL) (interlaced, 625 lines per frame, 50 fields per second), National Television Standards Committee (NTSC) (interlaced, 525 lines per frame, 59.94 fields per second), interlaced scan and progressive scan formats, 24 frames per second motion picture film, and the raft of new Advanced Television Systems Committee (ATSC) standard formats [1]. The work by Patti, Sezan and Tekalp extends [193, 146] to include an arbitrary input spatiotemporal sampling lattice. The form of the discrete observation model is unchanged from the SVPSF model proposed in [146]. A POCS formulation identical to that found in [146] is utilized for restoration. In a more recent paper [148], restoration of superresolution images from interlaced video sequences is addressed. This paper contains several extensions in terms of motion detection and analysis, utilizing both a global affine dominant motion model, estimated using gradient-based techniques as well as segmented local motion regions.

Full details of the work of Patti, Sezan and Tekalp may be found in [149] which consolidates and elaborates on the earlier papers [193, 146, 147]. The observation model incorporates a spatially-varying PSF which takes into account camera and scene motion, motion blurring resulting from nonzero aperture time, the physical dimensions of the low-resolution sensor elements and blurring caused by the imaging optics. Sensor noise is also considered. A new development in [149] is the inclusion of sampling of the continuous scene on an arbitrary space-time lattices. A detailed derivation of the motion model is provided as well as the techniques used for motion estimation. Restoration is via the POCS method discussed previously. The results of several experiments are presented to demonstrate the capabilities of the proposed method.

It is interesting to note that the IBP methods discussed in Section (3.3.3) are a special case of the POCS procedure where the data consistency constraint in Equation (3.48) requires equality rather than allowing an error tolerance.

The extension of observation models to include arbitrary local motion compounds the problem of restoration artifacts caused by unreliable motion estimates. Estimating local motion is generally less reliable than is the case for global motion. This is due simply to a lack of data. Global motion may often be reliably estimated since a large number of image pixels are utilized in the computation. In the case of local motion, however, an estimate must be computed using only a small local neighborhood of pixels, a consequence of which is a reduction in the reliability of the motion estimate (see Chapter (4) for an expanded discussion of this issue). In order to achieve high-quality superresolution restoration it is necessary to obtain motion estimates which are accurate and precise to subpixel resolution. It is imperative that the motion estimates be reliable, otherwise objectionable visual artifacts plague the restoration. In [52], Eren, Sezan and Tekalp extend [149] to consider means of improving the robustness of the superresolution restoration methods to errors in the motion estimation process. To achieve this, two new concepts are introduced, a *validity map* and a *segmentation map*. The validity map disables projections based on observations with inaccurate motion estimates, while the segmentation map enables object-based tracking and processing where more accurate object-based motion models may be used to improve the quality of the restoration. The segmentation map also allows restoration of specific objects in a manner similar to [89].

Bounding Ellipsoid Method

A variant of the POCS-based formulation using an ellipsoid to bound the constraint sets has been investigated by Tom and Katsaggelos [199, 200] and briefly mentioned by Elad and Feuer [47, 48]. Given a set of ellipsoidal constraint sets, a bounding ellipsoid is computed [101]. The centroid of this ellipsoid is taken as the superresolution image estimate. Since direct computation of this point is usually infeasible, an iterative procedure is used. It is interesting to note that this approach takes a form closely related to regularized methods. The observation model used in this work is similar to that proposed by Schultz and Stevenson [168].

Summary of Set Theoretic Restoration Methods

This discussion relates primarily to POCS-based methods but also includes a brief critique of the bounding ellipsoid method. The advantages of the POCS restoration framework include:

- Simplicity.
- A flexible spatial-domain observation model.
- The ability to include *a-priori* information in a very flexible framework.

The preceding discussion clearly demonstrates these points. POCS is not, however, without failings. Some of the disadvantages of POCS are:

- Nonuniqueness of the solution

  The feasible solution space for the POCS method is defined as the intersection of the convex constraint sets. Therefore, unless the intersection is a point, the solution is *nonunique*. This is a serious drawback. In the MAP formulations considered, a convex optimization was favored in order to ensure the existence and uniqueness of the solution. In POCS, this is usually not possible.

- Dependence of the solution on the initial guess

  Worse yet is the fact that the solution determined using the POCS iteration is dependent on the (arbitrary) initial estimate, as well as the order of application of the projections. POCS repeatedly projects the current estimate of the superresolution image onto the constraint sets. The solution is the first point in the vector space of possible solutions which satisfies all the constraint sets. When not using relaxed projections, the solution lies on the surface of the volume of intersection of the constraint sets.

- Computational cost

  POCS-based methods require considerable computation and a large number of iterations to achieve convergence. An analysis of POCS methods shows that without the use of relaxed projections or similar modifications, POCS can be very slow to converge as the solution is approached. For this reason POCS is commonly not run to convergence but to a point where the estimate is visually acceptable.

- Difficulties implementing the necessary projection operators

> A fundamental assumption in the POCS formulation is the existence of projection operators for each constraint set. These operators should project a point onto the closest point of the constraint set and should be idempotent. In practical situations, this is not always possible. Consider for example the convex constraint that a signal be bandlimited. The POCS framework requires an operator which projects the signal onto the convex set representing bandlimited signals. The projection operator for the bandlimited constraint set is, of course, an ideal low pass filter. Given that such a filter is not realizable, the projection operator cannot be idempotent. Under these conditions POCS theory is not applicable. A more detailed discussion of this problem is found in [154].

It may be noted that the bounding ellipsoid method does have a unique solution — the centroid of the bounding ellipsoid. The problem is that this solution is quite arbitrary. There is little reason why the centroid should be favored among all the possible solutions within the bounding ellipsoid. In contrast, in the MAP formulation, under suitable conditions which are easily satisfied in practice, the solution is unique and maximizes the meaningful *a-posteriori* probability.

3.3.6 Hybrid Methods

Some work has been done on combined ML/MAP/POCS approaches to superresolution restoration. In particular, the desirable characteristics of the MAP framework and those of the very flexible POCS method are combined, creating a hybrid optimization.

Elad and Feuer [47, 48], after reviewing several existing superresolution techniques, propose a hybrid ML/POCS method which uses the ML formulation to pose superresolution as a statistical inference problem, while utilizing projection-based constraints to effect regularization. The authors claim that their hybrid restoration algorithm is a generalization of the existing methods. It is important to note, however, that a combined statistical estimation approach with projected constraints was presented earlier by Schultz and Stevenson [166, 165, 167, 168]. In these papers, Schultz and Stevenson utilize a constrained MAP superresolution restoration algorithm, in which an alternating projections-based constraint and a MAP iteration is used.

The general idea of these hybrid methods is to maximize the *a-posteriori* probability or likelihood function (in the MAP and ML frameworks respectively) while ensuring that the solution remains within a constrained space specified using a set-theoretic formulation. In [166, 165, 167, 168], the constraint set ensures that downsampling of the superresolution image exactly matches the reference frame of the low-resolution image sequence.

Hybrid methods which combine statistical inference and projected constraints are very promising as they combine the most favorable characteristics of stochastic methods (optimal estimation theoretic solution, mathematical rigor and direct inclusion of *a-priori* information) and POCS-based approaches (powerful mechanism for inclusion of linear and nonlinear set theoretic *a-priori* constraints).

3.3.7 Optimal and Adaptive Filtering Methods

Several researchers have proposed inverse filtering approaches to superresolution restoration. Efforts in this area are briefly reviewed. These techniques are considered primarily for completeness, as several are suboptimal in terms of inclusion of *a-priori* constraints.

Jacquemod, Odet and Goute [96] propose a simple deconvolution restoration approach that assumes subpixel translational motion. A deconvolution filter suitable for the restoration of the merged observed images is determined. This approach is poorly suited to the incorporation of more general observation models and is limited in terms of inclusion of *a-priori* constraints.

Erdem, Sezan and Ozkan [51] have proposed a LMMSE filtering approach, the motion compensated multiframe Wiener filter, for restoration of image sequences degraded by LSI spatial blur and additive noise. A global translational motion model is assumed but motion blurring is not incorporated. Though the motion and degradation models are limited and nonlinear *a-priori* constraints are difficult to incorporate in their formulation, this paper is noteworthy in that *simultaneous* multiframe restoration is undertaken. A similar framework is discussed by Srinivas and Srinath [185]. In Chapter (7), a new simultaneous multiframe superresolution restoration framework using an estimation-theoretic Bayesian approach is introduced.

Techniques based on adaptive filtering, especially the Kalman filter, have also been applied to superresolution restoration. Patti and Sezan [150] and Elad and Feuer [46, 49] are examples. In [150] a motion compensated model Kalman filter capable of superresolution restoration under spatially varying blurs is proposed. Though the proposed Kalman filtering formulation is computationally efficient, it is in effect, still a LMMSE estimator. Nonlinear image modeling constraints which provide bandwidth extrapolation cannot easily be incorporated into this framework. Similar comments apply to the work of Elad and Feuer [46, 49] which is reduced to a LMMSE estimate of the superresolution image. Kim [106] dedicates a chapter of his doctoral dissertation to a similar Kalman filtering approach to superresolution restoration.

3.3.8 Tikhonov Regularized Methods

Due to the ill-posedness of the superresolution restoration problem, Hong, Kang and Katsaggelos [80, 79] have proposed the use of Tikhonov regularization [194] for superresolution restoration. This is a deterministic regularization approach utilizing regularization functionals to impose smoothness constraints on the space of feasible solutions. The regularization functional used in these papers is equivalent to a Gaussian MRF prior in the Bayesian MAP framework. Gaussian MRF priors are well known to produce overly smoothed restorations [25]. Indeed the poor performance of the Gaussian MRF prior is the reason why priors such as the HMRF and the Generalized Gaussian Markov Random Field (GGMRF) [25], which impose smoothness while still enabling edge restoration, were introduced. The Tikhonov regularization approach is a special case of the more general Bayesian framework under the assumptions of Gaussian noise and prior. The observation model used in [80, 79] is similar to the motion compensated subsampling matrix of [168].

3.3.9 Summary of Spatial-Domain Superresolution Restoration Methods

Superresolution restoration via the spatial-domain approach addresses many of the shortcomings of frequency-domain approaches:

- Motion models

  Spatial-domain methods, using the linear observation model of the form $\mathbf{Y} = \mathbf{H}\mathbf{f}$ are capable of including an almost unlimited range of motion models. Since the matrix $\mathbf{H}$ models observed image pixels as a linear combination of *any* configuration of pixels in the superresolution image $\mathbf{f}$, there is enormous flexibility in the formulation of the motion model. There is no limitation to global models as is the case with frequency-domain approaches. It is just as simple to include a local motion model as a global model using the spatial-domain formulation.

- Degradation models

  The system matrix **H** also allows the inclusion of linear degradations such as motion blurring resulting from a nonzero aperture time (modeled as spatial integration over the motion trajectory), spatially varying or invariant blurs, missing pixels and so on. It is extremely cumbersome, if not impossible to include such degradations using the frequency-domain superresolution restoration framework.

- Inclusion of spatial-domain *a-priori* knowledge for regularization

  As was discussed at length, the inclusion of *a-priori* information is necessary for the solution of ill-posed inverse problems such as superresolution restoration. Stochastic prior models as well as convex constraint sets provide simple, yet very powerful methods to incorporate *a-priori* knowledge into the restoration process. Working with spatial-domain constraints is intuitive and direct. Using MRF models defined using only local spatial interactions, it is possible to provide smoothness constraints while preserving the ability to reconstruct sharp edge features (impossible in the frequency domain).

- Powerful mechanism for bandwidth extrapolation

  The combination of data from multiple images, as well as the use of realistic, often nonlinear, *a-priori* constraints on the reconstructed image endow spatial-domain methods with a powerful mechanism for image bandwidth extrapolation. MRF models which utilize nonlinear penalty functions are especially useful as they are capable of introducing novel frequency information.

- Theoretical framework

  Stochastic methods, especially the MAP estimation method, are derived from a solid mathematical framework within which further theoretical developments can be made. Theory applying to the optimality of estimators is directly applicable, as well as bounds on errors and so on. This is a rich framework which is not available in the frequency-domain approach.

Spatial-domain methods do however come at some cost:

- Increased complexity

  Unlike the frequency-domain approach where restoration was a relatively simple process, the optimizations inherent in most spatial-domain methods are more complex than their frequency-domain counterparts.

- Increased computation

  The increased flexibility of spatial-domain methods tend to come at the cost of substantially increased computational requirements. This is especially true for methods which utilize nonconvex priors for which it is necessary to resort to slow simulated annealing, graduated nonconvexity, or other approaches for optimization in the presence of local extrema.

- Less intuitive superresolution mechanism

  In the frequency-domain formulation, the mechanism for superresolution restoration was abundantly clear — dealiasing of shifted, undersampled frames. For spatial-domain techniques the mechanisms underlying resolution enhancement are not as transparent, even though dealiasing and bandwidth extrapolation is achieved.

It should be clear from this discussion that to achieve superresolution restoration of image sequences containing anything more complex than global translational motion, spatial-domain techniques are the preferred and often the only viable approach.

The advantages and disadvantages of spatial-domain restoration techniques have been discussed in general terms, however the issue as to which of the spatial-domain techniques holds the most promise for the future has not been addressed. This is done in Section (3.4) below.

3.4 Comparisons and Conclusions

This section presents a summarized comparison of the main categories of multiframe superresolution restoration techniques. A comparison of the two most promising restoration approaches is presented. Some salient points regarding motion estimation vis-à-vis multiframe superresolution restoration are also provided and conclusions are drawn.

3.4.1 Comparison of Multiframe Superresolution Restoration Techniques

A comparison of the two main classes of superresolution restoration techniques, frequency-domain and spatial-domain, is found in Table (3.1). This table is divided into two sections. The upper portion deals with the formulation of the observation and degradation models, while the lower portion makes generalizations concerning the solution approaches. It is important to recognize that these are generalizations — exceptions do exist within the wide range of techniques already discussed. Attention is also focused on the most promising approaches within the respective classes of frequency-domain and spatial-domain methods. It should be clear from an examination of Table (3.1) that spatial-domain restoration methods are, in general, superior to frequency-domain methods, but this superiority comes at the cost of

TABLE 3.1

SUMMARY AND COMPARISON OF SPATIAL-DOMAIN AND FREQUENCY-DOMAIN SUPERRESOLUTION RESTORATION TECHNIQUES

	Frequency-Domain	**Spatial-Domain**
Observation model	Frequency-domain	Spatial-domain
Motion models	Global translation	Almost unlimited
Degradation model	Limited	LSI or LSV
Noise model	Limited	Very flexible Even spatially varying
SR mechanism	Dealiasing	Dealiasing BW extrapolation with *a-priori* constraints
Simplicity	Very simple	Generally complex
Computational cost	Low	High
A-priori constraints	Limited	Almost unlimited
Regularization	Limited	Excellent
Extensibility	Poor	Excellent
Performance	Good for specific applications	Good

increased computational and theoretical complexity.

Within the class of spatial-domain superresolution restoration methods, two major techniques stand out as most promising: the Bayesian MAP approach and the set theoretic POCS methods. These are compared in Table (3.2). A detailed discussion of the relative merits of each of these approaches was presented earlier. The Bayesian and set-theoretic approaches are somewhat complimentary and the best characteristics of each approach can be combined as discussed in Section (3.3.6).

3.4.2 Motion Estimation and Multiframe Superresolution Restoration

It should be clear from the discussions in this chapter that motion estimation is a critical component of any multiframe superresolution restoration framework. What

TABLE 3.2

QUALITATIVE COMPARISON OF BAYESIAN MAP AND POCS SUPERRESOLUTION RESTORATION TECHNIQUES

	Bayesian (MAP)	**POCS**
Theoretical framework	Rich	Limited
A-priori constraints	Prior probability density Projected constraints possible	Convex sets Project onto constraints
SR solution	Unique MAP estimate	Nonunique Volume of $\bigcap$ of sets
Optimization	Iterative Various descent methods Good convergence	Iterative Projection methods Often slow convergence
Computational cost	High	High
Complications	Optimization difficult for nonconvex priors	Projection operators potentially problematic

follows are several important observations regarding motion information and the motion estimation problem in the context of multiframe superresolution restoration.

- Subpixel-resolution motion

  Subpixel-resolution relative motion between observed frames contributes novel information which in turn provides constraints for the solution of the restoration inverse problem.

- Accuracy and precision of the motion estimates

  Motion estimates must be accurate and precise to subpixel resolution. This is obvious in light of the previous point.

- Density of motion estimates

  Motion estimates need not necessarily be dense (defined for all spatial locations) in order to effect superresolution restoration. It is more important that the motion estimates be accurate than dense. Intuitively, motion estimates computed in near-constant-valued regions are unlikely to be accurate, but are likewise of limited utility to the superresolution restoration algorithm since little resolution enhancement is possible in such regions. The most fruitful motion estimates may be found where they can be most reliably estimated, that is, in regions of high spatial variance.

- Regularized motion estimation

  Motion estimation is itself an ill-posed problem. It should come as no surprise that regularization can be successfully applied to the motion estimation problem.

- General motion models

  Several global motion models including translation, rotation, affine and projective, as well as various nonglobal models were discussed. Model-based or object-based motion estimation featured in later papers, allowing restoration of objects subject to partial occlusion, transparency, in motion, and so on.

In Chapter (4), the problem of motion estimation is discussed in considerable detail.

CHAPTER 4

MOTION ESTIMATION

4.1 Introduction

This chapter provides a broad introductory discussion of the challenging problem of motion estimation before delving into specific techniques. The estimation of motion in image sequences is often the first step required for such diverse applications as video compression, standards conversion, filtering, computer vision and restoration. As discussed in the preceding chapters, motion estimation also plays a central role in the context of multiframe superresolution restoration. Motion estimation is a vast field and an enormous variety of approaches to the problem may be found in the literature. Consequently, rather than trying to review the full breadth of the field, this discussion attempts instead to elucidate the fundamental character and formidable challenges of the motion estimation problem. Attention is then focused on three exemplary motion estimation frameworks which represent a sparse, but nevertheless representative sample of the most prominent motion estimation techniques.

4.2 Three-Dimensional Motion, Projected Motion and Optical Flow

The motion of objects in three-dimensional (3-D) space or 3-*D motion* may be described by a 3-D velocity field. The two-dimensional (2-D) projection on the image plane which corresponds to the 3-D motion is known as the *projected motion* or the 2-*D motion field* (see Figure (4.1)).

Often an estimate of the projected motion is desired — this is the *motion estimation* problem. The estimated motion is typically described using instantaneous velocity (flow) or displacement vector fields. Under the assumption of constant velocity motion between frames, the two descriptions are equivalent. More general descriptions of the motion field which take acceleration into account are also possible. The 2-D motion estimate may in turn be used to infer 3-D motion and structure using techniques from the computer vision literature [98].

In an imaging system, however, the only information available is the spatiotemporal variation of the light intensity incident at the focal plane. There is no direct access to the projected motion, let alone the 3-D motion. Instead, the spatiotemporal variation at the focal plane results from the interaction of the scene illumination with the objects in the scene, motion of the objects in the scene, as well as changes of camera extrinsic parameters (position, orientation) or intrinsic parameters (focal length, focus setting, etc.) [131]. It is the resulting light intensity variation projected at the focal plane which is recorded. Though this spatiotemporal intensity variation carries information about the projected motion (and therefore also information about the 3-D motion) it does *not* directly correspond to the 2-D velocity field. In particular, not all changes in the image intensity correspond to scene motion, nor does all scene motion result in image intensity variation. For example, changes in scene lighting result in image changes which do not correspond to any 3-D motion, while a featureless, uniformly illuminated disk undergoing axial rotation (a 3-D motion) does not result in any observable change in image intensity at the focal plane.

Despite these difficulties, using the time-varying intensity information it is nev-

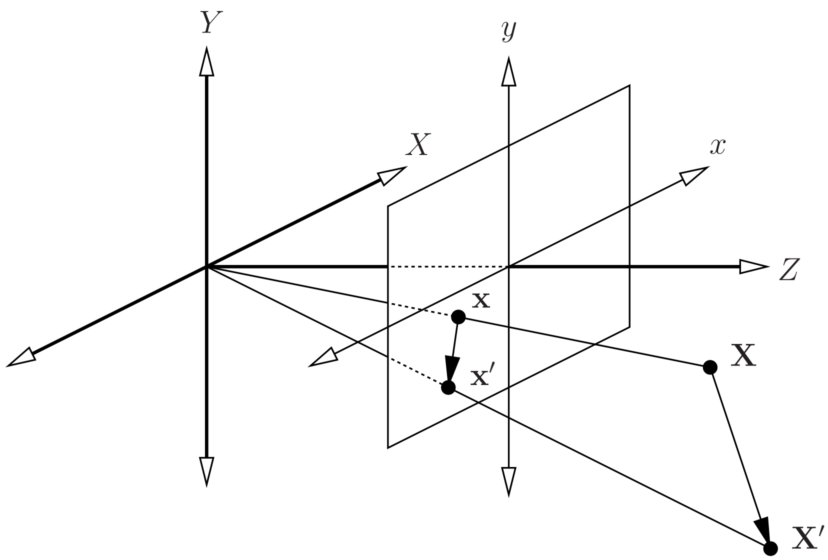


Figure 4.1. 3-D motion of a point in 3-D space and its projected 2-D motion at the image plane.

ertheless possible to construct an *approximation* of the projected motion, known as the *optical flow*.

4.3 The Motion Estimation Problem

Consider a point $\mathbf{X}(t) = [X(t), Y(t), Z(t)]^\top$ on a moving object in 3-D space. At time t, the camera system projects the 3-D point $\mathbf{X}(t)$ onto the camera focal plane at position $\mathbf{x}(t) = [x(t), y(t)]^\top$. Given two time instants t, τ with $t < \tau$ and corresponding image intensities $f_t(\mathbf{x})$ and $f_\tau(\mathbf{x})$, the position of the projection of the 3-D point on the image plane at time t, given by $\mathbf{x}(t)$, may be related with its position at time τ, given by $\mathbf{x}(\tau)$, in two ways (see Figure (4.2)):

1. The *forward motion*, $\mathbf{d}_{t,\tau}(\mathbf{x}) = \mathbf{x}(\tau) - \mathbf{x}(t)$, describes the displacement in the image plane of the projected 3-D point from time t to time τ. The forward motion estimate $\mathbf{d}_{t,\tau}(\mathbf{x})$ may be used for *backward prediction* where image values are predicted from a *future* reference frame using $f_t(\mathbf{x}) = f_\tau(\mathbf{x}+\mathbf{d}_{t,\tau}(\mathbf{x}))$.

2. The *backward motion*, $\mathbf{d}_{\tau,t}(\mathbf{x}) = \mathbf{x}(t) - \mathbf{x}(\tau)$, describes the displacement in the image plane of the projected 3-D point from time τ to time t. The backward motion estimate $\mathbf{d}_{\tau,t}(\mathbf{x})$ may be used for *forward prediction* where image values are predicted from a *past* reference frame using $f_\tau(\mathbf{x}) = f_t(\mathbf{x} + \mathbf{d}_{\tau,t}(\mathbf{x}))$.

Note that the definition of the forward and backward motion vectors presented here differs from that used in the video compression community. In that context, a forward motion vector is a motion vector that is used for forward prediction (motion compensation from a past reference frame) and a backward motion vector is a motion vector that is used for backward prediction (motion compensation from a future reference frame).

Assuming the velocity of the 2-D projection of the moving 3-D point is constant between frames, the 2-D instantaneous velocity (flow) $\mathbf{v}_t(\mathbf{x})$ and the displacement are related by $\mathbf{d}_{t,\tau}(\mathbf{x}) = \mathbf{v}_t(\mathbf{x})\Delta T$ where $\Delta T = \tau - t$ represents the time interval between frames. If, however, the apparent motion trajectory exhibits acceleration, the first order (constant velocity) representation will fail to model accurately the motion occurring between frames. This is illustrated in Figure (4.2) where the dashed line represents the first order approximation to the true motion trajectory. For certain applications, such as image sequence interpolation or standards conversion where it is necessary to reconstruct images between temporal sampling instants, performance is improved by utilizing higher order temporal motion models [145]. Where the notion of the displacement field and flow (velocity) field are interchangeable, the term *motion field* will be used.

Thus, the problem of motion estimation is that of determining the motion field (displacement, velocity, acceleration) associated with the observed image sequence.

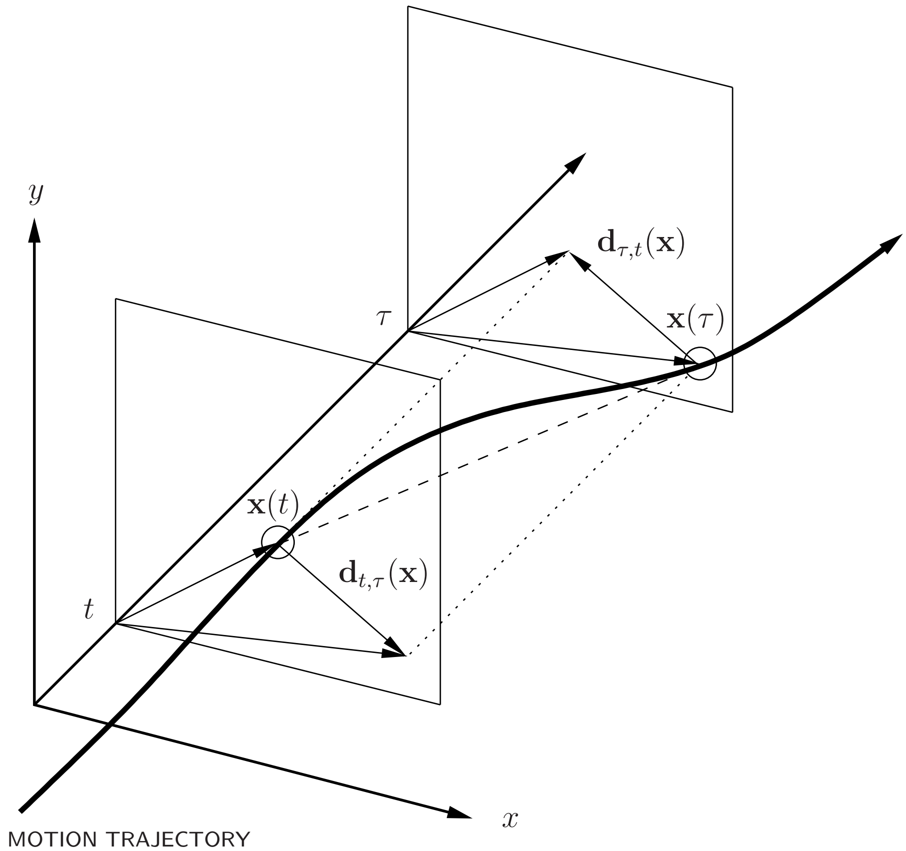


Figure 4.2. Motion trajectory showing forward displacement vector $\mathbf{d}_{t,\tau}(\mathbf{x})$ and backward displacement vector $\mathbf{d}_{\tau,t}(\mathbf{x})$.

4.4 Complications in Motion Estimation

4.4.1 Occlusion

Occlusion refers to the covering of objects due to motion. Translating objects will uncover background regions behind the path of motion and will cover regions in the path of the motion. Similarly, object rotation as well as camera movement can result in occlusions. A consequence of occlusion is that it is not possible to find correspondences for presently visible points which become covered in future frames. Additionally, regions that become uncovered have no motion vectors that point into them. This is illustrated in Figure (4.3). Occlusion complicates the process of motion estimation.

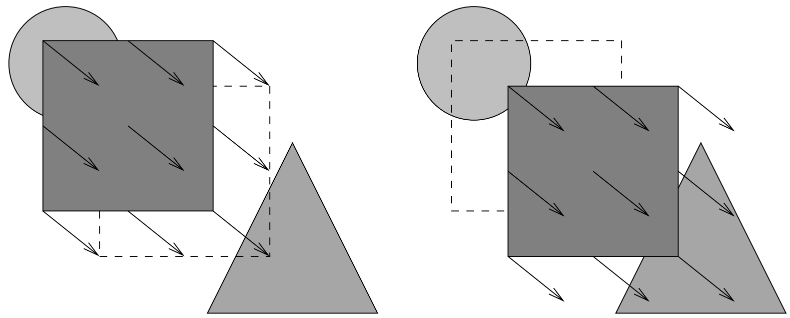


Figure 4.3. The occlusion problem.

4.4.2 The Aperture Problem

The *aperture problem* is a term used to describe a fundamental difficulty associated with the estimation of motion. The problem is illustrated in Figure (4.4) in which a uniformly shaded block undergoes translational motion in a north-easterly direction as indicated by the thin velocity vectors. The three smaller squares, labeled A, B and C, represent apertures through which the translating block might be viewed. Viewing the moving block through aperture A, one would only observe the uniform region moving upward, even though there is also a horizontal component of motion. Similarly at aperture B, only the horizontal motion component is perceived. At each of these apertures, only the motion component in the direction of the local image spatial gradient, the so-called *normal flow*, is discernible. At aperture C, however, there is sufficient local spatial variation to yield two distinct local image gradient components. As a result, correct recovery of the motion is possible from the observation at aperture C.

On the right half of Figure (4.4), the horizontal (u) and vertical (v) velocity components of the motion are represented on the uv-plane of *velocity space*. The observations at apertures A and B do not uniquely determine the motion, but instead

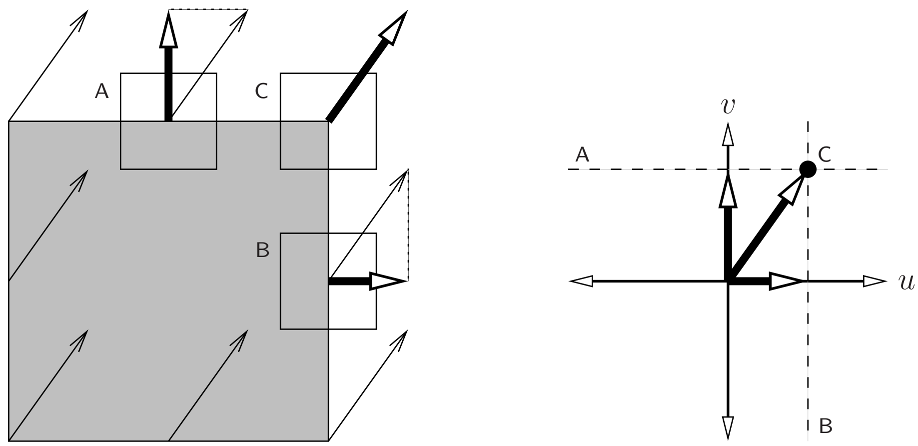


Figure 4.4. The aperture problem.

constrain the motion estimate to lie on a locus of points indicated by the dashed horizontal and vertical lines respectively. For aperture C, however, a unique location in the uv-plane is determined.

Although the term “aperture problem” is commonly used, it is something of a misnomer since the effect it describes has little to do with the observation aperture, but is uniquely a consequence of the lack of spatial variation in the moving scene. For an extreme example which illustrates this distinction, consider a scene which contains variation in only a single spatial dimension, an infinite vertically-oriented edge undergoing translational motion. In this case no aperture, no matter how large, can disambiguate the vertical component of the motion. Only the component of motion in the direction of the spatial gradient (that is, perpendicular to the infinite edge) can be computed.

In Section (4.6.1), which discusses the mathematical formulation of the optical flow equation, the aperture problem will be apparent in the algebraic equation which is used to compute the optical flow.

4.4.3 Aggregating Constraints to Remove Ambiguities

It is instructive to note that if the motion component constraints for apertures A and B in Figure (4.4) are combined, the correct motion can be determined. This suggests that the ambiguity associated with the aperture problem can be ameliorated if the motion is estimated using textured regions with sufficient spatial gradient information and/or if local motion estimates are aggregated. Both these approaches imply using information from larger spatial regions. Often this is achieved by imposing "smoothness" constraints to ensure that the motion field constraints are propagated to neighboring estimates to arrive at a smooth motion field which adequately fits the observed data.

Unfortunately, with the aggregation of data over larger areas comes the increased likelihood of a region containing more than one motion which can lead to erroneous estimates. Thus there is a trade-off on region size: larger regions are required to address the aperture problem, but regions must not be so large that they will contain multiple moving objects.

4.4.4 Difficulties Associated with Multiple Motions

As suggested in the previous section, the presence of multiple moving objects can lead to further difficulties in motion estimation. Consider the example (after Burt and Sperling [29] and Weiss [207]) in Figure (4.5) where two occluding objects undergo independent motion.

As before, local motion estimates from apertures A, B, D and E lead to ambiguous motion estimates in the form of constraint lines in velocity space. While estimators at apertures C and F lead to unambiguous estimates of the true motion of the objects, the local estimates from apertures G and H lead to unambiguous, but nevertheless spurious motion estimates as they do not correspond to the motion

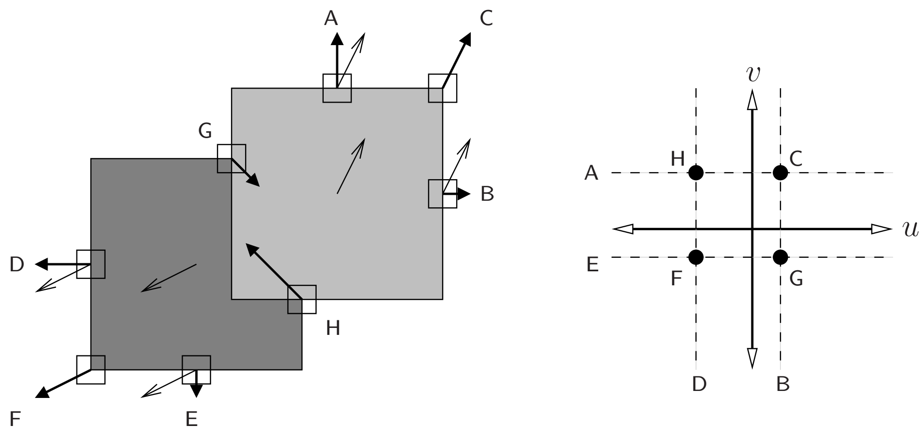


Figure 4.5. Multiple motions and occlusions leading to incorrect motion estimates.

of either of the moving objects.

In this example, techniques which globally aggregate motion estimates from the local estimators in order to remove local ambiguities will be biased by these spurious estimates which result from the combined effect of the two independent motions. The global smoothness approach will lead to a motion representation which is something of a "compromise" between the two physical motions and will not correspond to the true motion. This may be seen as a fundamental limitation of the smoothness approach which implicitly assumes a single underlying motion model. This is discussed further in Section (4.7.2).

In order to adequately address this problem, knowledge of the structure of the scene is required. In particular, mechanisms are required to identify the independently moving objects so as to best utilize the local motion constraints which are consistent with the object motion. Thus there is an interrelationship between the motion estimates and the *motion segmentation* that is required. Knowledge of the motion is assumed in order to identify the moving objects, and knowledge of the moving objects is used to estimate correctly the motion. Motion estimation in the presence of multiple moving objects represents the current frontier for research.

4.4.5 Motion Estimation as an Ill-Posed Problem

If displacement vectors (each consisting of a horizontal and a vertical component) are to be estimated for each point in the spatiotemporal sampling lattice, *twice as many unknowns as there are observations* must be estimated. There are even more unknowns to estimate in the case of motion models which account for acceleration. Thus the problem of motion estimation is inherently underconstrained.

The problem of motion estimation is more accurately described as an *ill-posed* problem. Hadamard [67] defined a problem as *well-posed* if its solution (i) exists, (ii) is unique, and (iii) exhibits continuous dependence on the data. An *ill-posed* problem is one which fails to meet these requirements.

In the case of motion estimation, all three requirements may fail to be satisfied. Existence of the solution may not be guaranteed due to occlusions. Lacking sufficient observation constraints, uniqueness of the solution may fail to be satisfied as discussed earlier. Finally, the motion estimates tend to be highly sensitive to noise in the observed image sequence thereby failing to satisfy the requirement of continuous dependence on the data.

To begin to address these problems, additional constraints are required. One approach is to utilize more data when estimating the motion parameters. This approach typically implies utilizing more than two frames for estimating the motion, as is necessary when estimating accelerated motion. Another approach is to include assumptions concerning the structure of the motion field. Naturally these approaches are not mutually exclusive. In the section that follows motion field representations which provide structural constraints are discussed.

4.5 Motion Field Representation

4.5.1 Modeling Three-Dimensional and Projected Motion

In Section (4.2) the concepts of 3-D motion, projected motion and optical flow were introduced. It was shown that the 3-D motion and the projected motion are related according to the geometry of the 3-D to 2-D projection determined by the imaging system. The choice of the projection model has a direct effect on the resulting 2-D projected motion of objects in the 3-D scene. By modeling the motion of objects in the 3-D scene it is possible, given the transformation describing the projection, to determine the motion of points projected in the image plane [131].

The motion of points, planar patches and more general surfaces in the 3-D scene result in motion of their projections on the camera focal plane. Given that 3-D motion is typically highly structured (e.g. rigid translations and rotations), it is feasible to model the projected motion which results from a given 3-D motion. Typically, motion of planar patches (and less frequently, quadratic patches) undergoing rigid translation and/or rotation is assumed and the resulting projected motion determined. Knowledge of the relationship between 3-D and 2-D motion has considerable benefits: it provides appropriate models for 2-D motion which may be used in estimating 2-D motion from time-varying image intensity information (the motion estimation problem) [189]. Also, relating 3-D and 2-D motion is essential for computer vision and other applications where there is a need to infer 3-D structure from image sequences [98].

A summary of commonly used 2-D motion models and the 3-D surface and 3-D motion models from which they are derived may be found in [189].

4.5.2 Parametric and Nonparametric Motion Models

Motion field representations may be divided into two broad categories, each having distinct advantages and disadvantages:

- Nonparametric motion field models

  In nonparametric motion field models, a representation of the motion field is typically sought on a finite set of points in the 2-D image plane indexed by $\mathbf{x}$. It is common to choose the set of points to correspond with the uniformly-spaced, discrete image sampling grid Λ.

  The primary advantage of this approach is that arbitrary motion fields may be represented (albeit on a finite sampling grid). The motion field may be interpolated to yield values between sampling points.

  The main disadvantage of the nonparametric representation is that it requires the estimation of a large number of motion parameters. Without additional assumptions on the spatial variation of the motion field, this may be impossible due to a lack of available observation constraints, as discussed previously. A common assumption is that the motion field is, in some well defined sense, "smooth". The large number of motion vectors make nonparametric models poorly suited to image sequence compression applications as considerable overhead is required for the transmission of the motion vectors which compose the discrete motion field. Nonparametric motion field representations are sometimes referred to as *dense*.

- Parametric motion field models

  An alternate approach is to use a *parametric* motion model which represents the motion field over some region of the image plane. Parametric models are typically continuous parameterized functions of the spatial location $\mathbf{x}$. Common parametric motion models use from 2 to 12 parameters. Once the parameters and the region of support of the model are determined, the model may be evaluated at any location $\mathbf{x}$ within the region, thus there is no need for interpolation.

Parametric models have the advantage of requiring relatively few model parameters to describe a potentially large region of the motion field. Data from larger regions of the image may be aggregated when estimating the model parameters. Since the number of model parameters is small, this tends to yield more reliable estimates. The parsimonious representation associated with parametric models makes them well suited for compression applications. Many of the parametric models used to describe the motion of regions are derived based on assumptions concerning the projection onto the 2-D image plane of moving planar or quadratic patches in 3-D space.

Parametric models do have drawbacks. It is not possible to represent arbitrary motion fields using commonly used parametric models without increasing the number of model parameters to be comparable with nonparametric models. As parametric models represent a curve fitting approach to modeling the motion field, they are prone to local error due to oversmoothing. Finally, for general motion fields, estimation of the region of support of the parametric model can be very difficult. Since the region of support of the nonparametric model is a point, this problem is not encountered.

In summary: nonparametric models represent the motion field by samples. Parametric models represent the motion field by parameterized regions.

The parametric and nonparametric motion field models described above are, in some advanced applications, extended to include dependence on the temporal variable t as well as the spatial variable $\mathbf{x}$ to yield the value of the motion field at $\mathbf{x}, t$.

4.5.3 Common Motion Models

The three most commonly used motion models are the two-parameter translational model, the six-parameter affine model and the eight-parameter perspective model. These models describe the variation of the displacement/flow as a function of the image plane spatial coordinates. Recall that the 2-D displacement vector $\mathbf{d}_{t,\tau}(\mathbf{x})$ describes the motion of a point $\mathbf{x}(t)$ to $\mathbf{x}(\tau)$ where $\mathbf{x} = (x, y)^\top$ is the spatial coordinate in the image plane. In describing the three motion models, explicit dependence on t and τ is dropped since the models apply equally to forward and backward motion fields.

- The two-parameter translational motion model,

$$\mathbf{d}(\mathbf{x}) = \mathbf{b}, \quad \mathbf{b} \in \mathbb{R}^2, \tag{4.1}$$

  is appropriate for modeling the orthographic projection of surfaces undergoing 3-D translation. It finds widespread use in motion compensated video compression applications where the model applies to blocks in the image plane. The translational model is also commonly used at each sampling location in dense motion representations. In this way arbitrary motion fields may be modeled.

- The six-parameter affine motion model,

$$\mathbf{d}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{2\times 2}, \mathbf{b} \in \mathbb{R}^2, \tag{4.2}$$

which includes the translational model as a special case, models the effects of orthographic projection of 3-D affine motion of planar patches in the 3-D scene. The affine model can describe rotation, but parallel lines in 3-D space are mapped to parallel lines under the projection assumptions.

- The eight-parameter perspective model,

$$\mathbf{d}(\mathbf{x}) = \frac{\mathbf{A}\mathbf{x} + \mathbf{b}}{\mathbf{c}^\top \mathbf{x} + 1} - \mathbf{x}, \quad \mathbf{A} \in \mathbb{R}^{2\times 2},\ \mathbf{b}, \mathbf{c} \in \mathbb{R}^2, \tag{4.3}$$

  is appropriate under the assumptions of perspective projection of 3-D affine motion of planar patches in the 3-D scene. The perspective model can accommodate more general quadrilateral deformations and includes the affine and translation models as special cases.

More complete reviews of parametric motion field models may be found in [189, 191, 176, 126].

4.5.4 Regions of Support for Motion Models

In this section the *region of support* for the motion models presented above is discussed. In particular, various partitions of the image plane into regions on which these parametric models apply are described. Denoting the image plane by $\mathcal{R}$, a *partition* is a set of regions $\{\mathcal{R}_i\}_1^N$ such that $\bigcup_{i=1}^N \mathcal{R}_i = \mathcal{R}$ and $\mathcal{R}_i \bigcap \mathcal{R}_j = \emptyset \; \forall \; i \neq j$.

- Global models

  For global models the partition consists of a single region $\mathcal{R}_1 \equiv \mathcal{R}$, that is, the region of support for global motion models is the entire image plane. Global models are particularly useful for describing camera translation, rotation or zooming on an unchanging scene. Applications which utilize global motion models include computer vision, image stabilization, camera calibration and video compression.

- Block-based models

  In block-based models the partition regions $\mathcal{R}_i$ are equal sized rectangular blocks. A parametric motion model applies for each block. Translational block motion models are used in all current video compression standards (MPEG and ITU H.26x). More general block motion models are also possible by applying the affine or perspective model to each block, though this is less common. Block-based models, though attractive for certain applications, are poorly suited to the task of accurately describing general motion fields. This is a consequence of the motion model being fixed for all locations within each block, and therefore unable to adequately represent motion field discontinuities or deviations which may be located within the block itself.

- Generalized block models (meshes)

  A generalization of the fixed size, regular block partition uses adaptive triangular or hierarchical block-based meshes. In generalized block models, the regions $\mathcal{R}_i$ are triangles or blocks of various sizes. The meshes are selected to model and adapt over time to the intensity and/or motion structures within the image sequence. Generalized block models have the advantage of a relatively small number of easily described regions while delivering considerably improved motion representation as compared with fixed block methods. Mesh-based models have been applied to video compression in the MPEG-4 standard.

- General regions

  Removing the restriction of regularly shaped regions leads to region-based motion models. The regions $\{\mathcal{R}_i\}_1^N$ may take on arbitrary shapes. This leads naturally to the question of how the regions are chosen. Since each region

$\mathcal{R}_i \subseteq \mathcal{R}$ represents the region of support of a motion model, $\mathcal{R}_i$ is defined to cover those parts of $\mathcal{R}$ where the model is appropriate. Thus the partition is a *segmentation* of the motion field into regions which exhibit similar 2-D motion. This partition structure is allowed to adapt over time to track the apparent motion.

Unfortunately, determining the regions $\mathcal{R}_i$ given the image sequence is a difficult undertaking. Determining the partition requires segmentation of the 2-D motion field, but in order to reliably estimate the motion field required for segmentation, some kind of region of support model is needed. This leads to approaches which use an initial motion estimation process followed by segmentation or, more recently, simultaneous motion estimation and segmentation.

Region-based motion representations often provide accurate and efficient motion representations and as a result are favored for very low bit-rate video encoding. The MPEG-4 standard relies heavily on region-based object representation. As with generalized block models, region-based models require that the shape of the regions be represented in some form.

- Points

  Earlier a distinction was drawn between parametric and nonparametric (dense) motion models. Given the preceding discussion on regions of support for motion models, it is instructive to think of nonparametric models as the extreme case of parametric models where the region of support for each motion model is a single point. Assuming that the image plane $\mathcal{R}$ is sampled on a finite lattice Λ consisting of N unique sampling locations, the partition of Λ consists of the set of N locations $\{\mathcal{R}_i\}_1^N$. Associated with each location $\mathcal{R}_i$ is a parametric motion model. Thus the total number of parameters used to describe the motion field on Λ is of the same order as the number of sample points. Typically a two-parameter motion model is used at each location which makes the tacit assumption of constant velocity along the motion trajectory between frames. As with any of the models discussed above, it is possible to include second order (acceleration) parameters if necessary.

4.6 Relating Motion and Image Intensities

In this section the relationship between the apparent motion and the observed image intensities is discussed. One of the most common assumptions made when attempting to determine optical flow is that the image intensity along a motion trajectory remains constant. This is referred to as the *brightness constancy assumption*. This assumption is implicit in a wide variety of motion estimation techniques even though the formulation of this constraint differs from technique to technique.

4.6.1 The Optical Flow Equation

In their seminal paper, Horn and Schunck [81] formalize the brightness constancy assumption in the Optical Flow Equation (OFE). Since the OFE is one of the most widely used models for the brightness constancy assumption, an in-depth treatment is provided. Denoting the time-varying intensity image as $I(x, y, t)$, let the variable s parameterize a motion trajectory. The assumption that the intensity remains constant along the motion trajectory may be described mathematically by

$$\frac{\mathrm{d}I}{\mathrm{d}s} = 0. \tag{4.4}$$

Applying the chain rule yields the *optical flow equation*

$$\frac{\partial I}{\partial x}\, u + \frac{\partial I}{\partial y}\, v + \frac{\partial I}{\partial t} = 0, \tag{4.5}$$

where $u = \partial x/\partial t$ and $v = \partial y/\partial t$ are the horizontal and vertical components of the optical flow respectively. It is common to write the OFE using the more compact vector notation as

$$\nabla I \cdot \mathbf{v} + \frac{\partial I}{\partial t} = 0, \tag{4.6}$$

where $\nabla I = [\partial I/\partial x \ , \ \partial I/\partial y]^\top$ and $\mathbf{v} = [u, v]^\top$. It is essential to realize that the OFE (Equation (4.6)) is a single equation with two unknown components u and v, thus it is not possible to determine the local motion without additional constraints. Only the component of the optical flow that is in the direction of the intensity gradient ∇I, called the *normal flow*, can be determined (see Figure (4.6)). The normal flow is the projection of $\mathbf{v}$ onto ∇I given by $\mathbf{v} \cdot \nabla I/\|\nabla I\|$. By manipulating Equation (4.6), it can be seen that the normal flow $\mathbf{v}_\perp$ is given by

$$\mathbf{v}_\perp \triangleq \frac{\mathbf{v} \cdot \nabla I}{\|\nabla I\|} = -\frac{\frac{\partial I}{\partial t}}{\|\nabla I\|}. \tag{4.7}$$

Since the OFE constrains only the normal component of the optical flow to equal $-\frac{\partial I}{\partial t}/\|\nabla I\|$, there exists a locus of flow vectors satisfying this constraint. This is indicated by the dashed line in Figure (4.6). The distance from the origin along the intensity gradient vector ∇I to this constraint line is $\frac{\partial I}{\partial t}/\|\nabla I\|$.

From the above discussion, it is clear that given the intensity gradient and temporal partial derivative only the normal component of the optical flow $\mathbf{v}_\perp$ can be estimated without additional constraints. This limitation is often referred to as the *aperture problem* (see Section (4.4)) and indicates that the problem is under-constrained. In order to overcome this limitation additional constraints must be formulated. These are typically obtained by aggregating observations and/or by imposing constraints on the structure of the flow field.

4.6.2 The Gradient Constancy Assumption

Changes in scene illumination generally violate the brightness constancy assumption discussed above, resulting in optical flow which does not correspond to actual

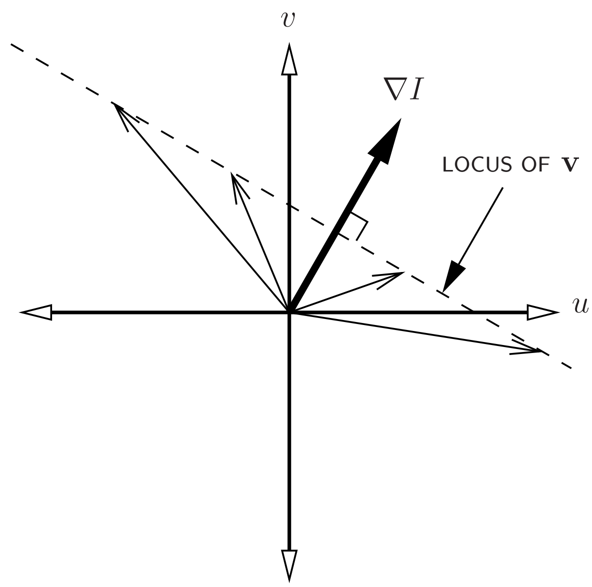


Figure 4.6. The normal flow constraint.

motion. To address this problem, the *gradient constancy assumption*

$$\frac{\mathrm{d}\nabla I}{\mathrm{d}s} = \mathbf{0} \tag{4.8}$$

has been formulated [134]. Unfortunately, this formulation, which assumes that the brightness gradient is locally constant, is violated under transformations such as scaling and rotation, thus limiting its applicability. Additionally, estimation of the second partial derivatives required in Equation (4.8) tends to be error-prone, resulting in less reliable motion estimates.

4.7 Selected Motion Estimation Frameworks

In this section three motion estimation frameworks are discussed in detail:

1. Block-based motion estimation methods

2. The classic Horn-Schunck approach based on the optical flow equation

3. A modern Bayesian approach which accounts for multiple motions

Out of the broad selection of approaches to motion estimation that exist in the literature, these three frameworks have been identified due to their pivotal significance: block-based motion estimation and motion compensation techniques form the heart of motion compensated video coding schemes such as MPEG and ITU-T H.26x and as such represent the most tangible and widespread application of motion estimation. The classic presentation of Horn and Schunck [81] pioneered the use of the optical flow equation for motion estimation and revolutionized the way in which the motion estimation problem was addressed. The Horn-Schunck approach and its successors have found widespread use in computer vision and image sequence processing applications where accurate motion estimation is a requirement. Modern Bayesian methods represent the current state of the art for simultaneous motion estimation and segmentation required for reliable estimation of multiple independently moving objects. These methods offer significant performance improvements over earlier approaches and lend themselves well to applications where accurate motion estimation is critical. These methods provide a framework for object-based motion representation and will thus enable emerging object-based image coding schemes.

4.7.1 Block-Based Motion Estimation

Block matching is the simplest motion estimation algorithm, but is nevertheless ubiquitous in video compression applications. The image is divided into N equal sized blocks $\{\mathcal{R}_i\}_1^N$ with dimensions K pixels by L pixels. Typical sizes for K and L used for compression applications are 8 or 16 pixels. The motion of each block is modeled by the two-parameter translational model in Equation (4.1). Constant velocity motion is assumed between frames (linear temporal model).

In compression applications, the goal is to reduce the cost of coding the residual error between the current frame to be coded and the motion compensated prediction derived from a reference frame. The residual error is minimized on a block by block basis. Denoting the intensity values of the current and reference frames by $f_c(\mathbf{n})$ and $f_r(\mathbf{n})$ respectively, the objective is to determine the motion vector $\mathbf{d}_{c,r}^{\mathcal{R}_i}$ which minimizes the residual error for block $\mathcal{R}_i$. Formally this is expressed as,

$$\mathbf{d}_{c,r}^{\mathcal{R}_i} = \arg \min_{\mathbf{d} \in \mathcal{S}^{\mathcal{R}_i}} \sum_{\mathbf{n} \in \mathcal{R}_i} \rho\big(f_c(\mathbf{n}) - f_r(\mathbf{n} + \mathbf{d})\big). \tag{4.9}$$

The spatial coordinates $\mathbf{n}$ form a discrete lattice, which is commonly the pixel sampling lattice. Subpixel-resolution block motion estimation can be achieved by interpolating between known pixel values. $\rho(x)$ is typically the function x^2 or $|x|$, yielding estimates that minimize the Mean Square Error (MSE) and the Mean Absolute Error (MAE) respectively. Other matching criteria involving nonlinear operators such as the median or maximum matching pixel counts are also possible. The argument of the minimization $\mathbf{d}$ is allowed to range over a set $\mathcal{S}^{\mathcal{R}_i}$, the so-called *search area* (see Figure (4.7)). The difference $f_c(\mathbf{n}) - f_r(\mathbf{n} + \mathbf{d})$ is known as the *displaced block difference* and the collection of such terms for all blocks $\{\mathcal{R}_i\}_1^N$ is called the Displaced Frame Difference (DFD).

Given the optimization implied by Equation (4.9) a procedure must be determined for finding the value of $\mathbf{d}_{c,r}^{\mathcal{R}_i}$ which minimizes the residual error (the objective function). This can be accomplished by exhaustively evaluating the optimization criterion for every location in the search area $\mathcal{S}^{\mathcal{R}_i}$. This is known as *full search.* This approach, though guaranteed to find the optimal value for $\mathbf{d}_{c,r}^{\mathcal{R}_i}$, is computationally expensive so cheaper but suboptimal alternatives have been investigated.

One such method, known commonly as the "three step search" is illustrated in Figure (4.8). The example presented is for a 15 pixel by 15 pixel search area. At the first level, the objective function is evaluated at the nine locations illustrated with large circles. The location with the lowest value of the objective function (shown as the large filled circle) is chosen as the starting location for the second iteration. The

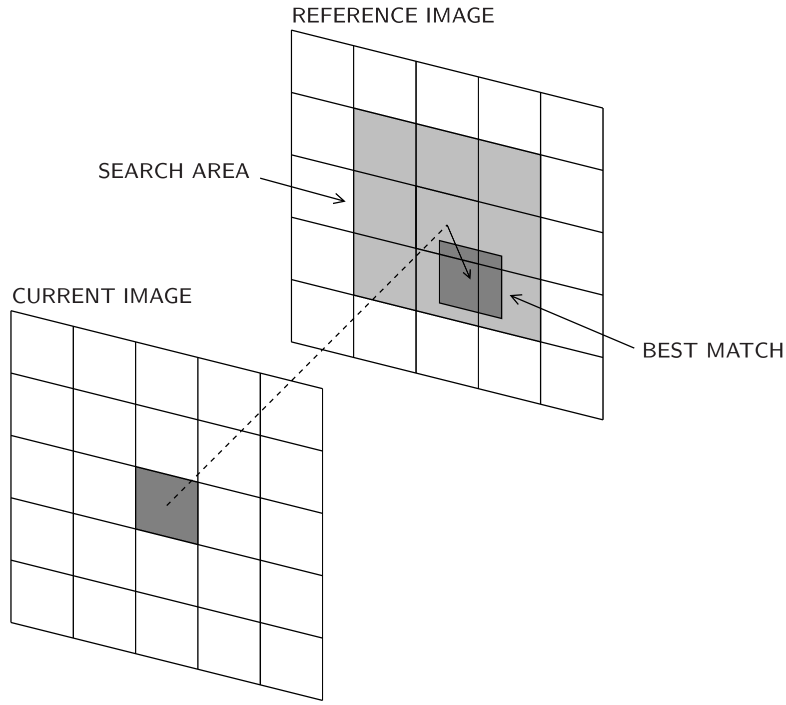


Figure 4.7. Block motion estimation.

objective function is then evaluated at the eight locations (illustrated with squares) surrounding the optimal point from the first iteration. Once again, the optimal location is found and the process is repeated at the third level to find the final estimate.

The three step search is a special case of more general “n-step” or “logarithmic” searches. It is important to realize that these fast search algorithms are based on the assumptions of a unique minimum of the objective function and that the objective function increases monotonically as the estimate moves away from the optimum. In reality the objective function may fail to meet these requirements and the estimate will be suboptimal. This is usually not too serious for compression applications where the goal is not the estimation of the true motion, but rather a motion estimate which will result in a small prediction residual. The motion estimate which achieves this need not necessarily correspond to the true motion.

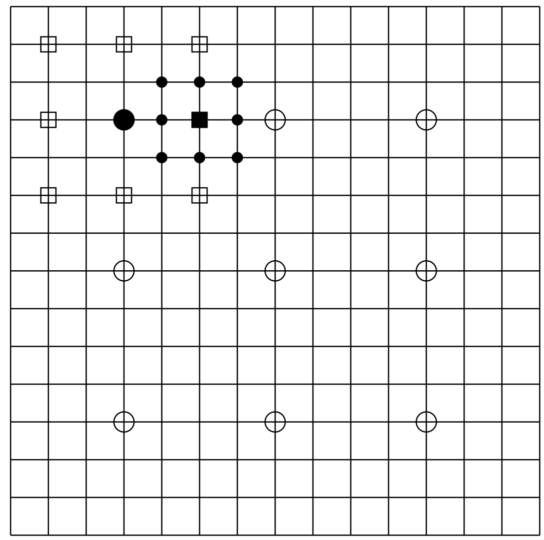


Figure 4.8. The three step search procedure.

Generalizations of the block matching algorithm to more complex parametric motion models are possible and include techniques used for mesh-based motion models. Hierarchical block motion estimation methods based on image pyramids have also been considered for improving the accuracy and computational performance of the basic block matching algorithm. For more details on these topics, see [176, 191, 189].

4.7.2 Motion Estimation using the Optical Flow Equation

In this section motion estimation approaches based on the optical flow equation (Equation (4.5)) are discussed. The presentation begins with the now-classic formulation of Horn and Schunck [81] and continues with a brief discussion of some extensions to, and the significance of, this ground-breaking work.

As discussed in Section (4.6.1), the OFE constrains the motion estimate (u, v) to lie on a line perpendicular to the image brightness gradient ∇I. Thus the local flow cannot be computed without additional constraints. Horn and Schunck addressed this problem by introducing a global smoothness constraint on the local motion estimates. The assumption made is that neighboring points on a moving object are likely to have similar velocities. Horn and Schunck suggested imposing a smoothness constraint which seeks to minimize the square of the magnitude of the optical flow gradients,

$$\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 \quad \text{and} \quad \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2. \tag{4.10}$$

The motion estimation problem is thus posed as the minimization of the error associated with the optical flow equation

$$\mathcal{E}_b = I_x u + I_y v + I_t \tag{4.11}$$

and the smoothness constraint given by

$$\mathcal{E}_c^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2, \tag{4.12}$$

where $I_x = \partial I/\partial x$, $I_y = \partial I/\partial y$ and $I_t = \partial I/\partial t$. Introducing the parameter α which gives the relative weight of each term, the total error to be minimized is given by

$$\mathcal{E}^2 = \iint \left(\mathcal{E}_b^2 + \alpha^2 \mathcal{E}_c^2\right) \,\mathrm{d}x\,\mathrm{d}y. \tag{4.13}$$

In practice, the minimization is performed on sampled data, so the integrals are replaced by summations and the partial derivatives are approximated from the sampled data. By applying the calculus of variations the optimal solution to the minimization problem can be shown to be the solution to the equations

$$\begin{aligned} I_x^2 u + I_x I_y v &= \alpha^2 \nabla^2 u - I_x I_t \\ I_x I_y u + I_y^2 v &= \alpha^2 \nabla^2 v - I_y I_t. \end{aligned} \tag{4.14}$$

Horn and Schunck solved this problem using a Gauss-Seidel iteration of the form

$$\begin{aligned} u^{(n+1)} &= \bar{u}^{(n)} - I_x \left[I_x \bar{u}^{(n)} + I_y \bar{v}^{(n)} + I_t\right] / \left(\alpha^2 + I_x^2 + I_y^2\right) \\ v^{(n+1)} &= \bar{v}^{(n)} - I_y \left[I_x \bar{u}^{(n)} + I_y \bar{v}^{(n)} + I_t\right] / \left(\alpha^2 + I_x^2 + I_y^2\right), \end{aligned} \tag{4.15}$$

where $\bar{u}$ and $\bar{v}$ are respectively the weighted averages of the horizontal and vertical motion estimates of the eight nearest spatial neighbors of each site.

The iterative procedure in Equation (4.15) has interesting characteristics. In regions of the image where the brightness gradient is zero, the velocity estimates are assigned to be the average of the surrounding estimates. Thus motion estimates propagate from areas of high spatial gradient to areas with small gradient, "filling in" uniform regions.

Estimation of the spatial and temporal gradient information required for application of the Horn-Schunck optical flow estimation algorithm is nontrivial. The original Horn-Schunck formulation used $2 \times 2 \times 2$-point spatiotemporal finite difference approximations for computing the necessary gradients. This simple approach is highly sensitive to noise and can lead to biases in the motion estimates. More sophisticated methods utilize smooth interpolation kernels or polynomial curve fitting methods for estimating the gradients.

The smoothness constraint in Equation (4.12) is applied globally. Though the constraint is appropriate for regions with similar motion, it leads to erroneous smoothing across motion boundaries, as is evident from the example illustrated in Figure (4.9). In an attempt to address this problem, directional-smoothness constraints have been formulated which suspend the smoothness constraint in the direction of the image gradient [134, 133]. Hildreth [78] proposed minimizing Equation (4.13) along object contours. None of these techniques adequately addresses the problems of occlusions.

The Horn-Schunck formulation, which includes the use of a smoothness constraint, is an early example of a *regularized* solution to an *ill-posed* problem. Recall that the problem of motion estimation is inherently ill-posed. In addition to other difficulties, the motion estimation problem typically fails to have a unique solution due to the aperture problem. By imposing a smoothness constraint, a new regularized problem is formulated to closely resemble the characteristics of the original ill-posed problem, but with a unique and well-behaved solution.

An example of the application of the Horn-Schunck method for computing optical flow is illustrated in Figure (4.9). In this example, a pair of occluding squares undergo independent translational motion with velocities $(u, v) = (1, 1)$ and $(u, v) = (-1, -1)$ respectively. The situation presented in this figure is deliberately analogous to the theoretical scenario presented in Figure (4.5). The resulting optical flow field computed using the Horn-Schunck method illustrates many of the characteristics of, and problems associated with, motion estimation methods utilizing a global smoothness constraint:

1. Propagation of local motion information:

   Imposing a local smoothness constraint results in the propagation of motion information from regions of high spatial variation where the optical flow equation yields an estimate of the normal flow, to regions with small spatial variation. In the regions where the estimated flow is ambiguous due to the aperture problem (the vertical and horizontal edges in this example), the component orthogonal to the normal flow is determined from the propagation of information from nearby regions where the flow is fully determined.

2. "Filling in" of areas where the spatial gradient is small:

   Related to the previous point, in regions where the spatial gradient is small, the smoothness constraint results in blank areas being "filled in" from surrounding regions with high spatial variation where the optical flow can be estimated.

3. Excessive smoothing across motion boundaries:

Examining Figure (4.9) it is clear that the global smoothness constraint results in excessive smoothing across the motion boundaries. Though the degree of smoothing may be controlled via the regularization parameter α, the optical flow fields computed for too small α (the underregularized case) are highly susceptible to noise in the data and do not exhibit the propagation of motion information to regions where the estimates are ambiguous.

4. Failure to accommodate multiple motions:

   The Horn-Schunck method, which uses a global smoothness constraint and includes no accommodation for multiple independent motion fails badly in the presence of multiple motions as illustrated in Figure (4.9). In this example, the motion field is exactly described by two rigidly translating regions with a static background. The Horn-Schunck approach, which includes no notion of segmented regions, instead models the motion field using a nonparametric motion field which is equivalent to a single elastic motion model. This result is clearly incompatible with the physical motion.

5. Optical flow which does not correspond to any physical motion:

   Recall Figure (4.5) in which the optical flow estimated at apertures G and H did not correspond to the motion of the objects in the scene, but were instead artifacts resulting from the occluding motion of the two independently translating objects. This effect is evident in the optical flow computed using the Horn-Schunck procedure illustrated in Figure (4.9). Because the Horn-Schunck procedure includes no model for the independence of the moving objects it cannot account for this effect.

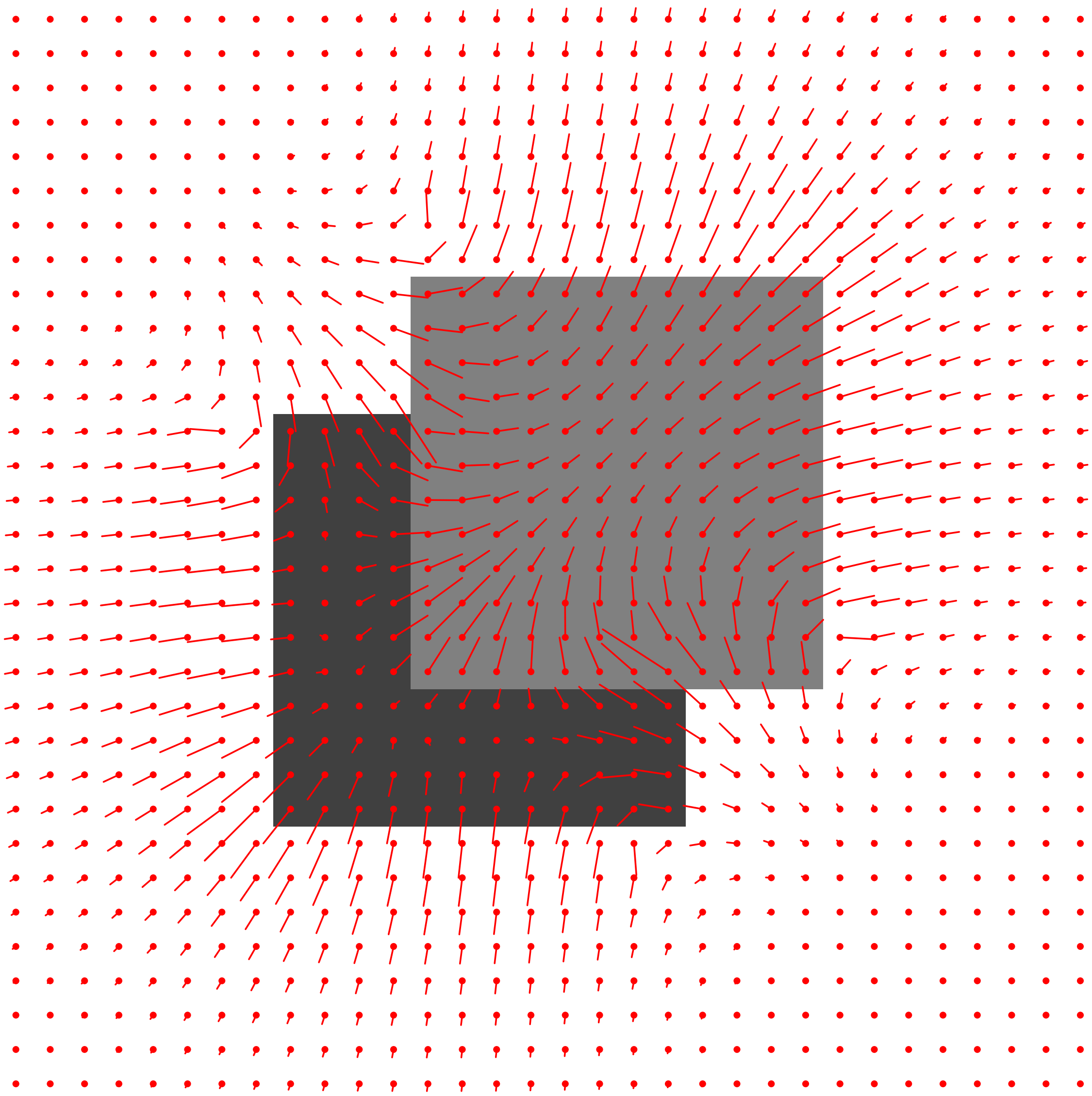


Figure 4.9. Optical flow computed using the Horn-Schunck method for occluding squares undergoing independent translational motion.

4.7.3 A Bayesian Method for Multiple Motions

Since the seminal work of Geman and Geman [60], Bayesian methods have become prominent in image and video processing. In this section, the Bayesian framework for simultaneous motion estimation and segmentation proposed by Chang, Tekalp and Sezan [31] is presented. Since the discussion is fully detailed, this section represents a complete example of the application of Bayesian techniques and thus may be considered to be an in-depth tutorial. The approach taken here is representative of many modern Bayesian approaches for solving problems in imaging. For another excellent example of the application of the Bayesian approach to motion estimation see the work of Konrad and Dubois [118].

In the presentation, lexicographically ordered vectors (row by row ordering of a 2-D image into a one-dimensional (1-D) vector) are represented in boldface, while elements of the vectors are in normal face. The motion vector field $\mathbf{d}$ for each frame is modeled as the sum of a parametric field $\mathbf{d}_p$ which is dependent on the segmentation labels $\mathbf{s}$, and a residual field $\mathbf{d}_r$ as

$$d(m,n) = d_p(m,n) + d_r(m,n). \tag{4.16}$$

It is assumed that there are N independently moving opaque objects in the scene, so the segmentation field $s(m,n)$ assumes values from the set $\{1, 2, \ldots, N\}$ and assigns each motion vector $d(m,n)$ to one of the N classes.

Given the current frame $\mathbf{f}_{\mathrm{c}}$ and a reference frame $\mathbf{f}_{\mathrm{r}}$, the authors propose computing the MAP estimate of the horizontal ($\mathbf{u}$) and vertical ($\mathbf{v}$) components of the motion field $\mathbf{d}$ and the segmentation field $\mathbf{s}$. The posterior probability is given by the expression, $\mathcal{P}(\mathbf{u}, \mathbf{v}, \mathbf{s} \,|\, \mathbf{f}_{\mathrm{c}}, \mathbf{f}_{\mathrm{r}})$. By applying Bayes' rule, the posterior probability can be expressed as

$$\mathcal{P}(\mathbf{u}, \mathbf{v}, \mathbf{s} \,|\, \mathbf{f}_{\mathrm{c}}, \mathbf{f}_{\mathrm{r}}) = \frac{\mathcal{P}(\mathbf{f}_{\mathrm{c}} \,|\, \mathbf{u}, \mathbf{v}, \mathbf{s}, \mathbf{f}_{\mathrm{r}})\mathcal{P}(\mathbf{u}, \mathbf{v} \,|\, \mathbf{s}, \mathbf{f}_{\mathrm{r}})\mathcal{P}(\mathbf{s} \,|\, \mathbf{f}_{\mathrm{r}})}{\mathcal{P}(\mathbf{f}_{\mathrm{c}} \,|\, \mathbf{f}_{\mathrm{r}})}. \tag{4.17}$$

The MAP estimate for $(\mathbf{u}, \mathbf{v}, \mathbf{s})$ corresponds to the values $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{s}$ which maximize the posterior probability. Noting that the denominator in Equation (4.17) is constant with respect to $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{s}$, the MAP estimate $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{s}})$ may be expressed as

$$(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{s}}) = \arg\max_{(\mathbf{u},\mathbf{v},\mathbf{s})} \mathcal{P}(\mathbf{f}_{\mathrm{c}} \,|\, \mathbf{u}, \mathbf{v}, \mathbf{s}, \mathbf{f}_{\mathrm{r}}) \mathcal{P}(\mathbf{u}, \mathbf{v} \,|\, \mathbf{s}, \mathbf{f}_{\mathrm{r}}) \mathcal{P}(\mathbf{s} \,|\, \mathbf{f}_{\mathrm{r}}). \tag{4.18}$$

Equation (4.18) consists of three probability expressions. The significance of each expression is interpreted in turn:

1. $\mathcal{P}(\mathbf{f}_{\mathrm{c}} \,|\, \mathbf{u}, \mathbf{v}, \mathbf{s}, \mathbf{f}_{\mathrm{r}})$

   The conditional probability density function (pdf) $\mathcal{P}(\mathbf{f}_{\mathrm{c}} \,|\, \mathbf{u}, \mathbf{v}, \mathbf{s}, \mathbf{f}_{\mathrm{r}})$ describes how well the current frame $\mathbf{f}_{\mathrm{c}}$ is modeled given the motion vector field components $\mathbf{u}$ and $\mathbf{v}$, the motion segmentation $\mathbf{s}$ and the reference frame $\mathbf{f}_{\mathrm{r}}$. More technically, this term describes the *likelihood* of observing the current image given the motion field, the segmentation map and the reference image. A Gibbs distribution with the energy function described by

   $$\mathcal{E}_1(\mathbf{f}_{\mathrm{c}} \,|\, \mathbf{u}, \mathbf{v}, \mathbf{s}, \mathbf{f}_{\mathrm{r}}) = \sum_{(m,n)} \left\| \epsilon(m,n) \right\|_2^2, \tag{4.19}$$

   where

   $$\epsilon(m,n) = f_{\mathrm{c}}(m,n) - f_{\mathrm{r}}\left(m + u(m,n), n + v(m,n)\right) \tag{4.20}$$

   is used to model this dependence. The term $\epsilon(m,n)$ is the residual error in the motion compensated prediction of the pixel $f_{\mathrm{c}}(m,n)$ from the reference frame $\mathbf{f}_{\mathrm{r}}$. Thus the Gibbs energy function $\mathcal{E}_1$ computes the energy of the DFD and represents the degree to which the brightness constancy constraint is satisfied for given $\mathbf{d}$, $\mathbf{s}$, $\mathbf{f}_{\mathrm{c}}$ and $\mathbf{f}_{\mathrm{r}}$. Since maximizing the posterior probability of a Gibbs distribution corresponds to minimizing the Gibbs energy function, the MAP solution ensures that the DFD is minimized. It is interesting to note that this formulation of the Gibbs energy for the likelihood function is equivalent to assuming that the pixel errors in the motion compensated prediction are IID zero mean Gaussian random variables with variance $\sigma^2 = 0.5$. Other models for the error include the longer-tailed Laplacian pdf (L^1 norm), generalized Gaussian pdf, as well as robust measures with outlier rejection.

2. $\mathcal{P}(\mathbf{u}, \mathbf{v} \,|\, \mathbf{s}, \mathbf{f}_{\mathrm{r}})$

   The second term, $\mathcal{P}(\mathbf{u}, \mathbf{v} \,|\, \mathbf{s}, \mathbf{f}_{\mathrm{r}})$ represents the conditional pdf of the motion vector field given the segmentation field and the reference frame. A Gibbs distribution is again used to model this dependence, this time with the Gibbs energy described by

$$\begin{aligned}\mathcal{E}_2(\mathbf{u}, \mathbf{v} \,|\, \mathbf{s}, \mathbf{f}_{\mathrm{r}}) = \alpha \sum_{(m,n)} \left\| d(m,n) - d_p(m,n) \right\|_2^2 \\ + \beta \sum_{(m,n)} \sum_{(i,j) \in \mathcal{N}_{(m,n)}} \left\| d(m,n) - d(i,j) \right\|_2^2 \delta\big(s(m,n) - s(i,j)\big). \quad (4.21)\end{aligned}$$

   Recall that $d(m,n) = d_p(m,n) + d_r(m,n)$, so $\left\| d(m,n) - d_p(m,n) \right\|_2^2 = \left\| d_r(m,n) \right\|_2^2$. Thus the first term measures the energy of the residual motion vector field $\mathbf{d}_r$. Since this term is to be minimized, this corresponds to a minimum norm residual field $\mathbf{d}_r$ while the parametric field $\mathbf{d}_p$ minimizes the DFD. The second term describes the energy associated with the motion vector field, modeled as a MRF. Smoothness is imposed between neighbors if and only if they are within the same labeled region (this is controlled by the multiplicative Kronecker delta term). α and β describe the relative weight given to each term.

3. $\mathcal{P}(\mathbf{s} \,|\, \mathbf{f}_{\mathrm{r}})$

   The third conditional probability term, $\mathcal{P}(\mathbf{s} \,|\, \mathbf{f}_{\mathrm{r}})$ models the dependence of the region labels on the reference image $\mathbf{f}_{\mathrm{r}}$. The authors choose not to model dependence on $\mathbf{f}_\mathrm{r}$, but instead use this term to model the structure of the segmentation field independent of $\mathbf{f}_\mathrm{r}$. In particular they chose a Gibbs distribution with a energy function given by

$$\mathcal{E}_3(\mathbf{s} \,|\, \mathbf{f}_\mathrm{r}) = \gamma \sum_{(m,n)} \sum_{(i,j)\in\mathcal{N}_{(m,n)}} V_2\big(s(m,n), s(i,j)\big) \tag{4.22}$$

where

$$V_2\big(s(m,n), s(i,j)\big) = \begin{cases} +1 & \text{if} \quad s(m,n) = s(i,j) \\ -1 & \text{otherwise} \end{cases} . \tag{4.23}$$

This is, once again, a MRF model with Equation (4.23) describing the two-site clique potential.

It is interesting to note that the second and third conditional probability terms, $\mathcal{P}(\mathbf{u}, \mathbf{v} \,|\, \mathbf{s}, \mathbf{f}_\mathrm{r})$ and $\mathcal{P}(\mathbf{s} \,|\, \mathbf{f}_\mathrm{r})$ are used to model *prior knowledge* of the properties of the solution. In particular, these terms model piecewise smoothness of the motion field and global smoothness of the segmentation field. When these pdf terms do not involve the observations, they are termed "priors" as they represent the a-priori knowledge of the distribution of the solution. The use of prior information to constrain the solution places Bayesian methods in the class of stochastic regularization techniques.

Combining the Gibbs energy functions in Equation (4.19), Equation (4.21) and Equation (4.22) yields the Gibbs posterior probability density

$$\mathcal{P}(\mathbf{u}, \mathbf{v}, \mathbf{s} \,|\, \mathbf{f}_\mathrm{c}, \mathbf{f}_\mathrm{r}) = \frac{1}{Z} \exp\left\{-\mathcal{E}_1(\mathbf{f}_\mathrm{c} \,|\, \mathbf{u}, \mathbf{v}, \mathbf{s}, \mathbf{f}_\mathrm{r}) - \mathcal{E}_2(\mathbf{u}, \mathbf{v} \,|\, \mathbf{s}, \mathbf{f}_\mathrm{r}) - \mathcal{E}_3(\mathbf{s} \,|\, \mathbf{f}_\mathrm{r})\right\} \tag{4.24}$$

where Z is a scaling constant called the *partition function.* Since the logarithm is an increasing function, the argument which maximizes Equation (4.24) also maximizes the logarithm of Equation (4.24), thus it is equivalent to maximize

$$\log \mathcal{P}(\mathbf{u}, \mathbf{v}, \mathbf{s} \,|\, \mathbf{f}_\mathrm{c}, \mathbf{f}_\mathrm{r}) = -\log Z - \mathcal{E}_1(\mathbf{f}_\mathrm{c} \,|\, \mathbf{u}, \mathbf{v}, \mathbf{s}, \mathbf{f}_\mathrm{r}) - \mathcal{E}_2(\mathbf{u}, \mathbf{v} \,|\, \mathbf{s}, \mathbf{f}_\mathrm{r}) - \mathcal{E}_3(\mathbf{s} \,|\, \mathbf{f}_\mathrm{r}), \tag{4.25}$$

or minimize

$$-\log \mathcal{P}(\mathbf{u},\mathbf{v},\mathbf{s}\,|\,\mathbf{f}_{\mathrm{c}},\mathbf{f}_{\mathrm{r}}) = \mathcal{E}_1(\mathbf{f}_{\mathrm{c}}\,|\,\mathbf{u},\mathbf{v},\mathbf{s},\mathbf{f}_{\mathrm{r}}) + \mathcal{E}_2(\mathbf{u},\mathbf{v}\,|\,\mathbf{s},\mathbf{f}_{\mathrm{r}}) + \mathcal{E}_3(\mathbf{s}\,|\,\mathbf{f}_{\mathrm{r}}), \tag{4.26}$$

having dropped the constant term $\log Z$ which does not alter the argument resulting in the minimum. Finally, expanding the terms involving the energy functions $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_3$, results in the problem of minimizing the energy

$$\begin{aligned}\mathcal{E}_{\mathbf{u},\mathbf{v},\mathbf{s}} = \sum_{(m,n)} \Bigg\{ &\left\|\epsilon(m,n)\right\|_2^2 \;+\; \alpha\, \left\|d(m,n) - d_p(m,n)\right\|_2^2 \\ +\;\; \beta &\sum_{(i,j)\in\mathcal{N}_{(m,n)}} \left\|d(m,n) - d(i,j)\right\|_2^2 \,\delta\big(s(m,n) - s(i,j)\big) \\ +\;\; \gamma &\sum_{(i,j)\in\mathcal{N}_{(m,n)}} V_2\big(s(m,n), s(i,j)\big) \Bigg\}.\end{aligned} \tag{4.27}$$

Finding the MAP estimate for the motion vector field $\mathbf{d}$ (with components $\mathbf{u}$ and $\mathbf{v}$) and the segmentation field $\mathbf{s}$ requires the minimization of the energy expression in Equation (4.27). The Gibbs energy in Equation (4.27) describes all the interacting constraints formulated through the conditional probability terms.

To review, the term $\left\|\epsilon(m,n)\right\|_2^2$ represents the energy of the DFD (the energy of the error in the motion compensated prediction) and ensures that the motion estimates are compatible with the observed data. The term scaled by α ensures that the parametric motion field $\mathbf{d}_p$ is responsible for modeling the majority of the variation of the total motion vector field $\mathbf{d}$, with the residual field $\mathbf{d}_r$ being of minimum energy. The term scaled by β ensures that the motion field $\mathbf{d}$ within each labeled region is smooth and the γ term ensures that the segmentation field $\mathbf{s}$ is smooth.

Being highly nonlinear, direct minimization of Equation (4.27) is a challenging task, so the authors propose a two step, iterative procedure:

1. Update the motion field $\mathbf{d}$ given the current estimate of the segmentation field

s. This requires minimizing

$$
\begin{aligned}
\mathcal{E}_{\mathbf{u},\mathbf{v}} = & \sum_{(m,n)} \big\|\epsilon(m,n)\big\|_2^2 \; + \; \alpha \sum_{(m,n)} \big\|d(m,n) - d_p(m,n)\big\|_2^2 \\
& + \; \beta \sum_{(m,n)} \sum_{(i,j)\in\mathcal{N}_{(m,n)}} \big\|d(m,n) - d(i,j)\big\|_2^2 \, \delta\big(s(m,n) - s(i,j)\big) \qquad (4.28)
\end{aligned}
$$

with respect to **d**. The Highest Confidence First (HCF) method [36] is used to perform this minimization.

2. Update the segmentation field **s** using the current estimate of the motion field **d**. This involves minimizing

$$
\begin{aligned}
\mathcal{E}_{\mathbf{s}} = & \; \alpha \sum_{(m,n)} \big\|d(m,n) - d_p(m,n)\big\|_2^2 \\
& + \; \beta \sum_{(m,n)} \sum_{(i,j)\in\mathcal{N}_{(m,n)}} \big\|d(m,n) - d(i,j)\big\|_2^2 \, \delta\big(s(m,n) - s(i,j)\big) \\
& + \; \gamma \sum_{(m,n)} \sum_{(i,j)\in\mathcal{N}_{(m,n)}} V_2\big(s(m,n), s(i,j)\big) \qquad (4.29)
\end{aligned}
$$

with respect to the segmentation field **s**. Terms involving the motion vector field are included as **d** and **s** are interdependent. The minimization is performed using the HCF approach. Once **s** is updated, the parametric motion field parameters are updated using a least squares estimate with outlier rejection based on the new segmentation field estimate.

The weight parameters α, β and γ are determined experimentally. As mentioned previously, the number of objects N is assumed to be known a-priori.

Some comments are in order regarding the MAP motion estimation framework. This detailed presentation demonstrates the typical structure of the Bayesian approach. First the problem is formulated in a probabilistic framework with modeling resulting in an a-posteriori probability expression which relates the unknowns to be estimated to the observations. Bayes' rule is invoked to break the posterior probability expression into likelihood and prior terms. Each of these terms is modeled according to knowledge of the problem domain and *a-priori* knowledge of the solution. Maximizing the *a-posteriori* probability implies an optimization problem. Often this optimization cannot be solved in closed form, so iterative methods are used.

In this section a Bayesian motion estimation approach that simultaneously performs motion estimation and segmentation was examined. This combined approach directly addresses the problem of optimal constraint aggregation to resolve the motion estimation ambiguity of the aperture problem. Local motion estimates are aggregated only if they are supported under a single motion hypothesis (a labeled region) and regions are labeled according to the structure of the motion field.

Recently, techniques based on robust estimation have come to the fore. Some extend the Bayesian framework by using robust, but nonconvex Gibbs energy functions [13, 14], while others have used layered representations and clustering techniques [205], and mixture models and Bayesian inference [207]. Most of these approaches are capable of estimating the number of moving objects N. Motion estimation remains a challenging and active research area.

4.8 Motion Models for Applications

It should be clear from the preceding discussions that the type of motion information that is appropriate for a given application is highly dependent on the specifics of that application. Or, put more succinctly, there is no single motion representation that is appropriate for all applications.

This is especially clear when considering compression applications, where the objective is not accurate motion representation per sé, but motion compensation to reduce video transmission bit-rate for a given video Coder/Decoder (CODEC) implementation complexity. Motion information is used in compression applications to remove temporal redundancy by predicting image values using knowledge of the motion of regions in previously transmitted frames. The motion information (and typically a residual error) is transmitted to enable the receiver to reconstruct the motion compensated frame. For compression, where the objective is reduced bit-rate transmission, accurate motion compensation is traded off against the cost of encoding both the motion information and the residual error. An additional requirement is that the motion compensation scheme be amenable to real-time and Very Large Scale Integration (VLSI) implementation. Though it may be possible to utilize a dense motion representation to achieve highly accurate motion compensation thereby reducing the residual error, the cost of transmitting a dense representation outweighs the gain in reduced residual error. For this reason, and for reduced computational complexity, video compression standards utilize parametric models, especially block-based and more recently, region-based and global motion methods. The motion models used are invariably translation only so as to minimize the cost of coding the motion information as well as to minimize decoder complexity. Despite the simplicity of the motion representations used for compression applications, this presently represents the complexity/performance optimum.

For many video processing applications, however, the situation is markedly different. In computer vision, video restoration, video tracking and similar applications, motion estimation is often the performance limiting factor. For this reason, emphasis is placed on the choice of appropriate motion models and the performance of the motion estimator applied. In restoration applications computational requirements often take second place to performance issues so sophisticated, often computationally intensive, motion estimation techniques are common. Dense motion representations as well as high-order parametric region-based methods are common for such applications.

CHAPTER 5

AN EMPIRICAL STUDY OF THE PERFORMANCE OF SELECTED BLOCK-MATCHING SUBPIXEL-RESOLUTION MOTION ESTIMATORS

5.1 Introduction

This chapter presents an empirical study of the performance of selected block-matching subpixel-resolution motion estimation algorithms under the adverse conditions of image undersampling and additive noise. The study is motivated by the requirement for reliable subpixel-resolution motion estimates for the motion compensated observation models used in multiframe superresolution restoration.

Idealized test functions representing a scene subject to translational motion are defined. These functions are subsampled and corrupted with additive noise and used as source data for various block-matching subpixel-resolution motion estimation techniques. Motion estimates computed from these data are compared with the known motion which enables an assessment of the performance of the motion estimators considered.

Subpixel-resolution motion estimation techniques are widely used in video coding for transmission and/or compression where half-pixel-resolution estimators are routinely used [86, 91]. A less well known, but more demanding application is in the emerging field of superresolution video and image restoration where a sequence of noisy, undersampled low-resolution observed frames are used to restore superresolved imagery [168, 70, 149, 48]. Accurate subpixel-resolution motion compensation is critical to the success of superresolution restoration algorithms since, (i) motion information provides essential constraints for the solution of the ill-posed superresolution restoration problem, (ii) erroneous motion estimates result in objectionable artifacts in the restoration, and (iii) the degree of resolution enhancement possible is dependent on the resolution and accuracy of the motion information [162].

Since the subpixel-resolution motion information required for superresolution restoration must be estimated from the noisy and undersampled observed data, questions concerning the performance of subpixel-resolution motion estimation techniques under these adverse conditions naturally arise. While it is tempting to apply increasingly higher resolution motion estimation methods in the hope of obtaining improved superresolution restoration, this work shows that this is impractical.

In order to demonstrate why this is the case, the "realistic best case" performance of commonly used block-matching subpixel-resolution motion estimators is examined. The reason for this is simple — to provided *typical upper bounds* on performance. It is pointless to address worst case performance since this is represented by the case of constant valued regions for which local motion estimation is meaningless. In order to evaluate the "realistic best case" performance, motion estimators are tested using synthetic data which are amenable to accurate motion estimation. These data are not optimal in the sense that they are the data for which the motion estimators considered yield the *best possible* accuracy. Rather, they are chosen so as to represent features of real-world imagery with the property that, in normal usage, the performance of the motion estimators considered are typically bounded above by the performance on the synthetic data.

It is important to note that this work only addresses the problem of local motion estimation. It does not consider improvements that are possible using constraints on the total motion field.

The development begins with the modeling of an imaging system consisting of a diffraction-limited optical system with a light-sensing FPA which is subject to additive noise. The synthetic scene data are generated and subjected to known translational motion. The imaging process is then simulated, using the synthetic scenes as input imagery, to generate observation data which resemble the output of a typical FPA camera system. Block-matching subpixel-resolution motion estimation techniques are applied to the simulated camera imagery resulting in motion estimates which can be compared with the *a-priori* known motion. This approach affords complete knowledge of the true motion and thus enables meaningful statistical evaluation of the performance of the motion estimators.

5.2 Imaging System Modeling

The imaging system modeled consists of a diffraction-limited optical system and a FPA camera operating at optical wavelengths. Models for these components are developed below.

5.2.1 Optical System Model

The optical system model assumes diffraction-limited optics with a circular exit pupil operating in the visible spectrum with an f-number ($f/\#$) of 2.8. The Modulation Transfer Function (MTF) for such a system can be shown to be of the form [63],

$$H_o(u,v) = \begin{cases} \dfrac{2}{\pi}\left[\cos^{-1}\left(\dfrac{\rho}{\rho_c}\right) - \dfrac{\rho}{\rho_c}\sqrt{1-\left(\dfrac{\rho}{\rho_c}\right)^2}\,\right], & \text{for } \rho < \rho_c \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

where $\rho = \sqrt{u^2+v^2}$ and u, v are spatial frequency variables. The radial cut-off frequency is given by,

$$\rho_c = \frac{1}{\lambda \cdot f/\#}. \tag{5.2}$$

The wavelength is chosen in the visible spectrum with $\lambda = 670$nm, resulting in a cut-off at approximately 533 lines per mm. The optical system MTF is shown in Figure (5.2)(a).

5.2.2 Focal Plane Array Model

Focal plane sensor arrays are ubiquitous in imaging applications and are therefore considered in this work. Typical examples include Charge-Coupled Device (CCD), Charge-Injection Device (CID) and Complementary Metal-Oxide Semiconductor (CMOS) sensor arrays. The FPA is modeled as a regular array of independent sensing elements or pixels which are assumed to be square with edge dimension X_s. Each pixel performs a spatiotemporal integration of the incident illumination over the light-sensitive active region of the pixel for the duration of the aperture time to yield a measurement. Several simplifying assumptions are made concerning the FPA: (i) the entire area of each pixel is light-sensitive, or equivalently, the pixel has 100% fill-factor (the ratio of light-sensing to light-insensitive area of the pixel), (ii) the spatial response over the active region is constant and (iii) the aperture time is of sufficiently short duration that no blurring occurs due to scene motion. The geometry of the simplified FPA is shown in Figure (5.1). In this work, the sampling aperture of each pixel is assumed to have dimensions of 9μm square which is a typical value for modern general purpose sensor arrays.

Using the above assumptions, the PSF of a FPA pixel may be modeled as,

$$h_{\mathrm{FPA}}(x, y) = \frac{1}{X_s^2} \operatorname{rect}\left(\frac{x}{X_s}\right) \operatorname{rect}\left(\frac{y}{X_s}\right) = \begin{cases} 1/X_s^2, & |x| \leq X_s/2 \text{ and } |y| \leq X_s/2 \\ 0, & \text{otherwise} \end{cases} . \tag{5.3}$$

The 2-D Fourier transform is given by,

$$H_{\mathrm{FPA}}(u, v) = \operatorname{sinc}(X_s u) \operatorname{sinc}(X_s v) = \frac{\sin(\pi X_s u)}{\pi X_s u} \cdot \frac{\sin(\pi X_s v)}{\pi X_s v}. \tag{5.4}$$

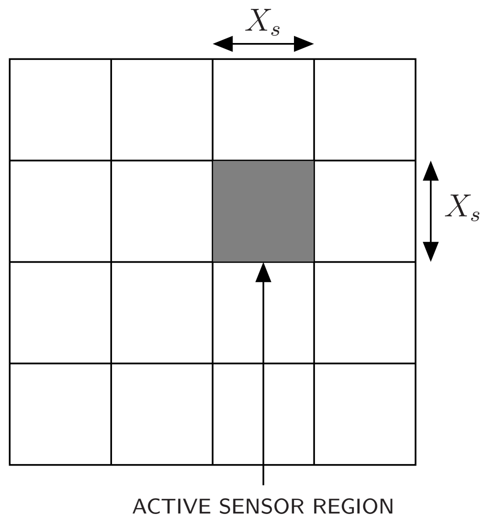


Figure 5.1. Geometry of the simplified focal plane array model.

A plot of the magnitude of Equation (5.4) is shown in Figure (5.2)(b). This plot also corresponds to the MTF since the peak at the origin has unit magnitude. Notice that the spatial integration performed over the active region of the sensor has the effect of a lowpass filter. Nulls of the spatial frequency response of the sampling aperture occur at integer multiples (excluding zero) of the sampling rate $1/X_s$.

5.2.3 Combined Focal Plane Array and Optical System Response

The MTF of the combined response of the optical system and a single sampling aperture of the FPA is found by taking the product of the optical system MTF and the sampling aperture MTF. The combined MTF is shown in Figure (5.2)(c). The PSF of a FPA pixel, taking into account the response of the optical system, is shown in Figure (5.2)(d).

5.2.4 Sampling Considerations

Next it is shown that images captured using typical FPA cameras (as modeled above) fail to meet the Nyquist criterion. It is assumed the scene is not bandlimited.

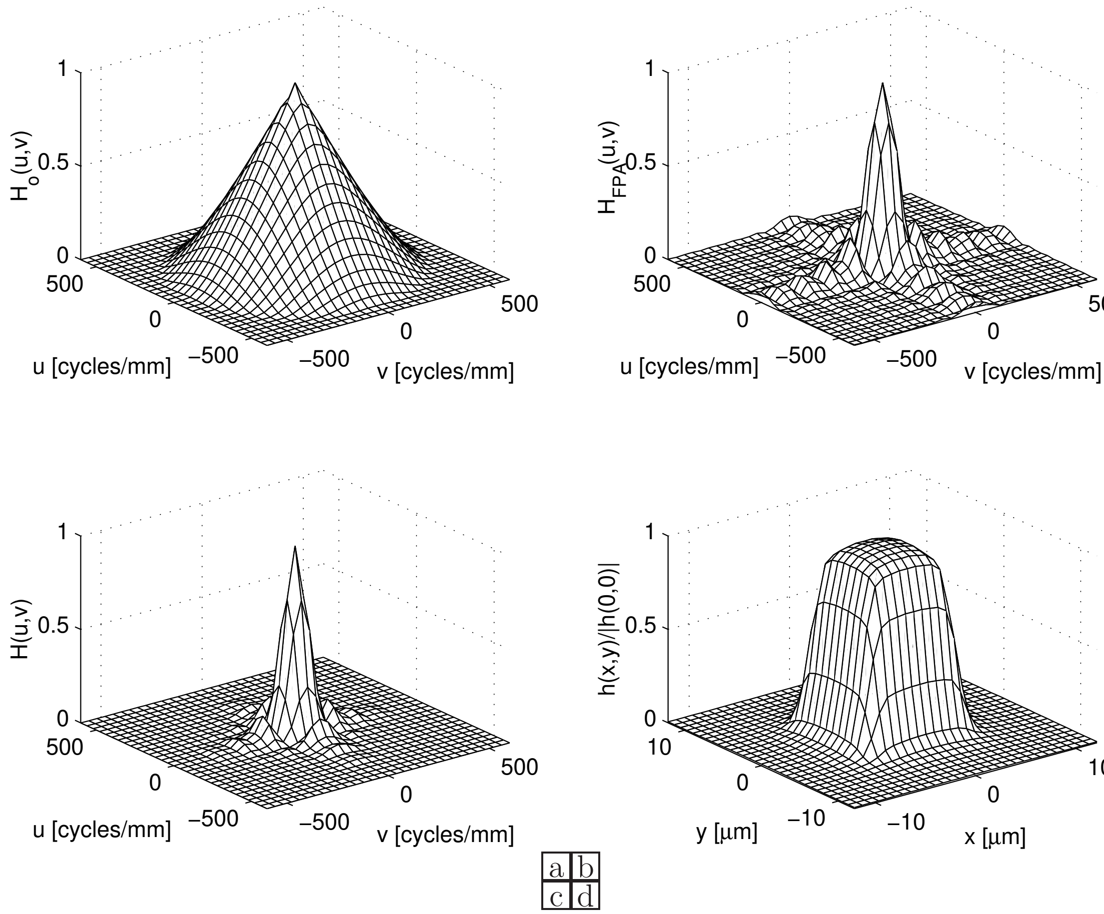


Figure 5.2. (a) MTF of optical system. (b) MTF of sampling aperture. (c) Combined MTF of optical system and sampling aperture. (d) Normalized PSF of combined optical system and sampling aperture.

Two sources of lowpass filtering exist in the imaging system: (i) the diffraction-limited optical system with radial frequency cut-off ρ_c in Equation (5.1) and (ii) the lowpass sinc filter resulting from the sampling aperture in Equation (5.4).

Consider first the optical system: It is clear from the optical system MTF shown in Figure (5.2)(a) that the optical system is a lowpass filter, albeit with a wide passband. If the optical system is to be utilized as an antialiasing filter, then sampling at the Nyquist rate requires approximately 1066 samples per mm, corresponding to a sensor spacing of approximately 0.94μm. This sampling density is impossible to achieve with current focal plane array technologies. Thus the optical system cannot be relied on to ensure that the image is sufficiently bandlimited for critical sampling.

Next consider the FPA: Nyquist rate sampling requires that the image be sampled at twice the highest frequency component present. The FPA sensor spacing determines the sampling rate at $f_s = 1/X_s = 1/9\mu\text{m} \approx 111$ samples per mm. Nyquist sampling requires that the image be bandlimited to $f_{\text{Nyquist}} = 1/2X_s \approx 55$ cycles per mm. Examining Figure (5.2), it is clear that this requirement is not met.

Though the spatial integration performed by the FPA sampling aperture is in effect a low pass filter, it should be noticed that the first zeros in the frequency response occur (assuming the size of the active area of each sensor is equal to the spacing between sensors) at $f = 1/X_s = 2f_{\text{Nyquist}}$. Thus the lowpass characteristic of the FPA sampling aperture is insufficient to ensure critical sampling. This argument does not even consider that the FPA aperture response contains multiple sidelobes before the optical system cut-off is reached, nor does it address the fact that for realistic FPAs, there is a "dead zone" between the active regions of adjacent sensors resulting in the first zeros of the FPA aperture frequency response occurring at higher spatial frequency.

In conclusion, for a diffraction-limited optical system and FPA configuration as described above, without additional lowpass filtering aliasing will occur. In the imaging system modeled in this work this undersampling is evident in Figure (5.2)(c), which shows the combined sensor and optical system MTF, in which the MTF is nonzero well beyond the Nyquist rate.

5.3 Synthesized Scene Data

5.3.1 Continuous-Space Scene Model

Evaluating the performance of motion estimation algorithms necessitates the use of data for which the motion is known *a-priori*. This precludes the use of uncalibrated real-world data sets for which the motion information is not accurately known. In the absence of calibrated sequences, scenes synthesized using known motion are utilized.

Since the objective is to consider the “realistic best case” performance of block-matching subpixel-resolution motion estimation techniques, attention is restricted to local *translational* motion. This is reasonable since (i) estimating translational motion is typically the “easiest” motion to estimate and is most likely to yield the “best case” performance, (ii) while real-world scene motion is obviously not restricted to pure translations, many classes of motion are reasonably approximated by local translational motion, (iii) translational block-motion estimators are quite common and therefore relevant, (iv) the methodology developed here may be applied to other classes of motion.

Thus far, the background has been developed in terms of the two spatial dimensions present in imaging systems. For the purposes of experimentation, however, the problem is restricted to the 1-D case. This simplification is justified since the extension to the 2-D case is straightforward and does not add considerable insights not gleaned from the 1-D treatment. The points this work addresses are most easily demonstrated using a 1-D model and remain readily applicable to the 2-D case.

To achieve the “realistic best case” a scene is chosen which is both realistic and likely to yield the best possible motion estimates. A natural choice which satisfies both requirements is a step edge — commonly found in real-world scenes and a near-optimal function for motion estimation. In order to make the test scene more realistic, texture is also considered. It is assumed that the step edge $s(x)$ has texture features which are simulated using an additive white Gaussian noise process $g(x)$, and so has frequency content beyond that passed by the optical system.

For the motion estimation tests, two underlying continuous-space signals are required: the original signal $f(x, t_1)$ at time t_1, and the shifted signal $f(x, t_2)$ at time t_2. In terms of the step edge and the texture model, the continuous-space scene functions are described by,

$$
\begin{aligned}
f(x, t_1) &= s(x) + g(x) \\
f(x, t_2) &= f(x - X_m, t_1) \\
&= s(x - X_m) + g(x - X_m),
\end{aligned}
\tag{5.5}
$$

where the translation parameter X_m is to be estimated using the motion estimation techniques.

5.3.2 Discrete-Space Approximation

For computational purposes, the continuous-space functions $f(x, t_1)$ and $f(x, t_2)$ must be approximated on a discrete space. This is achieved by sampling. Since the highest motion-estimation resolution considered in this work is $\frac{1}{20}$-th pixel (motion estimates quantized to $\frac{1}{20}$-th pixel increments), the sampling period used for the discrete approximation of the continuous-space functions is kept significantly smaller than $\frac{1}{20}X_s$. For the experiments presented here, a sampling rate of approximately 200 samples per FPA pixel is utilized. This rate may at a first glance appear excessively large, however it is necessary in order to obtain a statistically meaningful spread of motion estimation errors. This may be understood with the realization that this sampling rate effectively quantizes the *a-priori* motion to $\frac{1}{200}$-th pixel resolution.

The discrete-space step is first created, and the texture added. This results in the discrete-space approximations to the continuous-space scene functions of Equation (5.5) as,

$$\begin{aligned} f[k,t_1] &= s(k\Delta) + g(k\Delta) \\ f[k,t_2] &= f[k-d,t_1] \\ &= f(k\Delta - d\Delta, t_1) \\ &= f(k\Delta - X'_m, t_1) \\ &= s(k\Delta - X'_m) + g(k\Delta - X'_m), \end{aligned} \tag{5.6}$$

where Δ is the sampling rate used in the discrete-space approximation, k and d are integers, and $d\Delta = X'_m$ is the Δ-quantized approximation to X_m. For notational convenience X'_m shall hereafter be referred to simply as X_m.

5.4 Simulated Imaging Process

The imaging system is simulated in three steps: (i) lowpass filtering to approximate the optical system MTF of Equation (5.1), followed by (ii) numerical integration over the length of each FPA pixel to model the effect of the spatial integration performed by the FPA sensors (the PSF of Equation (5.3)) and (iii) addition of Gaussian noise to model sensor noise. Using the discrete-space approximation of the synthetic scene as the input to the simulated imaging process results in output which approximates that of a typical FPA camera observing the continuous-space synthetic scene.

The simulated imaging process applied to the discrete-space scenes $f[k,t_1]$ and $f[k,t_2]$ results in the observed discrete-space image data $\hat{f}[m,t_1]$ and $\hat{f}[m,t_2]$, in which each sample represents the output of a single FPA pixel. Samples over the discrete variable m represent a physical distance equal to X_s in continuous-space. Due to the effects of the optical system MTF, the sensor integration, the added sensor noise, and sampling, $\hat{f}[m,t_1]$ and $\hat{f}[m,t_2]$ are in general not equal to the samples of the original continuous-space signals $f(mX_s,t_1)$ and $f(mX_s,t_2)$, respectively.

It is important to realize that due to the spatial integration performed by the FPA pixels, the signals $\hat{f}[m, t_1]$ and $\hat{f}[m, t_2]$ are in general *not* shifted versions of each other. Equality will hold only when the true motion X_m is an integer multiple of the sampling period X_s. This is a consequence of the integration over shifted regions of the texture. After the addition of sensor noise independently to each signal however, equality can no longer hold.

An example of a realization of the signals $\hat{f}[m, t_1]$ and $\hat{f}[m, t_2]$ with a shift of 42μm or 4.67 pixels, is shown in Figure (5.4)(d). The FPA signal to noise ratio is set at 50dB.

5.5 Motion Estimation

Given the simulated FPA data derived from a synthetic scene, it is possible to proceed to the motion estimation tests. First the motion estimators used are briefly reviewed before the test procedure is discussed.

5.5.1 Motion Estimators Considered

Subpixel-resolution motion estimation is often achieved using standard block-matching motion estimation techniques performed on *interpolated* image data. Since standard block motion estimation is accurate in theory to 1 pixel, interpolating the image data $\hat{f}[m, t]$ by a factor of p such that $\hat{f}_I[\frac{i}{p}, t] = \hat{f}[m, t]$ for $i = mp$ and performing motion estimation on the interpolated data $\hat{f}_I[i, t]$ yields $\frac{1}{p}$-pixel-resolution motion estimates. This does not, however, mean to say that the motion estimated from the interpolated data are *accurate* to $\frac{1}{p}$-pixel. The raison d'être of this work is to determine the bounds on the accuracy that can be achieved using interpolation-based block matching motion estimation.

Block matching motion estimation methods utilize a similarity measure to determine the best match between the reference and displaced image data. Three matching criteria [68] are considered:

1. Sum of Absolute Differences (SAD)

$$\mathrm{SAD}(d) = \sum_{i \in \mathcal{B}} \left| \hat{f}_I\left[i, t_1\right] - \hat{f}_I\left[i + d, t_2\right] \right| \tag{5.7}$$

2. Mean Square Error (MSE)

$$\mathrm{MSE}(d) = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \left\{ \hat{f}_I\left[i, t_1\right] - \hat{f}_I\left[i + d, t_2\right] \right\}^2 \tag{5.8}$$

3. Normalized cross-Correlation Function (NCF)

$$\mathrm{NCF}(d) = \frac{\displaystyle\sum_{i \in \mathcal{B}} \hat{f}_I\left[i, t_1\right] \hat{f}_I\left[i + d, t_2\right]}{\left[\displaystyle\sum_{i \in \mathcal{B}} \hat{f}_I^2\left[i, t_1\right]\right]^{\frac{1}{2}} \left[\displaystyle\sum_{i \in \mathcal{B}} \hat{f}_I^2\left[i + d, t_2\right]\right]^{\frac{1}{2}}} \tag{5.9}$$

In the above, the summations are over $\mathcal{B}$, which is a predefined block containing $|\mathcal{B}|$ elements. The motion estimate for each of these criteria is the value d which optimizes the matching criteria for $d \in \mathcal{W}$, where $\mathcal{W}$ is some predefined search window of allowable motions. Since $\hat{f}_I[i, t]$ is interpolated by a factor of p from the original FPA data, the actual motion returned by the subpixel-resolution estimators is $\frac{d}{p}$ pixels, or in continuous-space $\frac{d}{p} X_s$.

For interpolation of the test signals, both linear and cubic spline interpolation techniques are explored.

5.5.2 Motion Estimator Test Procedure

The goal of motion estimation is to estimate the motion X_m between the signal at times t_1 and t_2 to the highest degree of accuracy possible. A $\frac{1}{p}$-pixel-resolution motion estimator yields estimates which are quantized to integer multiples of $\frac{1}{p}$. If the $\frac{1}{p}$-pixel-resolution motion estimator is perfect then the motion estimates will be accurate to $X_m \pm \frac{1}{2p}$. In reality, however, it cannot be assumed that resolution and accuracy are likely to be equivalent.

The purpose of this research is to find the highest motion estimation accuracy one can expect to achieve under realistic circumstances. Since the resolution of the motion estimator used provides no assurance as to the accuracy of the motion estimates, the largest p is chosen (thus yielding the highest possible motion estimation resolution) for which the resulting motion estimates are accurate to within some prespecified statistical bound.

It is thus reasonable to define the following criterion for accuracy: a motion estimator is accurate to $\frac{1}{p}$ pixels if α-% of motion estimation errors fall within the theoretical resolution bounds $(-\frac{1}{2p}, \frac{1}{2p})$. α should be chosen according to the acceptable level of error in a given application.

In order to characterize the statistical properties of the motion estimation errors, 5000 trials for each $1 \leq p \leq 20$ were made. Each trial consists of the following steps:

1. Choose the *a-priori* motion X_m uniformly on $(-wX_s, +wX_s)$ where $\pm w$ denotes the motion estimator search window size in pixels.

2. Choose X_δ uniformly on $\left(-\dfrac{X_s}{2}, +\dfrac{X_s}{2}\right)$.

3. Generate the scene data described in Section (5.3) to yield,

$$\begin{aligned} f[k,t_1] &= s(k\Delta + X_\delta) + g(k\Delta + X_\delta) \\ f[k,t_2] &= s(k\Delta - X_m + X_\delta) + g(k\Delta - X_m + X_\delta) \end{aligned}$$

4. Simulate the imaging process of Section (5.4) to determine the observed discrete-space image data $\hat{f}[m,t_1]$ and $\hat{f}[m,t_2]$.

5. Determine the estimate $\hat{X}_m$ by performing block-matching subpixel-resolution

motion estimation using the SAD, MSE and NCF similarity measures on the interpolated data $\hat{f}_I[i, t_1]$ and $\hat{f}_I[i, t_2]$.

6. Determine the motion estimation error $\epsilon = \hat{X}_m - X_m$.

These trials generate 5000 independent values of ϵ for each p and for each similarity measure. This number is sufficiently large to determine reliable statistics. X_δ is introduced to remove any possibility of bias resulting from the spatial location of the step edge in the reference signal $\hat{f}[m, t_1]$. Choosing the translational motion X_m within the range of the motion estimator search window ensures that it is possible for the motion estimators to find the correct translation.

Several variables have a potential impact on motion estimator performance, with the reciprocal resolution p being the most obvious one. Indeed, determining the largest p for a given α is a primary goal of this research. In the next section it is shown that the range $1 \leq p \leq 20$ is sufficient to demonstrate a leveling-off behavior in the performance of subpixel-resolution motion estimators. The next variable of interest is the similarity measure. Utilizing three similarity measures, SAD, MSE, and NCF enables a comparison of the performance of motion estimators using these measures. Another closely related variable is the type of interpolation used for the subpixel-resolution motion estimation. Linear and cubic spline interpolation techniques are explored. The final variable of interest is the block size $|\mathcal{B}|$ used in the block-matching in Equation (5.7), Equation (5.8) and Equation (5.9). The nominal $p = 1$ block size was $|\mathcal{B}| = 7$ pixels (effectively $6p+1$ pixels after interpolation). The $p = 1$ block size was varied from 7 to 25 pixels in 2 pixel increments to determine its effect.

5.6 Results

First, the empirically-determined accuracy of the subpixel-resolution motion estimators of increasing resolution is compared with the theoretical resolution limits.

Recall that a perfect $\frac{1}{p}$-pixel-resolution motion estimator has errors bounded by $\pm\frac{1}{2p}$. Since the *a-priori* motion is taken as a realization of a uniformly distributed random variable, it is expected that for a perfect $\frac{1}{p}$-pixel motion estimator, the error will be distributed uniformly on $(-\frac{1}{2p}, \frac{1}{2p})$.

Results of tests using the three matching criteria SAD, MSE, and NCF are shown in Figure (5.3)(a), (b) and (c). Motion estimators of resolution from 1 to $\frac{1}{20}$-th pixel ($1 \leq p \leq 20$) are tested. The abscissas show the reciprocal resolution p of the motion estimators and the ordinates reflect the experimentally observed errors in the motion estimates. The boxes bound 80% of the motion estimation errors, with the tails delimiting the remaining errors. The center bar in the boxes denotes the mean error of the motion estimates. The smooth decaying curves in the figure are plots of the error bounds for a perfect $\frac{1}{p}$-pixel-resolution motion estimator, $\pm\frac{1}{2p}$. As described in Section (5.5.2), 5000 randomly chosen scene motions were used for each value of p. Linear interpolation and a block size of 7 pixels were used.

Figure (5.3) indicates that if an 80% success rate is used as the criterion for satisfactory subpixel-resolution motion estimation, then the SAD, MSE and NCF measure motion estimators with $\frac{1}{4}$ to $\frac{1}{5}$-pixel-resolution yield the highest resolution for the given error requirement. The MSE measure estimator performs slightly better than both SAD and NCF. Notice that beyond the highest resolution meeting the 80% success rate there is little improvement in motion estimation accuracy.

Figure (5.3)(d) shows plots of the Cumulative Distribution Function (CDF) of the estimation errors committed by the MSE estimator for $1 \leq p \leq 20$. Since the *a-priori* scene motion is uniformly distributed, it is expected that the errors

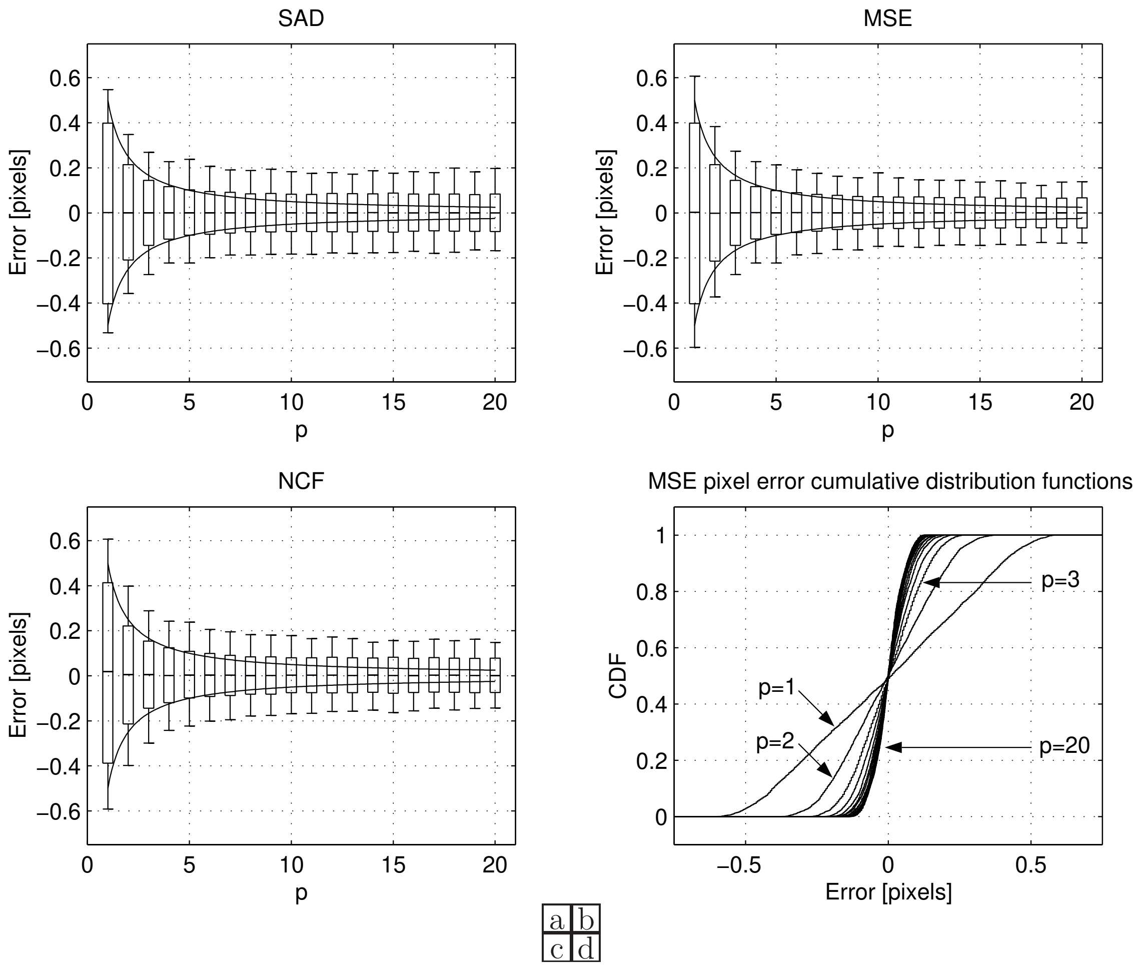


Figure 5.3. Motion estimation error (in pixels) as a function of increasing estimator resolution. (a) SAD (b) MSE (c) NCF (d) CDF of errors for MSE motion estimator as a function of interpolation factor.

for a perfect $\frac{1}{p}$-pixel motion estimator will be uniformly distributed in $(-\frac{1}{2p}, \frac{1}{2p})$; this would correspond to a CDF that is linear in $(-\frac{1}{2p}, \frac{1}{2p})$ with slope p. From Figure (5.3), it is evident that the CDF is approximately linear, with some rounding occurring at the edges due to errors outside of the $\pm\frac{1}{2p}$ bounds. For values of $p \leq 5$, the slope of the CDF is approximately p. However, for larger values of p, the CDF ceases to exhibit the near-optimal behavior, and approaches a slope (at the 0-error point) of roughly 7. This is indicative of the leveling-off of the errors evident in parts (a), (b), and (c) of Figure (5.3).

A natural issue to address is the performance of the motion estimators for all error percentages α within theoretical bounds. This is shown in Figure (5.4) (a), (b) and (c). These graphs are very useful since they provide bounds on the percentage of errors inside (or outside) the theoretical $\pm\frac{1}{2p}$ bounds for each of the $\frac{1}{p}$-pixel-resolution motion estimators discussed. These curves provide the highest resolution motion estimator for α-% of errors within the $\pm\frac{1}{2p}$ bounds, that is, the highest p for a given α.

The results presented utilize linear interpolation of the FPA data prior to motion estimation. It is interesting to note that there is almost no difference in the performance of the motion estimation techniques when using cubic spline interpolation. This does not imply that linear interpolation will in general deliver comparable performance to cubic spline or other more sophisticated interpolation techniques. The similarity in performance observed in these tests is probably a consequence of the choice of test signal. For other more general test functions, using cubic spline interpolation may provide better results.

Also considered is the effect of $\mathcal{B}$, the block over which the similarity measurement is performed. Varying the block size from 7 pixels through 25 pixels in 2 pixel increments yields little difference in motion estimator performance. This should

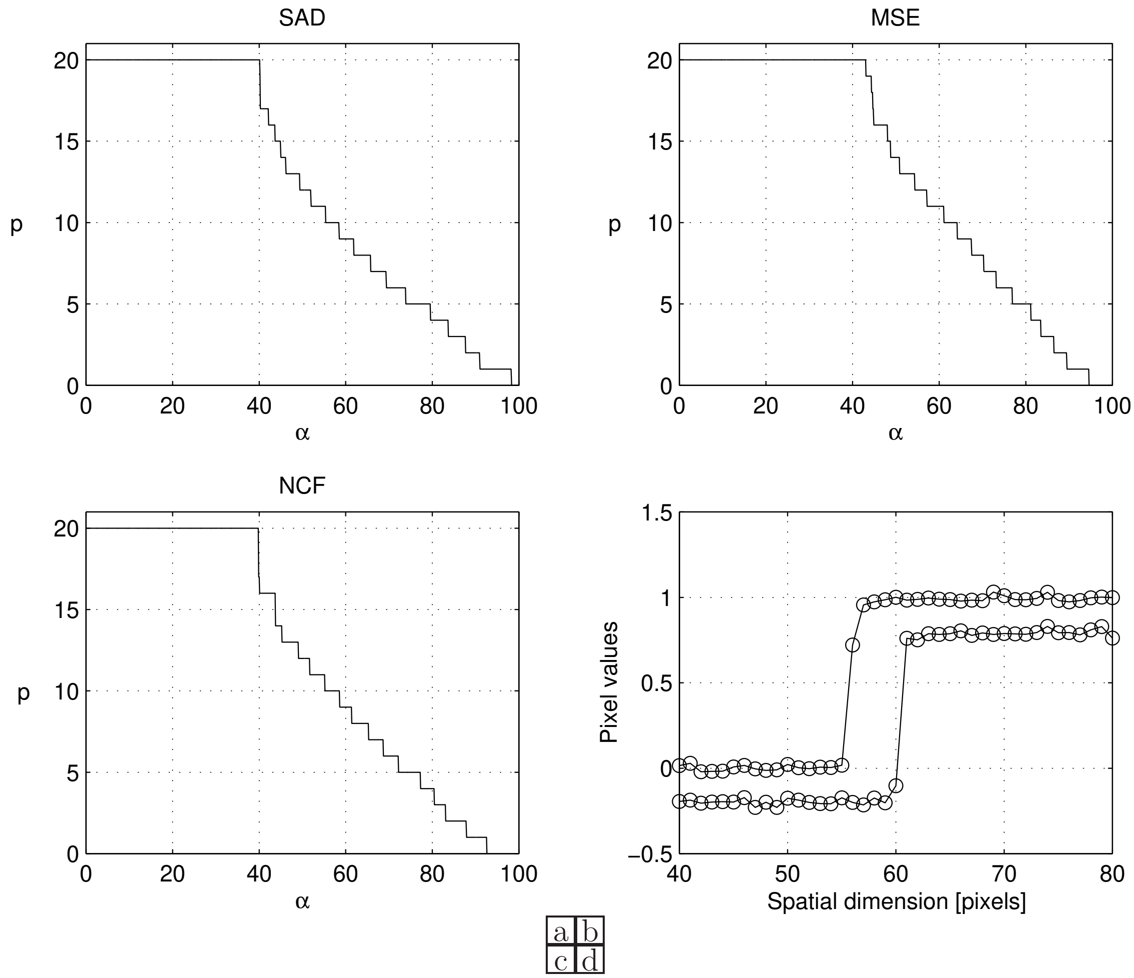


Figure 5.4. Highest p yielding α-% of motion estimates within theoretical limits. (a) SAD (b) MSE (c) NCF (d) Typical simulated FPA output for t_1 and t_2. One signal is displayed with a vertical offset for clarity.

come as no surprise given the step function scene model for which increasing block size will have little effect.

The previous two paragraphs illustrate that the choice of the step function scene model successfully isolates the effects of block size and interpolation method on the test procedure.

5.7 Conclusions

This chapter began by discussing a model of a typical imaging system, consisting of a diffraction-limited optical system with a focal plane sensor array. A 1-D synthetic scene was defined as a step function, with additive noise to represent texture. The test scene was subjected to translational motion. The imaging process was simulated, using the synthetic scenes as input imagery, to generate the discrete-space image data which closely resemble the output of a typical FPA camera system. Block-matching subpixel-resolution motion estimation using the SAD, MSE, and NCF similarity measures were compared. Both linear and cubic spline interpolation were used. The resolution of the motion estimators was varied from integer- to $\frac{1}{20}$-pixel.

The differences between the three motion estimation similarity measures were small, but noticeable. The motion estimator using the MSE measure performed best when requiring approximately 90% or less of the motion estimates to be within the theoretical bounds, whereas the motion estimator using the SAD measure was best for above 90%.

The primary result concerns the limits on the accuracy attainable by the subpixel-resolution motion estimators considered. It was shown that a perfect $\frac{1}{p}$-pixel-resolution motion estimator exhibits errors bounded within $\pm\frac{1}{2p}$ pixels, for $p \geq 1$. For the real-world motion estimators discussed in this work, the error approximately follows this theoretic bound for small p and levels off thereafter. That is, there is a value of the reciprocal resolution p beyond which no additional gains in accuracy are made. At this point, it becomes impractical to attempt further increases in p; increasing p then merely increases the computational complexity of the subpixel-resolution motion estimation, while the motion estimation accuracy no longer increases but rather remains approximately constant. The level of acceptable error in the motion estimates determines the highest resolution motion estimator which may be used. Graphs were presented to facilitate this selection.

CHAPTER 6

RAYTRACING AND THE CAMERA MATRIX

6.1 Introduction

This chapter explores algebraic relationships between the pinhole camera models utilized in two different imaging disciplines; *computer graphics*, which is concerned with the rendition of images from synthetic models, and *computer vision*, which addresses the problem of deriving models and measurements from observed images.

In the field of computer graphics, *raytracing* is a prominent method used to create visually realistic renditions of synthesized three dimensional scenes. Raytracing is a computational technique which simulates the physics of geometric optics and the interaction of light rays with surfaces in a effort to model the interaction of illumination with a 3-D scene model in an effort to produce visually highly realistic renditions of the scene. Through careful modeling of the physical processes governing the interaction between light rays, objects, light sources, atmospheric effects and so on, raytracing methods are capable of rendering images which achieve a high degree of realism.

Computer vision may be viewed as the inverse problem of computer graphics in that the data in the computer vision problem are real-world images and the objective is the computation of descriptions of the scene given the observed images. Classic examples of computer vision problems include computing shape from shading, texture or motion, recognizing objects in the scene, computing physical dimensions of objects from images, construction of models from images and so on. All these applications have in common the requirement to compute geometric properties of the scene from the observed image data.

This chapter explores of the relationship between the mathematical models of the pinhole camera which is central to both fields. The models used in the raytracing and computer vision problems will be developed from first principles and fundamental connections between the two approaches will be made explicit. While the development of the theory specific to each field is canonical, the relationship between the two approaches is not well described elsewhere. This work makes explicit these connections, thereby bridging the terminology and basic results for the pinhole camera models used in the two fields.

The results presented in this chapter derive from observations made when conducting the work presented in Chapter (8) in which both raytracing and computer vision principles were applied in connection with an investigation into linear observation models for multiframe superresolution restoration.

6.2 The Pinhole Camera

Figure (6.1) shows a pinhole camera which consists of an image forming pinhole which projects an image of objects in the scene onto the focal plane. The pinhole camera, while capable of producing beautiful images, suffers from two problems: severe blurring due to diffraction by the pinhole and low incident illumination at the focal place due to the small camera aperture [151]. Imaging optical systems are designed to mitigate these deficiencies while retaining (to a first order approximation) the basic geometrical optics properties of a pinhole camera. For this reason, the pinhole model is often used as an idealized representation of the properties of real optical systems. The convenient projective geometric properties of the pinhole

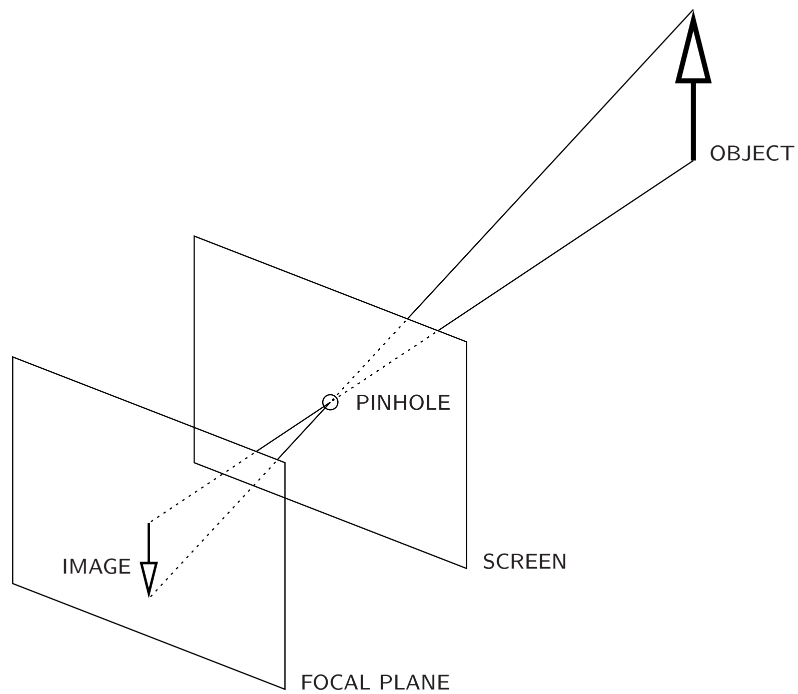


Figure 6.1. Image formation in a pinhole camera.

model are so attractive that even when real optical systems deviate from the ideal, the images produced by these systems are geometrically corrected so that the corrected images conform with the pinhole projection. This is especially important in computer vision applications where measurements are estimated from real-world imagery.

While real pinhole cameras take the form illustrated in Figure (6.1), for theoretical and computational purposes it is more convenient to abstract the pinhole camera model to a form in which the focal plane, which ordinarily lies behind the pinhole, is moved in front of the camera center (corresponding to the pinhole) so that the image forming optical rays pass through the virtual focal plane, as shown in Figure (6.2). Notice that in this model, the image formed in the focal is not subject to the inversion as in Figure (6.1).

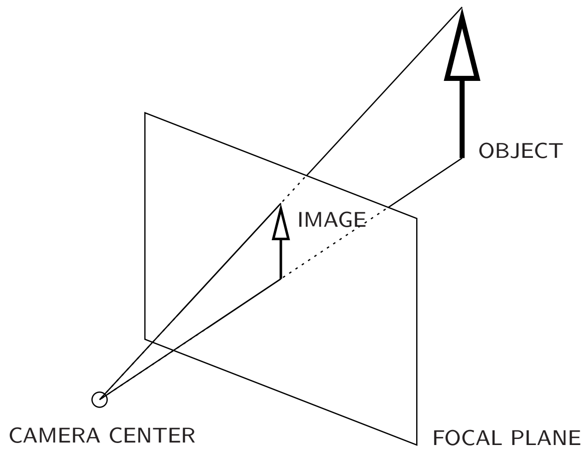


Figure 6.2. Pinhole camera model with virtual focal plane and camera center.

6.3 Computer Graphics and Raytracing

Given a geometric model of a (typically) 3-D scene, the field of computer graphics is concerned with the rendering of views of the scene as observed by a virtual camera. In raytracing, the approach for computing the camera image involves the simulation of the interaction of light rays with objects in the scene. In this chapter the model of the camera is of interest, rather than the mechanisms whereby optical rays interact with scene objects. The virtual camera, like its physical counterpart, is endowed with certain intrinsic characteristics (such as the focal length) as well as extrinsic characteristics which describe the camera position and orientation in the synthetic three dimensional scene space. The most commonly used camera model is the idealized pinhole camera discussed above.

6.3.1 A Camera Model for Raytracing

In this section the algebraic equations used to model the projection of optical rays in the scene are derived. By simulating the interaction of a large number of these light rays with objects in the three-dimension scene model, raytracing methods are capable of producing highly realistic renditions of scene from any camera position.

In practical raytracing implementations, somewhat counter-intuitively but for reasons of computational efficiency, the simulated ray trajectories originate not from the illumination source as would be expected. Instead, rays emanate from the camera center and are directed through the focal plane and proceed into the scene where they may encounter objects or illumination sources [56]. Rays are emitted in this fashion from the camera center though sample points (pixels) in the focal plane. The color and brightness of each pixel is determined by the interaction of the ray passing though that pixel with the objects that the ray may intersect in the 3-D scene. By simulating one or sometimes more rays for each pixel in the focal plane an image is constructed which corresponds to the view of an idealized pinhole camera. The simulated camera may be freely positioned and oriented with respect to the scene model by moving the camera center and by adjusting the position and orientation of the focal plane as illustrated in Figure (6.3). Vector geometry and linear algebra provide a convenient framework to describe the ray trajectories and the geometric transformations which operate on the camera and scene object coordinate systems.

In raytracing, it is convenient to describe the properties of the camera in a manner that simplifies the problem of projecting rays from the camera center into the 3-D scene. In particular, the following parameters are used to describe the geometric projection characteristics of the pinhole camera: (1) the camera center $\boldsymbol{c}$, (2) the principal, or optical axis $\hat{\boldsymbol{p}}$, (3) the horizontal axis of the focal plane $\hat{\boldsymbol{u}}$, (4) the vertical axis of the focal plane $\hat{\boldsymbol{v}}$, and (5) the focal length f. Vectors are typeset in boldface and unit vectors are denoted with a hat. In this description the simplifying assumption is made that $\hat{\boldsymbol{p}}$, $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ form an orthonormal set, which is the case for the most common cameras. The camera geometry and parameters

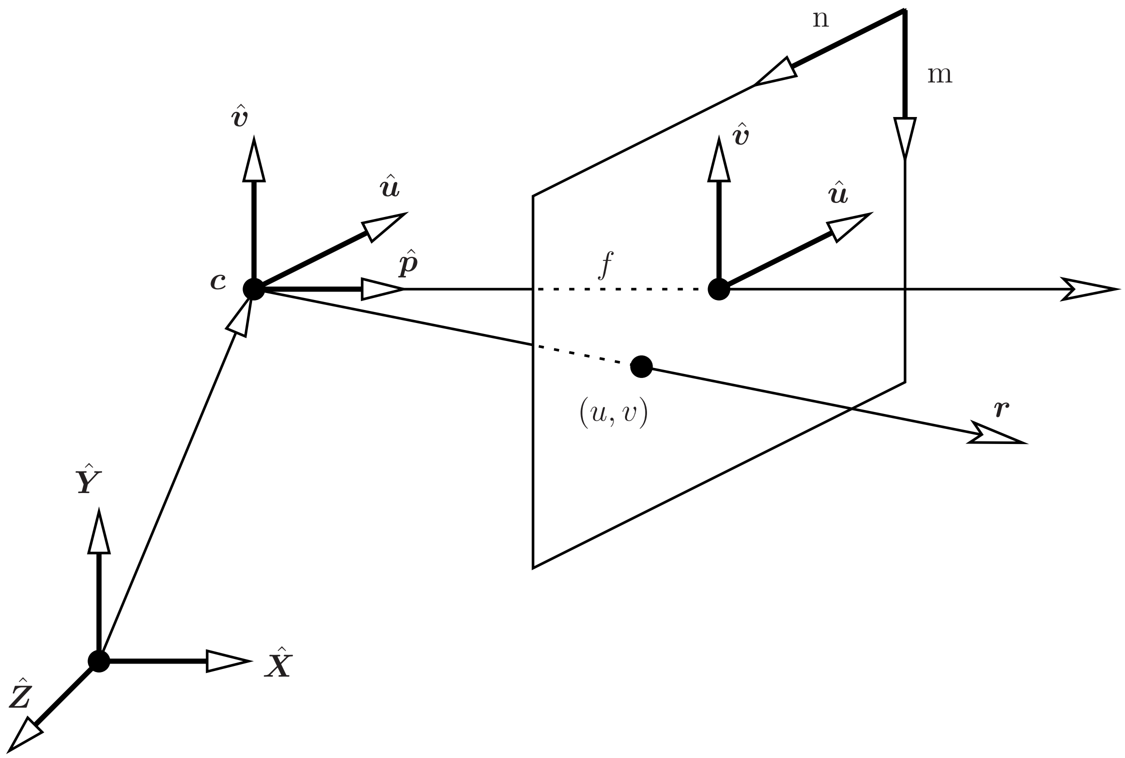


Figure 6.3. Pinhole camera geometry.

are illustrated in Figure (6.3) in which the camera is positioned in the 3-D world-coordinate system defined by the $\hat{\boldsymbol{X}}$, $\hat{\boldsymbol{Y}}$ and $\hat{\boldsymbol{Z}}$ axes. The principal axis $\hat{\boldsymbol{p}}$ intersects the focal plane at the *principal point* at the origin of the (u, v) focal plane coordinate system, a distance f from the camera center $\boldsymbol{c}$ along the optical axis.

As discussed above, raytracing involves casting rays originating at the camera center $\boldsymbol{c}$ toward locations in the focal plane and on through to the scene where the ray interacts with objects. Let $\boldsymbol{r}$ denote a ray originating at the camera center $\boldsymbol{c}$ and which passes through the location (u, v) in the focal plane coordinate system. Since the trajectory of the ray is linear, it may be described using its origin $\boldsymbol{c}$, a scalar parameter $\gamma \geq 0$ and a direction vector $\boldsymbol{d}$ in the linear equation,

$$\boldsymbol{r} = \boldsymbol{c} + \gamma \boldsymbol{d}. \tag{6.1}$$

Note that $\boldsymbol{d}$ in this expression is not, in general, a unit vector as will become clear from the discussion that follows. Referring to Figure (6.3), the direction of the ray given by the vector $\boldsymbol{d}$ may be described equivalently in terms of the camera focal plane axes $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ and the optical axis $\hat{\boldsymbol{p}}$ as,

$$\boldsymbol{r} = \boldsymbol{c} + \gamma\,[\, u\hat{\boldsymbol{u}} + v\hat{\boldsymbol{v}} + f\hat{\boldsymbol{p}}\,]\,. \tag{6.2}$$

In this equation u and v are the coordinates of the ray intersection with respect to the $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}})$ focal plane coordinate system. In Equation (6.2), with $u = v = 0$, the ray intersects the focal plane at the principal point for $\gamma = 1$. It is more convenient, however, to have the scalar parameter represent the distance along the principal axis from the camera center $\boldsymbol{c}$. This is easily achieved by introducing the parameter $\alpha \geq 0$ and rescaling so that,

$$\boldsymbol{r} = \boldsymbol{c} + \alpha \left[\, \frac{u}{f}\hat{\boldsymbol{u}} + \frac{v}{f}\hat{\boldsymbol{v}} + \hat{\boldsymbol{p}}\,\right]. \tag{6.3}$$

Now with $\alpha = f$ (the focal length) the ray intersects the focal plane as desired. Typically, rays are cast through a discrete set of locations on the focal plane in form of a regular lattice consisting of M rows by N columns of pixels. It is convenient therefore to specify the ray intersection with the focal plane in this pixel coordinate system. The ordering of the pixels is chosen so that the origin in pixel indices is located at the top left of the array when viewed from the camera center as shown in Figure (6.4). Note that the principal point, the location where the principal axis intersects the focal plane, need not be located at the center of the sampling array. The (u, v)-coordinates of the center of the pixel in row m and column n for $m \in \{0, 1, \ldots, M-1\}$ and $n \in \{0, 1, \ldots, N-1\}$ are then given by,

$$u = (n_0 - n)S_x \quad \text{and} \quad v = (m_0 - m)S_y, \tag{6.4}$$

where m_0 and n_0 are respectively the row and column locations of the principal point in pixel coordinates and S_x and S_y are the sample spacing in the horizontal

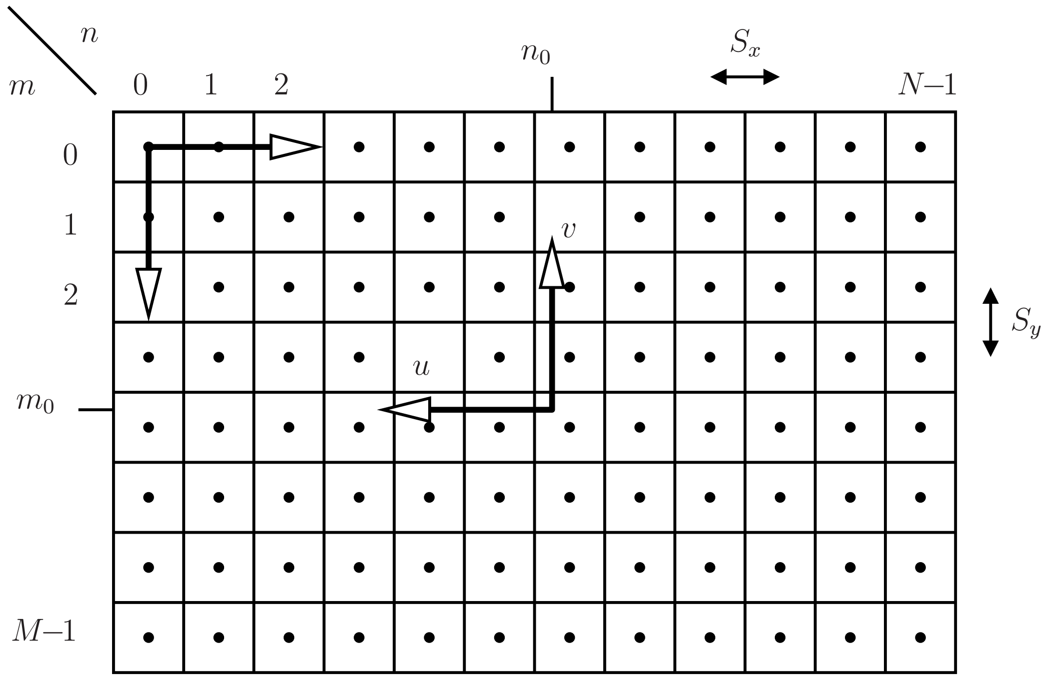


Figure 6.4. Focal plane array geometry.

and vertical directions respectively. Combining Equation (6.3) with Equation (6.4) yields,

$$\begin{aligned}\boldsymbol{r} &= \boldsymbol{c} + \alpha\left[\frac{n_0 - n}{f} S_x \hat{\boldsymbol{u}} + \frac{m_0 - m}{f} S_y \hat{\boldsymbol{v}} + \hat{\boldsymbol{p}}\right] \\ &= \boldsymbol{c} + \alpha\left[-n\frac{S_x}{f}\hat{\boldsymbol{u}} - m\frac{S_y}{f}\hat{\boldsymbol{v}} + \left\{n_0\frac{S_x}{f}\hat{\boldsymbol{u}} + m_0\frac{S_y}{f}\hat{\boldsymbol{v}} + \hat{\boldsymbol{p}}\right\}\right]. \qquad (6.5)\end{aligned}$$

Note that the terms surrounded by { } in Equation (6.5) are constant with respect to the pixel indices (n, m). The direction of the rays generated in a raster scan over the rows and columns of the FPA may thus be very efficiently computed using the increments $\frac{S_x}{f}\hat{\boldsymbol{u}}$ and $\frac{S_y}{f}\hat{\boldsymbol{v}}$.

6.3.2 Representing the Ray Equation in Matrix Form

It is instructive to restate the expression for the ray trajectory in Equation (6.3) in matrix-vector notation. For a given value of α, let the location of the ray $\boldsymbol{r}$

in world coordinates be given by $\boldsymbol{q} = [q_x\ q_y\ q_z]^\top$ and the camera center be $\boldsymbol{c} = [c_x\ c_y\ c_z]^\top$. It is then possible to write the ray trajectory in the form,

$$\begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix} = \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} + \alpha \begin{bmatrix} | & | & | \\ \hat{\boldsymbol{u}} & \hat{\boldsymbol{v}} & \hat{\boldsymbol{p}} \\ | & | & | \end{bmatrix} \begin{bmatrix} u/f \\ v/f \\ 1 \end{bmatrix}. \tag{6.6}$$

By using Equation (6.4) the ray trajectory can be expressed in terms of the focal plane pixel coordinate system as

$$\begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix} = \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} + \begin{bmatrix} | & | & | \\ \hat{\boldsymbol{u}} & \hat{\boldsymbol{v}} & \hat{\boldsymbol{p}} \\ | & | & | \end{bmatrix} \begin{bmatrix} -S_x/f & 0 & n_0 S_x/f \\ 0 & -S_y/f & m_0 S_y/f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha\, n \\ \alpha\, m \\ \alpha \end{bmatrix}. \tag{6.7}$$

Although this representation may at first appear somewhat peculiar, it will prove useful in the light of the material in the following section.

6.4 Computer Vision and the Camera Matrix

In raytracing, a ray cast through a given location on the focal plane may intersect with an object in the world coordinate system. Since the initial trajectory of a ray emanating from the camera center and propagating into the 3-D scene space ray is uniquely characterized by the 2-D location at which the ray intersects the focal plane, raytracing may be thought of as a mapping from a 2-D point in the focal plane to locations in the 3-D scene space.

The opposite problem, that of determining the image on the 2-D focal plane of a location in the 3-D world coordinate system, is addressed using a *camera matrix*. The camera matrix is a model for the geometric projection of a pinhole camera and is ubiquitous in computer vision literature. The mathematics describing the camera model using the camera matrix is greatly simplified by the use of homogeneous coordinates [74].

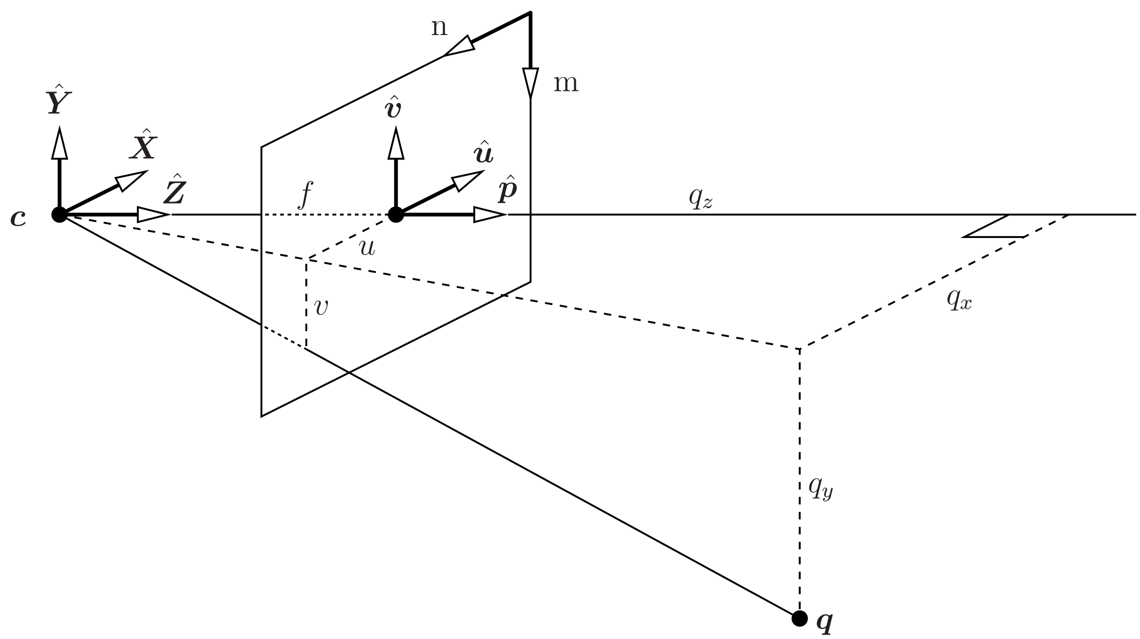


Figure 6.5. Camera in standard position.

6.4.1 The Camera Matrix

It is initially assumed that the camera center is located at the origin of the world coordinate system $\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}, \hat{\boldsymbol{Z}}$, with the camera coordinate system axes coincident with the world coordinate system axes. Thus the camera points in the positive $\hat{\boldsymbol{Z}}$ direction. It is also assumed that the camera focal plane is located a distance f from the camera center along the camera principal axis. This arrangement is illustrated in Figure (6.5).

Consider the point $\boldsymbol{q} = [q_x \; q_y \; q_z]^\top$ in the camera coordinate system. Examining similar triangles in Figure (6.5) yields the following expressions relating the focal plane coordinates corresponding to the point $\boldsymbol{q}$,

$$\frac{q_x}{q_z} = \frac{u}{f} \quad \text{and} \quad \frac{q_y}{q_z} = \frac{v}{f}. \tag{6.8}$$

Equivalently,

$$q_z u = f q_x \quad \text{and} \quad q_z v = f q_y. \tag{6.9}$$

These expressions may be written using homogeneous coordinates using a linear *projection matrix* as,

$$\begin{bmatrix} q_z u \\ q_z v \\ q_z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix}_{\mathsf{C}}, \tag{6.10}$$

where the subscript C draws attention to the fact that the coordinates of the point are relative to the camera coordinate system. This expression may be rewritten in terms of the row and column focal plane pixel coordinates using Equation (6.4) as

$$\begin{bmatrix} q_z S_x(n_0 - n) \\ q_z S_y(m_0 - m) \\ q_z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix}_{\mathsf{C}}, \tag{6.11}$$

which may in turn be rearranged into the more succinct form,

$$\begin{bmatrix} q_z n \\ q_z m \\ q_z \end{bmatrix} = \begin{bmatrix} -f/S_x & 0 & n_0 \\ 0 & -f/S_y & m_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix}_{\mathsf{C}}. \tag{6.12}$$

Next, the assumption that the camera and world-coordinate systems are coincident is relaxed so that the case where the camera is positioned and oriented arbitrarily with respect to the world coordinate system may be considered. In this case, the camera and world coordinate systems are related by at most a translation and a rotation. In particular, let the camera center be located at position $\boldsymbol{c}$ in the world coordinate frame. Let the orientation of the orthonormal camera coordinate axes be given by the unit vectors $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$ and $\hat{\boldsymbol{p}}$ in the world coordinate system, so that the camera axis vectors define the 3×3 rotation matrix $\boldsymbol{R} \triangleq [\ \hat{\boldsymbol{u}}\ \hat{\boldsymbol{v}}\ \hat{\boldsymbol{p}}\ ]$. Under these assumptions, a point $\boldsymbol{q}_{\mathsf{W}}$ in the world coordinate system is located at $\boldsymbol{q}_{\mathsf{C}}$ in the camera coordinate system where,

$$\boldsymbol{q}_{\mathsf{C}} = \boldsymbol{R}^{-1}(\boldsymbol{q}_{\mathsf{W}} - \boldsymbol{c}). \tag{6.13}$$

Note that since the rotation matrix $\boldsymbol{R}$ is unitary, $\boldsymbol{R}^{-1} = \boldsymbol{R}^{\top}$. Equation (6.13) can be represented using matrices and homogeneous coordinates as,

$$\begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix}_{\mathsf{C}} = \begin{bmatrix} - \hat{\boldsymbol{u}}^T - \\ - \hat{\boldsymbol{v}}^T - \\ - \hat{\boldsymbol{p}}^T - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \\ 1 \end{bmatrix}_{\mathsf{W}} . \tag{6.14}$$

Substituting Equation (6.14) into Equation (6.12) then yields an expression for the projection of points in the world coordinate system for an arbitrarily positioned and oriented camera as,

$$\begin{bmatrix} q_z n \\ q_z m \\ q_z \end{bmatrix} = \begin{bmatrix} -f/S_x & 0 & n_0 \\ 0 & -f/S_y & m_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} - \hat{\boldsymbol{u}}^T - \\ - \hat{\boldsymbol{v}}^T - \\ - \hat{\boldsymbol{p}}^T - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \\ 1 \end{bmatrix}_{\mathsf{W}} . \tag{6.15}$$

In Equation (6.15), the camera matrix is decomposed in a manner which reveals the position, orientation and projection characteristics of the camera. To ease notation in the sections ahead, the elements of the combined camera matrix in the decomposition of Equation (6.15) are denoted by,

$$\boldsymbol{P} = \boldsymbol{K}\boldsymbol{R}^{-1} \left[\, \mathbf{I} \mid -\boldsymbol{c} \,\right] \tag{6.16}$$

where $\boldsymbol{P}$ is the combined camera matrix, $\boldsymbol{K}$ represents the projection matrix of the camera in the standard position, $\boldsymbol{R}$ represents the orientation of the camera and $\boldsymbol{c}$ is the camera center.

For completeness, it is worth mentioning that some camera models include an additional, so called skew parameter in the matrix representing the focal plane coordinate system transformation. This parameter replaces the zero in the first row and second column of the coordinate transformation matrix and allows for focal plane arrays in which the sampling locations are on regular, but nonorthogonal lattice.

6.5 Relating the Raytracing and Camera Matrix Representations

Recall Equation (6.7) which describes the ray emanating from the camera center through the pixel in row m and column n in the focal plane. This equation may be rewritten as,

$$\begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix} - \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} = \alpha \begin{bmatrix} | & | & | \\ \hat{\boldsymbol{u}} & \hat{\boldsymbol{v}} & \hat{\boldsymbol{p}} \\ | & | & | \end{bmatrix} \begin{bmatrix} u/f \\ v/f \\ 1 \end{bmatrix}. \tag{6.17}$$

Using homogeneous coordinates the translation by $-\boldsymbol{c}$ may be achieved using matrix multiplication as,

$$\begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \\ 1 \end{bmatrix} = \begin{bmatrix} | & | & | \\ \hat{\boldsymbol{u}} & \hat{\boldsymbol{v}} & \hat{\boldsymbol{p}} \\ | & | & | \end{bmatrix} \begin{bmatrix} \alpha\, u/f \\ \alpha\, v/f \\ \alpha \end{bmatrix}. \tag{6.18}$$

Multiplying by the inverse of the rotation matrix $\boldsymbol{R}$ then yields,

$$\begin{bmatrix} - \hat{\boldsymbol{u}}^T - \\ - \hat{\boldsymbol{v}}^T - \\ - \hat{\boldsymbol{p}}^T - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha\, u/f \\ \alpha\, v/f \\ \alpha \end{bmatrix}. \tag{6.19}$$

Next, use the matrix transformation effecting the focal plane coordinate system transformation in Equation (6.4) to express Equation (6.19) in terms of the pixel coordinates as,

$$\begin{bmatrix} - \hat{\boldsymbol{u}}^T - \\ - \hat{\boldsymbol{v}}^T - \\ - \hat{\boldsymbol{p}}^T - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \\ 1 \end{bmatrix} = \begin{bmatrix} -S_x/f & 0 & n_0 S_x/f \\ 0 & -S_y/f & m_0 S_y/f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha\, n \\ \alpha\, m \\ \alpha \end{bmatrix}. \tag{6.20}$$

It is easy to show that the inverse of the matrix performing the transformation of the focal plane coordinate system is given by,

$$\begin{bmatrix} -S_x/f & 0 & n_0 S_x/f \\ 0 & -S_y/f & m_0 S_y/f \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -f/S_x & 0 & n_0 \\ 0 & -f/S_y & m_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6.21}$$

Premultiplying both sides of Equation (6.20) by this inverse yields the expression,

$$\begin{bmatrix} -f/S_x & 0 & n_0 \\ 0 & -f/S_y & m_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} - \hat{\boldsymbol{u}}^T - \\ - \hat{\boldsymbol{v}}^T - \\ - \hat{\boldsymbol{p}}^T - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha\, n \\ \alpha\, m \\ \alpha \end{bmatrix}. \tag{6.22}$$

Equation (6.22) is identical to Equation (6.15). Starting with the matrix-vector representation of the ray in Equation (6.3), the relationship between the raytracing camera model based on linear combinations of the camera axis vectors, as is typically used in the computer graphics community [56], and the standard projection matrix representation of the general pinhole camera in Equation (6.15) as is commonly found in the computer vision literature [74], has been shown. The two approaches are, not surprisingly, exact inverses of each other.

6.6 The Camera Matrix Decomposition and the Raytracing Equation

In computer graphics applications the camera matrix is often estimated from observed images or calibration targets in a process known as camera calibration. The camera calibration problem has received considerable attention and has progressed from requiring calibrated test targets to calibration from real-world scenery, known as self- or auto-calibration [202, 73, 76, 74]. No matter which approach is taken, the result of the calibration procedure is an estimate of the parameters of the 3×4 homogeneous camera matrix.

In Equation (6.15), however, the camera matrix is revealed as composed of three components which reveal the position, orientation and focal plane projection characteristics of the camera. Given a 3×4 camera matrix obtained from a calibration procedure, it is possible to obtain a description of the camera characteristics via a decomposition process.

The process of decomposing the camera matrix into its constituent parts is well known [74] and is achieved using a matrix decomposition which factors the camera matrix into the projection matrix $\boldsymbol{K}$ and the rotation matrix $\boldsymbol{R}^{-1}$ using the RQ matrix decomposition [62]. Recall from Equation (6.16) the camera matrix decomposition,

$$
\begin{aligned}
\boldsymbol{P} &= \boldsymbol{K}\boldsymbol{R}^{-1}\,[\,\mathbf{I}\,|-\boldsymbol{c}\,] \\
&= \boldsymbol{K}\,[\,\boldsymbol{R}^{-1}\,|-\boldsymbol{R}^{-1}\boldsymbol{c}\,] \\
&= [\,\mathbf{M}\,|-\mathbf{M}\boldsymbol{c}\,]
\end{aligned}
\tag{6.23}
$$

where $\mathbf{M} = \boldsymbol{K}\boldsymbol{R}^{-1}$ and $\boldsymbol{K}$ is upper-triangular and $\boldsymbol{R}^{-1}$ is orthogonal. The RQ decomposition is used to obtain $\boldsymbol{K}$ and $\boldsymbol{R}^{-1}$ from the left 3×3 submatrix $\mathbf{M}$ of the camera matrix $\boldsymbol{P}$ according to $\mathbf{M} = \boldsymbol{K}\boldsymbol{R}^{-1}$.

The camera center $\boldsymbol{c}$ may be easily determined after noticing that $\boldsymbol{P}\boldsymbol{c} = \mathbf{0}$ so that the camera center $\boldsymbol{c}$ may be found as the 4×1 (homogeneous) eigenvector describing the nullspace of $\boldsymbol{P}$ which is a rank-3 matrix. [55].

Once the decomposition of the matrix is obtained, the characteristics of the camera are completely known making it possible to move back and forth between the raytracing and camera matrix representations of the pinhole camera.

6.7 Summary

In this chapter two models of the pinhole camera, used in computer graphics raytracing and computer vision respectively, were developed from first principles. Although the two representations take quite different forms, it was shown that the models are in fact equivalent. The relationship between the two representations was explored through the use of the canonical decomposition of the camera matrix. It was shown how the parameters of the two models are related as well as how to switch between the two representations.

CHAPTER 7

SIMULTANEOUS MULTIFRAME SUPERRESOLUTION RESTORATION

7.1 Introduction

In this chapter a new simultaneous multiframe superresolution restoration technique utilizing spatiotemporal constraints and motion estimator confidence parameters is introduced. The ill-posed inverse problem of restoring superresolved imagery from the low-resolution, degraded observations is formulated as a statistical inference problem and a Bayesian Maximum A-posteriori Probability (MAP) approach is utilized for its solution. Motion estimator confidence parameters and temporal constraints are included in an attempt to obtain higher quality superresolution restoration with improved robustness to motion estimation errors.

As was discussed in Chapter (3), estimation-theoretic approaches using the Bayesian MAP framework for solving the superresolution problem have achieved high-quality restoration [168, 70] of single frame images from noisy, undersampled, low-resolution observed image sequences. This work utilizes the same estimation-theoretic framework as [168], however it departs from its predecessors in two main areas, both of which attempt to address shortcomings or overlooked aspects of earlier work:

1. Inclusion of temporal constraints on the restored sequence

   All existing MAP superresolution restoration approaches to *video* restoration utilize single frame restoration methods to a shifting subset of the observed frames similar to the method illustrated in Figure (2.1). This approach is attractive as it enables video restoration to be achieved using algorithms originally designed to restore single still images from image sequences. It is simple, intuitive and computationally attractive. A drawback, however, is that this approach provides no mechanism for imposing constraints among the restored frames, beyond those inherent in the single frame restoration algorithm.

To address this drawback, a new MAP superresolution video restoration algorithm is proposed which restores multiple superresolution images *simultaneously*, thereby enabling the inclusion of spatial *as well as temporal* constraints on the restored superresolution sequence.

2. Inclusion of motion estimator confidence parameters

   Also as discussed in Chapter (3), it is well known that erroneous motion estimates cause visually objectionable artifacts in the restorations computed using multiframe superresolution restoration algorithms. Multiframe restoration algorithms are highly sensitive to motion estimator errors since reliable subpixel motion estimates are required for successful registration and restoration of the observed sequence. Shah and Zakhor, in an attempt to address the problem of erroneous motion vectors, proposed a multiframe restoration procedure which maintains a set of candidate motion vectors and utilizes the "best" motion estimate [178, 179].

   In this work, a simple approach to reducing the effects of motion estimation errors is proposed. Motion compensated observations are weighted according to the level of confidence in the associated motion vectors. This is achieved by introducing a diagonal noise covariance matrix into the observation model.

7.2 Simultaneous Multiframe Restoration Problem Statement

Given a short video sequence consisting of P observed low-resolution frames with N_r rows by N_c columns of square pixels, the objective is to estimate temporally coincident superresolved images with dimensions qN_r rows by qN_c columns, $q \in \mathbb{N}$.

7.3 Observation Model

In this section, the linear observation model which relates the unknown superresolution image sequence to the observed low-resolution sequence is presented. The model is based on [168]. Superresolution and low-resolution images are represented as lexicographically ordered vectors $\{\boldsymbol{f}^{(k)}\}_1^P$ and $\{\boldsymbol{g}^{(l)}\}_1^P$ respectively, over the index set $\mathbb{P} = \{1, 2, \ldots, P\}$.

To enhance the clarity of the presentation, the complete imaging observation model is assumed to consist of two components: a model relating temporally coincident frames, and a model relating temporally noncoincident frames. These components, taken together, constitute the complete imaging model.

7.3.1 Temporally Coincident Observation Model

Consider the observation of the low-resolution frames $\boldsymbol{g}^{(l)}$ from the unknown, but temporally coincident superresolution frames $\boldsymbol{f}^{(l)}$. It is assumed that the frames are related by a *known* observation matrix $\boldsymbol{A}^{(l,l)}$ with additive noise accounting for measurement errors and uncertainties in the knowledge of the observation matrix,

$$\boldsymbol{g}^{(l)} = \boldsymbol{A}^{(l,l)}\boldsymbol{f}^{(l)} + \boldsymbol{n}^{(l,l)} \ , \quad l \in \mathbb{P}. \tag{7.1}$$

In Equation (7.1) the dimensions of $\boldsymbol{g}^{(l)}$ and $\boldsymbol{n}^{(l)}$ are $N_rN_c \times 1$, $\boldsymbol{f}^{(l)}$ is $q^2N_rN_c \times 1$ and $\boldsymbol{A}^{(l,l)}$ is $N_rN_c \times q^2N_rN_c$. The temporally coincident observation model in Equation (7.1) is illustrated (excluding the additive noise terms) in Figure (7.1). The structure of the linear observation equation is typical in image restoration problems

$$
\begin{array}{cccccc}
\boldsymbol{f}^{(1)} & \boldsymbol{f}^{(2)} & \cdots\cdots & \boldsymbol{f}^{(l)} & \cdots\cdots & \boldsymbol{f}^{(P)} \\
| & | & & | & & | \\
\boldsymbol{A}^{(1,1)} & \boldsymbol{A}^{(2,2)} & & \boldsymbol{A}^{(l,l)} & & \boldsymbol{A}^{(P,P)} \\
\downarrow & \downarrow & & \downarrow & & \downarrow \\
\boldsymbol{g}^{(1)} & \boldsymbol{g}^{(2)} & \cdots\cdots & \boldsymbol{g}^{(l)} & \cdots\cdots & \boldsymbol{g}^{(P)}
\end{array}
$$

Figure 7.1. Temporally coincident observation model.

but for the unequal dimensions of the unknown and observed vectors. Since the dimensions of the unknown vectors $\boldsymbol{f}^{(l)}$ are larger by a factor of q^2 than the observations $\boldsymbol{g}^{(l)}$, it is clear that Equation (7.1) is a highly underdetermined system of equations. Multiple solutions $\boldsymbol{f}^{(l)}$ satisfying Equation (7.1) exist, that is, $\boldsymbol{A}^{(l,l)}$ has a nontrivial nullspace (singular).

In Equation (7.1) the additive noise is modeled as independent, zero mean Gaussian noise with known per-pixel variances ${\sigma_j^{(l,l)}}^2$. The observed frames $\{\boldsymbol{g}^{(l)}\}_1^P$ are known with high confidence since they derive from direct observation of the scene via the known observation matrix $\boldsymbol{A}^{(l,l)}$. The variances ${\sigma_j^{(l,l)}}^2$ may therefore be chosen according to the imaging system Signal to Noise Ratio (SNR). Given the assumption of independence, the observation noise pdf may be expressed as,

$$
\mathcal{P}\Big(\boldsymbol{n}^{(l,l)}\Big) = \frac{1}{(2\pi)^{\frac{N_r N_c}{2}} |\boldsymbol{K}^{(l,l)}|} \exp\left\{-\frac{1}{2} {\boldsymbol{n}^{(l,l)}}^{\top} {\boldsymbol{K}^{(l,l)}}^{-1} \boldsymbol{n}^{(l,l)}\right\}, \tag{7.2}
$$

where $\boldsymbol{K}^{(l,l)} = \operatorname{diag}\left({\sigma_1^{(l,l)}}^2, {\sigma_2^{(l,l)}}^2, \cdots, {\sigma_{N_r N_c}^{(l,l)}}^2\right)$. Furthermore, it is assumed that the observation noise is independent over the observed frames $\{\boldsymbol{g}^{(l)}\}_1^P$.

7.3.2 Temporally Noncoincident Observation Model

To obtain additional constraints for estimating the superresolution sequence, unknown superresolution frames $\{\boldsymbol{f}^{(k)}\}_1^P$ are related to temporally noncoincident observed low-resolution frames. For each $\boldsymbol{f}^{(k)}$ assume it is possible to find a *motion* *compensating* observation matrix $\boldsymbol{A}^{(l,k)}$ and noise vector $\boldsymbol{n}^{(l,k)}$ such that,

$$\boldsymbol{g}^{(l)} = \boldsymbol{A}^{(l,k)}\boldsymbol{f}^{(k)} + \boldsymbol{n}^{(l,k)} \ , \quad l \in \mathbb{P} \setminus \{k\}. \tag{7.3}$$

The dimensions of the matrix and vectors remain the same as in Equation (7.1), however fundamental differences exist between Equation (7.1) and Equation (7.3) concerning the functions of the matrix $\boldsymbol{A}^{(l,k)}$ and the additive noise $\boldsymbol{n}^{(l,k)}$. In Equation (7.1), $\boldsymbol{A}^{(l,l)}$ models only the effects of spatial degradations inherent in the imaging system, whereas in Equation (7.3), $\boldsymbol{A}^{(l,k)}$ accounts for imaging degradations *and* compensates for motion occurring between frames k and l. In Equation (7.1), the additive noise term $\boldsymbol{n}^{(l,l)}$ accounts for observation noise and sensor PSF uncertainties. In the temporally noncoincident observation model, $\boldsymbol{n}^{(l,k)}$ accounts for these effects, as well as for pixels which are not observable in $\boldsymbol{g}^{(l)}$ from motion compensation of $\boldsymbol{f}^{(k)}$. Similar to Equation (7.1), the noise term $\boldsymbol{n}^{(l,k)}$ in Equation (7.3) is described by an independent, zero mean Gaussian random vector with per-pixel variances ${\sigma_j^{(l,k)}}^2$, however two specific cases are identified:

1. Pixels in $\boldsymbol{g}^{(l)}$ which are *unobservable* from motion compensation of $\boldsymbol{f}^{(k)}$.

   These pixels contribute no information toward determining the unknown superresolution image $\boldsymbol{f}^{(k)}$. The corresponding row of $\boldsymbol{A}^{(l,k)}$ contains all zeros, while the corresponding entry in $\boldsymbol{n}^{(l,k)}$ contains the unobservable pixel value. Since this value cannot be determined, the noise variance for the unobservable pixel is assigned the value $+\infty$. This ensures that no weight is given to the pixels in $\boldsymbol{g}^{(l)}$ which are not observable from $\boldsymbol{f}^{(k)}$ when estimating $\boldsymbol{f}^{(k)}$.

2. Pixels in $\boldsymbol{g}^{(l)}$ which are *observable* from motion compensation of $\boldsymbol{f}^{(k)}$.

   These pixels contribute valuable constraints for determining $\boldsymbol{f}^{(k)}$. It is well recognized [168] that reliable subpixel motion estimation is essential for high quality superresolution image restoration. A measure of the reliability of the motion estimates is therefore incorporated into the restoration by appropriate choice of the observed pixel noise variances ${\sigma_j^{(l,k)}}^2$. In this work the noise variances of the motion compensated observations are proportional to the Mean Absolute Difference (MAD) associated with the motion vector obtained from the hierarchical block matching motion estimation procedure [168]. Blocks for which the MAD is large typically represent less reliable motion estimates and therefore corresponding to large observation noise variances. Similarly, small MAD usually corresponds to more reliable motion estimates and small observation noise variance.

As in Equation (7.1) it is assumed that the observation noise is independent across pixels in the observed frames $\{\boldsymbol{g}^{(l)}\}_1^P$. The choice of an independent Gaussian random vector model for errors committed in motion estimation and for the values of the unobservable pixels, though arguably oversimplistic, nevertheless ensures computational tractability of the optimization problem discussed in Section (7.4). Under these assumptions, the noise pdf for $k, l \in \mathbb{P}, l \neq k$ is given by

$$\mathcal{P}\big(\boldsymbol{n}^{(l,k)}\big) = \frac{1}{(2\pi)^{\frac{N_r N_c}{2}} |\boldsymbol{K}^{(l,k)}|} \exp\left\{ -\frac{1}{2} {\boldsymbol{n}^{(l,k)}}^\top {\boldsymbol{K}^{(l,k)}}^{-1} \boldsymbol{n}^{(l,k)} \right\}. \qquad (7.4)$$

The motion compensating observation Equation (7.3) is sufficiently general to represent a wide variety of scene motion and imaging degradations through the appropriate choice of the entries of $\boldsymbol{A}^{(l,k)}$ which may be found by linearization of the continuous optical flow field and discretization of the continuous image intensities. The practical difficulty lies in the fact that $\boldsymbol{A}^{(l,k)}$ must be estimated from the observed (degraded) low-resolution sequence $\{\boldsymbol{g}^{(l)}\}_1^P$.

$$
\begin{tikzcd}[column sep=small]
 & & & \boldsymbol{f}^{(k)} \arrow[dll, dash] \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & & \\
 & \boldsymbol{A}^{(1,k)} \arrow[dl] & \boldsymbol{A}^{(2,k)} \arrow[dl] & \boldsymbol{A}^{(k,k)} \arrow[d] & \boldsymbol{A}^{(P,k)} \arrow[dr] & \\
\boldsymbol{g}^{(1)} & \boldsymbol{g}^{(2)} & \cdots\cdots & \boldsymbol{g}^{(k)} & \cdots\cdots & \boldsymbol{g}^{(P)}
\end{tikzcd}
$$

Figure 7.2. Combined observation model for $\boldsymbol{f}^{(k)}$.

7.3.3 Combined Observation Model

Figure (7.2) illustrates the combined temporally coincident and noncoincident observation model relating the low-resolution sequence $\{\boldsymbol{g}^{(l)}\}_1^P$ and a single super-resolution image $\boldsymbol{f}^{(k)}$.

The P equations relating $\{\boldsymbol{g}^{(l)}\}_1^P$ to $\boldsymbol{f}^{(k)}$ may be expressed in vector notation as

$$
\begin{bmatrix} \boldsymbol{g}^{(1)} \\ \boldsymbol{g}^{(2)} \\ \vdots \\ \boldsymbol{g}^{(P)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}^{(1,k)} \\ \boldsymbol{A}^{(2,k)} \\ \vdots \\ \boldsymbol{A}^{(P,k)} \end{bmatrix} \boldsymbol{f}^{(k)} + \begin{bmatrix} \boldsymbol{n}^{(1,k)} \\ \boldsymbol{n}^{(2,k)} \\ \vdots \\ \boldsymbol{n}^{(P,k)} \end{bmatrix}. \tag{7.5}
$$

These equations may be written compactly as

$$
\boldsymbol{g} = \boldsymbol{A}^{(*,k)} \boldsymbol{f}^{(k)} + \boldsymbol{n}^{(*,k)}. \tag{7.6}
$$

Combining the temporally coincident and noncoincident observation equations Equation (7.1) and Equation (7.3) for all $\{\boldsymbol{f}^{(k)}\}_1^P$ yields P^2 equations relating the super-resolution sequence $\{\boldsymbol{f}^{(k)}\}_1^P$ to the observed low-resolution images $\{\boldsymbol{g}^{(l)}\}_1^P$. Following the notation in Equation (7.6) this relationship may be written as,

$$
\begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{g} \\ \vdots \\ \boldsymbol{g} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}^{(*,1)} & & & \\ & \boldsymbol{A}^{(*,2)} & & \\ & & \ddots & \\ & & & \boldsymbol{A}^{(*,P)} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}^{(1)} \\ \boldsymbol{f}^{(2)} \\ \vdots \\ \boldsymbol{f}^{(P)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{n}^{(*,1)} \\ \boldsymbol{n}^{(*,2)} \\ \vdots \\ \boldsymbol{n}^{(*,P)} \end{bmatrix}. \tag{7.7}
$$

This may be expressed in the classic linear form,

$$\boldsymbol{G} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{N}. \tag{7.8}$$

Under the noise model assumptions described in Section (7.3.1) and Section (7.3.2) the pdf of the noise $\boldsymbol{N}$ is given by,

$$\mathcal{P}_{\boldsymbol{N}}(\boldsymbol{N}) = \frac{1}{(2\pi)^{\frac{P^2 N_r N_c}{2}} |\boldsymbol{K}|} \exp\left\{-\frac{1}{2}\boldsymbol{N}^\top \boldsymbol{K}^{-1}\boldsymbol{N}\right\}, \tag{7.9}$$

where $\boldsymbol{K} = \operatorname{diag}(\boldsymbol{K}^{(1,1)}, \boldsymbol{K}^{(2,1)}, \ldots, \boldsymbol{K}^{(P,1)}, \boldsymbol{K}^{(1,2)}, \ldots, \boldsymbol{K}^{(P,P)})$ and is itself diagonal.

7.4 Bayesian Estimation of Superresolution Images

Bayesian MAP estimation is used to estimate the superresolved images $\{\boldsymbol{f}^{(k)}\}_1^P$ given the observed data $\{\boldsymbol{g}^{(l)}\}_1^P$. The MAP estimate $\hat{\boldsymbol{f}}_{\text{MAP}}$ maximizes the *a-posteriori* probability $\mathcal{P}(\boldsymbol{f}|\boldsymbol{G})$ given by:

$$\begin{aligned} \hat{\boldsymbol{f}}_{\text{MAP}} &= \arg\max_{\boldsymbol{f}} \{\mathcal{P}(\boldsymbol{f}|\boldsymbol{G})\} \\ &= \arg\max_{\boldsymbol{f}} \left\{ \frac{\mathcal{P}(\boldsymbol{G}|\boldsymbol{f})\,\mathcal{P}(\boldsymbol{f})}{\mathcal{P}(\boldsymbol{G})} \right\}. \end{aligned} \tag{7.10}$$

The second expression in Equation (7.10) results from the application of Bayes' theorem. Taking the logarithm and noting that the maximum independent of $\mathcal{P}(\boldsymbol{G})$, yields

$$\hat{\boldsymbol{f}}_{\text{MAP}} = \arg\max_{\boldsymbol{f}} \{\log \mathcal{P}(\boldsymbol{G}|\boldsymbol{f}) + \log \mathcal{P}(\boldsymbol{f})\}. \tag{7.11}$$

It remains to determine the form of the likelihood function $\mathcal{P}(\boldsymbol{G}|\boldsymbol{f})$ and the prior $\mathcal{P}(\boldsymbol{f})$.

7.4.1 The Likelihood Function

Given the expression for the observation process in Equation (7.8), the conditional probability $\mathcal{P}(\boldsymbol{G}|\boldsymbol{f})$ is determined by the noise pdf in Equation (7.9) as,

$$\begin{aligned}\mathcal{P}(\boldsymbol{G}|\boldsymbol{f}) &= \mathcal{P}_{\boldsymbol{N}}(\boldsymbol{G}-\boldsymbol{A}\boldsymbol{f}) \\ &= \frac{1}{(2\pi)^{\frac{q^2 N_r N_c}{2}}|\boldsymbol{K}|}\exp\left\{-\frac{1}{2}(\boldsymbol{G}-\boldsymbol{A}\boldsymbol{f})^{\top}\boldsymbol{K}^{-1}(\boldsymbol{G}-\boldsymbol{A}\boldsymbol{f})\right\}. \end{aligned} \tag{7.12}$$

7.4.2 The Spatiotemporal Prior

One of the motivations for the approach taken in this work is the untapped potential for the inclusion of more powerful *a-priori* constraints via simultaneous estimation of the superresolution frames. In single frame superresolution restoration, the prior $\mathcal{P}(\boldsymbol{f})$ typically includes only *spatial* smoothness constraints. Effecting *simultaneous* multiframe restoration makes possible the inclusion of *spatiotemporal* constraints in the prior. This is achieved using a novel motion-compensated MRF model in which the Gibbs distribution energy is dependent on the pixel intensity variation along the motion trajectory as well as the on the usual local spatial interactions. The prior is chosen to be convex to ensure a computationally tractable optimization. The pdf $\mathcal{P}(\boldsymbol{f})$ of the MRF is, according to the Hammersley-Clifford theorem [11], given by a Gibbs distribution. For the development that follows, the term in the exponent of the Gibbs distribution is expanded as,

$$\mathcal{P}(\boldsymbol{f}) = \frac{1}{k_p}\exp\left\{-\frac{1}{\beta}\sum_{c\in\mathcal{C}}\rho_\alpha\left(\boldsymbol{\partial}_c\boldsymbol{f}\right)\right\}. \tag{7.13}$$

In Equation (7.13), k_p is a normalizing constant known as the partition function and β is the MRF "temperature" parameter (inspired by the Gibbs distribution in thermodynamics). The summation is over the set of all "cliques" $\mathcal{C}$ with $\boldsymbol{\partial}_c$ computing local spatiotemporal activity. The nonlinear spatial activity penalizing function $\rho_\alpha(x)$ is the Huber function [188],

$$\rho_\alpha(x) = \begin{cases} x^2 & |x| \leq \alpha \\ 2\alpha|x| - \alpha^2 & |x| > \alpha \end{cases} . \tag{7.14}$$

The clique structure determines the spatial and temporal interactions. Five clique types are divided into two classes:

1. Spatial activity is computed using finite difference approximations to the second directional derivatives (vertical, horizontal and two diagonal directions) in each superresolution image $\{\boldsymbol{f}^{(k)}\}_1^P$. In particular,

$$\begin{aligned}
\boldsymbol{\partial}_1 \boldsymbol{f}^{(k)} &= f^{(k)}[n_1-1, n_2] - 2f^{(k)}[n_1, n_2] + f^{(k)}[n_1+1, n_2] \\
\boldsymbol{\partial}_2 \boldsymbol{f}^{(k)} &= f^{(k)}[n_1, n_2-1] - 2f^{(k)}[n_1, n_2] + f^{(k)}[n_1, n_2+1] \\
\boldsymbol{\partial}_3 \boldsymbol{f}^{(k)} &= \tfrac{1}{2} f^{(k)}[n_1+1, n_2-1] - f^{(k)}[n_1, n_2] + \tfrac{1}{2} f^{(k)}[n_1-1, n_2+1] \\
\boldsymbol{\partial}_4 \boldsymbol{f}^{(k)} &= \tfrac{1}{2} f^{(k)}[n_1-1, n_2-1] - f^{(k)}[n_1, n_2] + \tfrac{1}{2} f^{(k)}[n_1+1, n_2+1] .
\end{aligned} \tag{7.15}$$

   For pixels on the image boundaries, where it is not possible to compute the above second derivatives, appropriate first derivatives are substituted.

2. Temporal smoothness constraints are imposed on the reconstructed superresolution sequence using a finite difference approximation to the second temporal derivative along the motion trajectory as,

$$\boldsymbol{\partial}_5 \boldsymbol{f}^{(k)} = f^{(k-1)}[n_1+\delta_1, n_2+\delta_2] - 2f^{(k)}[n_1, n_2] + f^{(k+1)}[n_1+\Delta_1, n_2+\Delta_2] . \tag{7.16}$$

   The tuples (δ_1, δ_2) and (Δ_1, Δ_2) represent the motion vectors into the previous and subsequent frames respectively. In the case of pixel coverings and uncoverings, where it is not possible to find the forward or backward motion vectors, the appropriate first order difference is utilized. In the case where neither forward or backward motion information is available, the temporal activity measure is not included in the computation of the MRF energy. These choices ensure that pixel values in the temporally adjacent frames are not affected by the covered or uncovered pixel thus assigning no penalty to temporal discontinuities where they are known to exist. Where correspondences are known to exist, the temporal derivatives may be weighted proportional to the reliability of the motion estimates thereby imposing a robust temporal smoothness constraint. At the sequence temporal boundaries, appropriate first temporal differences are utilized.

In Figure (7.3) the typical structure of the local interactions for the spatiotemporal MRF is illustrated. Shown are three frames in the restoration space. In each frame the local spatial interactions ($\boldsymbol{\partial}_1, \ldots, \boldsymbol{\partial}_4$) for a single pixel are shown (interaction between a pixel and its 8-neighbors). Temporal dependence is shown along a motion trajectory joining the pixels in each frame. The temporal clique interaction $\boldsymbol{\partial}_5$ for a given pixel in the center frame applies, in general, in both directions along this motion trajectory.

7.4.3 The Objective Function

Substituting the likelihood term in Equation (7.12) and the prior term in Equation (7.13) into Equation (7.11) and removing constants independent of $\boldsymbol{f}$ gives the objective function,

$$\begin{aligned} \hat{\boldsymbol{f}}_{\text{MAP}} &= \arg\max_{\boldsymbol{f}} \left\{ -\frac{1}{2}(\boldsymbol{G}-\boldsymbol{A}\boldsymbol{f})^\top \boldsymbol{K}^{-1}(\boldsymbol{G}-\boldsymbol{A}\boldsymbol{f}) - \frac{1}{\beta}\sum_{c\in\mathcal{C}} \rho_\alpha(\boldsymbol{\partial}_c \boldsymbol{f}) \right\} \\ &= \arg\min_{\boldsymbol{f}} \left\{ \frac{1}{2}(\boldsymbol{G}-\boldsymbol{A}\boldsymbol{f})^\top \boldsymbol{K}^{-1}(\boldsymbol{G}-\boldsymbol{A}\boldsymbol{f}) + \frac{1}{\beta}\sum_{c\in\mathcal{C}} \rho_\alpha(\boldsymbol{\partial}_c \boldsymbol{f}) \right\}. \qquad (7.17) \end{aligned}$$

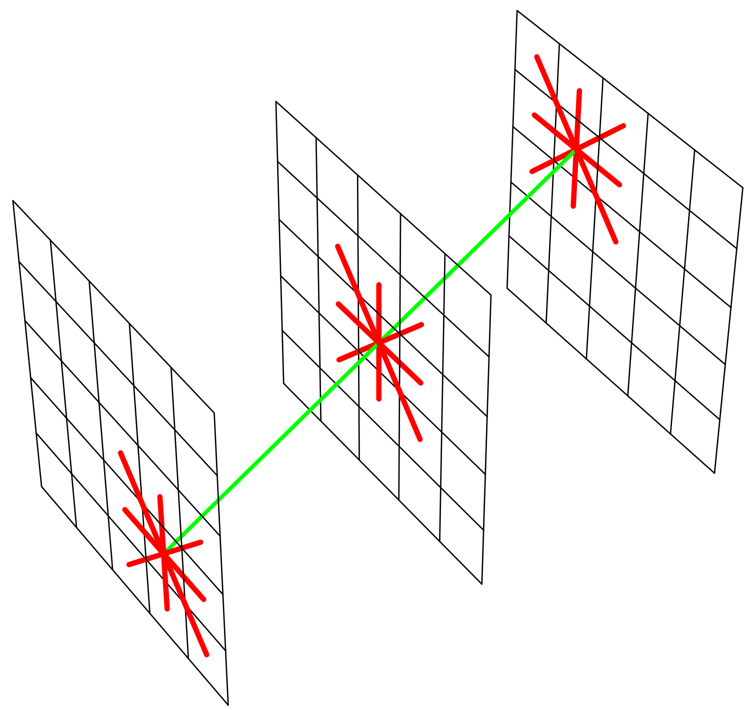


Figure 7.3. Local spatiotemporal structure of the motion-compensated Markov random field prior model.

7.4.4 Optimization

Determining the MAP estimate $\hat{\boldsymbol{f}}_{\text{MAP}}$ requires the minimization of the objective function in Equation (7.17). By the choice of a strictly convex prior in Section (7.4.2), existence and uniqueness of a global minimum, corresponding to MAP estimate $\hat{\boldsymbol{f}}_{\text{MAP}}$ in Equation (7.17) is assured. As a result, efficient gradient descent optimization methods [9] may be employed to yield the MAP estimate $\hat{\boldsymbol{f}}_{\text{MAP}}$. Despite the large number of unknowns, convexity ensures that the optimization remains tractable.

7.5 Examples

Tests were performed using the algorithm proposed in this chapter on a short sequence of a landmark at the University of Notre Dame. Ten low-resolution monochrome video frames of dimension $N_r = 120$ rows by $N_c = 160$ columns were used as input data. These images are shown in Figure (7.5).

The motion occurring between all pairs of frames of the observed low-resolution image sequence was estimated using the hierarchical subpixel-resolution motion estimator described in [168] yielding 1/4-pixel-resolution local translational motion estimates for each observed low-resolution pixel. Restoration was performed with $q = 4$ resulting in reconstructed images of dimension 480 rows by 640 columns. The reconstructed frame in temporal correspondence with the lower-left image in Figure (7.5) is shown in Figure (7.5). The zero-order hold image corresponding to this low-resolution frame is shown in Figure (7.5) for comparison with the restoration in Figure (7.5). In Figure (7.7), a region of interest has been selected from the restored image shown in Figure (7.5) and is contrasted with the cubic spline interpolation of the same region in the temporally-coincident low-resolution observed frame. The resolution improvement is apparent.

Evaluation of the restored video sequences computed using the proposed method demonstrate subtle but noticeable improvements in restoration quality. These improvements result from the inclusion of the motion estimation confidence parameter via the covariance matrix $\boldsymbol{K}$ and the inclusion of the temporal dependence in the motion-compensated MRF prior. Unfortunately it is difficult to illustrate these improvements without viewing the actual video.

Figure 7.4. A selection of observed frames from the Notre Dame dome sequence.

7.6 Conclusions and Extensions

In this chapter a new simultaneous multiframe superresolution video restoration procedure was presented. The approach utilizes Bayesian maximum *a-posteriori* estimation with a prior model which imposes spatiotemporal smoothness. Motion estimator confidence parameters are also incorporated into the proposed framework. The technique exhibits improved robustness to motion estimation errors through the inclusion of reliability measures for motion estimates as well as by imposing temporal constraints on the reconstructed image sequence.

The observation noise covariance matrix $\boldsymbol{K}$ was chosen to be diagonal. With minimal additional complexity it is possible to generalize $\boldsymbol{K}$ to an arbitrary positive definite covariance matrix, thereby allowing more realistic motion estimation error modeling.

Figure 7.5. Zero order hold of an observed low-resolution image.

Figure 7.6. Superresolution restoration.

Figure 7.7. Region of interest extracted from the cubic spine interpolation of an observed low-resolution image (left), and the corresponding region in the superresolution restoration (right).

CHAPTER 8

LINEAR MODELS FOR MULTIFRAME SUPERRESOLUTION RESTORATION UNDER NONAFFINE REGISTRATION AND SPATIALLY VARYING PSF

8.1 Introduction

It is hardly surprising that multiframe superresolution restoration techniques, which offer the prospect of enhanced resolution seemingly for "free," have been the focus of considerable attention in recent years. In a wide variety of imaging modalities, a sequence of related images derived from an underlying scene may be easily obtained. By utilizing the totality of the information comprising the observed image sequence simultaneously, multiframe restoration methods can exceed the resolving ability of classical single-frame restoration techniques. The mechanism whereby multiframe restoration methods can reverse the degrading effects of aliasing in imaging systems which undersample the observed scene is well understood [203]. Although these ideas are no longer new, their application to general imaging environments remains challenging.

At the heart of all multiframe restoration methods is an *observation model* which relates the observed image sequence measurements to the unknown restoration which is to be estimated. The observation model often incorporates a model for noise which may corrupt the observation process. Additionally, in modern restoration methods, it is common to incorporate *a-priori* knowledge regarding the characteristics of the solution to be estimated. The goal of the multiframe restoration process is the estimation of a superresolved image or sequence given the observed data, the observation model and whatever *a-priori* knowledge may be available.

As the restoration process seeks to "invert" the effects of the observation process, multiframe restoration is an example in the broad category of mathematical *inverse problems*. In practice the multiframe restoration problem is typically *ill-posed*. In order to address this problem, the technique of mathematical *regularization* is applied in order to achieve acceptable restoration.

This chapter demonstrates how to generalize the widely used linear observation model to accommodate nonaffine geometric image registration and spatially varying imaging system degradations. This effort represents a contribution in the application of multiframe restoration techniques to more general imaging systems and environments.

The presentations begins in Section (8.2) with a discussion on *image warping*, a technique commonly used in computer graphics in which digital images are resampled under geometric transformations. In the computer graphics literature, image warping is often considered synonymous with *image resampling*, despite the fact that the concept of resampling is more general than the (typically spatial) geometric transformations considered in computer graphics applications. Despite this mild abuse of terminology, these terms shall nevertheless be used interchangeably. The basic theory of image resampling is presented and after considering an idealized image resampling process, a realizable resampling system which is applicable to both affine and nonaffine geometric transformations is discussed.

In Section (8.3) a multiframe observation model is developed which is appropriate to the superresolution restoration problem. It is shown that there are close similarities between image resampling theory and the model proposed for multiframe restoration, with the result that ideas from image resampling may be utilized in application to multiframe restoration. A linear but spatially varying observation filter is developed which relates the multiple observed images to the unknown superresolution restoration.

A practical algorithm for computing the linear observation filter is presented in Section (8.4), followed by a more detailed discussion on the components which comprise the observation model in Section (8.5).

In Section (8.6) it is shown how the observation filter may be represented in a classical linear inverse problem formulation of the restoration problem. In Section (8.7) a Bayesian MAP estimation procedure is used to solve this inverse problem, using a MRF prior model. The optimization problem that results from the requirement to maximize the *a-posteriori* probability is convex and thus computationally quite feasible despite the typically large number of unknowns.

Section (8.8) presents an example of the application of these methods to the restoration of a superresolved frame from a short image sequence. The findings of this work are summarized in Section (8.9).

8.2 Image Resampling

Image resampling refers to methods for sampling discrete images under coordinate transformation. Using the terminology of the computer graphics community, image resampling or *image warping* is the process of geometrically transforming a sampled source image or *texture image* to a sampled output image or *warped image* using a coordinate transformation called a *warp*. Given the source image $f(\boldsymbol{u})$, defined only at samples $\boldsymbol{u} \in \mathbb{Z}^2$ and the continuous, invertible coordinate transformation or warp, $H : \boldsymbol{u} \mapsto \boldsymbol{x}$ the objective is to find the sampled output image $g(\boldsymbol{x})$ on the lattice $\boldsymbol{x} \in \mathbb{Z}^2$.

A naïve algorithm for image resampling might go as follows: for each pixel $\boldsymbol{x} \in \mathbb{Z}^2$ in the warped image g, let $g(\boldsymbol{x}) = f(H^{-1}(\boldsymbol{x}))$. Inherent to this approach are at least two problems: firstly, $H^{-1}(\boldsymbol{x})$ need not fall on sample points, thus necessitating some form of interpolation of the source image f; secondly, and more seriously, $H^{-1}(\boldsymbol{x})$ may undersample $f(\boldsymbol{u})$ resulting in aliasing which is manifested as objectionable artifacts in warped images in areas where the mapping results in minification. In the quest for realism in computer graphics rendering, addressing these artifacts spurred the development of image resampling theory as well as the development of many optimized techniques for fast, high-quality, texture rendering.

8.2.1 Conceptual Image Resampling Pipeline

Heckbert was one of the first researchers to address these complications in his now classic work [75]. Following Heckbert's exposition, an ideal image resampler may be considered to be composed of a four stage process, enumerated below and illustrated in Figure (8.1). In these equations, functions of a continuous variable are denoted with a subscript "c" and $*$ denotes convolution:

1. Reconstruct a continuous version $f_c(\boldsymbol{u})$ of the sampled texture image $f(\boldsymbol{u})$ with $\boldsymbol{u} \in \mathbb{Z}^2$ using the reconstruction kernel $r(\boldsymbol{u})$.

$$f_c(\boldsymbol{u}) = f(\boldsymbol{u}) * r(\boldsymbol{u}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^2} f(\boldsymbol{k}) \cdot r(\boldsymbol{u} - \boldsymbol{k}). \tag{8.1}$$

2. Use the inverse mapping $H^{-1}(\boldsymbol{x})$ which associates locations $\boldsymbol{x}$ in the warped output image with locations $\boldsymbol{u}$ in the texture image to find the continuous warped image $g_c(\boldsymbol{x})$.

$$g_c(\boldsymbol{x}) = f_c\left(H^{-1}(\boldsymbol{x})\right). \tag{8.2}$$

3. Apply the antialias prefilter $p(\boldsymbol{x})$ to the warped output image $g_c(\boldsymbol{x})$ to produce

a prefiltered image $g'_c(\boldsymbol{x})$ suitable for sampling.

$$g'_c(\boldsymbol{x}) = g_c(\boldsymbol{x}) * p(\boldsymbol{x}) = \int g_c(\boldsymbol{\alpha}) \cdot p(\boldsymbol{x} - \boldsymbol{\alpha}) \, \mathrm{d}\boldsymbol{\alpha}. \tag{8.3}$$

4. Sample the prefiltered image $g'_c(\boldsymbol{x})$ to produce the discrete output image $g(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathbb{Z}^2$.

$$g(\boldsymbol{x}) = g'_c(\boldsymbol{x}) \quad \text{for} \quad \boldsymbol{x} \in \mathbb{Z}^2. \tag{8.4}$$

8.2.2 Realizable Image Resampling Pipeline

Although the four-stage resampling pipeline described above is conceptually useful, in practice construction of the intermediate functions f_c, g_c and g'_c can be avoided. Starting with Equation (8.4) and working backward to Equation (8.1) it is possible to derive an expression for the warped output which does not require the intermediate steps.

$$\begin{aligned}
g(\boldsymbol{x})\Big|_{\boldsymbol{x} \in \mathbb{Z}^2} &= g'_c(\boldsymbol{x}) \quad \text{for} \quad \boldsymbol{x} \in \mathbb{Z}^2 \\
&= \int f_c\left(H^{-1}(\boldsymbol{\alpha})\right) \cdot p(\boldsymbol{x} - \boldsymbol{\alpha}) \, \mathrm{d}\boldsymbol{\alpha} \\
&= \int p(\boldsymbol{x} - \boldsymbol{\alpha}) \sum_{\boldsymbol{k} \in \mathbb{Z}^2} f(\boldsymbol{k}) \cdot r\left(H^{-1}(\boldsymbol{\alpha}) - \boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha}.
\end{aligned} \tag{8.5}$$

By defining

$$\rho(\boldsymbol{x}, \boldsymbol{k}) = \int p(\boldsymbol{x} - \boldsymbol{\alpha}) \cdot r\left(H^{-1}(\boldsymbol{\alpha}) - \boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha}, \tag{8.6}$$

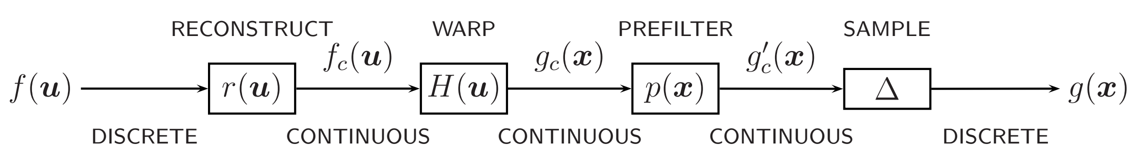


Figure 8.1. Heckbert's conceptual image resampling pipeline.

it is possible to write Equation (8.5) in the succinct form,

$$g(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{k})\rho(\boldsymbol{x},\boldsymbol{k}), \tag{8.7}$$

where $\rho(\boldsymbol{x},\boldsymbol{k})$ is a *spatially varying* resampling filter. Thus the warped image $g(\boldsymbol{x})$ may be computed using only a discrete filtering operation. The resampling filter of Equation (8.6) is described in terms of the warped reconstruction filter r with the integration performed over $\boldsymbol{x}$-space. With a change of variables $\boldsymbol{\alpha} = H(\boldsymbol{u})$, and integrating in $\boldsymbol{u}$-space the resampling filter can be expressed in terms of the warped prefilter p as,

$$\rho(\boldsymbol{x},\boldsymbol{k}) = \int p(\boldsymbol{x} - H(\boldsymbol{u})) \cdot r\left(\boldsymbol{u} - \boldsymbol{k}\right) \left|\frac{\partial H}{\partial \boldsymbol{u}}\right| \mathrm{d}\boldsymbol{u}, \tag{8.8}$$

where $\left|\frac{\partial H}{\partial \boldsymbol{u}}\right|$ is the determinant of the Jacobian of the transformation H. This form for the resampling filter will be useful in the sections ahead.

8.2.3 Affine and Nonaffine Geometric Transformations

In the case of an affine geometric transformation H, the Jacobian $\left|\frac{\partial H}{\partial \boldsymbol{u}}\right|$ is a constant. The resulting resampling filter $\rho(\boldsymbol{x},\boldsymbol{k})$ can be shown to be *space invariant* [75], thus radically simplifying the resampling process as the output image may be computed by a simple convolution with the resampling filter which may be accomplished efficiently using Fourier-domain methods. For nonaffine warps, however, the Jacobian is not constant and the associated resampling filter is spatially varying. Computing the warped output image then involves the application of a spatially varying resampling filter which is computationally far more costly than the affine case.

8.3 A Multiframe Observation Model

The multiframe observation model which is used in formulating the superresolution restoration problem is developed in this section. Comparing the observation model discussed here with the theory of image resampling developed above, will illuminate some common characteristics. The presentation begins by modeling a continuous-discrete process where P sampled, low-resolution, images are captured from an underlying continuous scene. The observed imagery may constitute a temporal image sequence obtained from a single (typically moving) camera, or may derive from multiple cameras observing the scene.

The objective is the estimation of the underlying scene, computed on a discrete, superresolution sampling lattice. This is achieved using a discrete-discrete observation model derived from the continuous-discrete model. Superresolution describes a process of bandwidth extrapolation — not merely an increase in the number of image samples. Superresolution restoration is possible since the totality of information available from the image sequence as a whole is typically greater than that available from any single image of the scene.

8.3.1 Continuous-Discrete Multiframe Observation Model

Let the observed sampled low-resolution image sequence consist of P images $g^{(i)}(\boldsymbol{x}),\ i \in \mathbb{P} \triangleq \{1, 2, \ldots, P\}$ derived from a continuous scene $f_c(\boldsymbol{u})$. The observed images are geometrically related to $f_c(\boldsymbol{u})$ via the coordinate transformations $H^{(i)} : \boldsymbol{u} \mapsto \boldsymbol{x}$ which account for the relative motion between the scene and the camera. Associated with each observation is a (possibly spatially varying) PSF $h^{(i)}$, which may differ from observation to observation. The PSFs model the lens and sensor response, defocus, motion blur and so on. Finally the images are sampled. The continuous-discrete multiframe observation process is illustrated in Figure (8.2) and is described by,

$$g^{(i)}(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} = \int h^{(i)}(\boldsymbol{x}, \boldsymbol{\alpha}) \cdot f_c\left(H^{(i)^{-1}}(\boldsymbol{\alpha})\right) \mathrm{d}\boldsymbol{\alpha}. \tag{8.9}$$

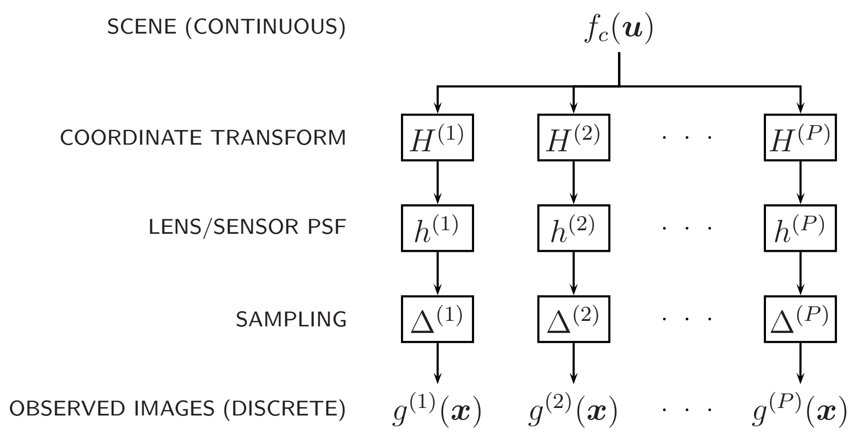


Figure 8.2. Continuous-discrete multiframe observation model.

8.3.2 Discrete-Discrete Multiframe Observation Model

The next objective is to obtain a discretized approximation of the continuous scene $f_c(\boldsymbol{u})$ on a high-resolution sampling lattice given the observed images $g^{(i)}$ and knowledge of the observation process characterized by $H^{(i)}$, $h^{(i)}$ and $\Delta^{(i)}$ for $i \in \mathbb{P}$. The continuous scene $f_c(\boldsymbol{u})$ may be approximated by interpolating the discrete samples $f(\boldsymbol{k})$, $\boldsymbol{k} \in \mathbb{Z}^2$ using an interpolation kernel h_r as,

$$f_c(\boldsymbol{u}) \approx \sum_{\boldsymbol{k} \in \mathbb{Z}^2} f(\boldsymbol{V}\boldsymbol{k}) \cdot h_r\left(\boldsymbol{u} - \boldsymbol{V}\boldsymbol{k}\right) \tag{8.10}$$

where $\boldsymbol{V}$ is the sampling matrix [45]

$$\boldsymbol{V} = \begin{bmatrix} 1/q_x & 0 \\ 0 & 1/q_y \end{bmatrix},$$

and $q_x\,, q_y \in \mathbb{N}$ are the horizontal and vertical magnification factors respectively of the superresolution image estimate. Combining Equation (8.9) with Equation (8.10) then yields the relationship between the observed frames and the discrete approxi-

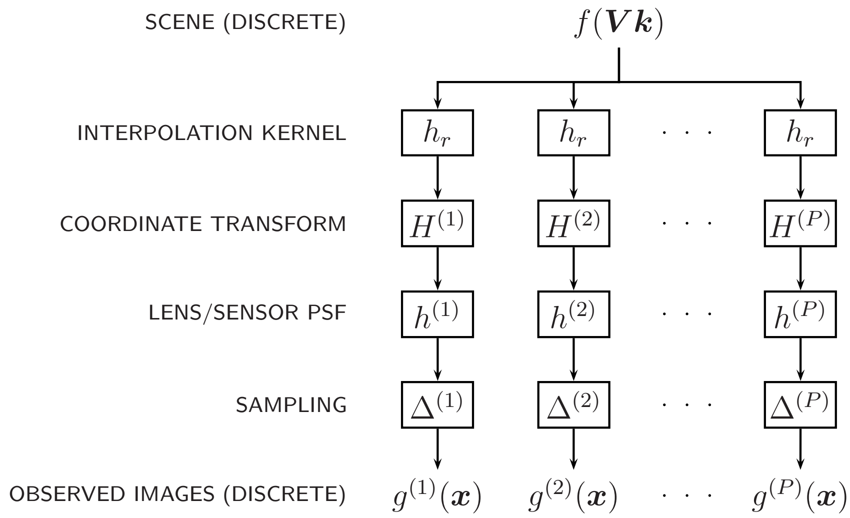


Figure 8.3. Discrete-discrete multiframe observation model.

mation of the scene as,

$$\begin{aligned} g^{(i)}(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} &= \int h^{(i)}(\boldsymbol{x},\boldsymbol{\alpha}) \cdot f_c\left(H^{(i)^{-1}}(\boldsymbol{\alpha})\right) \mathrm{d}\boldsymbol{\alpha} \\ &= \int h^{(i)}(\boldsymbol{x},\boldsymbol{\alpha}) \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{V}\boldsymbol{k}) \cdot h_r\left(H^{(i)^{-1}}(\boldsymbol{\alpha}) - \boldsymbol{V}\boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha}. \quad (8.11) \end{aligned}$$

These relationships are illustrated in Figure (8.3) which augments Figure (8.2) to include the interpolation kernel h_r.

8.3.3 Image Resampling and the Multiframe Observation Model

Comparing Equation (8.11) with the image resampling filter in Equation (8.5), which is reproduced below for convenience,

$$g(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathbb{Z}^2} = \int p(\boldsymbol{x}-\boldsymbol{\alpha}) \sum_{\boldsymbol{k}\in\mathbb{Z}^2} f(\boldsymbol{k}) \cdot r\left(H^{-1}(\boldsymbol{\alpha}) - \boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha},$$

it is apparent that there is a parallel between the image resampling equation and the multiframe observation model. The spatially varying PSF $h^{(i)}(\boldsymbol{x},\boldsymbol{\alpha})$ associated with the observation $g^{(i)}(\boldsymbol{x})$ replaces the spatially invariant image resampling prefilter $p(\boldsymbol{x}-\boldsymbol{\alpha})$ in the resampling equation, and the reconstruction filter $h_r(H^{(i)^{-1}}(\boldsymbol{\alpha})-\boldsymbol{V}\boldsymbol{k})$

TABLE 8.1

COMPARING IMAGE RESAMPLING WITH THE MULTIFRAME OBSERVATION MODEL

Image Resampling		Multiframe Observation Model	
Discrete texture	$f(\boldsymbol{u})$	Discrete scene estimate	$f(\boldsymbol{u})$
Reconstruction filter	$r(\boldsymbol{u})$	Interpolation kernel	$h_r(\boldsymbol{u})$
Geometric transform	$H(\boldsymbol{u})$	Scene/camera motion	$H^{(i)}(\boldsymbol{u})$
Antialias prefilter	$p(\boldsymbol{x})$	Observation SVPSF	$h^{(i)}(\boldsymbol{x}, \boldsymbol{\alpha})$
Warped output image	$g(\boldsymbol{x})$	Observed images	$g^{(i)}(\boldsymbol{x})$
Resampling filter	$\rho(\boldsymbol{x}, \boldsymbol{k})$	Observation filter	$\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$
$\int p(\boldsymbol{x} - H(\boldsymbol{u})) \cdot r(\boldsymbol{u} - \boldsymbol{k}) \left\lvert \frac{\partial H}{\partial \boldsymbol{u}} \right\rvert \mathrm{d}\boldsymbol{u}$		$\int h^{(i)}\left(\boldsymbol{x}, H^{(i)}(\boldsymbol{u})\right) \cdot h_r(\boldsymbol{u} - \boldsymbol{V}\boldsymbol{k}) \left\lvert \frac{\partial H^{(i)}}{\partial \boldsymbol{u}} \right\rvert \mathrm{d}\boldsymbol{u}$	

has a counterpart in the reconstruction filter $r(H^{-1}(\boldsymbol{\alpha}) - \boldsymbol{k})$ in the resampling equation. Noting these parallels, it is possible to follow the approach used in Section (8.2.2) and define the spatially varying filter,

$$\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k}) = \int h^{(i)}(\boldsymbol{x}, \boldsymbol{\alpha}) \cdot h_r\left(H^{(i)^{-1}}(\boldsymbol{\alpha}) - \boldsymbol{V}\boldsymbol{k}\right) \mathrm{d}\boldsymbol{\alpha}, \tag{8.12}$$

which allows Equation (8.11) to be written as,

$$g^{(i)}(\boldsymbol{x})\Big|_{\boldsymbol{x} \in \mathbb{Z}^2} = \sum_{\boldsymbol{k} \in \mathbb{Z}^2} f(\boldsymbol{V}\boldsymbol{k}) \cdot \rho^{(i)}(\boldsymbol{x}, \boldsymbol{k}). \tag{8.13}$$

Analogous to the resampling derivation, the change of variables $\boldsymbol{\alpha} = H^{(i)}(\boldsymbol{u})$ is applied so that the integration is over $\boldsymbol{u}$-space and the observation filter $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ can be expressed in terms of the warped PSF $h^{(i)}(\boldsymbol{x}, \boldsymbol{\alpha})$ as,

$$\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k}) = \int h^{(i)}\left(\boldsymbol{x}, H^{(i)}(\boldsymbol{u})\right) \cdot h_r(\boldsymbol{u} - \boldsymbol{V}\boldsymbol{k}) \left\lvert \frac{\partial H^{(i)}}{\partial \boldsymbol{u}} \right\rvert \mathrm{d}\boldsymbol{u}. \tag{8.14}$$

The parallels between image resampling and the multiframe observation model are summarized in Table (8.1).

8.4 Determining the Observation Filter

The linear multiframe observation process of Equation (8.13) relates the unknown scene f to the observed frames $g^{(i)}$ via the spatially varying observation filters $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$. The coefficients of the observation filters must be determined according to the coordinate transformations $H^{(i)}$, the lens and sensor characteristics $h^{(i)}$ and the sampling characteristics $\Delta^{(i)}$ for $i \in \mathbb{P}$. In this section a general-purpose algorithm for computing the filter coefficients is introduced. This is followed in Section (8.5) with a discussion on specific lens, sensor and coordinate transformation models and their integration into the observation filter using the algorithm presented here.

8.4.1 An Algorithm for Computing the Warped Pixel Response

An algorithm for computing the coefficients of the spatially varying observation filter of Equation (8.14) is presented in Figure (8.4). In order to obtain a reasonably efficient algorithm implementation some simplifying assumptions are made. It is assumed that the dimensions of the restored pixels are sufficiently small that the integral may be approximated by a weighted point sample. Also, it is assumed that the interpolation kernel is constant over the support of each restored pixel (a box function) so that its evaluation may be omitted, instead ensuring that the sum of the filter coefficients is unity. Note that the function of the back-projection step in the algorithm is to reduce the range of $\boldsymbol{k}$ over which the warped filter need be evaluated.

8.4.2 Graphical Illustration of the Computation of the Warped Pixel Response

The back-projection and projection process used to determine the filter response $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ in the algorithm discussed above is graphically illustrated in Figure (8.5).

for each observed image $\boldsymbol{g}^{(i)}$ {
    for each pixel $\boldsymbol{x}$ {
        back-project the boundary of $h^{(i)}(\boldsymbol{x}, \boldsymbol{\alpha})$ from $g^{(i)}(\boldsymbol{x})$ to ...
            ... the restored image space using $H^{(i)^{-1}}$
        determine a bounding region for the image of $h(\boldsymbol{x}, \boldsymbol{\alpha})$ under $H^{(i)^{-1}}$
        for each pixel indexed by $\boldsymbol{k}$ in the bounding region {
            set $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k}) = h^{(i)}\left(\boldsymbol{x}, H^{(i)}(\boldsymbol{u})\right) \cdot \left|\frac{\partial H^{(i)}}{\partial \boldsymbol{u}}\right|$ with $\boldsymbol{u} = \boldsymbol{V}\boldsymbol{k}$
        }
        normalize $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ so that $\sum_k \rho^{(i)}(\boldsymbol{x}, \boldsymbol{k}) = 1$
    }
}

Figure 8.4. Algorithm for computing the observation filter coefficients $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$.

First, an idealized scenario is considered, followed by an example showing actual filter coefficients computed for a realistic pixel model.

In Figure (8.5)(a), the support of the response of an idealized observed pixel is illustrated as a darkly-shaded region. The image of this region under H^{-1} is shown as the darkly-shaded region in the restored image space in Figure (8.5)(b). Also shown in Figure (8.5)(b), as a lightly-shaded region, is the smallest integer-pixel bounding region containing the image of the observed pixel response under H^{-1}. A single high-resolution pixel contained in the lightly-shaded bounding region in Figure (8.5)(b) is shown in white. The image of this high resolution pixel under H is illustrated by the small, white, warped region contained within the darkly-shaded region representing the observed pixel response in Figure (8.5)(a). The ratio of the areas of the high resolution pixel and its image under H is given by the inverse of the determinant of the Jacobian of the transformation H.

In Figure (8.5)(c) and Figure (8.5)(d) an example of the computation of the filter coefficients $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ is illustrated for a realistic pixel model and geometric transformation. Figure (8.5)(c) shows the spatial response of a sensor pixel characterized by a uniform spatial response and 100% fill-factor in a camera with a diffraction-limited optical system. Under a nonaffine projective transformation, the corresponding observation filter response is shown in Figure (8.5)(d). Notice the warping of the pixel response function which results from the projective transformation relating the views. Also note that the regular sampling lattice in the superresolution restoration space in Figure (8.5)(d) maps under H to the irregular lattice in the space of the observed pixel response in Figure (8.5)(c). This is most easily seen by examining the samples adjacent to the left and right edges of Figure (8.5)(c). The gray-level samples in Figure (8.5)(d) represent the values of the discrete observation filter associated with the observed pixel under the geometric transformation H.

8.5 Modeling the Observation Process

In this section, the components which comprise the observation filter $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ are examined in further detail:

1. The observation PSF

   A model for the imaging system PSFs is derived by considering a diffraction-limited optical system and focal plane sensor array.

2. The coordinate transformations

   In practice the coordinate warps result from relative scene/camera motion. The discussion considers a case where a moving camera views a static planar scene. This viewing geometry allows a simple computation of the mapping between points in each camera view.

3. The interpolation kernel

   For simplicity, a constant-valued box function is used as the interpolation kernel $h_r(\boldsymbol{u})$. This is equivalent to a zero-order hold.

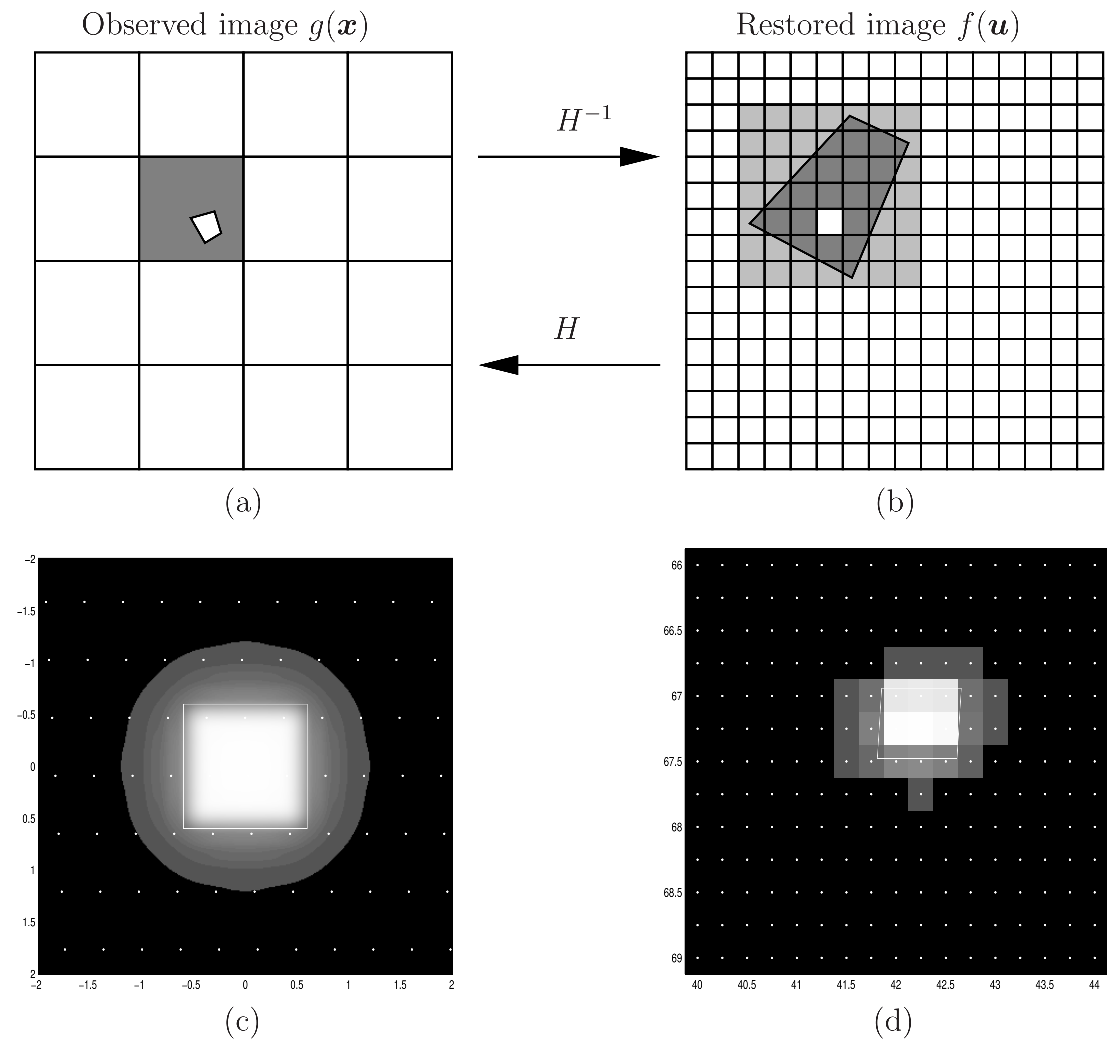


Figure 8.5. Determining the observation filter response (see text): (a) Support of the response of an idealized observed pixel. (b) Image of the idealized pixel response under the transformation H^{-1} and the associated bounding region. (c) Observed pixel response for realistic observed pixel (assumes diffraction-limited optics, spatially uniform pixel response with 100% fill factor). (d) Observation filter $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$ associated with the observed pixel under under the geometric transformation H.

8.5.1 Optical System Model

The optical system is assumed to be diffraction-limited, with the lens system having a circular exit pupil. The lens system f-number, that is, the ratio of focal length to aperture diameter, is given by N. The system is assumed to operate under incoherent, monochromatic illumination with wavelength λ. No lens aberrations are accounted for. Under these assumptions, the PSF associated with the optical system is radially symmetric and is given by,

$$h_o(r') = \left[2\frac{J_1(r')}{r'}\right]^2, \quad \text{with} \quad r' = (\pi/\lambda N)r \tag{8.15}$$

where $J_1(\cdot)$ is the Bessel function of the first kind and r' is the normalized radial distance computed from the true radial distance r [151, 64]. A plot of the optical system PSF for $\lambda = 550$nm, corresponding to the spectral color green, and a lens f-number of $N = 2.8$ is shown in Figure (8.6). The first zero ring of the Airy disk occurs at at $1.22\lambda N = 1.88\mu$m.

The optical system may be equivalently characterized by the Optical Transfer Function (OTF), which is the normalized frequency response of the optical system. The OTF of the diffraction-limited optical system described above is given by [64],

$$H_o(\rho') = \begin{cases} \frac{2}{\pi}\left[\cos^{-1}(\rho') - \rho'\sqrt{1-\rho'^2}\right], & \text{for } \rho' \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{8.16}$$

where $\rho' = \rho/\rho_c$ is a normalized radial spatial frequency. The normalized radial frequency ρ' is defined in terms of ρ, the true radial spatial frequency and $\rho_c = 1/\lambda N$ the radial spatial frequency cut-off which is a function of the wavelength of the illumination and the lens f-number. A plot of the OTF corresponding to the PSF above, is presented in Figure (8.7). In this example, the radial frequency cut-off $\rho_c = 1/\lambda N = 649.35$ lines/mm.

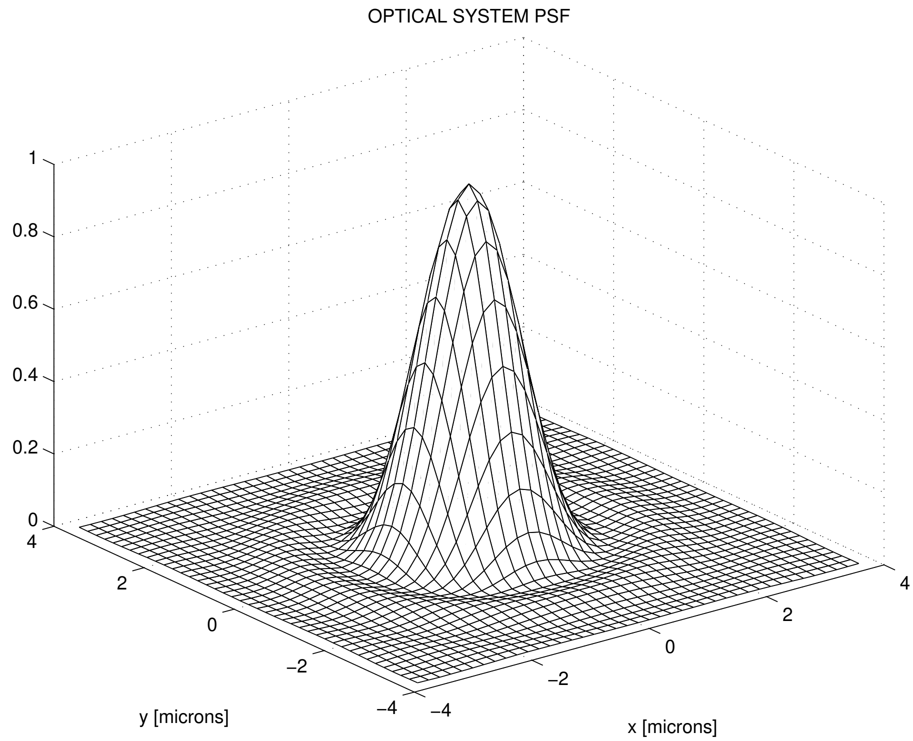


Figure 8.6. PSF of a diffraction-limited optical system with a circular exit pupil with f-number $N = 2.8$ and incoherent illumination with wavelength $\lambda = 550$nm.

It is important to realize that the simple, diffraction-limited optical system model presented here represents a best case model. Even the most superlative optical systems do not approach this of level of performance. This can be confirmed by consulting the MTF plots produced for commercially available lenses. Typically the MTF for real lenses taper off far more rapidly than the limits determined by diffraction. Nevertheless, the discussion illustrates the basic approach that can be taken for real systems, that is, a characterization of the optical system using the PSF or MTF.

8.5.2 Focal Plane Sensor Array Model

The image projected onto the focal plane by the optical system is measured by a FPA consisting of a regular lattice of independent light-sensing pixels. Each pixel

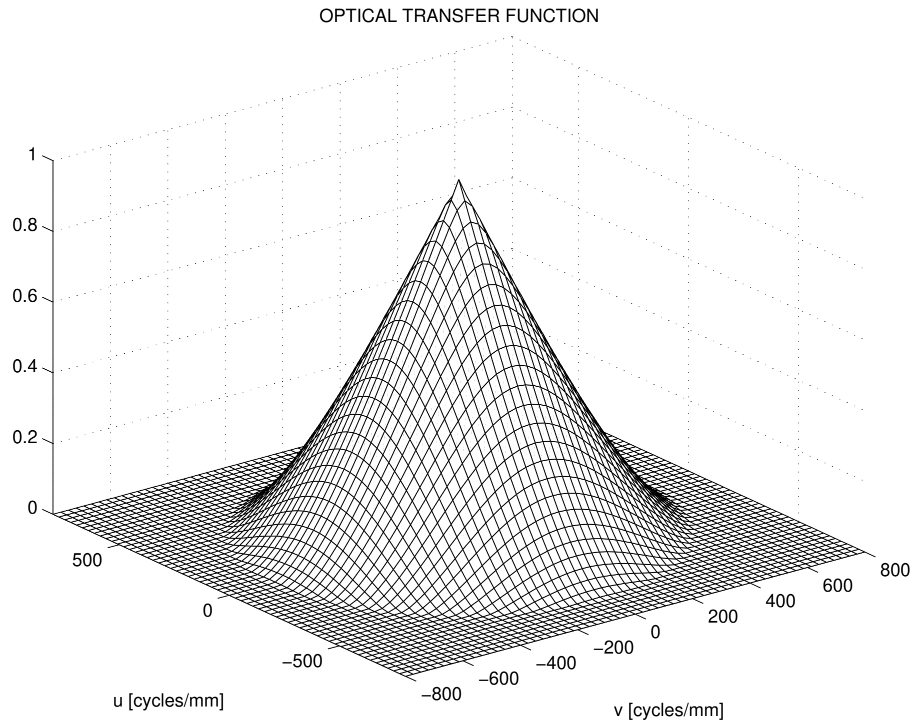


Figure 8.7. OTF of a diffraction-limited optical system with a circular exit pupil with f-number $N = 2.8$ and incoherent illumination with wavelength $\lambda = 550$nm.

integrates the incident illumination over the pixel's *active region* for the duration of the *aperture time*. For example, in a CCD incident photons striking the active area of a pixel results, through the photoelectric effect, in the generation of electronic charge which is accumulated in a charge "well" where it is held until it is read out. The amount of accumulated charge is proportional to the number of photons striking the pixel, or equivalently, to the incident illumination.

The sensor array is modeled as a regular tessellation of square pixels with dimension T. Each pixel integrates the incident illumination over its entire $T \times T$ region of support and the response to photons striking at any point across this region is assumed to be uniform. These assumptions make possible a simple but useful model of an ideal pixel. The PSF associated with such a pixel may be modeled as,

$$h_p(x,y) = \frac{1}{T^2}\,\mathrm{rect}\left(\frac{x}{T}\right)\mathrm{rect}\left(\frac{y}{T}\right) = \begin{cases} 1/T^2, & |x| \le T/2 \ \text{ and } \ |y| \le T/2 \\ 0, & \text{otherwise} \end{cases}. \quad (8.17)$$

The 2-D Fourier transform of Equation (8.17) is given by

$$H_p(\xi_1,\xi_2) = \mathrm{sinc}(T\xi_1)\cdot\mathrm{sinc}(T\xi_2) = \frac{\sin(\pi T\xi_1)}{\pi T\xi_1}\cdot\frac{\sin(\pi T\xi_2)}{\pi T\xi_2}. \quad (8.18)$$

Since the response $H_p(0,0) = 1$, this expression is also the OTF associated with the pixel.

In reality, sensor elements deviate significantly from the assumptions we made in this section. Firstly the "fill-factor", that is the ratio of light sensitive area of the pixel to the area of the full pixel is often not unity. This is because some portion of the pixel area is often required to contain read-out electronics. Additionally, the sensitivity of the pixel to incident illumination is often not uniform across the entire pixel area. This is especially true for devices which utilize micro-lenses in an attempt to artificially increase the pixel fill-factor. Finally, due to variations across the device substrate, the sensitivity of pixels across the device is not uniform. Nevertheless the model developed is useful as a first approximation to the true sensor response. Fortunately, incorporating measured sensor model responses into the multiframe observation model does not pose significant additional difficulty.

8.5.3 Combined Lens and Sensor Model

Convolving the PSFs associated with the lens and sensor yields the PSF for the combined system, $h_c = h_o * h_p$. A plot of a combined PSF is shown in Figure (8.8). A cross-sectional profile of this function is shown in Figure (8.9). It is instructive to notice that in the combined lens and sensor response characteristic that there is considerable "leakage" between adjacent pixels. This is evident in Figure (8.9) from

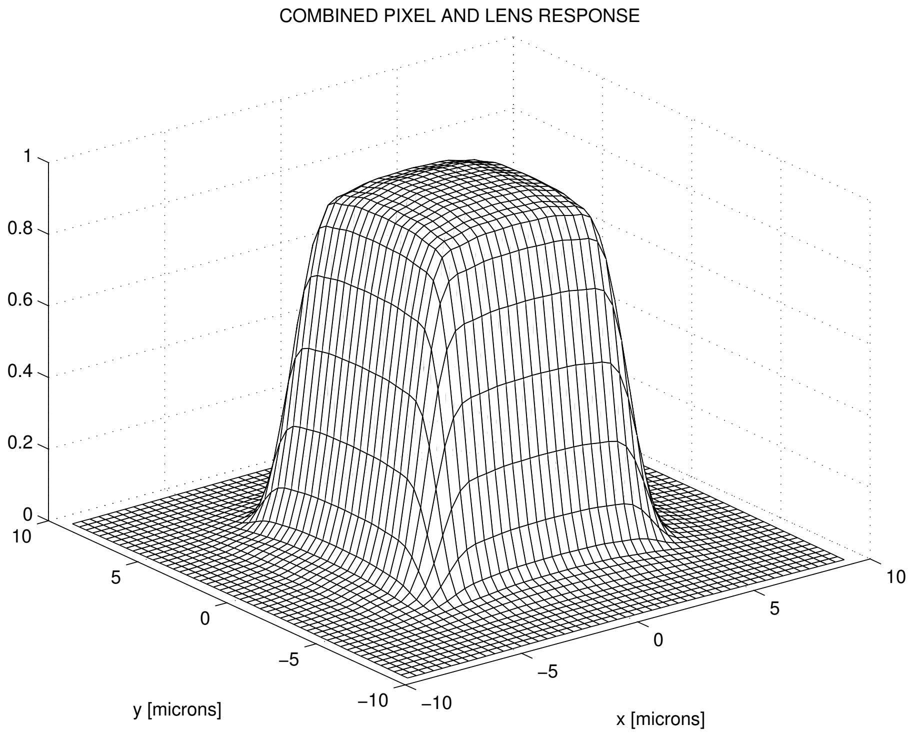


Figure 8.8. Combined lens/pixel point spread function with focal plane array pixel dimensions $9\mu\text{m} \times 9\mu\text{m}$, illumination wavelength $\lambda = 550\text{nm}$ (green), lens f-number $N = 2.8$.

which it can be clearly seen that the pixel response is significant even well outside the pixel boundaries. The response for real-world imaging systems will typically be considerably worse than these plots as discussed previously.

The combined lens and sensor MTF, H_c, is found by multiplying the optical system MTF and the sensor MTF expressions as $H_c = H_o \cdot H_p$.

8.5.4 Imaging Geometry

The coordinate transformations $H^{(i)}$ which relate the observed image coordinate systems to the scene coordinate system are considered here. In general, these mappings are computed by estimating the motion occurring between the observed frames. The warps need not be uniform but they must be locally invertible.

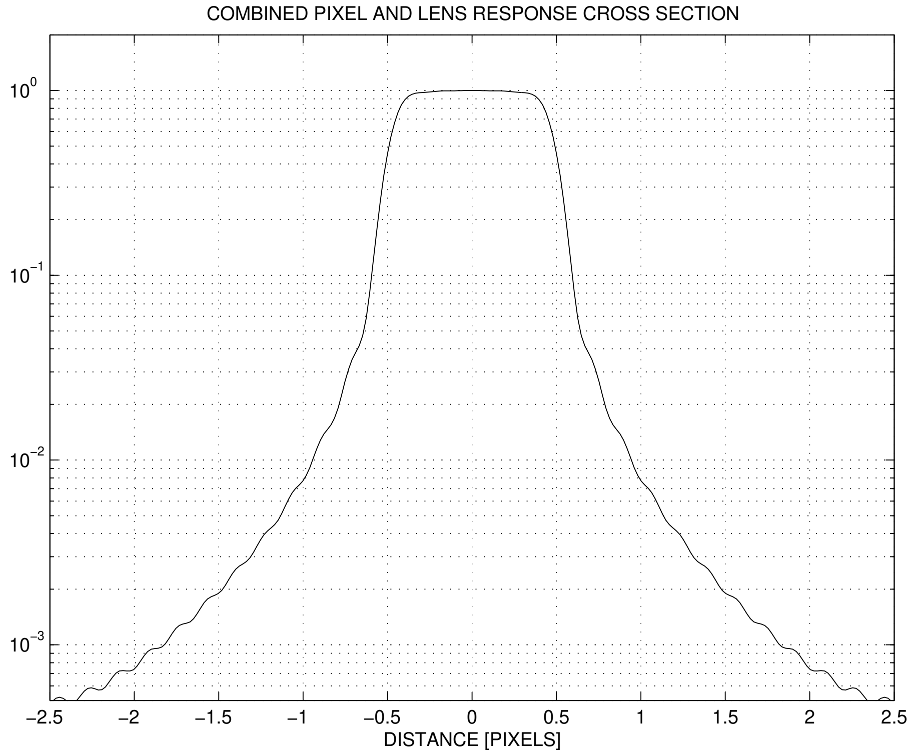


Figure 8.9. Profile of the combined lens/pixel point spread function with focal plane array pixel dimensions $9\mu\text{m} \times 9\mu\text{m}$, illumination wavelength $\lambda = 550\text{nm}$ (green), lens f-number $N = 2.8$.

In the example presented in Section (8.8), a 2-D "scene" is located in the x-y plane in 3-D space. Multiple simulated views of the planar scene are generated using a custom raytracer which models the lens and sensor effects discussed earlier. Since the calibration matrices of the cameras are known, it is possible to compute the homography induced by the scene plane [74]. This is a mapping which relates the coordinates of points in each image and is a projectivity. The coordinate transformation $H^{(i)} : \boldsymbol{u} \mapsto \boldsymbol{x}$ is thus of the form,

$$\begin{aligned} x &= \phi(u,v) = \frac{h_{11}u + h_{12}v + h_{13}}{h_{31}u + h_{32}v + 1} \\ y &= \psi(u,v) = \frac{h_{21}u + h_{22}v + h_{23}}{h_{31}u + h_{32}v + 1} \end{aligned} \tag{8.19}$$

and the determinant of the Jacobian $\left|\dfrac{\partial H^{(i)}}{\partial \boldsymbol{u}}\right|$ is given by $\begin{vmatrix} \frac{\partial\phi}{\partial u} & \frac{\partial\phi}{\partial v} \\ \frac{\partial\psi}{\partial u} & \frac{\partial\psi}{\partial v} \end{vmatrix}$.

8.6 The Linear Multiframe Observation Model

In Equation (8.13), samples $g^{(i)}(\boldsymbol{x})$ in the observed images are related to the sampled approximation of the scene $f(\boldsymbol{V}\boldsymbol{k})$ via a linear, but spatially varying, discrete filter kernel $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$. In practice the recorded images $g^{(i)}(\boldsymbol{x})$, $i \in \mathbb{P}$ in Equation (8.9) are known only for a finite subset of $\boldsymbol{x} \in \mathbb{Z}^2$, namely on a regular lattice $\boldsymbol{x} \in \{1, 2, \ldots, N_r\} \times \{1, 2, \ldots, N_c\}$ where N_r and N_c are the number of image rows and columns respectively. Similarly, the reconstruction of the original scene f in Equation (8.10) is computed on a lattice consisting of $q_y N_r$ rows and $q_x N_c$ columns. Each observed image consisting of $N_r N_c$ pixels can be conveniently represented as a $N_r N_c$-element column vector $\boldsymbol{g}^{(i)}$ using a row-major ordering of the observed data and likewise, the reconstructed image can be represented with a $q_y N_r q_x N_c$-element column vector $\boldsymbol{f}$.

Referring to Equation (8.13), the value of the j^{th} pixel in the observed image $\boldsymbol{g}^{(i)}$ is related to the unknown reconstruction $\boldsymbol{f}$ via a linear combination of the elements of the unknown $\boldsymbol{f}$ as,

$$\boldsymbol{g}_j^{(i)} = \sum_{k=1}^{q_y N_r q_x N_c} A_{jk}^{(i)} \boldsymbol{f}_k \tag{8.20}$$

where the matrix coefficients $A_{jk}^{(i)}$ are determined from the discrete kernel $\rho^{(i)}(\boldsymbol{x}, \boldsymbol{k})$. Note that the matrix containing the observation filter coefficients will be very sparse. Using matrix-vector notation, Equation (8.20) may be written as,

$$\boldsymbol{g}^{(i)} = \boldsymbol{A}^{(i)} \boldsymbol{f}. \tag{8.21}$$

Since P images are observed, it is convenient to stack the observation vectors $\boldsymbol{g}^{(i)}$

and matrices $\boldsymbol{A}^{(i)}$ defining,

$$\boldsymbol{g} \triangleq \begin{bmatrix} \boldsymbol{g}^{(1)} \\ \boldsymbol{g}^{(2)} \\ \vdots \\ \boldsymbol{g}^{(P)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{A} \triangleq \begin{bmatrix} \boldsymbol{A}^{(1)} \\ \boldsymbol{A}^{(2)} \\ \vdots \\ \boldsymbol{A}^{(P)} \end{bmatrix}$$

so that the observation process can be expressed in the classic form,

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}. \tag{8.22}$$

8.7 A Bayesian Framework for Restoration

Equation (8.22) represents the linear observation model relating the unknown image $\boldsymbol{f}$ to be estimated, to the observed images $\boldsymbol{g}$. Solving for $\boldsymbol{f}$ is a classic example of an inverse problem. It is well known [168] that the multiframe superresolution restoration problem is ill-posed, so special care must be taken when estimating a value for $\boldsymbol{f}$. In this work a Bayesian framework is used to compute a meaningful solution to Equation (8.22).

It is assumed that the observation process is corrupted by an additive noise process. While arguably not the most realistic choice, for reasons of convenience it is commonly assumed that the additive noise process is Gaussian with zero-mean. The linear observation process of Equation (8.22) is thus augmented to include a noise term,

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{n}. \tag{8.23}$$

The pdf of the zero-mean Gaussian noise term $\boldsymbol{n}$ is given by,

$$\mathcal{P}_{\boldsymbol{N}}(\boldsymbol{n}) = \frac{1}{(2\pi)^{\frac{PN_rN_c}{2}}|\boldsymbol{K}|} \exp\left\{-\frac{1}{2}\boldsymbol{n}^\top \boldsymbol{K}^{-1}\boldsymbol{n}\right\} \tag{8.24}$$

where $\boldsymbol{K}$ is the positive-definite noise covariance matrix.

Bayesian MAP estimation is used to estimate the superresolved image $\boldsymbol{f}$ given the observed data $\boldsymbol{g}$. The MAP estimate $\hat{\boldsymbol{f}}_{\mathrm{MAP}}$ maximizes the *a-posteriori* probability $\mathcal{P}(\boldsymbol{f}|\boldsymbol{g})$ as,

$$\begin{aligned}\hat{\boldsymbol{f}}_{\mathrm{MAP}} &= \arg\max_{\boldsymbol{f}} \{\mathcal{P}(\boldsymbol{f}|\boldsymbol{g})\} \\ &= \arg\max_{\boldsymbol{f}} \left\{ \frac{\mathcal{P}(\boldsymbol{g}|\boldsymbol{f})\,\mathcal{P}(\boldsymbol{f})}{\mathcal{P}(\boldsymbol{g})} \right\} \qquad (8.25)\end{aligned}$$

where the expression in Equation (8.25) results from the application of Bayes' theorem. Taking logarithms and noting that the maximum is independent of $\mathcal{P}(\boldsymbol{g})$, yields

$$\hat{\boldsymbol{f}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{f}} \{\log \mathcal{P}(\boldsymbol{g}|\boldsymbol{f}) + \log \mathcal{P}(\boldsymbol{f})\}. \qquad (8.26)$$

8.7.1 The Likelihood Function

Given the observation expression in Equation (8.23) the conditional probability $\mathcal{P}(\boldsymbol{g}|\boldsymbol{f})$ is determined by the noise pdf in Equation (8.24) as

$$\begin{aligned}\mathcal{P}(\boldsymbol{g}|\boldsymbol{f}) &= \mathcal{P}_{\boldsymbol{N}}(\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}) \\ &= \frac{1}{(2\pi)^{\frac{PN_rN_c}{2}}|\boldsymbol{K}|} \exp\left\{-\frac{1}{2}(\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f})^{\top}\boldsymbol{K}^{-1}(\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f})\right\}. \qquad (8.27)\end{aligned}$$

This expression is called the *likelihood function* given its interpretation as the likelihood of observing $\boldsymbol{g}$ given $\boldsymbol{f}$.

8.7.2 The Prior Probability Density

In the Bayesian framework it is assumed that prior knowledge is available regarding the probability density of the superresolution image random vector $\mathcal{P}(\boldsymbol{f})$. A MRF model is used to impose a probability density function on the space of possible superresolution restorations. MRFs are often used for image modeling applications as they provide a convenient framework for specifying a global probability density function by specifying only spatially local energy interactions [112, 38, 43, 44]. By the Hammersley-Clifford theorem [27, 11] the pdf of a MRF is given by the Gibbs distribution,

$$\mathcal{P}(\boldsymbol{f}) = \frac{1}{k_p} \exp\left\{-\frac{1}{\beta}\,\mathcal{E}(\boldsymbol{f})\right\}, \tag{8.28}$$

where $\mathcal{E}(\boldsymbol{f})$ measures the energy of the configuration $\boldsymbol{f}$, k_p is a normalizing constant called the *partition function* and β is the MRF "temperature" parameter (terminology carried over from the statistical physics literature). In the case of a MRF model, the energy $\mathcal{E}(\boldsymbol{f})$ is specified in terms of local interactions between sets of spatially neighboring lattice locations. The Gibbs distribution associated with the MRF used in this work takes the form,

$$\mathcal{P}\left(\boldsymbol{f}\right) = \frac{1}{k_p} \exp\left\{-\frac{1}{\beta} \sum_{c\in\mathcal{C}} \rho_T\left(\boldsymbol{\partial}_c \boldsymbol{f}\right)\right\} \tag{8.29}$$

where the summation is over the set of all "cliques" $\mathcal{C}$ which define the sets of interacting pixels, with $\boldsymbol{\partial}_c$ computing the local spatial activity for each clique. The nonlinear penalizing function $\rho_T(x)$ is taken to be the convex Huber penalty function [188],

$$\rho_T(x) = \begin{cases} x^2 & |x| \le T \\ 2T|x| - T^2 & |x| > T \end{cases}. \tag{8.30}$$

Local spatial activity is computed using finite difference approximations to the second directional derivatives (vertical, horizontal and two diagonal directions) in the image $\boldsymbol{f}$, specifically,

$$\begin{aligned}
\boldsymbol{\partial}_1\boldsymbol{f} &= f\left[n_1-1, n_2\right] - 2f\left[n_1, n_2\right] + f\left[n_1+1, n_2\right] \\
\boldsymbol{\partial}_2\boldsymbol{f} &= f\left[n_1, n_2-1\right] - 2f\left[n_1, n_2\right] + f\left[n_1, n_2+1\right] \\
\boldsymbol{\partial}_3\boldsymbol{f} &= \tfrac{1}{2}\, f\left[n_1+1, n_2-1\right] - f\left[n_1, n_2\right] + \tfrac{1}{2}\, f\left[n_1-1, n_2+1\right] \\
\boldsymbol{\partial}_4\boldsymbol{f} &= \tfrac{1}{2}\, f\left[n_1-1, n_2-1\right] - f\left[n_1, n_2\right] + \tfrac{1}{2}\, f\left[n_1+1, n_2+1\right].
\end{aligned} \tag{8.31}$$

For pixels on the image boundaries, where it is not possible to compute these finite differences, appropriate first order differences are substituted.

This nonlinear MRF image model imposes local image smoothness but nevertheless does allow for the formation of sharp edge features in the restored image [168].

8.7.3 The Objective Function

Substituting the likelihood term in Equation (8.27) and the prior in Equation (8.29) into Equation (8.26) and removing constants independent of $\boldsymbol{f}$ gives the objective function

$$\begin{aligned} \hat{\boldsymbol{f}}_{\mathrm{MAP}} &= \arg\max_{\boldsymbol{f}} \left\{ -\frac{1}{2}(\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f})^{\top}\boldsymbol{K}^{-1}(\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}) - \frac{1}{\beta}\sum_{c\in\mathcal{C}} \rho_T\left(\boldsymbol{\partial}_c\boldsymbol{f}\right) \right\} \\ &= \arg\min_{\boldsymbol{f}} \left\{ \frac{1}{2}(\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f})^{\top}\boldsymbol{K}^{-1}(\boldsymbol{g}-\boldsymbol{A}\boldsymbol{f}) + \gamma\sum_{c\in\mathcal{C}} \rho_T\left(\boldsymbol{\partial}_c\boldsymbol{f}\right) \right\} \end{aligned} \tag{8.32}$$

where $\gamma \triangleq \frac{1}{\beta} > 0$ is a user-adjustable *regularization parameter* reflecting the degree to which the prior is favored over the observed data. Equation (8.32) shows that the objective function consists of two competing terms: the first measures the extent to which the estimate $\boldsymbol{f}$ is compatible with the observed data $\boldsymbol{g}$ while the second measures the conformance of the estimate to prior expectations regarding the characteristics of the solution. The adjustable regularization parameter γ allows this balance to be controlled by the user. Methods for automatic selection of the regularization parameter have been studied [69], but in this work γ is chosen experimentally.

8.7.4 Optimization

Determining the MAP estimate $\hat{\boldsymbol{f}}_{\mathrm{MAP}}$ thus requires the minimization of the objective function in Equation (8.32). The term deriving from the likelihood function is convex in $\boldsymbol{f}$, but need not be *strictly* convex since the nullspace of the observation matrix $\boldsymbol{A}$ may be (and is typically) nontrivial. Thus, considering the likelihood term alone, there is no assurance of the uniqueness of a solution for $\hat{\boldsymbol{f}}_{\mathrm{MAP}}$. Since the prior is described by a sum of convex functions of the operators $\boldsymbol{\partial}_c$ with $c \in \mathcal{C}$, if the intersection of the nullspace of the operators $\boldsymbol{\partial}_c$ and the nullspace of the observation matrix $\boldsymbol{A}$ is trivial, then it is assured that the solution will be unique [69]. For the MRF prior used here, the second derivative spatial activity measure in the penalty functional does not have a trivial nullspace (the family of linear functionals incur identical cost) but in practice this rarely presents any difficulties. If necessary, it is possible to impose an additional regularizing term which chooses the minimum norm solution with an appropriate regularization parameter which may be driven arbitrarily close to zero. In summary therefore, under very mild conditions, the existence and uniqueness of a global minimum, corresponding to the MAP estimate $\hat{\boldsymbol{f}}_{\mathrm{MAP}}$ in Equation (8.32) can be assured and a simple gradient descent optimization [9] may be employed to compute the MAP estimate $\hat{\boldsymbol{f}}_{\mathrm{MAP}}$. Despite the large number of unknowns, the optimization is nevertheless tractable.

8.8 Examples

In this section we demonstrate the application of the superresolution restoration framework discussed earlier to the restoration of a superresolved image from a short test image sequence. We have utilized a completely synthetic simulated imaging environment in which the scene, camera and imaging geometry are all under our control.

8.8.1 Simulated Imaging Environment

The test "scene" consists of an algebraically-defined planar texture residing within the unit square of the x-y plane given by,

$$f(x, y) \propto \mathrm{rect}(x)\mathrm{rect}(y)(1 + \cos(40 \arctan(y/x))) \tag{8.33}$$

and consists of 40 radial "spokes" emanating from the origin. This image was deliberately chosen not to be bandlimited.

A short image sequence consisting of seven noninterlaced grayscale images are generated using raytracing methods which compute the images produced by a virtual camera which traverses the 3-D simulation space. The raytracing camera models the characteristics of a diffraction-limited optical system and the spatial integration performed by a focal plane sensor array, as described in Section (8.5). The intrinsic characteristics of the simulated camera system are enumerated in Table (8.2) and selected extrinsic parameters of the seven camera positions and gaze points may be found in Table (8.3). Figure (8.10) illustrates the imaging geometry with the camera positions and gaze points, as well as the seven simulated images corresponding to each camera position. Note the aliasing artifacts due to undersampling by the sensor array. Since the target scene is planar, under the pinhole camera model, the geometric relationship between observed frames is the projective transformation. Since the projective transformation is, in general, nonaffine, the observation filter relating the restoration to each of the observed frames is spatially varying.

TABLE 8.2

CAMERA INTRINSIC CHARACTERISTICS

Projection model	Ideal pinhole
Image array dimensions	128×128
Pixel dimensions	$9\mu\text{m} \times 9\mu\text{m}$
Camera focal length	10 mm
Camera f/number	2.8
Illumination wavelength	550 nm
Diffraction limit cutoff	649.351 cycles/mm
Sampling rate	111.111 samples/mm
Folding frequency	55.5556 cycles/mm

TABLE 8.3

CAMERA EXTRINSIC PARAMETERS

Camera center			Camera gaze point		
x	y	z	x	y	z
-3.0902	-5.0000	9.5106	0.0100	0.0050	0.0000
-1.0453	-5.0000	9.9452	0.0033	0.0017	0.0000
1.0453	-5.0000	9.9452	-0.0033	-0.0017	0.0000
3.0902	-5.0000	9.5106	-0.0100	-0.0050	0.0000
5.0000	-5.0000	8.6603	-0.0167	-0.0083	0.0000
6.6913	-5.0000	7.4315	-0.0233	-0.0117	0.0000
8.0902	-5.0000	5.8779	-0.0300	-0.0150	0.0000

8.8.2 Superresolution Restoration

The superresolution restoration corresponding to the last image in the sequence was computed. The restoration increases the sampling density by a factor of four in each dimension ($q_x = q_y = 4$), implying a sixteen-fold increase in the number of sample points in the restored image as compared with a single observed frame. Given that the restoration is computed from only seven observed frames, it is clear that the number of unknowns to be estimated is considerably greater than the number of linear constraints available from the observations. This immediately implies that the restoration process is ill-posed as the solution fails to be unique. This uniqueness problem is addressed through the use of the MRF prior in the Bayesian restoration framework which stabilizes the inverse problem.

In Figure (8.11) a cubic-spline interpolation of the seventh observed frame is shown, while Figure (8.12) illustrates the result of a superresolution restoration with the parameters $T = 1.5$, $\gamma = 0.05$ and the assumption of independent, identically distributed Gaussian noise with unit variance defining $\boldsymbol{K}$. A comparison of the two images clearly illustrates the considerable increase in resolution obtained using the

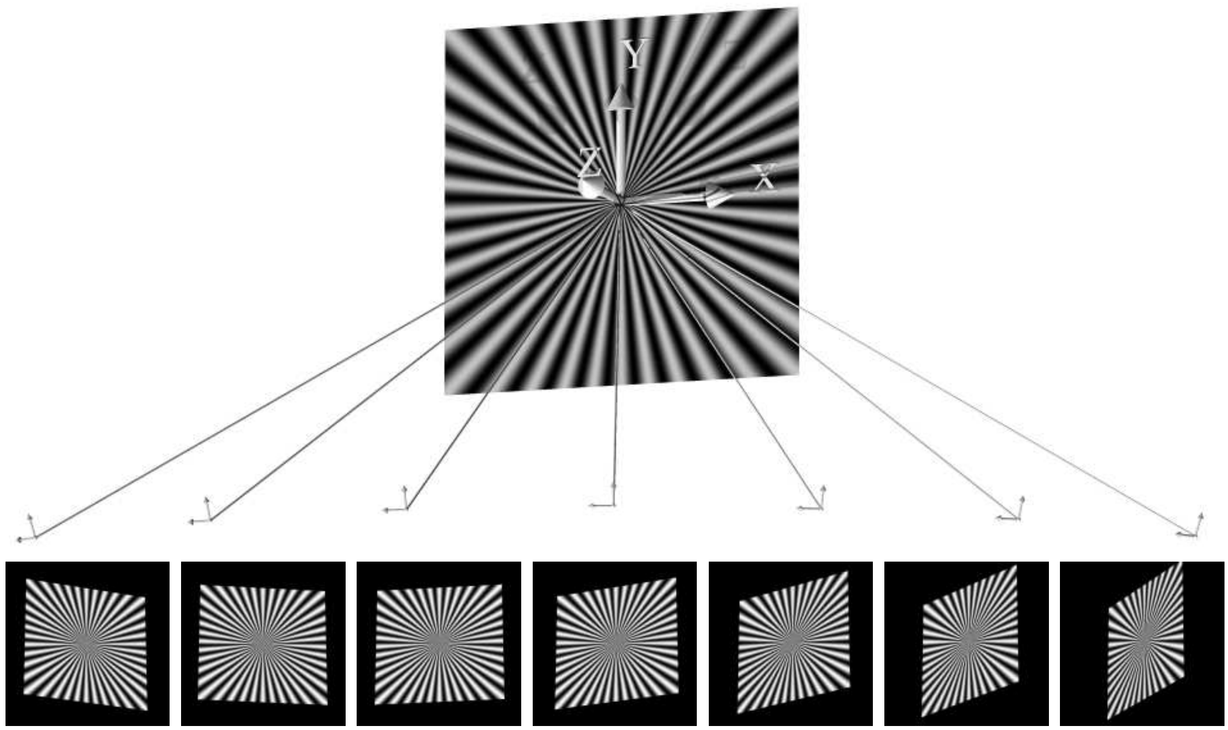


Figure 8.10. Simulated camera trajectory with the corresponding observed low-resolution images.

multiframe restoration method. Details near the center of the original image which were corrupted by aliasing have been restored in the superresolution image.

8.9 Conclusions

This chapter addressed the application of a linear observation model to the multiframe superresolution restoration problem under the assumptions of nonaffine image registration and spatially varying PSFs. Reviewing earlier work deriving from the computer graphics literature on image resampling theory, the relationship between the two seemingly disparate areas of research was established. It was shown that image resampling theory could be applied to the formulation of a linear but spatially-varying observation process that could account for nonaffine geometric transformations for image registration as well as for spatially varying degradations

Figure 8.11. Cubic spline interpolation.

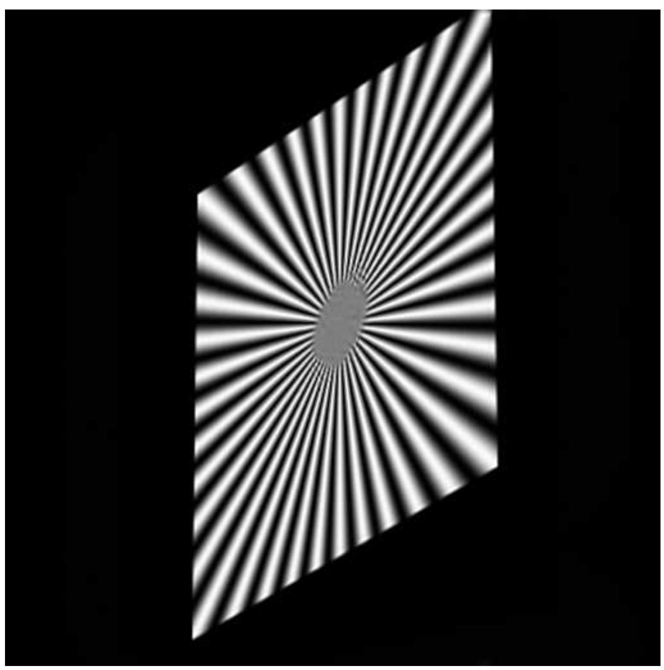


Figure 8.12. Multiframe superresolution restoration.

in the imaging system. The solution to the resulting linear inverse problem relating the degraded observations to the desired unknown restoration was tackled using a Bayesian maximum *a-posteriori* estimation formulation which utilizes prior information in the form of a Markov random field image model. The resulting optimization problem to obtain the MAP estimate was shown to be tractable due to the convexity of the objective function. The application of these methods to the restoration of a superresolved image from a short low-resolution aliased image sequence was successfully demonstrated.

CHAPTER 9

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

9.1 Summary of Research Contributions

In this section the main contributions presented in this dissertation are summarized.

The background information presented in Chapter (2), combined with the tutorial introduction to the problem of motion estimation in Chapter (4), provides a complete, yet accessible introduction for readers who are not familiar with the basic ideas and methods central to multiframe superresolution restoration.

Chapter (3) presents a comprehensive survey of the literature on multiframe superresolution restoration. This work is significant in that it covers virtually the entire literature and categorizes existing multiframe restoration methods according to their common characteristics in a systematic framework, drawing together research which utilizes a broad variety of approaches. At the time this survey was completed, no similar effort existed. None published since is as complete. A version of the survey presented in Chapter (3) and an abbreviated summary of it [18], are routinely cited by authors working in the field.

The empirical study of the performance of a class of block-matching subpixel motion estimators presented in Chapter (5) was motivated by the need to understand the performance of this commonly-used motion estimation method with an eye to the application of these methods to multiframe superresolution restoration. In particular, the study explored a framework for determining the performance which could be expected from these motion estimators since that would represent a limiting factor of the performance that could be achieved by any multiframe superresolution restoration framework based on the motion estimates. The author is not aware of any other studies which have considered the performance of motion estimators specifically in the context of multiframe superresolution restoration.

Chapter (6) presents a useful result which bridges the two fields of computer vision and computer graphics. Each field has independently developed models for the projective properties of a pinhole camera but use entirely different approaches specific to their respective requirements. In this work, it is proved that the two approaches are equivalent, with the derivations elucidating the exact relationship between these models. The discussion explores these relationships by considering the canonical decomposition of the camera matrix used by the computer vision community.

Chapter (7) represents a significant development in the area of multiframe superresolution restoration. While based on an existing Bayesian restoration framework, this work introduces for the first time temporal constraints in the restoration process. In order to achieve this, a new simultaneous multiframe superresolution restoration method was introduced, and its practicality was demonstrated. The proposed method is computationally tractable despite the extremely large number of unknown variables in the solution. The author anticipates that temporal constraints, similar to those pioneered in this work will become more commonly used in multiframe superresolution restoration frameworks.

Another novel contribution presented in this dissertation concerns the observation model relating the observed low-resolution images to the unknown superresolution restoration. In Chapter (8) it is shown how an observation matrix can be used to represent accurately the interframe motion in the image sequence as well as spatially varying PSFs. An algorithm for computing the matrix entries corresponding to a given motion model and SVPSF is presented. This work is particularly useful since it allows for the accommodation of very general nonaffine image registration and image degradations without a requirement to modify the basic restoration framework. Existing methods which utilize an observation matrix to represent a linear observation process can take advantage of the results presented in this chapter without any need to modify the restoration algorithm in any way.

9.2 Research Frontiers for Multiframe Superresolution Restoration

The problem of multiframe superresolution restoration has recently enjoyed a considerable increase in attention, mostly as a result of the potential that has been demonstrated by the existing body of work and the exciting prospects for the technology. The research presented in this dissertation has contributed to the advancement of multiframe superresolution restoration and as a secondary consequence, has elucidated various directions for future research to extend the state of the art. These include, motion estimation, observation and prior models, restoration frameworks and the theoretical foundations of superresolution restoration.

9.2.1 Motion Estimation

The importance of accurate subpixel-resolution motion information used to register the observed low-resolution frames was emphasized throughout this dissertation. Ideally, the motion estimation process should be able to estimate robustly and accurately the motion of multiple objects moving independently as is typically found in real-world scenes. Motion estimation is a difficult problem which has received considerable attention for many years, yet remains an open research topic. Recent developments, however, promise much higher performance using segmentation-based and model-based approaches, coupled with new estimation-theoretic formulations which are much more robust to noise.

One of the significant departures from the more traditional approaches to the estimation of optical flow has been the introduction of stochastic observation models. In particular the trend is toward formulations in terms of the conditional probability of the motion given the observed data [184, 183]. This has, in turn, led to the development of simultaneous motion estimation and segmentation approaches which utilize the EM framework [206, 207]. These approaches address the circular dependence of the motion estimate on the motion segmentation discussed earlier in this work. Segmentation-based methods are able to estimate the motion of multiple objects moving independently.

Chapter (5) demonstrates how the choice of a motion estimator can limit the precision of the motion vector field that can be achieved. This in turn limits the performance of any subsequent multiframe restoration process. Studies should extend the work of Chapter (5) to determine which motion estimators are best suited for multiframe superresolution restoration.

Almost no quantitative or qualitative work has been done to study the effects of motion estimation errors on superresolution restoration. While it is widely recognized that such errors lead to objectionable artifacts in the restoration, there is considerable scope for studies to characterize these artifacts. If these mechanisms can be understood, it is conceivable that detection of the restoration artifacts resulting from erroneous motion estimates could provide a means for correcting the motion estimation process. Furthermore, motion information must be estimated from the observed low-resolution image sequence which is usually undersampled and degraded. It is not known whether these adverse conditions have deleterious effects on the resulting motion estimates. This question should be addressed.

Perhaps one of the most interesting and difficult problems is that of simultaneously estimating both the motion information and the superresolution restoration. This problem has only been considered in very simple cases where the number of motion parameters is small [70, 195]. Recent developments in simultaneous motion estimation and segmentation [206, 207] may, however, be applicable to this problem.

An area with considerable potential concerns the application of kinematic models to objects segmented from the motion field. These models can be used as predictors of object motion and thus serve as prior models in regularized estimation frameworks. Such work would represent a generalization of the work presented in Chapter (7).

9.2.2 Observation and Prior Models

Since the observation model describes the relationship between the superresolution restoration and the observed data, any solution to the superresolution restoration inverse problem is limited by the accuracy of the observation model. It is therefore essential that the observation model accurately represent the processes which produce the observed data.

The linear observation model presented in Chapter (8) is sufficiently flexible to accommodate arbitrary optical system MTF characteristics. The MTF can be obtained through calibration or can be estimated from the observed image sequence itself. The latter approach is an open research problem, especially in the context of multiframe superresolution restoration. Likewise, estimating the SVPSF from the observed image sequence is an interesting research challenge.

Few of the existing multiframe superresolution restoration techniques directly address the problem of restoring multispectral or hyperspectral images. Typically superresolution methods assume single band images and make the implicit assumption that multiband images are processed one band at a time independently. This approach is suboptimal since it ignores the significant correlation that typically exists between spectral bands. This correlation can be used to advantage in the restoration process provided that the correlation relationship is faithfully represented in the observation model. The problem for color images captured using common video cameras is even more acute due to color subsampling at the FPA. This problem has received relatively little attention in the context of multiframe superresolution restoration with the exception of [178, 179, 54]. Addressing this problem requires (inter alia) that the characteristics of the color subsampling process be taken into account in the observation model.

Recently there has been growing interest in applying multiframe superresolution restoration methods to compressed video, in particular to streams compressed using the MPEG and ITU video compression standards [91, 92, 93, 94, 95]. To achieve this, it is necessary that the observation model accurately reflect the processes which yield the compressed stream. These include motion compensation, transform coding, color subsampling and quantization. Some examples of superresolution restoration applied to compressed video include [35, 53, 127, 128, 172, 66, 5, 6, 171]. It is anticipated that much more work will be done in this area given the demand for high-quality upsampling of Standard Definition Television (SDTV) content to High Definition Television (HDTV).

"Blind" restoration refers to methods which estimate the degradation process in addition to computing the restoration. This is in contrast to the usual methods which assume that the degradation process is known. Almost no work has been done in this area with the exception of [138, 139]. Developments in blind superresolution restoration would be very useful since the imaging degradation process is not typically known and must therefore be estimated or approximate assumptions must be made.

Careful choice of the image prior model can result in impressive improvements in restoration quality [10] and further developments in this area should be pursued.

The prior models used in this dissertation are *generic* in the sense that they are appropriate for an extremely broad class of images. It is, however, possible to use domain specific knowledge to formulate prior constraints which are highly specific to a given application. Examples include restoration of faces, text, astronomical objects and so on. By modeling these characteristics and using the model as a prior in the restoration process, the quality of the restoration can be improved dramatically. For example, "recognition-based" priors have been used to achieve impressive restoration of faces from highly undersampled image sequences [8]. Given the wide applicability of multiframe superresolution restoration methods, it is likely that the development of application-specific prior models will be an important research topic.

Along these lines, researchers have considered the use of trained models to represent prior image constraints [12, 140]. Knowing that the solution images fall within some class (such as text, faces, machined parts and so on), it is possible to estimate the parameters of the model using a corpus of data drawn from that class. The model is then incorporated in the restoration framework to provide *a-priori* constraints.

The key innovation of the simultaneous multiframe superresolution restoration framework introduced in Chapter (7) is the inclusion of temporal as well as spatial constraints in the restoration process. It remains, however, to explore new temporal models which extend the work presented here.

9.2.3 Restoration Frameworks

In this work, spatial-domain Bayesian methods were favored as the framework for formulating and solving the ill-posed superresolution restoration inverse problem. The Bayesian approach has gained prominence in an enormous range of applications as it provides a convenient and powerful framework for the inclusion of prior knowledge as well as for modeling dependence among related variables. It is anticipated that Bayesian methods will continue to be prominent in future developments in multiframe superresolution restoration [195].

In the review of multiframe superresolution restoration methods in Chapter (3), formations based on POCS featured prominently. In that chapter the significant advantages of these methods were highlighted. It is likely that projected constraints will be useful in hybrid methods which combine the best characteristics of POCS with those of the estimation-theoretic frameworks.

In Chapter (7) restoration was achieved by simultaneously estimating multiple superresolved images with temporal constraints applied in the restoration process. While this "batch processing" approach is successful in its own right, it also suggests an "online" or "sequential" method in which superresolved frames are computed as additional low-resolution observed data becomes available. Recently, sequential Monte Carlo methods, such as particle filtering, have attracted much attention. These model-based stochastic sampling methods provide a framework for estimation on multimodal conditional densities in the presence of deterministic drift. Since these methods are designed for sequential or time-evolving stochastic inference problems, they may be suitable for tackling the multiframe superresolution restoration problem.

Recently, especially in the context of graphical models [99], methods for approximate inference have been developed to address situations where exact inference is impossible or impractical. These methods may prove useful in multiframe superresolution restoration where the number of variables is very large, where observation and image models may be complex and exact inference may not be strictly necessary.

Earlier, in the discussion on potential developments in terms of observation models, the problem of blind restoration was mentioned. Blind multiframe superresolution restoration is likely to require extensions to existing restoration frameworks.

Similarly, in the discussion of possible developments in motion estimation the issue of simultaneous motion estimation and restoration was raised. Such an approach will likewise require further development of the standard restoration frameworks.

A hierarchy of image and video restoration techniques was presented in Chapter (2) and is illustrated in Figure (2.2). While this dissertation addresses spatial resolution enhancement exclusively, there remains scope for temporal as well as spatiotemporal resolution enhancement. Image sequence interpolation methods increase the frame rate by inserting new frames at time instants between the existing observed frames. The inserted frames are constructed from the existing data. So far, almost no work has been done to combine multiframe superresolution restoration methods which increase spatial resolution with these temporal interpolation methods.

Formulating multiframe superresolution restoration as a Bayesian MAP statistical inference problem leads to an optimization problem to find the mode of the posterior pdf. In this work, both the log-likelihood function derived from the observation model and the prior model were deliberately chosen to be convex so that the resulting optimization problem would have a unique solution which can be easily computed. Though the results achieved using this approach are impressive, the choice of convex models is quite restrictive. More sophisticated observation and prior models tend to have a multimodal pdf which result in a posterior density with local extrema. Consequently, advancements in multiframe superresolution restoration resulting from the use of more complex models will likely necessitate the use of more sophisticated optimization frameworks.

Undoubtedly, as the methods developed in this work and in the literature mature, a demand will emerge for hardware implementations delivering real-time performance for commercial products. Some of the applications of the technology which could benefit from fast hardware implementations are discussed in Chapter (1).

9.2.4 Theoretical Foundations

Despite the growing interest in the field of multiframe superresolution restoration, there has been surprisingly little work published on the theoretical foundations of these methods. Some early bounds on superresolution performance were given in [175, 173]. Some more recent theoretical results pertaining to the maximum achievable resolution increase are presented in [8]. A recent review of the standard mathematical methods used in the field may be found in [137]. On the whole, this is a wide open research area.

9.3 Conclusions

The future of multiframe restoration methods appears very bright indeed. Considering the expansive literature on classical single frame restoration methods born of literally decades of continuing interest in the restoration problem, there is every reason to believe that the multiframe extension will receive considerable attention in the research community. Apart from the obvious academic interest in extending existing methods in a new direction, the range of applications to which the technology can be applied is very broad.

The multiframe superresolution methods discussed in this work mostly involve computational post-processing of the observed low-resolution image sequence. This is extremely attractive since the performance of existing imaging systems can be enhanced without any requirement to modify these systems. Computational resources are becoming progressively more powerful and cheaper as microprocessors continue to follow Moore's law. This makes it feasible to implement algorithms which were previously prohibitive in terms of their computational complexity. The likely result will be the development of more accurate motion estimation methods, more realistic observation and prior models and more powerful restoration frameworks leading ultimately to improvements in restored image quality.

APPENDIX A

THE EXPECTATION MAXIMIZATION ALGORITHM

This appendix discusses the Expectation Maximization (EM) algorithm of Dempster, Laird and Rubin [42]. The approach taken follows that of an unpublished note by Stuart Russel. In order to ensure that the presentation is reasonably self-contained, some of the results on which the derivation of the algorithm is based are presented prior to the presentation of the main results. The EM algorithm has become a popular tool in statistical estimation problems involving incomplete data, or in problems which can be posed in a similar form, such as mixture estimation [129, 130]. The EM algorithm has also been used in various motion estimation frameworks [207] and variants of it have been used in multiframe superresolution restoration methods which combine motion estimation along the lines of [70].

A.1 Convex Functions

Definition 1 *Let f be a real valued function defined on an interval $I = [a, b]$. f is said to be* convex *on I if $\forall x_1, x_2 \in I, \lambda \in [0, 1]$,*

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

f is said to be strictly convex *if the inequality is strict. Intuitively, this definition states that the function falls below (strictly convex) or is never above (convex) the straight line (the secant) from points $(x_1, f(x_1))$ to $(x_2, f(x_2))$.*

Definition 2 *f is concave (strictly concave) if $-f$ is convex (strictly convex).*

Theorem 1 *If $f(x)$ is twice differentiable on $[a, b]$ and $f''(x) \geq 0$ on $[a, b]$ then $f(x)$ is convex on $[a, b]$.*

Proof: For $x \leq y \in [a,b]$ and $\lambda \in [0,1]$ let $z = \lambda y + (1-\lambda)x$. By definition, f is convex iff $f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda)f(x)$. Writing $z = \lambda y + (1-\lambda)x$, and noting that $f(z) = \lambda f(z) + (1-\lambda)f(z)$ we have that $f(z) = \lambda f(z) + (1-\lambda)f(z) \leq \lambda f(y) + (1-\lambda)f(x)$. By rearranging terms an equivalent definition for convexity, can be arrived at, namely, f is convex iff

$$\lambda\,[f(y) - f(z)] \geq (1-\lambda)\,[f(z) - f(x)] \tag{A.1}$$

By the mean value theorem, $\exists s, x \leq s \leq z$ s.t.

$$f(z) - f(x) = f'(s)(z - x) \tag{A.2}$$

Similarly, applying the mean value theorem to $f(y) - f(z)$, $\exists t, z \leq t \leq y$ s.t.

$$f(y) - f(z) = f'(t)(y - z) \tag{A.3}$$

Thus we have the situation, $x \leq s \leq z \leq t \leq y$. By assumption, $f''(x) \geq 0$ on $[a,b]$ thus

$$f'(s) \leq f'(t) \;\; \text{since } s \leq t. \tag{A.4}$$

We may rewrite $z = \lambda y + (1-\lambda)z$ in the form

$$(1-\lambda)(z - x) = \lambda(y - z). \tag{A.5}$$

Finally, combining the above we have,

$$\begin{aligned}
(1-\lambda)[f(z) - f(x)] &= (1-\lambda)f'(s)(z-x) && \text{by Equation (A.2)} \\
&\leq f'(t)(1-\lambda)(z-x) && \text{by Equation (A.4)} \\
&= \lambda f'(t)(y-z) && \text{by Equation (A.5)} \\
&= \lambda[f(y) - f(z)] && \text{by Equation (A.3).}
\end{aligned}$$

■

Proposition 1 $-\ln(x)$ *is strictly convex on* $(0, \infty)$.

Proof: With $f(x) = -ln(x)$, we have $f''(x) = \frac{1}{x^2} > 0$ for $x \in (0, \infty)$. By Theorem (1), $-\ln(x)$ is strictly convex on $(0, \infty)$. Also, by Definition (2) $\ln(x)$ is strictly concave on $(0, \infty)$. ■

The notion of convexity can be extended to apply to n points. This result is known as Jensen's inequality.

Theorem 2 (Jensen's inequality) *Let f be a convex function defined on an interval I. If $x_1, x_2, \ldots, x_n \in I$ and $\lambda_1, \lambda_2, \ldots, \lambda_n \geq 0$ with $\sum_{i=1}^{n} \lambda_i = 1$,*

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i)$$

Proof: For $n = 1$ this is trivial. The case $n = 2$ corresponds to the definition of convexity. To show that this is true for all natural numbers, we proceed by induction. Assume the theorem is true for some n then,

$$\begin{aligned}
f\left(\sum_{i=1}^{n+1} \lambda_i x_i\right) &= f\left(\lambda_{n+1} x_{n+1} + \sum_{i=1}^{n} \lambda_i x_i\right) \\
&= f\left(\lambda_{n+1} x_{n+1} + (1 - \lambda_{n+1}) \frac{1}{1 - \lambda_{n+1}} \sum_{i=1}^{n} \lambda_i x_i\right) \\
&\leq \lambda_{n+1} f\left(x_{n+1}\right) + (1 - \lambda_{n+1}) f\left(\frac{1}{1 - \lambda_{n+1}} \sum_{i=1}^{n} \lambda_i x_i\right) \\
&= \lambda_{n+1} f\left(x_{n+1}\right) + (1 - \lambda_{n+1}) f\left(\sum_{i=1}^{n} \frac{\lambda_i}{1 - \lambda_{n+1}} x_i\right) \\
&\leq \lambda_{n+1} f\left(x_{n+1}\right) + (1 - \lambda_{n+1}) \sum_{i=1}^{n} \frac{\lambda_i}{1 - \lambda_{n+1}} f\left(x_i\right) \\
&= \lambda_{n+1} f\left(x_{n+1}\right) + \sum_{i=1}^{n} \lambda_i f\left(x_i\right) \\
&= \sum_{i=1}^{n+1} \lambda_i f\left(x_i\right)
\end{aligned}$$

■

Since $\ln(x)$ is concave, we may apply Jensen's inequality to obtain the useful result,

$$\ln \sum_{i=1}^{n} \lambda_i x_i \geq \sum_{i=1}^{n} \lambda_i \ln(x_i). \tag{A.6}$$

This allows us to lower-bound a logarithm of a sum, a result that is used in the derivation of the EM algorithm.

Jensen's inequality provides a simple proof that the arithmetic mean is greater than or equal to the geometric mean.

Proposition 2

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \sqrt[n]{x_1 x_2 \cdots x_n}.$$

Proof: If $x_1, x_2, \ldots, x_n \geq 0$ then, since $\ln(x)$ is concave we have

$$\begin{aligned} \ln\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) &\geq \sum_{i=1}^{n} \frac{1}{n}\ln(x_i) \\ &= \frac{1}{n}\ln(x_1 x_2 \cdots x_n) \\ &= \ln(x_1 x_2 \cdots x_n)^{\frac{1}{n}} \end{aligned}$$

Thus, we have

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \sqrt[n]{x_1 x_2 \cdots x_n}$$

■

A.2 The Expectation-Maximization Algorithm

The EM algorithm is an efficient iterative procedure to compute the ML estimate in the presence of missing or hidden data. In ML estimation, we wish to estimate the model parameter(s) for which the observed data are the most likely.

Each iteration of the EM algorithm consists of two processes: The E-step, and the M-step. In the expectation, or E-step, the missing data are estimated given the observed data and current estimate of the model parameters. This is achieved using the conditional expectation, explaining the choice of terminology. In the M-step, the likelihood function is maximized under the assumption that the missing data are known. The estimate of the missing data from the E-step are used in lieu of the actual missing data.

Convergence is assured since the algorithm is guaranteed to increase the likelihood at each iteration.

A.2.1 Derivation of the EM-algorithm

Let $\mathbf{X}$ be random vector which results from a parameterized family. We wish to find θ such that $\mathcal{P}(\mathbf{X}|\theta)$ is a maximum. This is known as the Maximum Likelihood (ML) estimate for θ. In order to estimate θ, it is typical to introduce the *log likelihood function* defined as,

$$L(\theta) = \ln \mathcal{P}(\mathbf{X}|\theta). \tag{A.7}$$

The likelihood function is considered to be a function of the parameter θ given the data $\mathbf{X}$. Since $\ln(x)$ is a strictly increasing function, the value of θ which maximizes $\mathcal{P}(\mathbf{X}|\theta)$ also maximizes $L(\theta)$.

The EM algorithm is an iterative procedure for maximizing $L(\theta)$. Assume that after the n^{th} iteration the current estimate for θ is given by θ_n. Since the objective is to maximize $L(\theta)$, we wish to compute an updated estimate θ such that,

$$L(\theta) > L(\theta_n) \tag{A.8}$$

Equivalently we want to maximize the difference,

$$L(\theta) - L(\theta_n) = \ln \mathcal{P}(\mathbf{X}|\theta) - \ln \mathcal{P}(\mathbf{X}|\theta_n). \tag{A.9}$$

So far, we have not considered any unobserved or missing variables. In problems where such data exist, the EM algorithm provides a natural framework for their inclusion. Alternately, hidden variables may be introduced purely as an artifice for making the maximum likelihood estimation of θ tractable. In this case, it is assumed that knowledge of the hidden variables will make the maximization of the likelihood function easier. Either way, denote the hidden random vector by $\mathbf{Z}$ and a given realization by $\mathbf{z}$. The total probability $\mathcal{P}(\mathbf{X}|\theta)$ may be written in terms of the hidden variables $\mathbf{z}$ as,

$$\mathcal{P}(\mathbf{X}|\theta) = \sum_{\mathbf{z}} \mathcal{P}(\mathbf{X}|\mathbf{z}, \theta)\mathcal{P}(\mathbf{z}|\theta). \tag{A.10}$$

We may then rewrite Equation (A.9) as,

$$L(\theta) - L(\theta_n) = \ln \sum_{\mathbf{z}} \mathcal{P}(\mathbf{X}|\mathbf{z}, \theta)\mathcal{P}(\mathbf{z}|\theta) - \ln \mathcal{P}(\mathbf{X}|\theta_n). \tag{A.11}$$

Notice that this expression involves the logarithm of a sum. In Section (A.1) using Jensen's inequality, it was shown that,

$$\ln \sum_{i=1}^{n} \lambda_i x_i \geq \sum_{i=1}^{n} \lambda_i \ln(x_i)$$

for constants $\lambda_i \geq 0$ with $\sum_{i=1}^{n} \lambda_i = 1$. This result may be applied to Equation (A.11) which involves the logarithm of a sum provided that the constants λ_i can be identified. Consider letting the constants be of the form $\mathcal{P}(\mathbf{z}|\mathbf{X}, \theta_n)$. Since $\mathcal{P}(\mathbf{z}|\mathbf{X}, \theta_n)$ is a probability measure, we have that $\mathcal{P}(\mathbf{z}|\mathbf{X}, \theta_n) \geq 0$ and that $\sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X}, \theta_n) = 1$ as required.

Then starting with Equation (A.11) the constants $\mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n)$ are introduced as,

$$\begin{aligned}
L(\theta) - L(\theta_n) &= \ln \sum_{\mathbf{z}} \mathcal{P}(\mathbf{X}|\mathbf{z},\theta)\mathcal{P}(\mathbf{z}|\theta) - \ln \mathcal{P}(\mathbf{X}|\theta_n) \\
&= \ln \sum_{\mathbf{z}} \mathcal{P}(\mathbf{X}|\mathbf{z},\theta)\mathcal{P}(\mathbf{z}|\theta) \cdot \frac{\mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n)}{\mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n)} - \ln \mathcal{P}(\mathbf{X}|\theta_n) \\
&= \ln \sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n) \left(\frac{\mathcal{P}(\mathbf{X}|\mathbf{z},\theta)\mathcal{P}(\mathbf{z}|\theta)}{\mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n)} \right) - \ln \mathcal{P}(\mathbf{X}|\theta_n) \\
&\geq \sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n) \ln \left(\frac{\mathcal{P}(\mathbf{X}|\mathbf{z},\theta)\mathcal{P}(\mathbf{z}|\theta)}{\mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n)} \right) - \ln \mathcal{P}(\mathbf{X}|\theta_n) \\
&= \sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n) \ln \left(\frac{\mathcal{P}(\mathbf{X}|\mathbf{z},\theta)\mathcal{P}(\mathbf{z}|\theta)}{\mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n)\mathcal{P}(\mathbf{X}|\theta_n)} \right) \\
&\triangleq \Delta(\theta|\theta_n) \qquad \text{(A.12)}
\end{aligned}$$

Equivalently we may write,

$$L(\theta) \geq L(\theta_n) + \Delta(\theta|\theta_n) \qquad \text{(A.13)}$$

and for convenience define,

$$l(\theta|\theta_n) \triangleq L(\theta_n) + \Delta(\theta|\theta_n)$$

so that the relationship in Equation (A.13) can be made explicit as,

$$L(\theta) \geq l(\theta|\theta_n).$$

We have now a function, $l(\theta|\theta_n)$ which is bounded above by the likelihood function $L(\theta)$. Additionally, observe that,

$$\begin{aligned}
l(\theta_n|\theta_n) &= L(\theta_n) + \Delta(\theta_n|\theta_n) \\
&= L(\theta_n) + \sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n) \ln \frac{\mathcal{P}(\mathbf{X}|\mathbf{z},\theta_n)\mathcal{P}(\mathbf{z}|\theta_n)}{\mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n)\mathcal{P}(\mathbf{X}|\theta_n)} \\
&= L(\theta_n) + \sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n) \ln \frac{\mathcal{P}(\mathbf{X},\mathbf{z}|\theta_n)}{\mathcal{P}(\mathbf{X},\mathbf{z}|\theta_n)} \\
&= L(\theta_n) + \sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n) \ln 1 \\
&= L(\theta_n), \qquad \text{(A.14)}
\end{aligned}$$

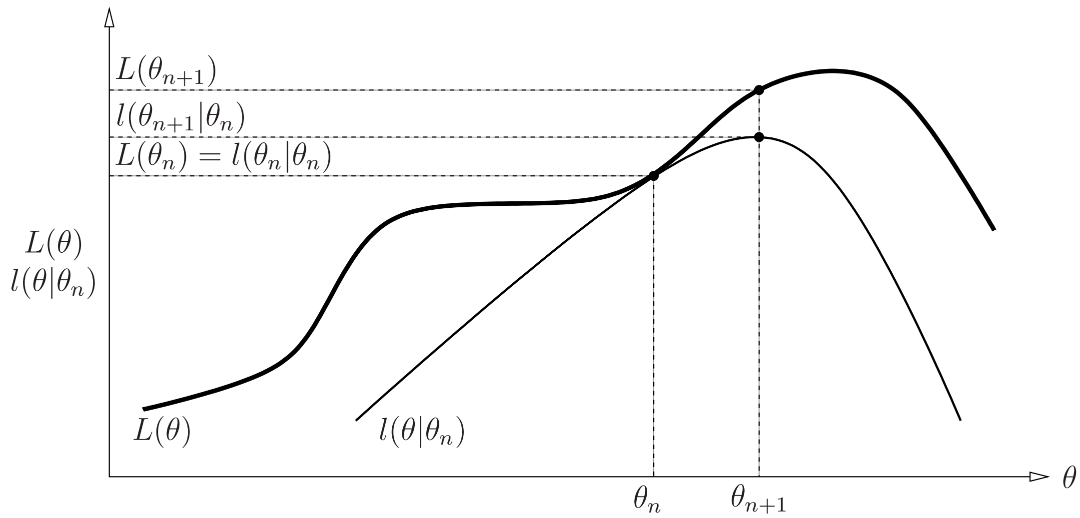


Figure A.1. Graphical interpretation of a single iteration of the EM algorithm: The function $L(\theta|\theta_n)$ is upper-bounded by the likelihood function $L(\theta)$. The functions are equal at $\theta = \theta_n$. The EM algorithm chooses θ_{n+1} as the value of θ for which $l(\theta|\theta_n)$ is a maximum. Since $L(\theta) \geq l(\theta|\theta_n)$ increasing $l(\theta|\theta_n)$ ensures that the value of the likelihood function $L(\theta)$ is increased at each step.

so for $\theta = \theta_n$ the functions $l(\theta|\theta_n)$ and $L(\theta)$ are equal.

Our objective is to choose a values of θ so that $L(\theta)$ is maximized. We have shown that the function $l(\theta|\theta_n)$ is bounded above by the likelihood function $L(\theta)$ and that the value of the functions $l(\theta|\theta_n)$ and $L(\theta)$ are equal at the current estimate for $\theta = \theta_n$. Therefore, any θ which increases $l(\theta|\theta_n)$ in turn increase the $L(\theta)$. In order to achieve the greatest possible increase in the value of $L(\theta)$, the EM algorithm calls for selecting θ such that $l(\theta|\theta_n)$ is maximized. We denote this updated value as θ_{n+1}. This process is illustrated in Figure (A.1).

Formally we have,

$$
\begin{aligned}
\theta_{n+1} &= \arg\max_{\theta} \{l(\theta|\theta_n)\} \\
&= \arg\max_{\theta} \left\{ L(\theta_n) + \sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n) \ln \frac{\mathcal{P}(\mathbf{X}|\mathbf{z},\theta)\mathcal{P}(\mathbf{z}|\theta)}{\mathcal{P}(\mathbf{X}|\theta_n)\mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n)} \right\} \\
& \quad \text{Now drop terms which are constant w.r.t. } \theta \\
&= \arg\max_{\theta} \left\{ \sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n) \ln \mathcal{P}(\mathbf{X}|\mathbf{z},\theta)\mathcal{P}(\mathbf{z}|\theta) \right\} \\
&= \arg\max_{\theta} \left\{ \sum_{\mathbf{z}} \mathcal{P}(\mathbf{z}|\mathbf{X},\theta_n) \ln \mathcal{P}(\mathbf{X},\mathbf{z}|\theta) \right\} \\
&= \arg\max_{\theta} \left\{ \mathrm{E}_{\mathbf{Z}|\mathbf{X},\theta_n} \{ \ln \mathcal{P}(\mathbf{X},\mathbf{z}|\theta) \} \right\} \qquad \text{(A.15)}
\end{aligned}
$$

In Equation (A.15) the expectation and maximization steps are apparent. The EM algorithm thus consists of iterating the:

1. *E-step*: Determine the conditional expectation $\mathrm{E}_{\mathbf{Z}|\mathbf{X},\theta_n}\{\ln \mathcal{P}(\mathbf{X},\mathbf{z}|\theta)\}$

2. *M-step*: Maximize this expression with respect to θ.

At this point it is fair to ask what has been gained given that we have simply traded the maximization of $L(\theta)$ for the maximization of $l(\theta|\theta_n)$. The answer lies in the fact that $l(\theta|\theta_n)$ takes into account the unobserved or missing data $\mathbf{Z}$. In the case where we wish to estimate these variables the EM algorithms provides a framework for doing so. Also, as alluded to earlier, it may be convenient to introduce such hidden variables so that the maximization of $L(\theta|\theta_n)$ is simplified given knowledge of the hidden variables. (as compared with a direct maximization of $L(\theta)$)

A.2.2 Convergence of the EM Algorithm

The convergence properties of the EM algorithm are discussed in detail by McLachlan and Krishnan [129]. In this section we discuss the general convergence of the algorithm. Recall that θ_{n+1} is the estimate for θ which maximizes the difference $\Delta(\theta|\theta_n)$. Starting with the current estimate for θ, that is, θ_n we had that $\Delta(\theta_n|\theta_n) = 0$. Since θ_{n+1} is chosen to maximize $\Delta(\theta|\theta_n)$, we then have that $\Delta(\theta_{n+1}|\theta_n) \geq \Delta(\theta_n|\theta_n) = 0$, so for each iteration the likelihood $L(\theta)$ is nondecreasing.

When the algorithm reaches a fixed point for some θ_n the value θ_n maximizes $l(\theta)$. Since L and l are equal at θ_n if L and l are differentiable at θ_n, then θ_n must be a stationary point of L. The stationary point need not, however, be a local maximum. In [129] it is shown that it is possible for the algorithm to converge to local minima or saddle points in unusual cases.

A.2.3 The Generalized EM Algorithm

In the formulation of the EM algorithm described above, θ_{n+1} was chosen as the value of θ for which $\Delta(\theta|\theta_n)$ was maximized. While this ensures the greatest increase in $L(\theta)$, it is however possible to relax the requirement of maximization to one of simply increasing $\Delta(\theta|\theta_n)$ so that $\Delta(\theta_{n+1}|\theta_n) \geq \Delta(\theta_n|\theta_n)$. This approach, to simply increase and not necessarily maximize $\Delta(\theta_{n+1}|\theta_n)$ is known as the Generalized Expectation Maximization (GEM) algorithm and is often useful in cases where the maximization is difficult. The convergence of the GEM algorithm can be argued as above.

APPENDIX B

THE FOURIER TRANSFORM OF UNIFORM TRANSLATIONAL MOTION

B.1 Uniform Translational Motion

Let $I_0(x, y)$ denote an intensity image with continuous spatial variables x, y. Assume that the intensity image undergoes uniform translational motion with horizontal velocity component V_x and vertical velocity component V_y. The resulting spatiotemporal intensity function $I(x, y, t)$ is related to the still frame $I_0(x, y)$ according to,

$$I(x, y, t) = I_0(x - V_x t - x_0, y - V_y t - y_0), \tag{B.1}$$

where x_0, y_0 denote the horizontal and vertical displacement of $I(x, y)$ at $t = 0$. For convenience it is assumed that $x_0 = y_0 = 0$ so,

$$I(x, y, t) = I_0(x - V_x t, y - V_y t). \tag{B.2}$$

B.2 The Fourier Domain Representation of Uniform Translational Motion

The simple case of uniform translational motion admits a closed-form spatiotemporal frequency domain description which provides useful insight into the motion problem. Let ω_x, ω_y and ω_t denote the two spatial frequency and one temporal frequency variable respectively. We seek an expression for $\hat{I}(\omega_x, \omega_y, \omega_t) = \mathcal{F}\{I(x, y, t)\}$, where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform operation.

From Equation (B.2) we have $I(x, y, t) = I_0(x - V_x t, y - V_y t)$ so,

$$
\begin{aligned}
\hat{I}(\omega_x, \omega_y, \omega_t) &= \mathcal{F}\{I(x, y, t)\} \\
&= \iiint_{-\infty}^{+\infty} I(x, y, t)\; e^{-i(\omega_x x + \omega_y y + \omega_t t)}\; \mathrm{d}x\; \mathrm{d}y\; \mathrm{d}t \\
&= \iiint_{-\infty}^{+\infty} I_0(x - V_x t, y - V_y t)\; e^{-i(\omega_x x + \omega_y y + \omega_t t)}\; \mathrm{d}x\; \mathrm{d}y\; \mathrm{d}t \\
& \text{Letting } u = x - V_x t \text{ and } v = y - V_y t, \\
&= \iiint_{-\infty}^{+\infty} I_0(u, v)\; e^{-i(\omega_x(u + V_x t) + \omega_y(v + V_y t) + \omega_t t)}\; \mathrm{d}u\; \mathrm{d}v\; \mathrm{d}t \\
&= \iiint_{-\infty}^{+\infty} I_0(u, v)\; e^{-i(\omega_x u + \omega_x V_x t + \omega_y v + \omega_y V_y t + \omega_t t)}\; \mathrm{d}u\; \mathrm{d}v\; \mathrm{d}t \\
&= \int_{-\infty}^{+\infty} \left\{ \iint_{-\infty}^{+\infty} I_0(u, v)\; e^{-i(\omega_x u + \omega_y v)}\; \mathrm{d}u\; \mathrm{d}v \right\} e^{-i(\omega_x V_x t + \omega_y V_y t + \omega_t t)}\; \mathrm{d}t \\
&= \int_{-\infty}^{+\infty} \hat{I}_0(\omega_x, \omega_y) e^{-i(\omega_x V_x t + \omega_y V_y t + \omega_t t)}\; \mathrm{d}t \\
&= \hat{I}_0(\omega_x, \omega_y) \int_{-\infty}^{+\infty} e^{-i(\omega_x V_x + \omega_y V_y) t} e^{-i\omega_t t}\; \mathrm{d}t \\
&= \hat{I}_0(\omega_x, \omega_y)\, \delta(\omega_x V_x + \omega_y V_y + \omega_t).
\end{aligned} \tag{B.3}
$$

Equation (B.3) is a product of two terms:

- $\hat{I}_0(\omega_x, \omega_y)$

  This is the Fourier transform of the still image $I_0(x, y)$. In the spatiotemporal frequency space $\omega_x, \omega_y, \omega_t$, $\hat{I}(\omega_x, \omega_y)$ is constant for all ω_t.

- $\delta(\omega_x V_x + \omega_y V_y + \omega_t)$

  This term describes a plane in spatiotemporal frequency space. The orientation of the plane in spatiotemporal frequency space is determined by the horizontal and vertical motion velocities. The normal vector of the plane is given by $[V_x, V_y, 1]^\top$.

Thus the spectral energy of the spatiotemporal intensity $I(x, y, t)$ is constrained to lie in a plane in spatiotemporal frequency space. In particular, the product in

Equation (B.3) constrains the spectral energy to the plane

$$[\omega_x, \omega_y, \omega_t][V_x, V_y, 1]^\top = \omega_x V_x + \omega_y V_y + \omega_t = 0,$$

with the value at a point $\omega_x, \omega_y, \omega_t$ in the plane given by the value of $\hat{I}_0(\omega_x, \omega_y)$.

B.2.1 Temporal Aliasing

One interesting consequence of the Fourier domain expression for uniform translational motion is that an image, spatially sampled so as to satisfy Nyquist's criterion can result in frequency domain aliasing when the signal undergoes translational motion.

B.3 Selected Fourier Transform Properties

B.3.1 Differentiation Properties

Let $f(x)$ and $f(\omega)$ be a Fourier transform pair, then we may express $f(x)$ in terms of $F(\omega)$ using the inverse Fourier transform relation,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{iwx} \, \mathrm{d}\omega.$$

Differentiating both sides yields,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x} \left(f(x) \right) &= \frac{\mathrm{d}}{\mathrm{d}x} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{iwx} \, \mathrm{d}\omega \right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} i\omega F(\omega) e^{iwx} \, \mathrm{d}\omega, \end{aligned}$$

from which we note that,

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) \Longleftrightarrow i\omega F(\omega). \tag{B.4}$$

Similarly, starting with the Fourier transform formula,

$$F(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} \mathrm{d}x$$

and differentiating with respect to ω yields the result,

$$\frac{\mathrm{d}}{\mathrm{d}\omega}F(\omega) \Longleftrightarrow -ixf(x) \tag{B.5}$$

or equivalently, $i\frac{\mathrm{d}}{\mathrm{d}\omega}F(\omega) \Longleftrightarrow xf(x)$.

In general, with $\boldsymbol{x}, \boldsymbol{\omega} \in \mathbb{R}^N$, if $f(\boldsymbol{x}) \Longleftrightarrow F(\boldsymbol{\omega})$ represents a Fourier transform pair then, for $k \in \{1, 2, \ldots, N\}$ and $p_k \in \{0, 1, \ldots, N\}$ such that

$$1 \leq \sum p_k \triangleq \sum_{k=1}^{N} p_k \leq N$$

then,

$$\frac{\partial^{\Sigma p_k}}{\prod_{k=1}^{N} \partial x_k^{p_k}} f(\boldsymbol{x}) \Longleftrightarrow \prod_{k=1}^{N} (i\omega_k)^{p_k} F(\boldsymbol{\omega}) \tag{B.6}$$

and

$$\frac{\partial^{\Sigma p_k}}{\prod_{k=1}^{N} \partial \omega_k^{p_k}} F(\boldsymbol{\omega}) \Longleftrightarrow \prod_{k=1}^{N} (-ix_k)^{p_k} f(\boldsymbol{x}). \tag{B.7}$$

B.3.2 Directional Derivatives

It is also useful to determine Fourier transform relations for directional derivatives. Let $f(\boldsymbol{x}) \Longleftrightarrow F(\boldsymbol{\omega})$ represent a Fourier transform pair with $\boldsymbol{x}, \boldsymbol{\omega} \in \mathbb{R}^N$. Let $\hat{\boldsymbol{u}}$ be a unit vector in $\mathbb{R}^N$ and let $\mathcal{D}_{\hat{\boldsymbol{u}}}$ denote the derivative operator in the direction $\hat{\boldsymbol{u}}$. With $\mathcal{F}\{\cdot\}$ representing the Fourier transform we have,

$$\begin{aligned}
\mathcal{F}\{\mathcal{D}_{\hat{\boldsymbol{u}}} f(\boldsymbol{x})\} &= \mathcal{F}\{[\hat{\boldsymbol{u}} \cdot \nabla] f(\boldsymbol{x})\} \\
&= \mathcal{F}\left\{\sum_{k=1}^{N} u_k \frac{\partial}{\partial x_k} f(\boldsymbol{x})\right\} \\
&= \sum_{k=1}^{N} u_k\, \mathcal{F}\left\{\frac{\partial}{\partial x_k} f(\boldsymbol{x})\right\} \\
&= \sum_{k=1}^{N} u_k \{i w_k F(\boldsymbol{\omega})\} \\
&= i[\hat{\boldsymbol{u}} \cdot \boldsymbol{\omega}] F(\boldsymbol{\omega}).
\end{aligned}$$

We thus we conclude that,

$$\mathcal{D}_{\hat{\boldsymbol{u}}} f(\boldsymbol{x}) \Longleftrightarrow i[\hat{\boldsymbol{u}} \cdot \boldsymbol{\omega}] F(\boldsymbol{\omega}). \tag{B.8}$$

We arrive at a similar relation for the directional derivative in the Fourier domain as,

$$\mathcal{D}_{\hat{\boldsymbol{u}}} F(\boldsymbol{\omega}) \Longleftrightarrow -i[\hat{\boldsymbol{u}} \cdot \boldsymbol{x}] f(\boldsymbol{x}). \tag{B.9}$$

For higher order directional derivatives, we note that $\mathcal{D}^n_{\hat{\boldsymbol{u}}} = [\hat{\boldsymbol{u}} \cdot \nabla]^n$ so that,

$$\begin{aligned} \mathcal{F}\left\{\mathcal{D}^n_{\hat{\boldsymbol{u}}} f(\boldsymbol{x})\right\} &= \mathcal{F}\left\{[\hat{\boldsymbol{u}} \cdot \nabla]^n f(\boldsymbol{x})\right\} \\ &= \mathcal{F}\left\{\left[\sum_{k=1}^{N} u_k \frac{\partial}{\partial x_k}\right]^n f(\boldsymbol{x})\right\}. \end{aligned}$$

Here the expression in the brackets is an n^{th} order polynomial containing all combinations of mixed partial derivatives. Using Equation (B.6) the Fourier transform of the product of each of these terms with $f(\boldsymbol{x})$ may be expressed in the frequency domain and the resulting terms collected to yield the result,

$$\mathcal{D}^n_{\hat{\boldsymbol{u}}} f(\boldsymbol{x}) \Longleftrightarrow i^n[\hat{\boldsymbol{u}} \cdot \boldsymbol{\omega}]^n F(\boldsymbol{\omega}). \tag{B.10}$$

Similarly,

$$\mathcal{D}_{\hat{\boldsymbol{u}}} F(\boldsymbol{\omega}) \Longleftrightarrow (-i)^n[\hat{\boldsymbol{u}} \cdot \boldsymbol{x}]^n f(\boldsymbol{x}). \tag{B.11}$$

B.3.3 The Fourier Transform of a Gaussian

In this section we show that the Fourier transform of a Gaussian is also Gaussian. First we will require the result that the Gaussian density integrates to unity,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x = 1. \tag{B.12}$$

This result may be shown as follows,

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \mathrm{d}y \quad \text{with} \quad y = 0 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \mathrm{d}x \ \mathrm{d}y \\
&= \int_{0}^{2\pi} \int_{0}^{\infty} \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \ r \ \mathrm{d}r \ \mathrm{d}\theta \\
&= \int_{0}^{2\pi} -\frac{1}{2\pi} \left[\int_{0}^{\infty} e^{-\frac{r^2}{2\sigma^2}} \cdot -\frac{r}{\sigma^2} \mathrm{d}r \ \right] \mathrm{d}\theta \\
&= \int_{0}^{2\pi} -\frac{1}{2\pi} \left[e^{-\frac{r^2}{2\sigma^2}} \right]_{0}^{\infty} \mathrm{d}\theta \\
&= \int_{0}^{2\pi} \frac{1}{2\pi} \ \mathrm{d}\theta \\
&= 1.
\end{aligned}$$

We will use this result to show that the Fourier transform of a Gaussian in also a Gaussian (up to a scale factor) but the variance σ^2 of which is equal to the *reciprocal* of the variance of the original Gaussian. Specifically, the Fourier transform relation we wish to show is given by,

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \Longleftrightarrow e^{-\frac{\omega^2}{2/\sigma^2}}. \tag{B.13}$$

Let $F(\omega)$ denote the Fourier transform of the function $f(x)$. With $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ we have,

$$\begin{aligned}
F(\omega) &= \int_{-\infty}^{\infty} f(x) e^{-i\omega x} \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} e^{-i\omega x} \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2} x^2 - i\omega x} \mathrm{d}x
\end{aligned}$$

We may apply the technique of completing the square whereby we rewrite a quadratic equation $ax^2 + bx + c$ in the alternate form,

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right) \tag{B.14}$$

which, after some algebraic manipulation, yields

$$\begin{aligned}
F(\omega) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left(x+i\omega\sigma^2\right)^2-\frac{1}{2}\omega^2\sigma^2}\mathrm{d}x \\
&= e^{-\frac{1}{2}\omega^2\sigma^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}\left(x+i\omega\sigma^2\right)^2}\mathrm{d}x \\
&= e^{-\frac{\omega^2}{2/\sigma^2}}.
\end{aligned}$$

The integral in the second to last line is equal to unity by our earlier result.

APPENDIX C

ALGEBRAIC SOLUTION FOR QUADRATIC AND TIKHONOV-REGULARIZED FORMS

C.1 Preliminaries

C.1.1 Differentiable Scalar Functions

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a scalar function of the vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^\top$. The gradient vector is the vector of partial derivatives,

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}. \tag{C.1}$$

C.1.2 The Gradient of a Scalar Product

Let $f = \boldsymbol{a}^\top \boldsymbol{x}$ where $\boldsymbol{a}$ and $\boldsymbol{x}$ are n-vectors, then $\nabla f(\boldsymbol{x}) = \boldsymbol{a}$. This is seen by expanding the inner product $f(\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x}$ as $f(\boldsymbol{x}) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$, and then differentiating to find $\nabla f(\boldsymbol{x}) = [a_1, a_2, \ldots, a_n]^\top = \boldsymbol{a}$.

Let $f = \boldsymbol{y}^\top \boldsymbol{A} \boldsymbol{x}$ with m-vector $\boldsymbol{y}$ and $m \times n$ matrix $\boldsymbol{A}$, then $\nabla f(\boldsymbol{x}) = \boldsymbol{A}^\top \boldsymbol{y}$. This may be seen by noting that $\boldsymbol{y}^\top \boldsymbol{A}$ is an n-vector and applying the result for the scalar product above.

C.1.3 The Gradient of a Quadratic Form

If $f = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ with n-vector $\boldsymbol{x}$ and real-symmetric $n \times n$ matrix $\boldsymbol{A}$, then $\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}$. We can show this by writing f in terms of the on-diagonal and off-diagonal products,

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j \\ &= \sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{i \neq j}^{n}\sum^{n} a_{ij} x_i x_j, \end{aligned}$$

then differentiating f with respect to x_k for $k \in \{1, 2, \ldots, n\}$ yields,

$$\begin{aligned} \frac{\partial f(\boldsymbol{x})}{x_k} &= 2a_{kk}x_k + 2\sum_{i \neq k} a_{ki} x_i \\ &= 2\sum_{i=1}^{n} a_{ki} x_i, \end{aligned}$$

so that

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}.$$

C.2 Minimization of Quadratic Forms

Here we consider the problem of finding $\hat{\boldsymbol{x}}$ which minimizes the objective function with quadratic form,

$$f(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2,$$

with $m \times n$ real matrix $\boldsymbol{A}$. This may be rewritten as,

$$\begin{aligned} f(\boldsymbol{x}) &= (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) \\ &= (\boldsymbol{y}^\top - \boldsymbol{x}^\top \boldsymbol{A}^\top) \cdot (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) \\ &= \boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} \\ &= \boldsymbol{y}^\top \boldsymbol{y} - 2\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}. \end{aligned}$$

Noting that $\boldsymbol{A}^\top A$ is real-symmetric and differentiating the objective function with respect to $\boldsymbol{x}$ yields the gradient vector,

$$\nabla f(\boldsymbol{x}) = -2\boldsymbol{A}^\top \boldsymbol{y} + 2\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}$$

A necessary requirement for $\hat{\boldsymbol{x}}$ to be a minimum of $f(\boldsymbol{x})$ is that $\nabla f(\hat{\boldsymbol{x}}) = 0$. In this case we have that,

$$\boldsymbol{A}^\top \boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^\top \boldsymbol{y}$$

and assuming that $\boldsymbol{A}^\top \boldsymbol{A}$ is invertible,

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1}\boldsymbol{A}^\top \boldsymbol{y}.$$

The expression $(\boldsymbol{A}^\top \boldsymbol{A})^{-1}\boldsymbol{A}^\top$ is known variously as the generalized inverse, pseudoinverse, or Moore-Penrose inverse of $\boldsymbol{A}$.

C.3 Optimization for Tikhonov Regularized Problems

In Tikhonov regularized solutions to inverse problems we are faced with the problem of finding $\hat{\boldsymbol{x}}$ which minimizes the objective function containing a quadratic form and regularizing functional,

$$f(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|^2,$$

where $\lambda > 0$ is the regularization parameter. This may be rewritten as,

$$\begin{aligned} f(\boldsymbol{x}) &= (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + \lambda \boldsymbol{x}^\top \boldsymbol{x} \\ &= (\boldsymbol{y}^\top - \boldsymbol{x}^\top \boldsymbol{A}^\top)(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + \lambda \boldsymbol{x}^\top \boldsymbol{x} \\ &= \boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{x} \\ &= \boldsymbol{y}^\top \boldsymbol{y} - 2\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{x}. \end{aligned}$$

Noting that $\boldsymbol{A}^{\top}A$ is real-symmetric and differentiating the objective function with respect to $\boldsymbol{x}$ yields the gradient vector,

$$\begin{aligned}\nabla f(\boldsymbol{x}) &= -2\boldsymbol{A}^{\top}\boldsymbol{y} + 2\boldsymbol{A}^{\top}\boldsymbol{A}\boldsymbol{x} + 2\lambda\boldsymbol{x} \\ &= 2\boldsymbol{A}^{\top}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) + 2\lambda\boldsymbol{x}\end{aligned}$$

A necessary requirement for $\hat{\boldsymbol{x}}$ to be a minimum of $f(\boldsymbol{x})$ is that $\nabla f(\hat{\boldsymbol{x}}) = 0$. In this case we have that,

$$(\boldsymbol{A}^{\top}\boldsymbol{A} + \lambda\boldsymbol{I})\,\hat{\boldsymbol{x}} = \boldsymbol{A}^{\top}\boldsymbol{y}$$

and assuming that $\boldsymbol{A}^{\top}\boldsymbol{A} + \lambda\boldsymbol{I}$ is invertible,

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^{\top}\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^{\top}\boldsymbol{y}.$$

APPENDIX D

COMMON PENALTY FUNCTIONS

In this appendix, several commonly used penalty functions are presented. These functions appear in various contexts. In robust estimation or statistics they are used to create estimators which are less susceptible to the influence of outliers [82]. Some of these penalty functions find use in the formulation of Markov random field models for images as their modified response to outliers is well suited to modeling image data which invariably includes discontinuities [60, 188, 25]. Still others are used for robust, nonconvex optimization problems using graduated nonconvexity [15].

In the terminology used in robust statistics, the first derivative of the penalty function is called the influence function. Expressions for the penalty functions, the corresponding influence functions and plots of both are presented in the following pages. For parameterized penalty functions, plots are presented for a representative range of the parameters described below the plot.

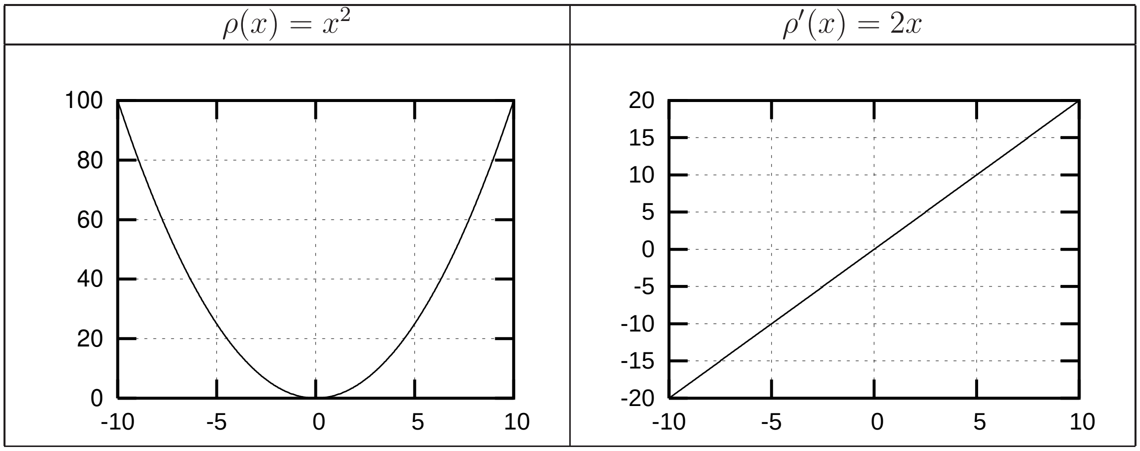


Figure D.1. Quadratic (Gaussian) penalty function and its influence function.

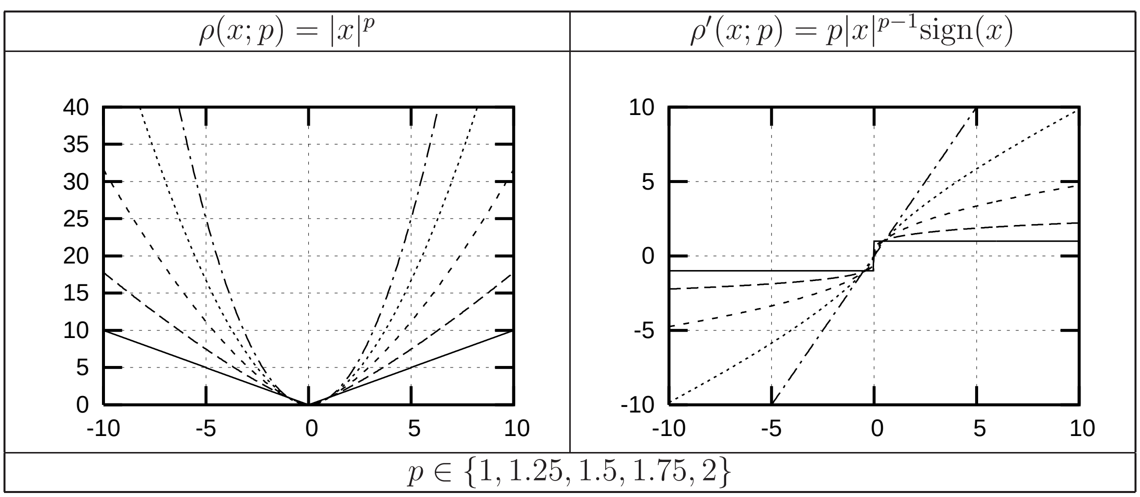
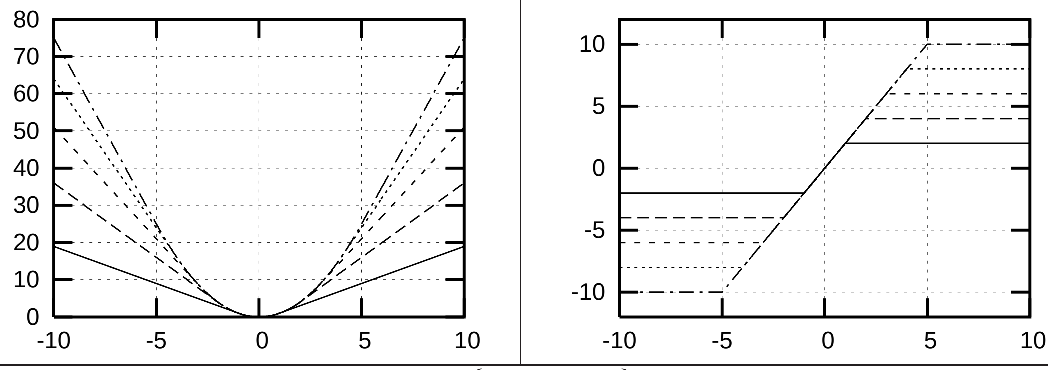


Figure D.2. Generalized Gaussian penalty function and its influence function.

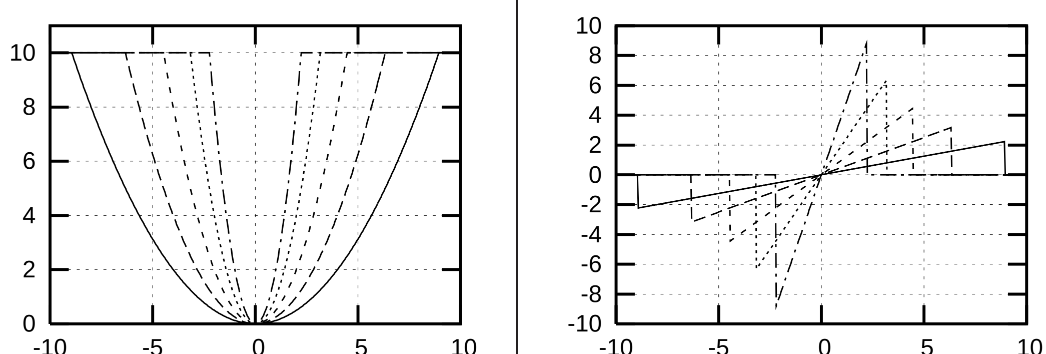
$\rho(x;T) = \begin{cases} x^2, & \lvert x\rvert \le T, \\ 2T\lvert x\rvert - T^2, & \text{else.} \end{cases}$	$\rho'(x;T) = \begin{cases} -2T, & x < -T, \\ 2x, & \lvert x\rvert \le T, \\ 2T, & x > T. \end{cases}$



$T \in \{1, 2, 3, 4, 5\}$

Figure D.3. Huber penalty function and its influence function.

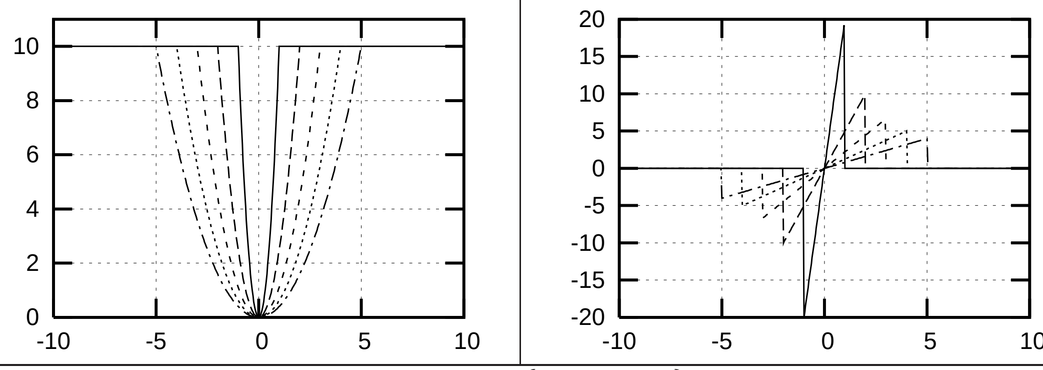
$\rho(x;\lambda,\alpha) = \begin{cases} \lambda x^2, & \lvert x\rvert \le \sqrt{\frac{\alpha}{\lambda}}, \\ \alpha, & \text{else.} \end{cases}$	$\rho'(x;\lambda,\alpha) = \begin{cases} 2\lambda x, & \lvert x\rvert \le \sqrt{\frac{\alpha}{\lambda}}, \\ 0, & \text{else.} \end{cases}$



$\alpha = 10, \lambda \in \{0.125, 0.25, 0.5, 1.0, 2.0\}$

Figure D.4. Blake and Zisserman penalty function and its influence function.

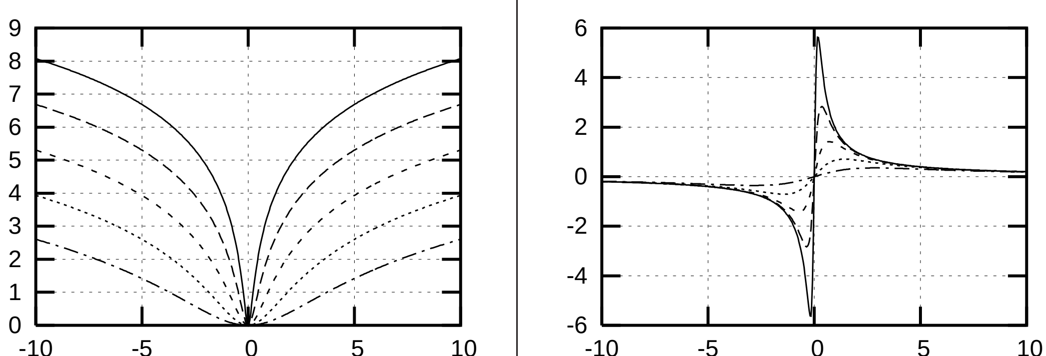
$\rho(x;T,\alpha) = \begin{cases} \frac{\alpha}{T^2}x^2, & \lvert x\rvert \le T, \\ \alpha, & \text{else.} \end{cases}$	$\rho'(x;T,\alpha) = \begin{cases} \frac{2\alpha}{T^2}x, & \lvert x\rvert \le T, \\ 0, & \text{else.} \end{cases}$



$\alpha = 10, T \in \{1, 2, 3, 4, 5\}$

Figure D.5. Modified Blake and Zisserman penalty function and its influence function. Modified form using threshold parameter T.

$\rho(x;\sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right)$	$\rho'(x;\sigma) = \frac{2x}{2\sigma^2+x^2}$



$\sigma \in \{0.125, 0.25, 0.5, 1.0, 2.0\}$

Figure D.6. Lorentzian penalty function and its influence function.

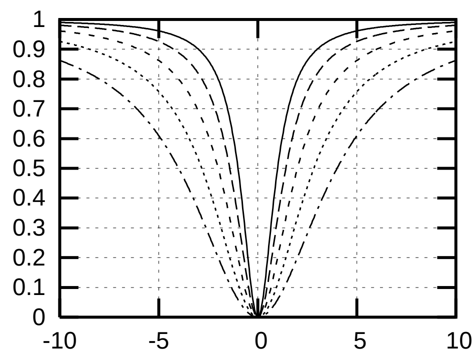
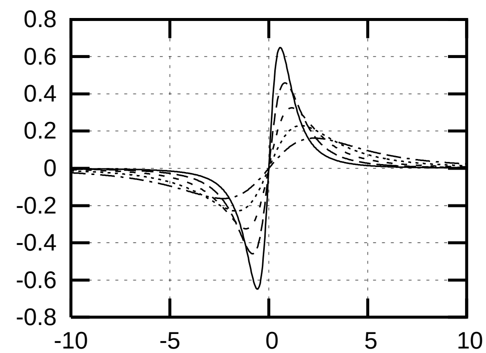
<table>
<tr><td>$\rho(x;\sigma) = \frac{x^2}{\sigma+x^2}$</td><td>$\rho'(x;\sigma) = \frac{2x\sigma}{(\sigma+x^2)^2}$</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">$\sigma \in \{1, 2, 4, 8, 16\}$</td></tr>
</table>

Figure D.7. Geman-McClure penalty function and its influence function.

BIBLIOGRAPHY

[1] Advanced Television Systems Committee, Washington, D.C., *ATSC Standard: Digital Television Standard, Revision B with Amendments 1 and 2* (2003).

[2] K. Aizawa, T. Komatsu and T. Saito, Acquisition of very high resolution images using stereo cameras. In *Visual Communications and Image Processing*, volume 1605 of *Proceedings of the SPIE*, pages 318–328 (1991).

[3] K. Aizawa, T. Komatsu and T. Saito, A scheme for acquiring very high resolution images using multiple cameras. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 289–292, San Francisco, CA (March 1992).

[4] Y. Altunbasak and A. Patti, A fast method of reconstructing high-resolution panoramic stills from MPEG-compressed video. In *IEEE Signal Processing Society 1998 Workshop on Multimedia Signal Processing*, pages 99–104, Los Angeles, California (December 1998).

[5] Y. Altunbasak and A. Patti, A maximum a posteriori estimator for high resolution video reconstruction from MPEG video. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 649–652, Vancouver, Canada (September 2000).

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

[7] H. Andrews and B. Hunt, *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, NJ (1977).

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

[9] M. Bazaraa, H. Sherali and C. Shetty, *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, second edition (1993).

[10] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*. Institute of Physics Publishing (1998).

[11] J. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B*, 36: 192–236 (1974).

[12] C. Bishop, A. Blake and B. Marthi, Super-resolution enhancement of video. In C. Bishop and B. Frey, editors, *Proceedings Artificial Intelligence and Statistics*, Society for Artificial Intelligence and Statistics, Key West, FL (2003).

[13] M. Black and P. Anandan, The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1): 75–104 (January 1996).

[14] M. Black and A. Jepson, Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10): 972–986 (October 1996).

[15] A. Blake and A. Zisserman, *Visual Reconstruction*. The MIT Press, Cambridge, MA (1987).

[16] S. Borman, M. Robertson and R. Stevenson, Block-matching sub-pixel motion estimation from noisy, under-sampled frames – an empirical performance evaluation. In K. Aizawa, R. Stevenson and Y. Zhang, editors, *Visual Communications and Image Processing '99*, volume 3653 of *Proceedings of the SPIE*, pages 1442–1451 (January 1999).

[17] S. Borman and R. Stevenson, Spatial resolution enhancement of low-resolution image sequences: A comprehensive review with directions for future research. Technical report, Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana, USA (July 1998).

[18] S. Borman and R. Stevenson, Super-resolution from image sequences – a review. In *Proceedings of the 1998 Midwest Symposium on Circuits and Systems*, pages 374–378, IEEE, Notre Dame, IN, USA (August 1998).

[19] S. Borman and R. Stevenson, Simultaneous multi-frame MAP super-resolution video enhancement using spatio-temporal priors. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 469–473, Kobe, Japan (October 1999).

[20] S. Borman and R. Stevenson, Image resampling and constraint formulation for multi-frame super-resolution restoration. In C. Bouman and R. Stevenson, editors, *Computational Imaging*, volume 5016 of *Proceedings of the SPIE*, San Jose, CA, USA (January 2003).

[21] S. Borman and R. Stevenson, Linear models for multi-frame super-resolution restoration under non-affine registration and spatially varying PSF. In C. Bouman and E. Miller, editors, *Computational Imaging II*, volume 5299 of *Proceedings of the SPIE*, San Jose, CA, USA (January 2004).

[22] N. Bose, H. Kim and H. Valenzuela, Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes. In *Proceedings of the IEEE conference on Acoustics, Speech and Signal Processing*, volume 5, pages 269–272, Minneapolis, MN (April 1993).

[23] N. Bose, H. Kim and H. Valenzuela, Recursive total least squares algorithm for image reconstruction from noisy, undersampled multiframes. *Multidimensional Systems and Signal Processing*, 4(3): 253–268 (July 1993).

[24] N. Bose, H. Kim and B. Zhou, Performance analysis of the TLS algorithm for image reconstruction from a sequence of undersampled noisy and blurred frames. In *Proceedings of the IEEE International Conference on Image Processing*, volume III, pages 571–575, Austin, TX (November 1994).

[25] C. Bouman and K. Sauer, A generalized Gaussian image model for edge preserving MAP estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(3): 296–310 (July 1993).

[26] R. Bracewell, *The Fourier Transform and its Applications*. McGraw Hill, New York, second revised edition (1986).

[27] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer-Verlag (1999).

[28] J. Brown, Multichannel sampling of low-pass signals. *IEEE Transactions on Circuits and Systems*, 28(2): 101–106 (1981).

[29] P. Burt and G. Sperling, Time, distance and feature trade-offs in visual apparent motion. *Psychological Review*, 88(2): 171–195 (1981).

[30] M. Carlotto and V. Tom, Interpolation of two-dimensional surfaces using the Gerchberg algorithm. *SPIE*, 359: 226–232 (1982).

[31] M. Chang, A. Tekalp and M. Sezan, Simultaneous motion estimation and segmentation. *IEEE Transactions on Image Processing*, 6(9): 1326–1333 (September 1997).

[32] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz and R. Hanson, Super-resolved surface reconstruction from multiple images. Technical Report FIA-94-12, NASA Ames Research Center, Moffet Field, CA 1994 (1994).

[33] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz and R. Hanson, Super-resolved surface reconstruction from multiple images. In G. R. Heidbreder, editor, *Maximum Entropy and Bayesian Methods*, pages 293–308, Kluwer, Santa Barbara, CA (1996).

[34] P. Cheeseman, B. Kanefsky, J. Stutz and R. Kraft, Subpixel resolution from multiple images. *Lunar and Planetary Science*, XXV: 241–242 (March 1994).

[35] D. Chen and R. Schultz, Extraction of high-resolution frames from MPEG image sequences. In *IEEE International Conference on Image Processing*, Chicago, IL (October 1998).

[36] P. Chou and C. Brown, The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, 4: 185–210 (1990).

[37] J. Clark, M. Palmer and P. Laurence, A transformation method for the reconstruction of functions from non-uniformly spaced samples. *IEEE Transactions on Acoustics Speech and Signal Processing*, 33(4): 1151–1165 (October 1985).

[38] G. Cross and A. Jain, Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(1): 25–39 (January 1983).

[39] C. Davila, Recursive total least squares algorithms for adaptive filtering. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1853–1856, Toronto, Canada (1991).

[40] C. Davila, Efficient recursive total least squares algorithms for FIR adaptive filtering. *IEEE Transactions on Signal Processing*, 42(2): 268–280 (February 1994).

[41] G. de Villiers, B. McNally and E. Pike, Positive solutions to linear inverse problems. *Inverse Problems*, 15(2): 615–635 (April 1999).

[42] A. Dempster, N. Laird and D. Rubin, Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1): 1–38 (November 1977).

[43] H. Derin and H. Elliot, Modelling and segmentation of noisy and textured images using gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(1): 39–55 (January 1987).

[44] R. Dubes and A. Jain, Random field models in image analysis. *Journal of Applied Statistics*, 16(2): 131–164 (1989).

[45] D. Dudgeon and R. Mersereau, *Multidimensional Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ (1984).

[46] M. Elad and A. Feuer, Super-resolution restoration of continuous image sequence using the LMS algorithm. In *Proceedings of the 18th IEEE Conference in Israel*, Tel-Aviv, Israel (March 1995).

[47] M. Elad and A. Feuer, Super-resolution reconstruction of an image. In *Proceedings of the 19th IEEE Conference in Israel*, pages 391–394, Jerusalem, Israel (November 1996).

[48] M. Elad and A. Feuer, Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6(12): 1646–1658 (December 1997).

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

[50] H. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*. Kluwer Academic Publishers (1996).

[51] A. Erdem, M. Sezan and M. Ozkan, Motion-compensated multiframe wiener restoration of blurred and noisy image sequences. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 293–296, San Francisco, CA (March 1992).

[52] P. Eren, M. Sezan and A. Tekalp, Robust, object-based high-resolution image reconstruction from low-resolution video. *IEEE Transactions on Image Processing*, 6(10): 1446–1451 (October 1997).

[53] K. Erickson and R. Schultz, Empirical limitations on super-resolution enhancement from MPEG compressed video. In *Proceedings of the IEEE International Conference on Image Processing*, Kobe, Japan (October 1999).

[54] S. Farsiu, M. Elad and P. Milanfar, Multi-frame demosaicing and super-resolution from under-sampled color images. In C. Bouman and E. Miller, editors, *Computational Imaging II*, volume 5299 of *Proceedings of the SPIE*, San Jose, CA, USA (January 2004).

[55] O. Faugeras, *Three-Dimensional Computer Vision – A Geometric Viewpoint*. MIT Press, Cambridge, Massachusetts (1993).

[56] J. Foley, A. van Dam, S. Feiner and J. Hughes, *Computer Graphics – Principles and Practice*. Addison-Wesley Publishing Company, second edition in c edition (1996).

[57] R. Franke, Scattered data interpolation: Tests of some methods. *Mathematics of Computation*, 38(157): 181–200 (January 1982).

[58] R. Franke, Smooth interpolation of scattered data by local thin plate splines. *Computers and Mathematics with Applications*, 8(4): 273–281 (1982).

[59] B. Frieden and H. Aumann, Image reconstruction from multiple 1-D scans using filtered localized projection. *Applied Optics*, 26(17): 3615–3621 (September 1987).

[60] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Patern Analysis and Machine Intelligence*, 6(6): 721–741 (June 1984).

[61] G. Golub and C. V. Loan, An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17: 883–893 (1980).

[62] G. Golub and C. V. Loan, *Matrix Computations*. Johns Hopkins University Press, second edition (1989).

[63] J. Goodman, *Introduction to Fourier Optics*. McGraw-Hill (1968).

[64] J. Goodman, *Introduction to Fourier Optics*. McGraw-Hill, second edition (1996).

[65] J. Green and B. Hunt, Super-resolution in a synthetic aperture imaging system. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages 865–868, Santa Barbara, CA (October 1997).

[66] B. Gunturk, Y. Altunbasak and R. Mersereau, Super-resolution reconstruction of compressed video using transform-domain statistics. *IEEE Transactions on Image Processing*, 13(1): 33–43 (January 2004).

[67] J. Hadamard, *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. Yale University Press, New Haven, CT (1923).

[68] H. Hang and Y. Chou, Motion estimation for image sequence compression. In H. Hang and J. Woods, editors, *Handbook of Visual Communications*, chapter 5, pages 147–188, Academic Press (1995).

[69] P. Hansen, *Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion*. SIAM (1998).

[70] R. Hardie, K. Barnard and E. Armstrong, Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 6(12): 1621–1633 (December 1997).

[71] R. Hardie, K. Barnard, J. Bognar, E. Armstrong and E. Watson, High resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. *Optical Engineering*, 37(1): 247–260 (January 1998).

[72] R. Hardie, T. Tuinstra, J. Bognar, K. Barnard and E. Armstrong, High Resolution Image Reconstruction From Digital Video with Global and Non-Global Scene Motion. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages 153–156, Santa Barbara, CA (October 1997).

[73] R. Hartley, Self-calibration of stationary cameras. *International Journal of Computer Vision*, 22(1): 5–23 (February 1997).

[74] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press (2000).

[75] P. Heckbert, *Fundamentals of Texture Mapping and Image Warping*. Master's thesis, University of California at Berkeley, Berkeley, CA, USA (June 1989).

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

[77] G. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. Academic Press, New York (1980).

[78] E. Hildreth, Computations underlying the measurement of visual motion. *Artificial Intelligence*, 23(3): 309–354 (August 1984).

[79] M. Hong, M. Kang and A. Katsaggelos, An iterative weighted regularized algorithm for improving the resolution of video sequences. In *Proceedings of the IEEE International Conference on Image Processing*, volume II, pages 474–477, Santa Barbara, CA (October 1997).

[80] M. Hong, M. Kang and A. Katsaggelos, A regularized multichannel restoration approach for globally optimal high resolution video sequence. In J. Biemond and E. Delp, editors, *Visual Communications and Image Processing*, volume 3024 of *Proceedings of the SPIE*, pages 1306–1316, San Jose, CA (February 1997).

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

[82] P. Huber, *Robust Statistics*. John Wiley & Sons, New York (1981).

[83] B. Hunt, Image super-resolution: Emerging prospects. In *Applications of Digital Image Processing XIV*, volume 1567 of *Proceedings of the SPIE*, pages 600–608, San Diego, CA (July 1991).

[84] B. Hunt, Quantifying the super-resolution capabilities of the CLEAN image processing algorithm. In *Applications of Digital Image Processing XV*, Proceedings of the SPIE, pages 300–306, San Diego, CA (July 1992).

[85] B. Hunt, Super-resolution of images: Algorithms, principles, performance. *International Journal of Imaging Systems and Technology*, 6(4): 297–308 (1995).

[86] International Telecommunication Union, *Draft Recommendation H.263+* (January 1998).

[87] M. Irani and S. Peleg, Super resolution from image sequences. In *Proceedings of the 10th International Conference on Pattern Recognition*, volume 2, pages 115–120, Atlantic City, NJ (June 1990).

[88] M. Irani and S. Peleg, Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53(3): 231–239 (May 1991).

[89] M. Irani and S. Peleg, Motion analysis for image enhancement: Resolution, occlusion and transparency. *Journal of Visual Communications and Image Representation*, 4: 324–335 (December 1993).

[90] M. Irani, B. Rousso and S. Peleg, Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1): 5–16 (February 1994).

[91] ISO/IEC JTC1 IS 11172-2, *Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbits/s* (1993).

[92] ISO/IEC JTC1 IS 13818, *Generic coding of moving pictures and associated audio* (1994).

[93] ISO/IEC JTC1 IS 14496, *Generic coding of audio-visual objects* (1999).

[94] ITU-T Recommendation H.261, *Video codec for audiovisual services at* p $\times$ *64 kbits* (1993).

[95] ITU-T Recommendation H.263, *Video coding for low bit rate communication* (1998).

[96] G. Jacquemod, C. Odet and R. Goutte, Image resolution enhancement using subpixel camera displacement. *Signal Processing*, 26(1): 139–146 (1992).

[97] A. Jain, *Fundamentals of Digital Image Processing*. Prentice-Hall, Englewood Cliffs, NJ (1989).

[98] T. Jebara, A. Azarbayejani and A. Pentland, 3D structure from 2D motion. *IEEE Signal Processing Magazine*, pages 66–84 (May 1999).

[99] M. Jordan, editor, *Learning in Graphical Models*. MIT Press, Cambridge, MA (1999).

[100] E. Kaltenbacher and R. Hardie, High-resolution infrared image reconstruction using multiple low resolution aliased frames. In *Proceedings of the IEEE National Aerospace Electronics Conference (NAECON)*, volume 2, pages 702–709, Dayton, OH (May 1996).

[101] A. Katsaggelos, A multiple input image restoration approach. *Journal of Visual Communication and Image Representation*, 1(1): 93–103 (September 1990).

[102] A. Katsaggelos, editor, *Digital Image Restoration*. Springer-Verlag (1991).

[103] D. Keightley and B. Hunt, A rigid POCS extension to a Poisson superresolution algorithm. In *IEEE International Conference on Image Processing*, volume II, pages 508–511, Washington, D.C. (October 1995).

[104] J. Keller, Inverse problems. *American Mathematical Monthly*, 83(2): 107–118 (February 1976).

[105] D. Keren, S. Peleg and R. Brada, Image sequence enhancement using subpixel displacements. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 742–746 (June 1988).

[106] H. Kim, *High Resolution Image Reconstruction from Undersampled Multiframes*. Ph.D. thesis, Pennsylvania State University (May 1994).

[107] S. Kim and N. Bose, Reconstruction of 2-D bandlimited discrete signals from non-uniform samples. *Proceeding of the IEE*, 137: 197–204 (1990).

[108] S. Kim, N. Bose and H. Valenzuela, Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(6): 1013–1027 (1990).

[109] S. Kim and W. Su, Recursive high-resolution reconstruction of blurred multiframe images. In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, volume 4, pages 2977–2980, Toronto, Canada (May 1991).

[110] S. Kim and W. Su, Recursive high-resolution reconstruction of blurred multiframe images. *IEEE Transactions on Image Processing*, 2: 534–539 (October 1993).

[111] S. Kim and W. Su, Subpixel accuracy image registration by spectrum cancellation. In *Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing*, volume V, pages 153–156, Minneapolis, MN (April 1993).

[112] R. Kindermann and J. Snell, *Markov Random Fields and Their Applications*. American Mathematical Society, Providence, Rhode Island (1980).

[113] S. Kirkpatrick, C. Gellat, Jr. and M. Vecchi, Optimization by simulated annealing. *Science*, 220(4598): 671–680 (May 1983).

[114] A. Kirsch, *An introduction to the mathematical theory of inverse problems*, volume 120 of *Applied mathematical sciences*. Springer (1996).

[115] T. Komatsu, K. Aizawa, T. Igarashi and T. Saito, Signal-processing based method for acquiring very high resolution images with multiple cameras and its theoretical analysis. *IEE Proceedings I*, 140: 19–25 (February 1993).

[116] T. Komatsu, K. Aizawa and T. Saito, Very high resolution image acquisition via reconstruction using multiple lower-resolution cameras various in their pixel apertures. In *Proceedings of the International Workshop on HDTV*, pages 581–588 (1992).

[117] T. Komatsu, T. Igarashi, K. Aizawa and T. Saito, Very high resolution imaging scheme with multiple different aperture cameras. *Signal Processing: Image Communication*, 5: 511–526 (December 1993).

[118] J. Konrad and E. Dubois, Bayesian estimation of vector motion fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9): 910–927 (1992).

[119] R. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*. Kluwer Academic Publishers (1991).

[120] L. Landweber, An iterative formula for Fredholm integral equations of the first kind. *American Journal of Mathematics*, 73: 615–624 (1951).

[121] R. Lenz and U. Lenz, Calibration of a color CCD camera with 3000x2300 picture elements. In *Close-range photogrammetry meets machine vision*, volume 1395 of *Proceedings of the SPIE*, pages 104–111, Zurich, Switzerland (September 1990).

[122] A. Lorette, H. Shekarforoush and J. Zerubia, Super-resolution with adaptive regularization. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages 169–172, Santa Barbara, CA (October 1997).

[123] B. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision. *IEEE Transactions on Acoustics Speech and Signal Processing*, 33(4): 1151–1165 (October 1985).

[124] S. Mann, Compositing multiple pictures of the same scene. In *Proceesings of the 46th annual IS&T conference*, Cambridge, MA (May 1993).

[125] S. Mann and R. Picard, Virtual bellows: Constructing high quality stills from video. In *Proceedings of the IEEE International Conference on Image Processing*, pages 363–367, Austin, TX (1994).

[126] S. Mann and R. Picard, Video orbits of the projective group: A simple approach to featureless estimation of parameters. *IEEE Transactions on Image Processing*, 6(9): 1281–1295 (September 1997).

[127] J. Mateos, A. Katsaggelos and R. Molina, Resolution enhancement of compressed low resolution video. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages 1919–1922, Istanbul, Turkey (2000).

[128] J. Mateos, A. Katsaggelos and R. Molina, Simultaneous motion estimation and resolution enhancement of compressed low resolution video. In *IEEE International Conference on Image Processing*, volume 2, pages 653–656, Vancouver, Canada (September 2000).

[129] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. John Wiley & Sons, New York (1996).

[130] G. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, New York (2000).

[131] A. Mitiche, *Computational Analysis of Visual Motion*. Plenum Press, New York (1994).

[132] M. Nadar, B. Hunt and P. Sementilli, Multigrid techniques and wavelet representations in image super-resolution. In *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, volume 2308, pages 1244–1255, Chicago, IL (September 1994).

[133] H.-H. Nagel, On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33(3): 299–324 (November 1987).

[134] H.-H. Nagel and W. Enkelmann, An investigation of smoothness constraints for the estimation of displacement vector fields from images sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(5): 565–593 (1986).

[135] Y. Nakazawa, T. Komatsu and T. Saito, High resolution image acquisition based on temporal integration with hierarchical estimation of image warping. In *Proceedings of IEEE International Conference on Image Processing*, volume III, pages 244–247, Washington, DC (1995).

[136] Y. Nakazawa, T. Saito, T. Sekimori and K. Aizawa, Two approaches for image-processing based high resolution image acquisition. In *Proceedings of the IEEE International Conference on Image Processing*, volume III, pages 147–151, Austin, TX (1994).

[137] M. Ng and N. Bose, Mathematical analysis of super-resolution methodology. *IEEE Signal Processing Magazine*, 20(3): 62–74 (May 2003).

[138] N. Nguyen, P. Milanfar and G. Golub, Blind superresolution with generalized cross-validation using Gauss-type quadrature rules. In *Thirty-third Asilomar Conference on Signals, Systems and Computers* (October 1999).

[139] N. Nguyen, P. Milanfar and G. Golub, Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement. *IEEE Transactions on Image Processing*, 10(9): 1299–1308 (September 2001).

[140] S. of Images Based on Local Correlations, Frank M. candocia and Jose C. principe. *IEEE Transactions on Neural Networks*, (2): 372–380 (March 1999).

[141] P. Oskoui-Fard and H. Stark, Tomographic image reconstruction using the theory of convex projections. *IEEE Transactions on Medical Imaging*, 7(1): 45–58 (March 1988).

[142] A. Papoulis, Generalized sampling expansion. *IEEE Transactions on Circuits and Systems*, 24(11): 652–654 (November 1977).

[143] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, third edition (1991).

[144] A. Patti and Y. Altunbasak, Super-resolution image estimation for transform coded video with application to MPEG. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 179–183, Kobe, Japan (October 1999).

[145] A. Patti, M. Sezan and A. Tekalp, Digital video standards conversion in the presence of accelerated motion. *Signal Processing: Image Communication*, 6(3): 213–227 (June 1994).

[146] A. Patti, M. Sezan and A. Tekalp, High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages 343–347, Austin, TX (1994).

[147] A. Patti, M. Sezan and A. Tekalp, High resolution standards conversion of low resolution video. In *Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing*, volume 4, pages 2197–2200, Detroit, MI (1995).

[148] A. Patti, M. Sezan and A. Tekalp, Robust methods for high quality stills from interlaced video in the presence of dominant motion. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(2): 328–342 (April 1997).

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

[150] A. Patti, A. Tekalp and M. Sezan, A new motion compensated reduced order model Kalman filter for space-varying restoration of progressive and interlaced video. *IEEE Transactions on Image Processing*, 7(4): 543–554 (April 1998).

[151] F. Pedrotti and L. Pedrotti, *Introduction to Optics*. Prentice Hall, second edition (1992).

[152] S. Peleg, D. Keren and L. Schweitzer, Improving image resolution by using subpixel motion. *Pattern Recognition Letters*, 5(3): 223–226 (March 1987).

[153] L. Ping and W. Yu-Mei, Image resolution improvement by swing CCD array imager in two dimensions and its signal processing. In *Visual Communications and Image Processing*, volume 2308 of *Proceedings of the SPIE*, pages 915–920, Chicago, IL (September 1994).

[154] S. Reeves and S. Eddins, Comments on "Iterative procedures for reduction of blocking effects in transform image coding". *IEEE Transactions on Circuits and Systems for Video Technology*, 3(6): 439–440 (December 1993).

[155] R. Roberts and C. Mullis, *Digital Signal Processing*. Addison-Wesley (1987).

[156] M. Robertson, S. Borman and R. Stevenson, Dynamic range improvement through multiple exposures. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 159–163, Kobe, Japan (October 1999).

[157] M. Robertson, S. Borman and R. Stevenson, Estimation-theoretic approach to dynamic range improvement using multiple exposures. *Journal of Electronic Imaging*, 12(2): 219–228 (April 2003).

[158] T. Saito, T. Komatsu and K. Aizawa, An image processing algorithm for a super high definition imaging scheme with multiple different-aperture cameras. In *Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing*, volume 5, pages 45–48, Adelaide, Australia (April 1994).

[159] K. Sauer and J. Allebach, Iterative reconstruction of band-limited images from nonuniformly spaced samples. *IEEE Transactions on Circuits and Systems*, 34(12): 1497–1506 (December 1987).

[160] A. Sawchuk, Space-variant image motion degradation and restoration. *Proceedings of the IEEE*, 60(7): 854–861 (July 1972).

[161] R. Schultz, L. Meng and R. Stevenson, Subpixel motion estimation for multiframe resolution enhancement. In J. Biemond and E. Delp, editors, *Visual Communications and Image Processing*, volume 3024 of *Proceedings of the SPIE*, pages 1317–1328, San Jose, CA (February 1997).

[162] R. Schultz, L. Meng and R. Stevenson, Subpixel motion estimation for super-resolution image sequence enhancement. *Journal of Visual Communication and Image Representation, special issue on High-Fidelity Media Processing*, 9(1): 38–50 (March 1998).

[163] R. Schultz and R. Stevenson, Improved definition image expansion. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume III, pages 173–176, San Francisco, CA (March 1992).

[164] R. Schultz and R. Stevenson, A Bayesian approach to image expansion for improved definition. *IEEE Transactions on Image Processing*, 3(3): 233–242 (1994).

[165] R. Schultz and R. Stevenson, Extraction of high-resolution frames from video sequences. In *Proceedings of the IEEE Workshop on Nonlinear Image Processing*, Neos Manos, Greece (June 1995).

[166] R. Schultz and R. Stevenson, Improved definition video frame enhancement. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages 2169–2172, Detroit, MI (May 1995).

[167] R. Schultz and R. Stevenson, Video resolution enhancement. In *Image and Video Processing III*, Proceedings of the SPIE, pages 23–34, San Jose, CA (February 1995).

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

[169] R. Schultz and R. Stevenson, Motion-compensated scan conversion of interlaced video sequences. In *Image and Video Processing IV*, volume 2666 of *Proceedings of the SPIE*, pages 107–118, San Jose, CA (February 1996).

[170] R. Schultz and R. Stevenson, Bayesian estimation of subpixel-resolution motion fields and high-resolution video stills. In *Proceedings of the IEEE International Conference on Image Processing*, volume III, pages 62–65, Santa Barbara, CA (October 1997).

[171] C. Segall, R. Molina and A. Katsaggelos, High-resolution images from low-resolution compressed video. *IEEE Signal Processing Magazine*, 20(3): 37–48 (May 2003).

[172] C. Segall, R. Molina, A. Katsaggelos and J. Mateos, Reconstruction of high-resolution image frames from a sequence of low-resolution and compressed observations. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 1701–1704, Orlando, Florida (USA) (May 2002).

[173] P. Sementilli, B. Hunt and M. Nadar, Analysis of the limit to superresolution in incoherent imaging. *Journal of the Optical Society of America A*, 10(11): 2265–2276 (November 1993).

[174] P. Sementilli, M. Nadar and B. Hunt, Poisson MAP super-resolution estimator with smoothness constraint. In *Neural and stochastic methods in image and signal processing II*, volume 2032 of *Proceedings of the SPIE*, pages 2–13, San Diego, CA (July 1993).

[175] P. Sementilli, M. Nadar and B. Hunt, Empirical evaluation of a bound on image superresolution performance. In *Image reconstruction and restoration*, volume 2302 of *Proceedings of the SPIE*, pages 178–187, San Diego, CA (July 1994).

[176] M. Sezan and R. Lagendijk, editors, *Motion analysis and image sequence processing*. Kluwer Academic Publishers (1993).

[177] M. Sezan and H. Stark, Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion. *IEEE Transactions on Medical Imaging*, MI-3(2): 91–98 (June 1984).

[178] N. Shah and A. Zakhor, Multiframe spatial resolution enhancement of color video. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages 985–988, Lausanne, Switzerland (September 1996).

[179] N. Shah and A. Zakhor, Resolution enhancement of color video sequences. *IEEE Transactions on Image Processing*, 8(6): 879–885 (June 1999).

[180] H. Shekarforoush, M. Berthod and J. Zerubia, 3D super-resolution using generalized sampling expansion. In *Proceedings of the IEEE International Conference on Image Processing*, volume II, pages 300–303, Washington, DC (October 1995).

[181] D. Sheppard, B. Hunt and M. Marcellin, Super-resolution of imagery aquired through turbulent atmosphere. In *Asilomar Conference on Signals, Systems and Computers*, pages 81–85, Washington, DC (1997).

[182] D. Sheppard, B. Hunt and M. Marcellin, Iterative multiframe super-resolution algorithms for atmospheric turbulence degraded imagery. *Journal of the Optical Society of America A*, 15(5): 978–992 (May 1998).

[183] E. Simoncelli, *Distributed Representation and Analysis of Visual Motion*. Ph.D. thesis, Massachusetts Institute of Technology (1993).

[184] E. Simoncelli, E. Adelson and D. Heeger, Probability distributions of optical flow. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 310–315, IEEE Computer Society, Mauii, Hawaii (1991).

[185] C. Srinivas and M. D. Srinath, A stochastic model-based approach for simultaneous restoration of multiple misregistered images. In *Visual Communications and Image Processing*, volume 1360 of *Proceedings of the SPIE*, pages 1416–1427, Lausanne, Switzerland (October 1990).

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

[187] R. Stevenson and S. Borman, Image sequence processing. In R. Driggers, editor, *Encyclopedia of Optical Engineering*, pages 840–879, Marcel Dekker, New York (2003).

[188] R. Stevenson, B. Schmitz and E. Delp, Discontinuity preserving regularization of inverse visual problems. *IEEE Transactions on Systems, Man and Cybernetics*, 24(3): 455–469 (March 1994).

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

[190] W. Su and S. Kim, High-resolution restoration of dynamic image sequences. *International Journal of Imaging Systems and Technology*, 5(4): 330–339 (1994).

[191] A. Tekalp, *Digital Video Processing*. Prentice Hall, Upper Saddle River, NJ (1995).

[192] A. Tekalp, *Digital Video Processing*, chapter 17. Prentice Hall, Upper Saddle River, NJ (1995).

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

[194] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. V. H. Winston & Sons, Washington, D.C. (1977).

[195] M. Tipping and C. Bishop, Bayesian image super-resolution. In S. Becker, S. Thrun and K. Obermeyer, editors, *Advances in Neural Information Processing Systems*, volume 15, pages 1303–1310 (2002).

[196] B. Tom and A. Katsaggelos, Multi-channel image identification and restoration using the expectation maximization algorithm. In *Applications of Digital Image Processing XVII*, volume 2298 of *Proceedings of the SPIE*, pages 316–331, San Diego, CA (July 1994).

[197] B. Tom and A. Katsaggelos, Reconstruction of a high resolution image from multiple degraded mis-registered low resolution images. In *Visual Communications and Image Processing*, volume 2308 of *Proceedings of the SPIE*, pages 971–981, Chicago, IL (September 1994).

[198] B. Tom and A. Katsaggelos, Reconstruction of a high-resolution image by simultaneous registration, restoration and interpolation of low-resolution images. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 539–542, Washington, DC (1995).

[199] B. Tom and A. Katsaggelos, An iterative algorithm for improving the resolution of video sequences. In *Visual Communications and Image Processing*, volume 2727 of *Proceedings of the SPIE*, pages 1430–1438, Orlando, FL (March 1996).

[200] B. Tom and A. Katsaggelos, Resolution enhancement of video sequences using motion compensation. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages 713–716, Lausanne, Switzerland (September 1996).

[201] B. Tom, A. Katsaggelos and N. Galatsanos, Reconstruction of a high resolution image from registration and restoration of low resolution images. In *Proceedings of the IEEE International Conference on Image Processing*, volume III, pages 553–557, Austin, TX (1994).

[202] R. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4): 323–344 (August 1987).

[203] R. Tsai and T. Huang, Multiframe image restoration and registration. In R. Y. Tsai and T. S. Huang, editors, *Advances in Computer Vision and Image Processing*, volume 1, pages 317–339, JAI Press Inc. (1984).

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

[205] J. Wang and E. Adelson, Representing moving images with layers. *IEEE Transactions on Image Processing*, 3(5): 625–638 (September 1994).

[206] Y. Weiss, Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *1997 Conference on Computer Vision and Pattern Recognition*, pages 520–526, Puerto Rico (June 1997).

[207] Y. Weiss, *Bayesian motion estimation and segmentation*. Ph.D. thesis, Massachusetts Institute of Technology (May 1998).

[208] S. Yeh and H. Stark, Iterative and one-step reconstruction from nonuniform samples by convex projections. *Journal of the Optical Society of America A*, 7: 491–499 (1990).

[209] D. Youla and H. Webb, Image restoration by the method of convex projections: Part 1 – Theory. *IEEE Transactions on Medical Imaging*, MI-1(2): 81–94 (October 1982).

[210] P. Yuen and B. Hunt, Constraining information for improvements to MAP image super-resolution. In *Image resonstruction and restoration*, volume 2302 of *Proceedings of the SPIE*, pages 144–155, San Diego, CA (July 1994).