MOTION-FREE SUPER-RESOLUTION

by

Subhasis Chaudhuri

Joshi Manjunath

Springer

MOTION-FREE SUPER-RESOLUTION

MOTION-FREE SUPER-RESOLUTION

by

Subhasis Chaudhuri

Manjunath V. Joshi

Springer

Subhasis Chaudhuri Manjunath V. Joshi

Motion-Free Super-Resolution

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

ISBN-10: 0-387-25890-6
ISBN-13: 978-0387-25587-3
e-ISBN-10: 0-387-25908-2
e-ISBN-13: 978-0387-25908-6
Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11054764 (HC) / 11427018 (eBook)

springeronline.com

To

Dearest Sucharita.
SC.

My Parents, Smita, Ninad and Nidhi (Figure 8.14).
MVJ.

Contents

Preface

It has been nearly four years since one of us edited the book 'Super-Resolution Imaging' which was published by Kluwer Academic Publishers. The research area of super-resolution imaging has witnessed a further growth since then. We see a number of papers appearing in various conferences and journals on a regular basis. We also observe that researchers are now concentrating on finding the performance bounds of different methods of image super-resolution. This is a sign that this topic is getting a wider acceptance among the researchers in computer vision and that the field is gradually maturing.

With the explosion of Internet technology and graphics engines, digital images are now everywhere. The image capturing tools are all pervading - in our pockets to inside a satellite. The imaging applications have also grown and many such applications demand an availability of high resolution images. However, all such images are not *picture perfect*. They may be lacking sufficient details in the picture. This requires that these images be super-resolved for improved details. How to achieve this is what constitutes the research area of image super-resolution.

The task of image super-resolution requires an availability of several low resolution observations of a scene. Each observation provides some additional information about the scene, and when these are fused together we obtain a high resolution description of the scene. Most of the researchers prefer using a moving camera to capture the scene and use the available motion cue. Although this is a very natural way of generating additional observations, the most difficult task here is to estimate a dense motion field between two frames at a subpixel accuracy. Hence we ask the question if it is at all possible to generate these additional observations without introducing any relative motion among them. This would alleviate the problem of having to establish the feature correspondences. We explore the applicability of cues other than the motion cue in super-resolving a scene. This justifies the title of this monograph.

Unlike in the area of motion-based super-resolution, the amount of published literature in this area is almost insignificant. Hence the current state of research in motion-free super-resolution is still in its primordial stage. We cannot claim that the contents of the monograph would relate to a firm recommendation of new technologies for immediate absorption by the industry. Rather, this book is meant to serve as a fodder for new research ideas in this area.

The book is addressed to a broad audience. It should be of great value to both practitioners and researchers in the area of image processing and computer vision. All topics have been covered with sufficient details so that there is no specific pre-requisite. A basic familiarity with the area of image processing should suffice. Hence the students may find this monograph useful as a reference book. We have provided a large number figures to help understand the topics well.

We would very much appreciate receiving comments and suggestions from the readers.

IT Bombay,
October 2004

Subhasis Chaudhuri
Manjunath V. Joshi

Acknowledgments

The authors are greatly indebted to Dr. Deepu Rajan at NTU, Singapore for his help in developing the contents of chapter 3. Thanks are due to C. V. Jiji at IIT Bombay for his help in building up the contents of chapter 7. The authors have also benefited from comments and suggestions received from the following people: Prof. S. C. Sahasrabudhe, Prof. Vikram Gadre, Prof. Sumantra Dutta Roy, Prof. A. N. Rajagopalan, Prof. Andrew Zisserman, Prof. Bertrand Zavidovique, Prof. Eckehard Steinbach, Rajashekhar, Vinay P. Namboodiri, V. Srikrishnan, Malay Nema and Dr. Shankar Chatterjee.

The authors wish to gratefully acknowledge partial financial assistances received under the following project grants: *Swarnajayanti Fellowship* under DST, Curriculum Development Programme at IIT Bombay, and Super-resolution projects under MHRD and ISRO.

The authors are also thankful to the staff of Springer, particularly Jennifer Evans, Paul Farrel, Anne Murray, Wayne Wheeler and Painter Anna, for their very enthusiastic cooperation and assistance.

The second author wish to thank the Principal and the management members of KLS Gogte Institute of Technology for their help and support.

Finally the authors wish to express their deep sense of gratitude to their family members for their constant love, encouragement and patience.

IIT Bombay,
October 2004

Subhasis Chaudhuri
Manjunath V. Joshi

1 Introduction

Digital pictures today are all around us - on the web, on digital versatile discs (DVDs), on satellite systems; they are everywhere. Having these pictures in a digital form allows us to manipulate them the way we want them. Digital image processing helps us to enhance the features of interest and to extract useful information about the scene from the enhanced image. Initial ideas on image processing were used back in 1920 for cable transmission of pictures. Since majority of the information received by a human being is visual, it was felt that integrating the ability to process visual information into a system would enhance its overall utility. Work on using computer techniques for improving the quality of images obtained from a space probe began at the Jet Propulsion Laboratory in 1964 when pictures of the moon transmitted by Ranger 7 were processed by a computer to correct various types of image distortion inherent in the on-board television camera [1]. The field of image processing has grown considerably during the past few decades with the improvement in the size, speed, and cost effectiveness of the digital computers. Today the advancement in image processing hardware as well as in software is so much so that one can purchase an entire image processing system off the shelf.

The field of image processing has several applications. Some of them include areas such as medical imaging, satellite imagery, image transmission, industrial inspection, surveillance, *etc.* In medical imaging, processing of images helps the doctors to make a correct diagnosis. To distinguish objects from similar ones such as detection of changes along the coast lines from satellite imagery is useful for natural resource management. In the post 9/11 era, there has been a massive boost in the research area of visual surveillance.

In almost all electronic imaging applications, images with a high resolution are desired. There is always a demand for better quality images. Availability of high quality images is crucial for several computer vision applications. With the high resolution imaging, one could obtain a better classification of regions in a multi-spectral image, a more accurate localization of a tumor in a medical image, or a more pleasing view in a high definition television (HDTV); But the resolution of an image is dependent on the sensor or the image acquisition device and a high resolution sensor is often very expensive. Also, the available camera resolution may not always suffice for a given application. Thus one has to look for image processing methods to increase the resolution.

1.1 What is Image Resolution?

Perhaps the most important technical concept to understand in imaging literature is the word resolution. Resolution is a fundamental issue in judging the quality of various image acquisition or processing systems. In its simplest form, *image resolution* is defined as the smallest discernible or measurable detail in a visual presentation. In optics the resolution of a device is determined by measuring the modulation transfer function (MTF) or the optical transfer function (OTF) which represents the response of the system to different spatial frequencies. MTF is not only used to give the resolution limit at a single point, but also to characterize the system to an arbitrary input [2]. Researchers in digital image processing and computer vision classify resolution into three different types.

- *Spatial Resolution:* An image is made up of small picture elements called pixels. Spatial resolution refers to the spacing of the pixels in an image and is measured in pixels per unit length. The higher the spatial resolution, the more are the pixels in an image. High spatial resolution allows a clear perception of sharp details and subtle color transitions in an image. In case an image with high levels of details is not represented by a spatially dense set of pixels, the image is said to suffer from aliasing artifacts. For an output device such as a printer the spatial resolution is expressed in dots per inch (dpi).
- *Brightness Resolution:* Also known as gray-level resolution, it refers to the number of brightness levels or gray-levels used to represent a pixel. The brightness resolution increases with the number of quantization levels used. A monochrome image is usually quantized using

256 levels with each level represented by 8 bits. For a color image, at least 24 bits are used to represent one brightness level, *i.e.*, 8 bits per color plane (red, green, blue). It should be noted that the number of gray value quantization levels is also intrinsically related to the spatial sampling rate. If the camera sensor has fewer quantization levels, it should have a much higher spatial sampling rate to capture the scene intensity. This idea is quite similar to that of delta modulation used in communication systems and to that of dithering used in half-tone printing.

- *Temporal Resolution:* It represents the frame rate or the number of frames captured per second. Higher the temporal resolution, lesser is the flicker observed. The lower limit on the temporal resolution is proportional to the amount of motion occurred between two consecutive frames. The typical frame rate for a pleasing view is about 25 frames per second or above.

Another kind of resolution of interest is the spectral resolution and it refers to the frequency or spectral resolving power of a sensor that gives the bandwidth of the light (or electro-magnetic wave) frequencies captured by the sensor. It is defined as the smallest resolvable wavelength difference by the sensor. The spectral resolution plays an important role in satellite imaging. In this monograph, the term resolution unequivocally refers to the spatial resolution, enhancement of which is the subject matter of this book. Further, we do not explore the inter-relationship between the brightness and spatial resolutions in this monograph.

1.2 Need for Resolution Enhancement

An image sensor or camera is a device which converts optical energy into an electrical signal. Modern imaging sensors are based on the charge-coupled device (CCD) technology, which consists of an array of photo-detector elements or pixels that have a voltage output proportional to the incident light [3]. The number of detector elements decide the spatial resolution of the camera. Higher the number of detector elements, more is the resolution. A sensor with less number of detector elements produces a low resolution image, giving blocky effect. This is because when a scene is photographed with a low resolution camera, it is sampled at a low spatial sampling frequency, causing aliasing effect. One could think of reducing the size of the photo-detector elements, thereby increasing the density and hence the sampling rate. But as the pixel size decreases the amount of light incident on each pixel also decreases and this causes a shot noise [4, 5], which degrades the image quality. Increasing the pixel density increases the resolution but also causes shot noise. Thus there exists a limitation on the size of a pixel in a sensor and the optimal size is estimated to be about $40\mu m^2$. The current image sensor technology has almost reached this level.

Another approach to increase the resolution is to increase the wafer size which leads to an increase in the capacitance [6]. This approach is not effective since an increase in capacitance causes a decrease in charge transfer rate. This limitation causes the image of a point light source to be blurred. Also there is distortion due to aliasing because of a low sampling rate for a low resolution sensor. Moreover in some applications like satellite imagery, the physical constraints make the sensor unrealizable for a high resolution. Thus there is a need for developing post acquisition signal processing techniques to enhance the resolution. These techniques being post processing methods applied on the low resolution images, they offer flexibility as well as cost benefit since there is no additional hardware cost involved. However, the increased computational cost may be the burden that an user has to bear.

1.3 Super-Resolution Concept

The low resolution representation resulting from the lower spatial sampling frequency produces distortion in the image due the to loss of high frequency components. This causes loss of important information such as edges and textures. Also a degradation occurs due to the sensor point spread function (PSF), and optical blurring due to camera motion or out-of-focus. Thus an image captured with a low resolution camera suffers from aliasing, blurring and presence of noise. *Super-resolution* (SR) refers to the process of producing a high spatial resolution image from several low resolution images, thereby increasing the maximum spatial frequency and removing the degradations that arise during the image capturing process using a low resolution camera. In effect, the super-resolution process extrapolates the high frequency components and minimizes aliasing and blurring.

As already mentioned, one way to increase the sampling rate is to reduce the pixel size, thereby increasing the pixel density. But an increase in pixel density causes shot noise and hence the distortion. Also the cost of sensor increases with the increase in pixel density. Hence the sensor modification is not always a practical solution for increasing the resolution. Thus we resort to image processing techniques to enhance the resolution. The advantage here is that there is no additional hardware cost involved and also it offers a flexibility such as region of interest super-resolution. One of the approaches towards this end is simple image interpolation that can be used to increase the size of the image. But the quality of the interpolated image is very much limited due to the use of a single, aliased low resolution image. Also the single image interpolation is a highly ill-posed problem since there may exist infinitely many upsampled or expanded images which are consistent with the original data. A single image interpolation cannot recover the high frequency components lost or degraded due to the low resolution sampling. Some progress can be achieved by convolving the image with a filter designed to boost the higher frequency components. Unfortunately this also amplifies any noise in the image and degrades the quality. Hence the image interpolation methods are not considered as super-resolution techniques.

In order to obtain super-resolution we must look for nonredundant information among the various frames in an image sequence. The most obvious method for this seems to be to capture multiple low resolution observations of the same scene through subpixel shifts due to the camera motion. These subpixel shifts can occur due to the controlled motion in imaging systems, *e.g.*, a landsat satellite captures images of the same area on the earth every eighteen days as it orbits around it. The same is true for uncontrolled motion, *e.g.*, movements of local objects or vibrating imaging systems. If the low resolution image shifts are integer units, then there is no additional information available from subsequent low resolution observations for super-resolution purposes. However, if they have subpixel shifts then each low resolution aliased frame contains additional information that can be used for high resolution reconstruction. Such a technique makes the ill-posed nature of the problem to a better-posed one, as more data is available from multiple frames.

Many researchers often term the process of super-resolving a scene as super-resolution restoration. It may be mentioned here that super-resolution differs from a typical image restoration problem wherein the image formation model (discussed in the next section) does not consider the decimation process *i.e.*, the aliasing which is inherently present in the low resolution observations. Thus the size of the restored image is the same as that of the observed image for image restoration while it is dependent on the decimation factor for a super-resolved image.

1.4 Super-Resolution Technique

The success of any super-resolution reconstruction method is based on the correctness of the low resolution image formation model that relates the original high resolution image to the observed images. The most common model used is based on observations which are shifted, blurred and decimated (aliased) versions of the high resolution image. The observation model relating a high resolution image to low resolution video frames is shown in Figure 1.1.

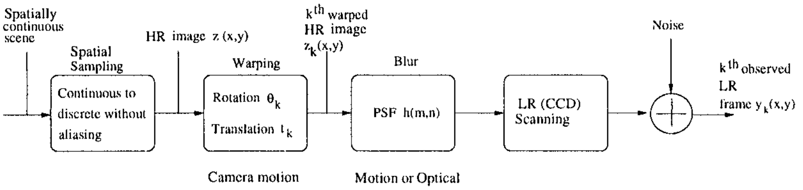


Fig. 1.1. Observation model relating a high resolution image to the observed low resolution frames for a static scene and a moving camera. Here HR and LR stand for high resolution and low resolution, respectively.

Let us assume that the scene is static and the camera is slowly moving. Let us further assume that the depth variation in the scene is negligible compared to its distance from the camera so that the perspective distortion due to camera motion can be neglected. In the figure $z(x, y)$ is the desired high resolution image which is obtained by sampling the spatially continuous scene at a rate greater than or equal to the Nyquist rate. Here the assumption is that the continuous scene is bandlimited. The camera motion at the k^{th} time instant during the exposure is modeled as pure rotation θ_k and translation t_k. Next, the blurring which may be caused by the optical system or due to relative motion between the camera and the scene, can be modeled as linear space invariant or linear space variant. One can select an appropriate point spread function (PSF) for the blur. These warped and blurred high resolution images undergo a low resolution scanning, *i.e.*, subsampling or decimation, followed by noise addition, yielding the low resolution observations.

Most of the super-resolution methods proposed in the literature use motion between the observed frames as a cue for estimating the high resolution image. This being the most intuitive approach for super-resolution, is based on a three-stage algorithm consisting of registration, interpolation and restoration. The registration step is used to find the relative motion between the frames with a subpixel accuracy. The assumption here is that all the pixels from the available frames can be mapped back onto the reference frame based on the motion vector information. Or in other words, there is no occlusion which is usually true if the depth variation on the scene is planer. Next, the interpolation onto a uniform grid is done to obtain a uniformly spaced upsampled image. Once the upsampled image on uniformly spaced grid points is obtained, restoration is applied to remove the effects of aliasing and blurring and to reduce noise. The restoration can be performed by using any deconvolution algorithm that considers the presence of an additive noise. A scheme for constructing the high resolution frame from multiple low resolution frames is shown in Figure 1.2 [7]. Here the low resolution observations $y_1, y_2 \cdots, y_p$ are used as input to the motion estimation module. The registered images are then interpolated onto a high resolution grid, which is then post-processed through restoration to generate a super-resolved image.

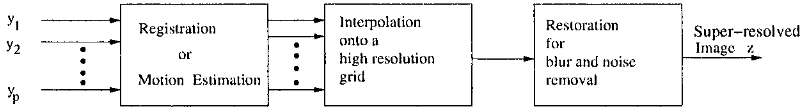


Fig. 1.2. Scheme for super-resolution from multiple subpixel shifted observations.

1.5 Tour of the Book

Nonredundant information among the low resolution frames is the key to super-resolution. Each low resolution frame provides a different "look" of the same scene. In order to get nonredundant information from different frames, most of the multi-frame methods use motion as a cue, and the super-resolution restoration is obtained by using the scheme shown in Figure 1.2. Here the motion information serves as a cue to solve the super-resolution problem. However, this method being a 2D dense feature matching technique, it requires an accurate registration or motion estimation. But the task of accurate registration is dependent on the observed image data and it requires that the images should not contain any degradation. Also for better restoration, the shifts between the images, *i.e.*, the registration must be accurately known. Thus, the registration and restoration are interdependent and it is a difficult task to obtain an accurate registration. It requires a considerable computational burden to obtain an accurate registration. Thus the performance of motion-based super-resolution algorithms will ultimately be limited by the effectiveness of motion estimation and modeling. For a proper super-resolution, the estimated motion field should be of high resolution. However, these high resolution fields should be estimated from the low resolution image data. Another problem of concern with motion-based super-resolution methods is that they do not consider the 3D structure of the scene being imaged although such information is inherently available. Since the structure of an object is embedded in the images in various forms such as stereo disparity, it limits the quality of the super-resolved image and its applicability in 3D computer vision problems. Also the motion-based super-resolution methods assume that the low resolution observations are all at the same spatial resolution.

Theoretically, nonredundant information about the scene can also be obtained by using different camera parameters or with different lighting conditions while capturing the scene without effecting a relative scene motion [8]. To this end researchers have explored the possibility of using cues other than the motion cue for super-resolution purposes. In [9], Rajan and Chaudhuri have successfully demonstrated the use of blur as a cue for super-resolving the intensity field. A similar idea of using the blur for image super-resolution was also proposed by Elad and Feuer [10]. The usefulness of the approach lies in the fact that there is no relative motion between the camera and the scene. Hence there is no requirement of image registration. This motivates us to use cues other than the motion cue for super-resolution. To this end we consider multiple frames of the same scene in which each frame does contain some additional information, although there is no relative shift between frames. The primary question we ask is that can we have image super-resolution without having to register images? We demonstrate in this monograph that it is, indeed, possible to perform image super-resolution without having to use the motion cue. We call this class of super-resolution techniques as motion-free super-resolution method. We discuss various techniques that we have developed over the last few years under this class. This justifies the title of the book. A quick summary of the various topics discussed in this monograph is as follows.

- In chapter 2, we discuss the current literature on super-resolution for both motion-based and motion-free methods.
- In chapter 3, we consider the case where the observations are blurred, *i.e.*, we use the blur cue instead of the motion cue. A super-resolution technique is presented in which a sequence of blurred, decimated and noisy versions of an ideal high resolution image is used to generate a super-resolved image. The depth related defocus blur provides the cue. This is a natural cue in any real aperture (non-pin-hole) imaging system. Multiple observations can be obtained by varying the camera parameters. We not only generate the super-resolved intensity map but the unknown depth map is also recovered at a finer grid. The super-resolved image and the depth map expressed in terms of the space variant blur parameter, are individually modeled as separate Markov random fields (MRFs). The maximum *a posteriori* (MAP) estimate of these fields *i.e.*, the super-resolved image and the depth are recovered through optimization of an appropriate cost function. Since the blur is related to unknown scene depth at a point, in effect, we solve the space varying blind deconvolution problem in this chapter.
- Next we consider the use of photometric cue for super-resolution, and explore the possibility of using the same for the estimation of both super-resolved image and the depth map. The observations are images captured under different light source positions keeping both the camera and the object stationary. We obtain the super-resolved image and the spatially enhanced structure simultaneously. In addition, we recover the super-resolved albedo of the surface. Since there is no relative motion between the camera and the scene there is no correspondence problem. The high resolution image is obtained not only for a particularly given light source position but also for an arbitrary virtual light source direction. In addition we can also perform a high resolution rendering of a scene. Our work here is initially based on the generalized interpolation scheme for super-resolution of the image intensity map proposed in [11]. However this method fails to achieve good results as it does not consider several issues while utilizing the photometric cue. No contextual constraints are used

in [11], which are very much necessary in the interpretation of the visual information. We model the high resolution image, the structure of the scene and the albedo of the surface as separate Markov random fields in order to take care of the contextual dependency. In practice the assumed reflectance model may differ significantly from the true model and this may lead to errors while reconstructing the surface. In order to circumvent this problem we reproject the reconstructed high resolution image on the low resolution observation so that they match well. We also use the surface integrability constraint which has to be satisfied by any physically valid surface. An optimization technique which incorporates the different constraints is developed to solve the problem.

- It was mentioned earlier that the super-resolution (SR) problem is equivalent to image restoration along with image upsampling. Most of the motion-based SR methods assume the blur PSF to be known. In order to show how the photometric stereo can also be used for blind restoration, we solve the problem of simultaneous estimation of scene structure along with restoration of the images from blurred photometric observations in chapter 5. In the existing literature on shape from shading the researchers have treated the problem of shape estimation without considering the blur introduced by the camera. They assume a pin-hole model that inherently implies that there is no camera blur during observations. However, when one captures the images with a camera, the degradation in the form of blur and noise is often present in these observed images. The blur could happen due to a variety of reasons such as improper focus setting or camera jitter. It is natural that the variations in image intensity due to camera blur affects the estimates of the surface shape. Thus, the estimated shape differs from the true shape in spite of possibly having the knowledge of the true surface reflectance model. This limits the applicability of these techniques in 3D computer vision problems. This motivates us to restore the images as well, while recovering the structure. We estimate the different fields (surface gradients, albedo, and image intensity) when the blur is unknown. Since the camera blur is not known, in addition, we estimate the point spread function (PSF) of the blur which caused the degradation. Thus the problem can be classified as a joint blind restoration and surface recovery problem. We show that the entire problem can be expressed as a simple problem of regularization and can be solved iteratively using existing mathematical tools.

- In the recent years theoretical as well as practical advances have been made in the field of computer vision by using new techniques which involve learning. Learning is basically used to enhance the performance of a system. The ability to learn certain aspects or characteristics of a scene from an image plays an important role in solving many of the computer vision problems. Learning based methods have been applied in variety of areas such as segmentation, feature extraction and object recognition. Researchers have applied learning based methods for solving super-resolution problem as well. Thus after having considered the photometric cue for super-resolution we show how the super-resolution reconstruction can be obtained using a learning based approach. Here a single low resolution image serves as the observation, but we assume the availability of a number of high resolution training images of various different kinds of scenes. We observe that edges in an image are the regions where the high frequency components are restricted. An attempt to upsample an image shows a noticeable degradation at edges due to blurring. Hence we consider edge primitives at the low resolution observation and try to learn them locally at the higher resolution from the high resolution training data set. Since the wavelets are best suited for analyzing a signal with discontinuities locally at different scales, we make use of the edge representation in the wavelet domain for the purpose of learning the missing high frequency components in the unknown high resolution image. The unknown wavelet coefficients at finer scales of the high resolution image are learnt from the training set and the image thus obtained in the wavelet domain is used for further regularization to remove possible blockiness in the the super-resolved image.
- In the previous chapter we learnt the high frequency details at a given location locally from the high resolution database. In chapter 7, we explore the usefulness of learning features globally from the training set. This requires that we restrict the input image to a given class of object (say, a face or a fingerprint image) and that the training data to the same class. We use a principal component analysis (PCA)-based method for image super-resolution for a class of images using the concept of generalized interpolation defined in chapter 4. The image is decomposed into principal components to

obtain eigen-images. A set of low resolution images is used for the same. The given low resolution image is then projected onto these eigen-images to obtain eigen coefficients. The eigen-images are interpolated using a suitable interpolation technique and the linear combination using the eigen coefficients yields the desired high resolution image. Unlike learning of the features locally, the global learning process is extremely fast, as the learnt eigen-images are all precomputed, however the domain of applicability is restricted to a given class of object only.

- Finally in chapter 8 we develop a motion-free super-resolution technique using zoom as a cue. Researchers have used zoom cue to solve computer vision problems which include depth estimation [12, 13, 14], minimization of view degeneracies [15], and zoom tracking [16]. We show that the zoom cue can also be used to solve the super-resolution problem by using the existing mathematical tools. This is because the amount of aliasing differs with zooming. When one captures the images with different zoom settings, the least zoomed entire area of the scene is represented by a limited number of pixels, *i.e.*, it is sampled with a very low spatial frequency and the most zoomed image at a higher sampling rate. Thus one can use zoom as an effective cue for generating a high resolution image at the lesser zoomed area of a scene. Our approach generates a super-resolved image of the entire scene although only a part of the observed scene has multiple observations. In effect what we do is as follows. If the wide angle view corresponds to a field of view of α^o, and the most zoomed view corresponds to a field of view of β^o (where $\alpha > \beta$), we generate a picture of the α^o field of view at a spatial resolution comparable to β^o field of view.

  The observations here are the images of the same scene captured with different zoom settings. We consider the most zoomed observation as the super-resolved one and obtain the super-resolution of the entire scene which is at a low resolution. There are no spatial shifts between the observations but the area of the scene captured is different with different zoom settings. We not only obtain the super-resolution for known integer zoom factors, but also for unknown arbitrary zooms. This is done by estimating the zoom factors among the different observations. The zoom factors are estimated by using a hierarchical cross-correlation technique. We model the super-resolved image as a Markov random field (MRF) and a maximum *a posteriori* (MAP) estimation technique is used to obtain the super-resolution. Since our objective is to reconstruct the high frequency details by data fusion *i.e.*, increase the sampling density by using a number of observations, we should naturally attempt to preserve discontinuities (*i.e.*, those features that carry high frequency information). To this end we consider the preservation of these discontinuities in the form of sudden changes in the intensity values by using appropriate line fields.

After having considered the case where the prior MRF parameters are selected on an adhoc basis, our next step is to learn these parameters so that the computational time complexity can be very much reduced as one need not spend time in choosing the appropriate model parameters manually on trial and error basis. Thus, learning represents the next part of our proposition. Since, we capture images of a static scene with different zoom settings, the most zoomed observation has the highest resolution. We learn the parameters of the super-resolved image from the most zoomed observation and use the same to super-resolve the rest of the part in the least zoomed entire scene. The high resolution field is modeled as a homogeneous MRF. The learnt field parameter set is then used as priors while super-resolving the observations. We make use of the MAP formulation with MRF prior to derive the cost function and the minimization is done using the gradient descent approach. However, the estimation of the MRF model parameters is a difficult task as most of the methods are computationally expensive. We use a relatively faster learning algorithm known as the maximum pseudo-likelihood (MPL) estimator to estimate the model parameters.

Although the priors in the form of MRF model parameters constitute a most general statistical model, and capture the local dependencies very well, the computational burden goes up drastically when one needs to use a larger neighborhood structure in order to capture the spatial dependency well. This motivates us to use a simultaneous autoregressive model (SAR) as the prior, which is a linear model. Although this represents a weaker model, the associated computational requirement is negligible. Here we use a larger neighborhood to capture the local dependency. An iterative maximum likelihood (ML) estimator is used for SAR parameter estimation. A suitable regularization scheme is employed to obtain the high resolution image with the SAR model prior.

- The book concludes in chapter 9 which also includes future issues for further research in the area of super-resolution.

We have carried out extensive experiments on real as well as simulated images. The results obtained show perceptual improvements as well as quantifiable gains in terms of mean squared error (MSE) or peak signal to noise ratio (PSNR). Wherever appropriate we have provided these figures of merit to demonstrate the usefulness of the methods discussed. We also highlight the demerits of these methods so that the practitioners in this area can have a better insight into these methods as regards their applicability.

During the course of evolution in research in this specific area, it is quite natural that some of these works have earlier been reported in a few conferences and journals. This monograph derives parts of its contents from these publications [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27].

2

Research on Image Super-Resolution

Many researchers have tackled the super-resolution reconstruction problem for both still images and video. Although the super-resolution reconstruction techniques for video are often extensions to still image super-resolution, many different approaches have also been proposed. In general, based on the type of cues used, the super-resolution methods can be classified into two categories: motion-based techniques and the motion-free approaches. Motion-based techniques use the relative motion between different low resolution observations as a cue in estimating the high resolution image, while motion-free super-resolution techniques may use cues such as blur, zoom, and shading. These methods do not require observations with relative motion among them. Some researchers have also attempted to solve the super-resolution reconstruction problem without considering any specific cue, but by using an ensemble of images as a training set in order to learn the required information for resolution enhancement.

Different methods to obtain super-resolution include nonuniform interpolation approach, frequency domain approach, and regularization based reconstruction technique which may be either deterministic or stochastic. Few other existing approaches include projection onto convex sets, iterative back projection method, adaptive filtering method, *etc.* Most of the super-resolution techniques discussed in the literature are based on the motion cue, *i.e.*, using the subpixel shifts among the observations. A few researchers have also tackled the super-resolution problem without using the motion cue. In this chapter we review the literature on super-resolution reconstruction for motion-based as well as for motion-free techniques. A comprehensive survey on super-resolution imaging can also be found in [28, 29].

2.1 Motion-Based Super-Resolution

The super-resolution idea was first proposed by Tsai and Huang [30]. They used the frequency domain approach to demonstrate the ability to reconstruct a single improved resolution image from several downsampled, noise free versions of it. A frequency domain observation model was defined for this problem which considered only the globally shifted versions of the same scene. Their approach is based on the following principles:

- the shifting property of the Fourier transform,
- the aliasing relationship between the continuous Fourier transform of the original image and the discrete Fourier transform (DFT) of the observed low resolution frames, and
- the assumption that the original high resolution image is bandlimited.

Kim *et al.* discuss a recursive algorithm, also in the frequency domain, for the restoration of super-resolution images from noisy and blurred observations [31]. They consider the same blur and noise characteristics for all the low resolution observations. Their recursive approach combines the two steps of filtering and reconstruction. The filtering operation on the registered images compensates for the degradation and noise, and the reconstruction step estimates the image samples on a high resolution grid in order to obtain the super-resolved image. Kim and Su [32] consider different amounts of blur for each low resolution image and used the Tikhonov regularization to obtain the solution of an inconsistent set of linear equations.

The disadvantage with the frequency domain approach lies on the restrictions imposed on the observation model. One may consider only a translational motion and a linear space invariant (LSI) blur. Also, since the data is uncorrelated in the frequency domain, it is difficult to apply apriori knowledge about the data for the purpose of regularization. Nonetheless, it was a good beginning and very soon researchers started looking at the problem in the spatial domain also. Needless to say, researchers have also explored the use of other types of image transforms to achieve super-resolution. For example, a discrete cosine transform (DCT) based method instead of DFT has been proposed by Rhee and Kang [33].

A minimum mean squared error approach for multiple image restoration, followed by interpolation of the restored images into a single high resolution image has been presented in [34]. Ur and Gross use the Papoulis and Brown generalized sampling theorem [35],[36] to obtain an improved resolution picture from an ensemble of spatially shifted observations [37]. These shifts are assumed to be known by the authors. A recursive total least squares method for super-resolution reconstruction to reduce the effects of registration error is discussed in [38]. All the above super-resolution restoration methods are restricted either to a globally uniform translational displacement between the measured images, or an LSI blur, and a homogeneous additive noise.

A different approach to the super-resolution restoration problem was suggested by Peleg and his co-authors [39, 40, 41], based on the iterative back projection (IBP) method adapted from computer aided tomography. This method starts with an initial guess of the output image, projects the temporary result to the measurements (simulating them), and updates the temporary guess according to this simulation error. A back projection kernel determines the contribution of the error to the reconstructed image at each iteration. The disadvantage of IBP is that it has no unique solution as it does not attempt to involve prior constraints. A set theoretic approach to the super-resolution restoration problem was suggested in [42]. The main result there is the ability to define convex sets which represent tight constraints on the image to be restored. Having defined such constraints it is straightforward to apply the projections onto convex sets (POCS) method, which was originally suggested by Stark and Oskoui [4]. The POCS based approach describes an alternative way to incorporating the prior knowledge about the solution into the super-resolution reconstruction process. According to this method, the solution is restricted to be a member of a closed convex set that is defined as a set of vectors which satisfy a user specified property. If the constraint sets have nonempty intersection, then a solution can be found by alternately projecting onto the convex sets. All these methods mentioned above are not restricted to having a specific motion characteristic. They can handle smooth motion, linear space variant blur, and non-homogeneous additive noise.

Ng *et al.* develop a regularized constrained total least squares (RCTLS) solution to obtain a high resolution image in [43]. They consider the presence of perturbation errors of displacements around the ideal subpixel locations in addition to sensor noise. The superiority of the approach over conventional least squares based approach is substantiated through examples. The analysis of the effect of displacement errors on the convergence rate of the iterative approach for solving the transform based preconditioned system of equations during high resolution image reconstruction with multiple sensors has been carried out in [44]. It is established that the use of MAP, L_2-norm or H_1-norm based regularization functional leads to a linear convergence of the conjugate gradient descent method in terms of the displacement errors caused by the imperfect subpixel localization. Bose *et al.* [45] point out the important role of the regularization parameter and suggest the use of a constrained least squares (CLS) method for super-resolution reconstruction which generates the optimum value of the regularization parameter, using the L-curve method [46].

In [47] the authors use a maximum *a posteriori* (MAP) framework for jointly estimating the registration parameters and the high resolution image for severely aliased observations. They use an iterative, cyclic coordinate-descent optimization technique to update the registration parameters. A similar idea of joint estimation applied to infrared imagery is presented in [48]. The high resolution estimate of the image is obtained by minimizing a regularized cost function based on the observation model. It is also shown that with a proper choice of tuning parameter, the algorithm exhibits robustness in presence of noise. Both the gradient descent and the conjugate gradient descent optimization techniques are used to minimize the cost function. An expectation maximization (EM) based algorithm solved in the frequency domain in order to simultaneously estimate the super-resolved image, the blur and the registration parameters is described in [49]. All these methods alternately estimate the high resolution image and the motion fields for an improved accuracy.

A MAP estimator with Huber-Markov random field (HMRF) prior is described by Schultz and Stevenson in [50] for improving the image resolution. Here a discontinuity preserving stabilizing functional is used for the preservation of edges. In HMRF, an edge preserving potential function is used to define the prior constraint. The potential function is given by

$$U(x) = \begin{cases} x^2, \text{ if } |x| \leq \alpha \\ 2\alpha|x| - \alpha^2, \quad \text{otherwise} \end{cases}$$

where x is the finite difference approximation of the first order derivative of the image at each pixel. HMRF is an example of a convex but nonquadratic prior. The purpose of making the prior linearly increasing beyond the threshold $|x| > \alpha$ is to partly reduce the rate of growth in the cost function when there is an edge between two pixels. The idea is quite similar to the concept of an M-estimator prevalent in the area of robust regression analysis. In the paper two separate algorithms have been derived: a constrained optimization method for a noise free image reconstruction, and an unconstrained optimization algorithm for image data containing Gaussian noise. The gradient projection algorithm has been used to minimize the cost derived from a noise free case and a gradient descent optimization is used for the noise corrupted case. Till date this method is probably the most popular one among the researchers as we notice that most of the currently proposed approaches compare their performances with the results obtained with an HMRF prior. Since all these methods claim a superiority over the HMRF based method, it is probably safe to state that the HMRF method, indeed, yields a reasonably accurate result.

In many resolution enhancement applications, the blurring process *i.e.*, the point spread function (PSF) of the imaging system, is not known. Nguyen *et al.* [51] propose a technique for parametric blur identification and regularization based on the generalized cross-validation (GCV) theory. The idea of cross-validation is to divide the data set into two parts; one part is used to construct an approximate solution, and the other is used to validate that approximation. They propose approximation techniques based on the Lanczos algorithm and Gauss quadrature theory for reducing the computational complexities of GCV. They solve a multivariate nonlinear minimization problem for the unknown parameters. They have also proposed circulant block preconditioners to accelerate the conjugate gradient descent (CG) method while solving the Tikhonov-regularized super-resolution problem [52]. Preconditioning is a process used to transform the original system into one with the same solution, but which can be solved more quickly by the iterative solver. They use specific preconditioners such that the preconditioned system has eigenvalues clustered around unity which makes CG method to converge rapidly.

Elad and Feuer [10] propose a unified methodology for super-resolution restoration from several geometrically warped, blurred, noisy and down-sampled observations by combining maximum likelihood (ML), MAP and POCS approaches. The proposed super-resolution approach is general but assumes explicitly a linear space variant blur, and an additive Gaussian noise. In addition to the motion-based super-resolution the authors also discuss the condition for motion-free super-resolution imaging when the observations are captured with different amounts of defocus blur even when both the camera and the object are stationary. This issue will be taken up in the next section. An adaptive filtering approach to super-resolution restoration is described by the same authors in [53] using the least mean squares (LMS) and the pseudo-recursive least squares (RLS) algorithms. Both the methods have been demonstrated with and without regularization. They exploit the properties of the operations involved in their previous work [10] and develop a fast super-resolution algorithm in [54] for a purely translational motion and space invariant blur, assuming them to be the same for all the images. The approach consists of deblurring and measurement fusion which is shown to be solvable using a non-iterative algorithm. Similarly two fast non-iterative algorithms for image super-resolution based on Choleskey decomposition have been developed by Jorge and Ferreira [55]. They use the spatial domain formulation and the frequency domain approach. The spatial domain approach leads to a set of linear equations for the unknown pixels, while the frequency domain approach leads to equations for the unknown DFT coefficients. An additional inverse Fourier transform is used to obtain the required image while working in the frequency domain.

A computationally fast super-resolution algorithm based on the preconditioner using the motion adaptive relaxation parameters is considered in [56]. The proposed algorithm can be implemented in real time by updating the motion compensated low resolution frame at each time instant by using the preconditioner which increases the converges rate. Thus the speed up operation is achieved through system preconditioning as discussed earlier. This method can be applied to a general image sequence with differently moving objects, thus can handle local variations in the motion parameters. Farsiu *et al.* propose a fast and robust super-resolution algorithm based on L_1-norm for both data fitting term and the prior term and show that it performs better with and even without the outliers present in the data [57]. The robustness is achieved by limiting the contribution of the highly erroneous outlier data through the use of L_1-norm. Quite naturally, we may replace the L_1-norm by any appropriate weight function $W(x)$ as it is commonly done in M-estimator. The authors in [58] investigate the performance of super-resolution algorithms using different potential functions such as convex, nonconvex, bounded, and the unbounded as a prior in the cost function and compare their performance on synthetic and real images.

They evaluate the performances of three different potential functions given below proposed, respectively, by Charbonnier [59], Hebert and Leahy [60], and Geman and Reynolds [61].

$$\begin{aligned} U(x) &= 2\sqrt{(1+x^2)} - 2, \\ U(x) &= \log(1+x^2), \quad \text{and} \\ U(x) &= \frac{x^2}{1+x^2}, \end{aligned}$$

where x is the finite difference approximation of the first order derivative of the image at each pixel. Different optimization methods have been used for each prior model.

Edges are typically the most important features in an image. For a homogeneous region, any kind of interpolation technique for image upsampling would suffice. However, one must be careful while upsampling the regions having edges as we would like them to be sharp in the high resolution data. Chiang and Boult [62] use edge models and a local blur estimate to develop an edge-based super-resolution algorithm. An image consistent reconstruction algorithm is used which gives the exact solution for some input function which, according to the sensor model, would have generated the measured input. Rather than obtaining the super-resolution by fusion of all the images together they choose one of the images from the image sequence and then fuse together all the edges from the other images. This requires that the reference image be re-estimated and scaled up based on the edge models and local blur estimation. Thus they mitigate the problem arising due to illumination variation during image capture since the edge positions are less sensitive to lighting variations. They have also applied image warping to reconstruct a high resolution image [63] which is based on a concept called integrating resampler [64] that warps the image subject to some constraints. Here the upsampled images are combined using the median, and the resultant image is convolved to remove blur, with a high pass filter. Similarly, a robust median-based estimator is used in an iterative process to achieve the super-resolution in [65]. This approach discards the measurements which are inconsistent with the imaging model, thus increasing the resolution even in regions having the outliers.

An image super-resolution technique based on the wavelet domain hidden Markov tree (HMT) model as a prior is proposed by Zhao *et al.* [66]. The wavelet domain HMT characterizes the statistical properties of the real image. Here the authors use the motion cue, but unlike using the Huber-MRF prior, they use the HMT prior. They formulate the problem as a constrained optimization problem and solve it using a cyclic optimization procedure.

All the methods discussed so far do use the motion cue for super-resolution, and in order to do that they need to actually compute the motion parameters. At this point some researchers may feel that since most of the available video sequences are already MPEG compressed, decompression of the video and then motion estimation is a wastage of time. The MPEG data already has the motion vectors in the bit stream. Can these motion vectors be used without fully decompressing the MPEG data? This problem of recovering a high resolution image from a sequence of DCT compressed images is addressed in [67]. It may be noted that the MPEG motion vectors may not always give us the true motion field. Also, the motion vectors are not dense. It is specified over a macro-block. So the authors recover the high resolution image using an iterative method considering the effects of quantization (residual) noise as well as registration errors, both modeled as zero mean additive Gaussian noise. A regularization functional is introduced not only to reflect the relative amount of registration error but also to determine the regularization parameter. Segall *et al.* estimate the high resolution image as well as subpixel displacements from compressed image observations [68]. They formulate the problem in a Bayesian framework and use the iterative cyclic coordinate descent approach for the joint estimation. Here the pixel intensities are no longer the observations, instead motion vectors and quantized transform coefficients are provided to the recovery algorithm.

There have been very few publications in the area of quantifying the performance of motion-based super-resolution methods. Lin and Shum determine the fundamental limits of reconstruction-based super-resolution algorithms and obtain the super-resolution limits from the conditioning analysis of the coefficient matrix [69]. They prove that fundamental limits do exist for reconstruction based super-resolution algorithms where a number of low resolution, subpixel displaced frames are used to estimate a high resolution image. They discuss two extreme cases and find that the practical limit for magnification is 1.6, if the registration and the noise removal is not good enough.

Let us now discuss some of the application specific super-resolution schemes. There has been an effort in the area of astrophysics for improving the image resolution of celestral objects. In [70] authors use a series of short-exposure images taken concurrently with a corresponding set of images of a guidestar and obtain a maximum-likelihood estimate of the undistorted image. Yang and Parvin [71] compute the dense map of feature velocities from lower resolution data and project them onto the corresponding high resolution data. The proposed technique is applied to measurement of sea surface temperature. The super-resolution principle has been applied to the face recognition systems as well in [72, 73]. They apply the super-resolution technique after dimensionality reduction to a set of inaccurate feature vectors of a subject, and their reconstruction algorithm estimates the true feature vector. Authors in [74] have proposed a MAP estimator based on the Huber prior for enhancing text images. The authors map the problem as that of a total variation and super-resolve the text. They consider images of scenes for which the point to point image transformation is a planar projective one.

It is now worth digressing a bit to look into the problem of image mosaicing. Mosaicing works on the principle that there are overlapping regions in the successive images so that interest points can be recovered in these regions and subsequently matched to compute the homography. Once the homography is computed, images are stitched together to obtain a high field of view mosaic. But while stitching these images across the overlapping regions, we throw away the additional information available from multiple views as redundant. This apparently redundant information is, however, the ideal cue for image super-resolution. The complementary set of information can be used for super-mosaicing purposes, [75] *i.e.,* to build a high resolution mosaic. An efficient super-resolution algorithm with application to panoramic mosaics has been proposed by Zomet and Peleg [76]. The method preserves the geometry of the original mosaic and improves spatial resolution. Capel and Zisserman have proposed a technique for automated mosaicing with super-resolution zoom in which a region of the mosaic can be viewed at a resolution higher than any of the original frames by fusing information from several views of a planar surface in order to estimate its texture [77]. Similarly, in [75], Bhosle *et al.* use the motion cue for super-resolution of a mosaic. They use the overlap among the observed images to increase the spatial resolution of the mosaic and to reduce the noise. In order to illustrate this, we show in Figure 2.1 a panoramic mosaic of a building constructed from 36 overlapped observations. The corresponding super-mosaic is displayed in Figure 2.2. One can notice an improvement in bringing out some of the finer details here.

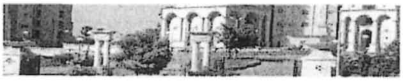


Fig. 2.1. Example of a low resolution panoramic mosaic.

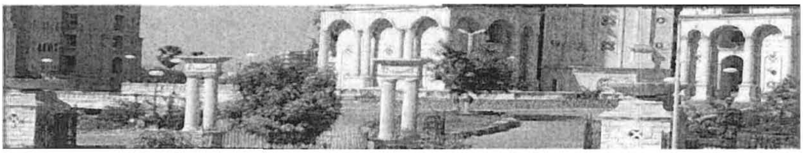


Fig. 2.2. Illustration of a super-mosaic constructed from the same set of observations used in obtaining Figure 2.1.

Now we discuss some of the research efforts in super-resolving a video sequence. Most of the super-resolution algorithms applicable to video are extensions of their single frame counterpart. Authors in [78] describe a complete model of video acquisition with an arbitrary input sampling lattice and a non-zero exposure time. They use the theory of POCS to reconstruct super-resolution still images or video frames from a low resolution time sequence of images. They restrict both the sensor blur and the focus blur to be constant during the exposure. Their video formation model includes an arbitrary space time lattice in order to obtain the sampled video signal. A hierarchical block matching algorithm is used to estimate the nonuniform translational motion between the low resolution images and the reference image. The motion model is incorporated into the video formation model to establish a linear space variant (LSV) relationship between the low resolution images and the desired super-resolved image at an arbitrary time t. By appropriately setting the values of t, a single super-resolved still image or a super-resolved video is reconstructed. Eren *et al.* extended the technique in [78] to scenes with multiple moving objects by introducing the concepts of validity maps and segmentation maps and by using the POCS framework [79]. The validity map disables projections based on observations with inaccurate motion information for a robust reconstruction whenever there is error in motion estimation. The segmentation map enables an object-based processing where a more accurate motion model can be utilized to improve the quality of reconstructed images.

In [80] a technique for robust deinterlacing for creating high quality stills from an interlaced video is presented. A method for motion compensated deinterlacing that combines a motion trajectory filter for removing the dominant motion such as camera zoom, pan and jitter, with motion detection to remove artifacts caused by independently moving objects has been discussed. The motion detection method employs an adaptive thresholding scheme that simultaneously suppresses aliasing artifacts and artifacts caused by independently moving objects.

Schultz and Stevenson use the hierarchical block matching algorithm to estimate the subpixel displacement vectors and then solve the problem of estimating the high resolution frame given a low resolution sequence by formulating it as a Bayesian MAP estimation with Huber-Markov random field (HMRF) prior, resulting in a constrained optimization problem with a unique minimum [81]. The super-resolution video enhancement technique proposed by Shah and Zakhor consider the fact that the motion estimates used in the reconstruction process will be inaccurate [82]. To this end their algorithm finds a set of candidate motion estimates instead of a single motion vector for each pixel, and then both the luminance and the chrominance values are used to compute the dense motion field with subpixel accuracy. The high resolution frame is restored subsequently by a method based on the Landweber algorithm.

Researchers have also used appropriate smoothness constraints over successive frames. Hong *et al.* define a multiple input smoothing convex functional and use it to obtain a globally optimal high resolution video sequence [83]. An iterative algorithm for resolution enhancement of a monochrome or a color video sequence using motion compensation has been presented in [84]. The choice of which motion estimator to use versus how the final estimates are obtained is weighed to see which issue is more critical in improving the estimated high resolution sequence. A single motion field is estimated using the three color fields. They use two different approaches for motion estimation, which recover the motion in two steps. In the first step, a displacement vector field (DVF) is estimated for each channel. In the second step, these three DVFs are combined via data fusion (merging the individual motion fields) to yield a single DVF. The straightforward examples of data fusion are the use of a prespecified vector corresponding to a particular color channel or the vector mean or the vector median. The estimated high resolution images using the block matching motion estimators have been compared to those obtained by using a pixel recursive scheme.

Altunbasak *et al.* [85] have proposed a motion-compensated, transform domain super-resolution procedure for creating high quality video or still images that directly incorporates the transform domain quantization information by working in the compressed bit stream. They apply this new formulation to MPEG-compressed video. In [86], a method for simultaneously estimating the high resolution image frames and the corresponding motion fields from a compressed low resolution video sequence is presented. The algorithm incorporates knowledge of the spatio-temporal correlation between low and high resolution images to estimate the original high resolution sequence from the degraded low resolution observation. The idea has been further extended to introduce additional high resolution frames in between two low resolution input frames to obtain a high resolution, slow motion sequencing of a given video [87]. The authors develop the above system for the purpose of *post-facto* video surveillance, *i.e.*, to find what exactly had happened from the stored video.

Authors in [88] propose a high-speed super-resolution algorithm using the generalization of Papoulis' sampling theorem for multichannel data with applications to super-resolving video sequences. They estimate the point spread function (PSF) for each frame and use the same for super-resolution. Borman and Stevenson [89] present a MAP approach for multi-frame super-resolution of a video sequence using the spatial as well as temporal constraints. The spatio-temporal constraint is imposed by using a motion trajectory compensated MRF model, in which the Gibbs distribution is dependent on pixel variation along the motion trajectory.

Most of the research works discussed so far assume that the low resolution image formation model illustrated in Figure 1.1, is indeed correct. Model uncertainties are not considered. In [90] the authors consider the problem of super-resolution restoration of the video, considering the model uncertainties caused by the inaccurate estimates of motion between frames. They use a Kalman filter based approach to solve the problem. For MPEG compressed data, quantization noise adds upto the uncertainties. Gunturk *et al.* propose a Bayesian approach for the super-resolution of MPEG-compressed video sequence considering both the quantization noise and the additive noise [91].

We observe that additional temporal data is used to improve the spatial resolution. Is it then possible to use additional spatial data (read high resolution image) to improve the temporal resolution? Or, in other words, can the concepts of resolution in space and time be fused together? This issue is discussed next. Shechtman *et al.* [92] construct a video sequence of high space-time resolution by combining information from multiple low resolution video sequences of the same dynamic scene. They used video cameras with complementary properties like low-frame rate but high spatial resolution and high frame-rate but low spatial resolution. They show that by increasing the temporal resolution using the information from multiple video sequences spatial artifacts such as motion blur can be handled without the need to separate static and dynamic scene components or to estimate their motion. To constrain the solution and provide numerical stability they use a space-time regularization term to impose the smoothness on the solution. A directional (or steerable) space-time regularization term applies smoothness only in directions where the derivatives are low, and does not smooth the space-time edges, thus preserving spatial edges as well as minimizing the motion blur due to the finite exposure time.

2.2 Motion-Free Super-Resolution

In the previous section we have discussed many different methods that use motion as the cue to generate the high frequency details. All these methods require a dense point correspondence among frames. Any error in establishing the correspondence affects the quality of super-resolution. Although the bulk of the work on super-resolution does use motion cue, of late, there has been work on using other possible cues. Motion-free super-resolution techniques try to obtain the spatial enhancement by using the cues which do not involve a motion among low resolution observations, thus avoiding the correspondence problem. One may expect an improved result since there would be no correspondence. However, we must find out what other cues can possibly be used as a substitute for the motion cue to bring in the high frequency details. We need to study how useful are these cues and what additional difficulties do they introduce during the super-resolution process. Another issue that comes out is how should we compare the performances of these methods with those of the motion-based methods. We simply cannot compare the methods as the data generation process is very different in both the cases. Further, the volume of work in this area is still quite small. Use of cues other than motion is the subject matter of this monograph. Before we discuss some of the specific methods in subsequent chapters, we begin reviewing some of the existing techniques in motion-free super-resolution.

Use of different amounts of blur is probably the first attempt in the direction towards motion-free super-resolution. In order to understand the problem let us take an example in 1D data. Let $f(n)$ be the unknown high resolution data, $g(m)$ be the observed data, $h_1(n)$ and $h_2(n)$ be the known finite impulse response (FIR) blurring kernels. Here the indices m and $n = 2m$ stand for the low and high resolution grids, respectively. We assume the decimation module to give us the average of the two adjacent pixels as the low resolution value. In order to explain the usefulness of the blur cue, let us further assume that the blur kernels are given by

$$h_1(n) = a_{11}\delta(n) + a_{12}\delta(n-1)$$

$$h_2(n) = a_{21}\delta(n) + a_{22}\delta(n-1),$$

where $\delta(n)$ is the delta function. Let us further assume that there is no observation noise. Then, neglecting boundary conditions,

$$g_1(m) = 0.5[a_{11}f(2m+1) + (a_{11}+a_{12})f(2m) + a_{12}f(2m-1)]$$

$$g_2(m) = 0.5[a_{21}f(2m+1) + (a_{21}+a_{22})f(2m) + a_{22}f(2m-1)]$$

Since the filter parameters are known the above two equations can easily be solved to obtain the high resolution data, provided the two blur kernels are linearly independent. Here we have $2m$ number of observations g_1 and g_2 and $2m$ number of unknowns in the high resolution signal f.

Hence we observe that it is, indeed, possible to use the differential blur as a cue for super-resolution. Definitely, there will be issues of sensor noise, availability of sufficient number of observations, smoothness of the reconstructed image, *etc.* This calls for the use of regularizing priors to solve the restoration problem.

A MAP-MRF based super-resolution technique has been proposed by Rajan *et al.* in [93]. Here the authors consider an availability of decimated, blurred and noisy versions of a high resolution image which are used to generate a super-resolved image. A known blur acts as a cue in generating the high resolution image. They model the high resolution image as an MRF to serve as a prior for regularization. In chapter 3 we shall relax the assumption of the known blur and extend it to deal with an arbitrary space-varying defocus blur for super-resolution purposes. Recently, Rajagopalan and Kiran [94] have proposed a frequency domain approach for estimating the high resolution image using the defocus cue. They derive the Cramer-Rao lower bound (CRLB) for the covariance of the error in the estimate of the super-resolved image and show that the estimate becomes better as the relative blur increases.

A scheme for image high resolution from several blurred observations by imposing a periodic grating with various absorptions in the object field is proposed in [95]. This method is based on the solution of a Fredholm's integral equation of the first kind. The method can be employed in different fields such as microscopy and for signal and image transmission under conditions of heavy blur. The super-resolution here is based on an interference of spatial frequencies of the object and the grating.

There has also been an effort in using a functional decomposition approach for super-resolution. One such example is the use of generalized interpolation [96]. Here a space containing the original function is decomposed into appropriate subspaces. These subspaces are chosen so that the rescaling operation preserves properties of the original function. On combining these rescaled sub-functions, they get back the original space containing the scaled or zoomed function. Here the photometric information is used as the cue. The authors in [18] proposed a multi-objective super-resolution technique for super-resolving both the intensity field and the structure using blur and shading as cues. It is shown in the paper that the use of the blur and the shading cues can be combined under a common mathematical framework. All these methods discussed thus far assume the availability of multiple observations of the same scene under different camera or lighting conditions. However, at times one may have to do with a single observation. What if you are given a low resolution image of a suspected criminal? Can this picture be super-resolved?

Researchers have also attempted to solve the super-resolution problem by using learning based techniques. These methods try to recognize the local features in a low resolution image and then retrieve the most likely high frequency information from the given training samples. In this book, these methods are also classified under motion-free super-resolution as the new information required for predicting the high resolution image is obtained from the training images rather than from the subpixel shifts among low resolution observations. Authors in [97] describe image interpolation algorithms which use a database of training images to create plausible high frequency details in zoomed images. They propose a learning framework called VISTA - Vision by Image/Scene TrAining. By blurring and down-sampling sharply defined images they construct a training set of sharp and blurred images. These are then incorporated into a Markov network to learn their relationship. A Bayesian belief propagation allows to find the maximum of the posterior probability.

A quite natural extension to the above is to use the best of the both world - information from multiple observations as discussed earlier and the priors learnt from a given high resolution training data set. Capel and Zisserman have proposed a super-resolution technique from multiple views using learnt image models [98]. Their method uses learnt image models either to directly constrain the ML estimate or as a prior for a MAP estimate. To learn the model, they use principal component analysis (PCA) applied to a face image database. Researchers have also attempted to combine the motion cue with the learning based method for super-resolution restoration. Pickup *et al.* [99] combine the motion information due to subpixel displacements as well as motion-free information in the form of learning of priors to propose a domain specific super-resolution using the sampled texture prior. They use training images to estimate the density function. Given a small patch around any particular pixel, they learn the intensity distribution for the central pixel by examining the values at the centers of similar patches available in the training data. The intensity of the original pixel to be estimated is assumed to be Gaussian distributed with mean equal to the learnt pixel value and obtain the super-resolution by minimizing a cost function.

There has also been some effort on applying an output feedback while super-resolving the images. If the purpose of super-resolution is to recognize a face, a character or a fingerprint, then the partially super-resolved image is first matched to a database to extract the correct match and then this information can be used to enhance the prior for further improving the image quality. In [100] Baker and Kanade develop a super-resolution algorithm by modifying the prior term in the cost to include the results of a set of recognition decisions, and call it as recognition-based super-resolution or hallucination. Their prior enforces the condition that the gradient of the super-resolved image should be equal to the gradient of the best matching high resolution training image. The learning of the prior is done by using a pyramidal decomposition.

An image analogy method applied to super-resolution is discussed by Hertzmann *et al.* in [101]. They use the low resolution and the high resolution versions of a portion of an image as the training pairs which are used to specify a "super-resolution" filter that is applied to a blurred version of the entire image to obtain an approximation to the high resolution original image. Here the emphasis is in learning the local statistics at a finer details. Candocia and Principe [102] address the ill-posedness of the super-resolution problem by assuming that the correlated neighbors remain similar across scales, and this apriori information is learnt locally from the available image samples across scales. When a new image is presented, a kernel that best reconstructs each local region is selected automatically and the super-resolved image is reconstructed by a simple convolution operation.

So far all these learning based methods are restricted to dealing with enhancing a still frame only. A learning based method for super-resolution enhancement of a video has been proposed by Bishop *et al.* [103]. Their approach builds on the principle of example based super-resolution for still images proposed by Freeman *et al.* [97]. They use a learnt data set of image patches capturing the relationship between the middle and the high spatial frequency bands of natural images and use an appropriate prior over such patches. A key concept there is the use of the previously enhanced frame to provide part of the training set for super-resolution enhancement of the current frame.

Having discussed the current research status in super-resolution imaging, we concentrate on a few specific ways of achieving motion-free super-resolution. These methods are discussed in detail in the subsequent chapters.

3

Use of Defocus Cue

This chapter introduces a technique to simultaneously estimate the depth map and the focused image of a scene, both at a super-resolution, from its defocused low resolution observations. The super-resolution technique has hitherto been restricted mostly to the intensity domain. We extend the scope of super-resolution imaging to acquire depth estimates at high spatial resolution simultaneously. Given a sequence of low resolution, blurred and noisy observations of a static scene, the problem is to generate a dense depth map at a resolution higher than one that can be generated from the observations as well as to estimate the true high resolution focused image. This is definitely an ill-posed problem and hence we need a proper regularization. Both the depth and the image are modeled as separate Markov random fields (MRF) to provide the necessary prior and a maximum *a posteriori* estimation method is used to recover the high resolution fields. Similar to the motion cue, defocus cue is the most natural cue in a real aperture imaging system, *i.e.*, a lens with a finite aperture. We demonstrate the use of defocus cue in super-resolving an image. Since there is no relative motion between the scene and the camera, as is the case with most of the super-resolution and structure recovery techniques, we do away with the correspondence problem. We explain the method and demonstrate its applicability through experimental results.

3.1 Introduction

It was Pentland who first suggested that measuring the amount of blurring at a given point in the image could lead to computing the depth at the corresponding point in the scene, provided the parameters of the lens system like aperture, focal length and lens-to-image plane distance are known [104]. Given two images of a scene recorded with different camera settings, we obtain two constraints on the spread parameters of the point spread function corresponding to the two images. One of the constraints is obtained from the geometry of image formation while the other is obtained from the intensity values in the defocused images. These two constraints are simultaneously solved to determine distances of objects in the scene [105].

In this chapter, we expand the scope of super-resolution to include high resolution depth recovery in a scene, in addition to restoring intensity values. As mentioned in chapter 1, one of the degradations in a low resolution image is the sensor related blur which appears as a consequence of the low resolution point spread function of the camera. Blurring can also arise due to relative motion between the camera and the scene. Note that all subsequent discussions in this chapter assume that the lens has a finite aperture and we cannot assume a pin-hole model of the camera. In the case of real aperture imaging, we know that the blur at a point is a function of the depth in the scene at that point. Thus, we notice that blur is a *natural* cue in a low resolution image formed by a real-aperture camera. We exploit this blur to recover the depth map through the depth from defocus formulation. We demonstrate how the depth map can be estimated at a resolution higher than one that can be normally extracted from such observations. We may call such a dense spatial depth map as the *super-resolved depth*. In addition to this, we show how to simultaneously estimate the true, high resolution focused image of the scene. This process may be called super-resolved, space varying restoration. Thus this is also a problem of blind restoration. The two stage process of identifying the blur and deconvolving the observed image with the corresponding PSF performs unsatisfactorily in the presence of noise [106]. In this chapter, we demonstrate that these two tasks, namely, the super-resolved depth recovery and the super-resolved, space varying image restoration, can be combined through the interplay of two separate Markov random fields (MRFs) - one representing the depth map and the other representing the intensity field.

3.2 Depth from Defocus

In the era of automatic cameras, we often forget that a typical camera has parameters like focus, aperture and shutter. Unfortunately, we must use a manual camera in this chapter. Since we assume that there is no relative motion between the scene and the camera, the shutter speed does not play any role. But we do vary the other two parameters in the work explained in this chapter. Most of the works on computer vision assume that the camera model is a pin-hole one and hence everything in the scene is always sharply in focus. This is no longer true in a real aperture imaging system.

Since the degree of defocus is a function of lens setting and the depth of the scene, it is possible to recover the depth at a point if the amount of blur can be estimated, provided the lens setting is known. An out-of-focus point light source images into a blurred circle [107], whose radius is described by a blur parameter σ defined as

$$\sigma = \varrho \; \zeta \; v \; (\frac{1}{F_l} - \frac{1}{v} - \frac{1}{u}), \tag{3.1}$$

where F_l is the focal length, u is the distance of the object point from the lens, v is the distance between the lens and the image detector, ζ is the radius of the lens aperture and ϱ is a camera constant that depends on its optics and CCD array resolution. The above relationship is valid primarily in geometric optics and when the lens suffers from no aberrations. In the literature, we encounter two kinds of blur, viz. the Gaussian blur and the pillbox (circular) blur. We have used the Gaussian blur here for computational ease, although the circular blur will work equally as well. As a matter of fact any single parameter class of PSF can be handled under the current formulation.

Figure 3.1 illustrates the formation of the image of an object point as a circular patch due to the dislocation of the image plane from the focusing plane. Since the depth at various points in the scene may be varying continuously, σ would also vary all over the image accordingly. The shift-varying PSF of the optical system is modeled as a circularly symmetric 2D Gaussian function

$$h(i,j;m,n) = \frac{1}{2\pi[\sigma(m,n)]^2} \exp(-\frac{i^2+j^2}{2[\sigma(m,n)]^2}). \tag{3.2}$$

It is quite well known that the blurring due to the lens may be modeled by a linear operator [108]. Hence, the superposition theorem

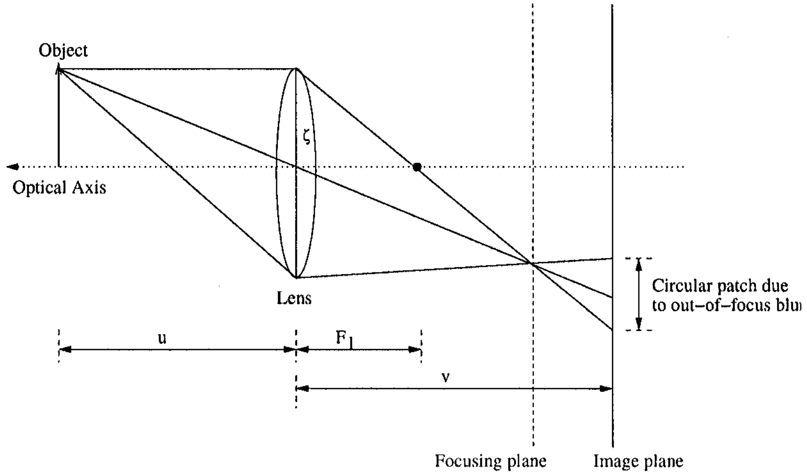


Fig. 3.1. Illustration of out-of-focus image formation.

holds with respect to the light distribution on the image plane. Since the PSF is a function of the depth, the defocusing process is linear, but *shift-variant*. Shift-invariance may, however, be assumed for subimages over which the depth is nearly constant.

Several approaches have been proposed in the literature for recovering the depth from defocused images. We describe here the popular approach due to Subbarao [109]. In this scheme, two defocused images of the scene are obtained by choosing different sets of lens parameters. For two different lens settings, we get

$$\sigma_k = \varrho\ \zeta_k\ v_k\ (\frac{1}{F_{lk}} - \frac{1}{v_k} - \frac{1}{u}),\quad k = 1, 2. \tag{3.3}$$

Eliminating the unknown object distance u from the above equations, we obtain the relation

$$\sigma_1 = \alpha\sigma_2 + \beta, \tag{3.4}$$

where $\alpha = \frac{\zeta_1 v_1}{\zeta_2 v_2}$ and $\beta = \varrho\zeta_1 v_1 \left(\frac{1}{F_{l1}} - \frac{1}{v_1} - \frac{1}{F_{l2}} + \frac{1}{v_2}\right)$.

The above equation which gives a relation between σ_1 and σ_2 in terms of the known camera parameters, plays a central role in depth recovery.

Given two defocused images, a local region $g_k(x, y)$ about the location (x, y) in the k^{th} observed image can be expressed as

$$g_k(x, y) = h_k(x, y) * f(x, y) \tag{3.5}$$

Where the operator $*$ represents convolution, $f(x,y)$ is the corresponding local region in the focused (pin-hole equivalent) image of the scene while $h_k(x,y)$ is the PSF corresponding to the depth of the scene at that point in the k^{th} defocused image. In Eq. (3.5) it is assumed that the depth is constant over the local region. If the depth is not constant within the subimage, then the method would give the average depth. Let $\hat{F}(w_x,w_y)$, $\hat{G}_k(w_x,w_y)$, and $\hat{H}_k(w_x,w_y)$ be the Fourier transforms of $f(x,y)$, $g_k(x,y)$ and $h_k(x,y)$, respectively. Hence, from Eq. (3.5) we have

$$\hat{G}_k(w_x,w_y) = \hat{H}_k(w_x,w_y)\hat{F}(w_x,w_y).$$

Dividing $\hat{G}_1(w_x,w_y)$ by $\hat{G}_2(w_x,w_y)$, the unknown $\hat{F}(w_x,w_y)$ can be eliminated. Since the PSF can be approximated by the Gaussian function, we obtain

$$\frac{\hat{G}_1(w_x,w_y)}{\hat{G}_2(w_x,w_y)} = \exp\left[-\frac{1}{2}({w_x}^2+{w_y}^2)({\sigma_1}^2-{\sigma_2}^2)\right].$$

Taking the logarithm on both sides and rearranging terms, we get

$${\sigma_1}^2-{\sigma_2}^2 = \frac{-2}{{w_x}^2+{w_y}^2}\log\frac{\hat{G}_1(w_x,w_y)}{\hat{G}_2(w_x,w_y)}.$$

For some (w_x,w_y), the righthand side of this equation can be computed from the given image pair. Therefore, ${\sigma_1}^2-{\sigma_2}^2$ can be estimated from the given observations. Measuring the Fourier transform at a single point (w_x,w_y) is, in principle, sufficient to obtain the value of ${\sigma_1}^2-{\sigma_2}^2$, but a more robust estimate can be obtained by taking the average over some domain in the frequency space. Let the estimated average be C, which is given by

$$C = \frac{1}{\mathcal{A}}\int\int_A \frac{-2}{{w_x}^2+{w_y}^2}\log\frac{\hat{G}_1(w_x,w_y)}{\hat{G}_2(w_x,w_y)}dw_x dw_y. \qquad (3.6)$$

Where A is the region in the (w_x,w_y) space not containing the singularities, if any, and $\mathcal{A}$ is the area of A. Therefore, from the observed images, we get the following constraint

$${\sigma_1}^2-{\sigma_2}^2 = C. \qquad (3.7)$$

While Eq. (3.4) gives a relation between σ_1 and σ_2 in terms of the camera parameters, the above equation gives an estimate of the relative blur between the defocused images. Eqs. (3.4) and Eq. (3.7) together constitute two equations in two unknowns. From these equations, we get

$$(\alpha^2 - 1){\sigma_2}^2 + 2\alpha\beta\sigma_2 + \beta^2 = C. \tag{3.8}$$

In Eq. (3.8), we have a quadratic equation in σ_2 that can be easily solved. Given the lens parameters, one can then estimate the depth u from the value of σ_2 using Eq. (3.1). The procedure is now repeated at every pixel location to obtain the complete depth map. Thus, the depth of the scene is determined from only two observations obtained with different camera parameter settings.

For the general situation, where the depth at various points in the scene may be varying continuously, σ would also vary all over the image. The transformation from the focused image to the defocused image is still linear but no longer shift-invariant. Hence, the intensity at the location (x, y) in the k^{th} defocused image is given by

$$g_k(x, y) = \int\int f(t, \tau) h_k(t, \tau; x, y) dt d\tau, \tag{3.9}$$

where $f(x, y)$ is the focused image of the scene, and $h_k(t, \tau; x, y)$ is the space varying PSF corresponding to the k^{th} defocused image. Thus, in its generality, the recovery of depth from defocused images is a space-variant blur identification problem.

Early investigations of the DFD problem were carried out by Pentland [107], where he compared two images locally, one of which was formed with a pin-hole aperture and then recovered the blur parameter through deconvolution in the frequency domain. In [109], Subbarao removed the constraint of one image being formed with a pin-hole aperture by allowing several camera parameters like aperture, focal length and lens-to-image plane distance to vary simultaneously. Prasad *et al.* [110] formulate the DFD problem as a 3D image restoration problem. The defocused image is modeled as the combinatorial output of the depths and intensities of the volume elements (voxels) of an opaque 3D object. Klarquist *et al.* propose a maximum likelihood (ML) based algorithm that computes the amount of blur as well as the deconvolved images corresponding to each sub-region [111]. In [112], Gökstorp estimates blur by obtaining local estimates of the instantaneous frequency, amplitude and phase using a set of Gabor filters in a multi-resolution framework. In [113], Watanabe and Nayar describe a class of broadband rational operators which are capable of providing invariance to the scene texture. Schechner and Kiryati address the similarities in DFD and stereo techniques on the basis of the geometric triangulation principle in [114]. An active ranging device that uses an optimized illumination pattern to obtain an accurate and high resolution depth map is described by Nayar *et al.* in [115]. In [116], Rajagopalan *et al.* use the complex spectrogram and the pseudo-Wigner distribution for recovering the depth within the framework of the space-frequency representation of the image. In [117], they extend this approach to impose smoothness constraints on the blur parameter to be estimated and use a variational approach to recover the depth. In [118], a MAP-MRF framework is used for recovering the depth as well as the focused image of a scene from two defocused images. However, the recovered depth map and the scene image are at the same resolution as the observations. Other techniques for depth recovery and issues related to optimal camera settings are described in [105]. The DFD problem can also be viewed as a special case of the space-variant blur identification problem since eventually it is the blur at a point that acts as the cue for determining the depth in the scene at that point.

In this chapter, our aim is not only to recover the depth from defocused images, but also to do so at a higher spatial resolution, besides generating the super-resolved image of the scene. Thus, given a sequence of low resolution blurred observations of size $M_1 \times M_2$, we wish to generate a dense depth map of size, say $rM_1 \times rM_2$, where r is the upsampling factor. Clearly, by doing this, we get a more accurate description of the depth in the scene, which eventually leads to a better performance of the computer vision task at hand.

Since we are discussing about the super-resolution of both the intensity image and the depth map, let us look at some of the prior work on high resolution depth estimation. Shekarforoush *et al.* use MRFs to model the images and obtain a 3D high resolution visual information (albedo and depth) from a sequence of displaced low resolution images [119]. The effect of sampling a scene at a higher rate is acquired by having interframe subpixel displacements. But they do not consider blurred observations. Cheeseman *et al.* describe another Bayesian approach for constructing a super-resolution surface by combining information from a set of images of the given surface [120]. Their model includes registration parameters, the PSF and camera parameters that are estimated first and subsequently the surface reconstruction is carried out. In both these cases, the issue of registration has to be addressed since they involve camera displacement. As discussed earlier, errors in registration are reflected on the quality of the super-resolved image generated as well as on the depth estimate. Hence if we can avoid any relative motion between the camera and the scene, we would be able to do away with the correspondence problem. This is precisely what is achieved by resorting to using the defocus cue as it is commonly done in the depth from defocus approach. However, the restoration problem becomes a space varying one and the accuracy would depend on how well we can estimate the blur parameter.

3.3 Low Resolution Imaging Model

We now discuss the formation of a low resolution image from a high resolution description of the scene. Note that the problem we solve here is actually the inverse. A preliminary level of discussion was carried out in chapter 1. Now we make it specific to the problem. Once again we remind that through the choice of the term super-resolved depth or the intensity, we mean the enhancement in the spatial resolution and not that of the quantization levels of the depth or the intensity map.

A high resolution image of the scene is formed by the camera optics and this image is defocused due to varying depth components in the scene. The defocused scene is now sensed optically by the low resolution CCD elements. A sensor noise is now added to these measurements and one obtains the observed image.

Let us refer to Figure 3.2 for an illustration of the above model. Let $z(k,l)$ and $\sigma(k,l)$ be the true high resolution intensity and blur (parameterized) maps, respectively. Due to depth related defocus, one obtains the blurred but high spatial resolution intensity map $z^{'}(k,l)$. The low resolution image sensor plane is divided into $M_1 \times M_2$ square sensor elements and $\{g(i,j)\}$, $i = 0,\ldots,M_1-1$ and $j = 0,\ldots,M_2-1$ are the low resolution intensity values. For a decimation ratio of r, the high resolution grid will be of size $rM_1 \times rM_2$. The forward process of obtaining $\{g(i,j)\}$ from $\{z^{'}(k,l)\}$ is written as [50]

$$g(i,j) = \frac{1}{r^2} \sum_{k=ri}^{r(i+1)-1} \sum_{l=rj}^{r(j+1)-1} z^{'}(k,l) \tag{3.10}$$

i.e., the low resolution intensity is the average of the high resolution intensities over a neighborhood of r^2 pixels. This decimation model simulates the integration of light intensity that falls on the high resolution detector. It should be mentioned here that the above relationship assumes that the entire area of the pixel is used for light sensing and nothing is used for electrical wiring or insulation. Thus, we assume the fill-factor for the CCD array to be unity.

The process of blurring the high resolution image $z(k,l)$ due to defocus is modeled by

$$z^{'}(k,l) = \sum_i \sum_j z(i,j)h(k,l;i,j) \tag{3.11}$$

where $z^{'}(\cdot)$ is the defocused version of the high resolution image and $h(\cdot;\cdot)$ is the space variant blurring function given in the previous section.

The space varying blurring function is dependent only on a single blur parameter σ. However, in the present context, this blur describes a high resolution representation of the depth field compared to the spatial resolution at which the scene is observed. The addition of white Gaussian noise at the CCD sensor completes the low resolution observation model and is illustrated in Figure 3.2.

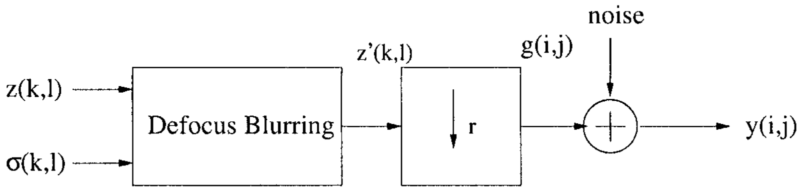


Fig. 3.2. Low resolution defocused image formation from high resolution image. Here the symbol $\downarrow r$ denotes the decimation by a factor of r.

We now define the super-resolution problem in a restoration framework. There are K observed images $y_m(i,j)$, $m = 1, 2, \cdots, K$, each of size $M_1 \times M_2$. These images are defocused, decimated and noisy versions of a single high resolution image $z(k,l)$ of size $rM_1 \times rM_2 = N_1 \times N_2$. If $\mathbf{y}_m$ is the $M_1M_2 \times 1$ lexicographically ordered vector containing pixels from the low resolution image $y_m(i,j)$, then a vector $\mathbf{z}^{'}$ of size $r^2M_1M_2 \times 1$ containing pixels of the high resolution defocused image can be formed by placing each of the $r \times r$ pixel neighborhoods sequentially so as to maintain the relationship between a low resolution pixel and its corresponding high resolution pixel. This relationship can also be represented through a decimation matrix D of size $M_1 M_2 \times r^2 M_1 M_2$ consisting of values of $\frac{1}{r^2}$ at r^2 locations in each row (using a proper reordering of $\mathbf{z}'$). For a decimation factor of r, $N_1 = rM_1$ and $N_2 = rM_2$ decimation matrix D has the form [50]

$$D = \frac{1}{r^2} \begin{bmatrix} 11\ldots1 & & & \mathbf{0} \\ & 11\ldots1 & & \\ & & \ddots & \\ & \mathbf{0} & & 11\ldots1 \end{bmatrix}. \tag{3.12}$$

As an example, consider an observation of size 2×2. For a decimation factor of $r = 2$ the size of $\mathbf{z}'$ becomes 4×4. Now with lexicographically ordered $\mathbf{z}'$ of size, say 16×1, the D matrix is of size 4×16 and can be written as (after reordering of $\mathbf{z}'$)

$$D = \frac{1}{4} \begin{bmatrix} 1111000000000000 \\ 0000111100000000 \\ 0000000011110000 \\ 0000000000001111 \end{bmatrix}. \tag{3.13}$$

The D matrix can also be expressed without reordering of $\mathbf{z}'$ as

$$D = \frac{1}{4} \begin{bmatrix} 1100110000000000 \\ 0011001100000000 \\ 0000000011001100 \\ 0000000000110011 \end{bmatrix}. \tag{3.14}$$

This decimation model simulates the integration of light intensity that falls on the high resolution detector. We use the same decimation model throughout in the book.

Let H be the blur matrix corresponding to the space variant blurring function $h(k,l;i,j)$ in Eq. (3.11). We reiterate that this function is governed by the high resolution blur $\sigma(k,l)$. Thus, for the m^{th} low resolution observation, the blur matrix H_m is a function of $\sigma_m(k,l)$, where the earlier notation is modified to include a subscript for the observation number. The blur field is now lexicographically ordered to obtain a vector $\mathbf{s}_m$. The blur matrix $H_m = H(\mathbf{s}_m)$ corresponding to the m^{th} observation can now be formed but due to the space varying nature of the blur, it does not possess a block Toeplitz structure. The image formation model is now compactly written as

$$\mathbf{y}_m = DH(\mathbf{s}_m)\mathbf{z} + \mathbf{n}_m, \quad m = 1, \ldots, K \tag{3.15}$$

where $H(\mathbf{s}_m)$'s are the high resolution space varying blurring matrix (PSF) of size $r^2M_1M_2 \times r^2M_1M_2$ and D is the decimation matrix. Here $\mathbf{n}_m$ is the $M_1M_2 \times 1$ noise vector which is zero mean i.i.d, and hence the multivariate noise probability density function is given by

$$P(\mathbf{n}_m) = \frac{1}{(2\pi{\sigma_n}^2)^{\frac{M_1M_2}{2}}} \exp\left\{-\frac{1}{2{\sigma_n}^2}\mathbf{n}_m^T\mathbf{n}_m\right\}, \tag{3.16}$$

where ${\sigma_n}^2$ denotes the variance of the noise process and K is the number of low resolution observations. Thus the model consists of a collection of low resolution images, each of which differs from the others in the blur matrix, which is akin to changing the focus of a stationary camera looking at a stationary scene.

3.4 MRF Model of Scene

Since the restoration problem is an ill-posed problem, we plan to model the scene in such a way that the model parameters can be used as priors for the purpose of obtaining a regularized solution. In the area of image processing and computer vision, stochastic modeling plays an important role. The Markov random field (MRF) provides a convenient and consistent way of modeling context dependent entities such as image pixels, depth of the object and other spatially correlated features. This is achieved through characterizing mutual influence among such entities over a spatial neighborhood using the MRF modeling. Until the equivalence between the MRF and the Gibbs random field (GRF) was discovered, the power of MRF as a spatial interaction model was not fully exploited. The Gibbs distribution was introduced in 1925 by Ising [121], where he used the same to model the molecular interaction in ferromagnetic materials. The difficulties involved in the use of MRF, described by the conditional distribution are now eliminated because the joint distribution is readily available with the GRF characterization. A GRF describes the global properties of an image, while an MRF is described in terms of local properties. The practical use of MRF models is largely ascribed to the equivalence between MRFs and Gibbs distributions (GRF) established by Hammersley and Clifford [122].

We now briefly introduce the concept of MRF for completeness purposes. Consider a lattice L described by a square array of pixels $\{0 \leq (i,j) \leq N-1\}$. A random field has a joint probability distribution for an N^2 dimensional vector $\mathbf{z}$, which contains the random variable

z_t as the 'label' at site t. The label could be gray values, pattern classes, *etc.* A collection of subsets of a lattice L defined as

$$\mathcal{N} = \{\mathcal{N}_{i,j} : \ (i,j) \in L, \mathcal{N}_{i,j} \subset L\}$$

is a neighborhood system on L, if and only if the neighboring relationship has the following properties.

- A site is not a neighbor to itself: $(i,j) \notin \mathcal{N}_{i,j}$
- The neighboring relationship is mutual: if $(k,l) \in \mathcal{N}_{i,j}$ then $(i,j) \in \mathcal{N}_{k,l}$ for any $(i,j) \in L$.

A hierarchically ordered sequence of neighborhood systems that are commonly used in the context of image modeling consists of $\mathcal{N}^1, \mathcal{N}^2, \cdots$ neighborhood systems. In general $\mathcal{N}^m$ is called the m^{th} order neighborhood. In the first order neighborhood system, every site has four neighbors as shown in Figure 3.3(a) where (i,j) denotes the site considered and 0's its neighbors. In the second order neighborhood system there are eight neighbors for every site as shown in Figure 3.3(b). The pair $(L, \mathcal{N})$ constitutes a graph where L contains the nodes and $\mathcal{N}$ determines the link between the nodes. A *clique* c for $(L, \mathcal{N})$ is defined as a subset of sites in L in which all pairs of sites are mutual neighbors. Cliques can occur as singletons, doublets, triplets and so on. The cliques corresponding to the first order neighborhood and the second order neighborhood system are shown in Figure 3.3(c).

Let Z be a random field over an arbitrary $N \times N$ lattice of sites $L = \{(i,j) | 0 \leq i,j \leq N-1\}$. From the Hammersley-Clifford theorem [123] which proves the equivalence of an MRF and a GRF, we have

$$P(Z = \mathbf{z}) = \frac{1}{Z_z} e^{-U(\mathbf{z})},$$

where $\mathbf{z}$ is a realization of Z, Z_z is the partition function given by $\sum_z e^{-U(\mathbf{z})}$ and $U(\mathbf{z})$ is the energy function given by

$$U(\mathbf{z}) = \sum_{c \in C^z} V_c^z(\mathbf{z}).$$

$V_c^z(\mathbf{z})$ denotes the potential function of clique c and C^z is the set of all cliques.

Let us now model the scene intensity by an MRF. The lexicographically ordered high resolution image $\mathbf{z}$ satisfying the Gibbs density function is now written as

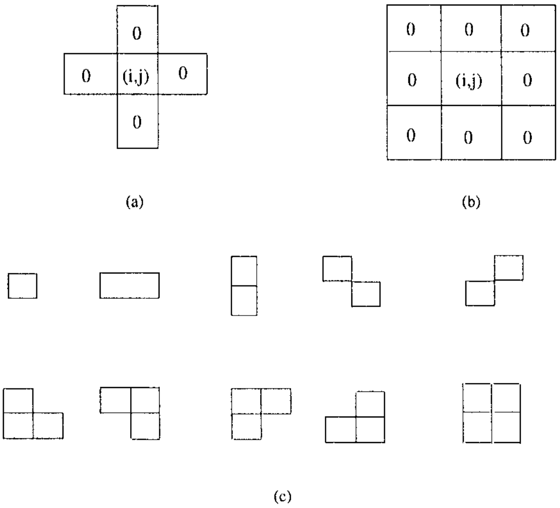


Fig. 3.3. (a) First and (b) second order neighborhood, and (c) cliques associated with (b).

$$P(\mathbf{z}) = \frac{1}{Z_z} \exp\left\{ - \sum_{c \in C^z} \mathrm{V}_c^z(\mathbf{z}) \right\}. \tag{3.17}$$

In order to employ a simple and fast minimization technique like gradient descent, it is desirable to have a convex energy function. To this end we can consider pair wise cliques on a first order neighborhood and impose a quadratic cost which is a function of finite difference approximations of the first order derivative at each pixel location, *i.e.*,

$$\sum_{c \in C^z} V_c^z(\mathbf{z}) = \frac{1}{\lambda} \sum_{k=1}^{N-1} \sum_{l=1}^{N-1} [(z_{k,l} - z_{k,l-1})^2 + (z_{k,l} - z_{k-1,l})^2], \tag{3.18}$$

where λ represents an appropriate weight for the prior and as the penalty for departure from the smoothness in $\mathbf{z}$.

The MRF prior serves as a contextual constraint to regularize the solution. These constraints push the reconstruction towards a smooth entity. Although this helps to stabilize the minimization process, it flattens the entity to be super-resolved causing distortions across discontinuities. The solution to this is to use a prior which preserves the discontinuities. Unlike in the one dimensional signals, it is well known that in images, pixels with significant change in intensities carry important information. In order to incorporate provisions for detecting such discontinuities, Geman and Geman [124] introduced the concept of line fields located on a dual lattice. The horizontal line field element ${l^z}_{i,j}$ connecting site (i, j) to $(i-1, j)$ aids in detecting a horizontal edge while the vertical line field element ${v^z}_{i,j}$ connecting site (i, j) to $(i, j-1)$ helps in detecting a vertical edge. We have chosen ${l^z}_{i,j}$ and ${v^z}_{i,j}$ to be binary variables over the line fields $\mathcal{L}^z$ and $\mathcal{V}^z$. The on-state of the line-process variable indicates that a discontinuity, in the form of a high gradient, is detected between neighboring points, *e.g.*,

$$\begin{aligned} {l^z}_{i,j} &= 1 \text{ if } |z_{i,j} - z_{i-1,j}| > \theta_1 \\ &= 0 \text{ else} \end{aligned} \tag{3.19}$$

Similarly

$$\begin{aligned} {v^z}_{i,j} &= 1 \text{ if } |z_{i,j} - z_{i,j-1}| > \theta_2 \\ &= 0 \text{ else,} \end{aligned} \tag{3.20}$$

where θ_1 and θ_2 are appropriate threshold values for declaring a discontinuity. It may be mentioned here that we make use of $\theta_1 = \theta_2 = \theta$ in all our experiments, *i.e.*, the thresholds for detecting the horizontal and vertical edges are kept the same. Each turn-on of a line process variable is penalized by a quantity γ_z so as to prevent spurious discontinuities. Else a trivial solution would declare a discontinuity at every location. Thus the energy function for the random process Z with discontinuity fields $\mathcal{L}^z$ and $\mathcal{V}^z$ is obtained by modifying Eq. (3.18)

$$\begin{aligned} U(\mathbf{z}, \mathbf{l}, \mathbf{v}) &= \sum_{c \in \mathcal{C}^z} {V_c}^z(\mathbf{z}, \mathbf{l}, \mathbf{v}) \\ &= \sum_{i,j} [\mu_z \{(z_{i,j} - z_{i,j-1})^2 (1 - {v^z}_{i,j}) + (z_{i,j+1} - z_{i,j})^2 (1 - {v^z}_{i,j+1}) \\ &+ (z_{i,j} - z_{i-1,j})^2 (1 - {l^z}_{i,j}) + (z_{i+1,j} - z_{i,j})^2 (1 - {l^z}_{i+1,j})\} \\ &+ \gamma_z \{{l^z}_{i,j} + {l^z}_{i+1,j} + {v^z}_{i,j} + {v^z}_{i,j+1}\}] \\ &= \sum_{i,j} [\mu_z e_{zs} + \gamma_z e_{zp}], \end{aligned} \tag{3.21}$$

where e_{zs} and e_{zp}, respectively, are the smoothness term and the penalty term necessary to prevent occurrence of spurious discontinuities. Here μ_z represents the penalty for departure from the smoothness.

We use this particular energy function in our studies in order to preserve discontinuities in the restored image. Any other form of energy function can also be used without changing the solution modality discussed here.

3.5 Modeling the Scene Depth

In the previous section, we showed that the image intensity can be modeled by an MRF. In [125] and [126], the depth of a scene is modeled as an MRF. This is justified because the change in depth of a scene is usually gradual and hence the depth can be said to exhibit a local dependency. Since the space varying blurring parameter σ in Eq. (3.1) is a function of the scene depth, we expect it to exhibit similar local dependencies and model it by an MRF. Let S be a random field defined over the same high resolution lattice L that denotes the defocus blur $\sigma(k,l)$ at that point. The advantage of defining the field over the blur parameter σ instead of over the depth values is that it reduces the mathematical complication over defining the blur matrix H in Eq. (3.15). Once σ is found out, the depth can be obtained using the known camera parameters.

The lexicographically ordered high resolution blur (or depth) field $\mathbf{s}$ satisfy the Gibbs density function

$$P(\mathbf{s}) = \frac{1}{Z_s} \exp\{-\sum_{c \in \mathcal{C}^s} V_c^s(\mathbf{s})\}.$$

Here Z_s is the corresponding partition function. We may use the same clique set $\mathcal{C}^z$ as in the case of the intensity field $\mathbf{z}$ or select a different set of cliques $\mathcal{C}^s$. We are now free to select any potential function function $V_c^s(\mathbf{s})$ as we deem fit for the problem. In this particular example we consider the same potential function as it was chosen for the intensity field. Thus the energy function for the random field S with corresponding discontinuity fields $\mathcal{L}^s$ and $\mathcal{V}^s$ is given by

$$\begin{aligned} U(\mathbf{s},\mathbf{l},\mathbf{v}) &= \sum_{c \in \mathcal{C}^s} V_c{}^s(\mathbf{s},\mathbf{l},\mathbf{v}) \\ &= \sum_{i,j} [\mu_s\{(s_{i,j} - s_{i,j-1})^2(1 - v^s{}_{i,j}) + (s_{i,j+1} - s_{i,j})^2(1 - v^s{}_{i,j+1}) \\ &+ (s_{i,j} - s_{i-1,j})^2(1 - l^s{}_{i,j}) + (s_{i+1,j} - s_{i,j})^2(1 - l^s{}_{i+1,j})\} \end{aligned}$$

$$+ \gamma_s\{l^s{}_{i,j} + l^s{}_{i+1,j} + v^s{}_{i,j} + v^s{}_{i,j+1}\}]$$
$$= \sum_{i,j}[\mu_s e_{ss} + \gamma_s e_{sp}]. \tag{3.22}$$

Here the superscript (or the subscript) s denotes that the fields (or parameters) correspond to the space varying blur.

It is worth noting that we are modeling the high resolution intensity and the depth (blur) fields as MRF and not the low resolution observations. This is due to the fact that the property of the MRF does not percolate across the scale [127]. Hence the low resolution observations are not constrained to be MRFs.

3.6 Super-Resolution Restoration

We now explain how a MAP estimation of the dense intensity and blur fields can be obtained. The MRF models serve as the priors for the MAP estimation. We have modeled both the high resolution image $z(k,l)$ and the blur process $\sigma(k,l)$ as separate Markov random fields which are used as a prior. Let S and Z denote the random fields corresponding to the high resolution space-variant blur parameter $\sigma(k,l)$ and the high resolution focused image $z(k,l)$ over the $rM_1 \times rM_2$ lattice of sites L, respectively. We assume that S can take B_s possible levels and Z can take B_z possible levels. Although the fields S and Z are actually continuous, the blur field is quantized to reduce the number of acceptable configurations in the combinatorial minimization while the intensity field is usually quantized to 256 gray levels. One may use a nonlinear quantization scheme for the levels of S for better results, but this is not pursued in this exercise. The *a posteriori* conditional joint probability of S and Z is given by $P(S = \mathbf{s}, Z = \mathbf{z} | Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K)$ where the Y_m's denote the random fields corresponding to the m^{th} observed image. From Bayes' rule,

$$P(S = \mathbf{s}, Z = \mathbf{z} | Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K)$$
$$= \frac{P(Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K | S = \mathbf{s}, Z = \mathbf{z}) P(S = \mathbf{s}, Z = \mathbf{z})}{P(Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K)}.$$

The random fields S and Z are assumed to be statistically independent in this study as they refer to two independent processes, namely the depth and intensity processes. However, the assumption of statistical independence of the two fields S and Z may not be always valid. In many cases the intensity and the depth maps are related, for example, in shape from shading or texture applications where the shading may depend on perspective effects or object geometry. In absence of any knowledge of the cross-covariance matrix between the two fields, we assume them to be independent. Since the denominator in the above equation is not a function of $\mathbf{s}$ or $\mathbf{z}$, the maximum *a posteriori* (MAP) problem of simultaneous estimation of high resolution space-variant blur identification and super-resolved image can be posed as:

$$\begin{aligned}(\mathbf{z},\mathbf{s}) = \arg\ \max_{\mathbf{s},\mathbf{z}} P[(Y_1 = \mathbf{y}_1,\ldots,Y_K = \mathbf{y}_K | S = \mathbf{s}, Z = \mathbf{z}) \\ \times P(S = \mathbf{s})P(Z = \mathbf{z})].\end{aligned}$$

Note that the random fields S and Z are high resolution while the observations are low resolution. Since S and Z are both modeled as MRFs, the priors $P(S = \mathbf{s})$ and $P(Z = \mathbf{z})$ have a Gibbs distribution given by

$$P(S = \mathbf{s}) = \frac{1}{Z_s}\exp\{-\sum_{c\in C^s} V_c^s(\mathbf{s})\} \tag{3.23}$$

and

$$P(Z = \mathbf{z}) = \frac{1}{Z_z}\exp\{-\sum_{c\in C^z} V_c^z(\mathbf{z})\} \tag{3.24}$$

where Z_s and Z_z are normalizing constants known as partition functions, $V_c(.)$ is the clique potential and C^s and C^z are the set of all cliques in S and Z, respectively. All these terms have been explained in the previous two sections. Thus the posterior energy function to be minimized is obtained by taking the log of posterior probability and by assuming the sensor noise to be independent and identically distributed (i.i.d) Gaussian

$$U(\mathbf{s},\mathbf{z}) = \sum_{m=1}^{K} \frac{\| \mathbf{y}_m - DH(\mathbf{s}_m)\mathbf{z} \|^2}{2{\sigma_n}^2} + \sum_{c\in C^s} V_c^s(\mathbf{s}) + \sum_{c\in C^z} V_c^z(\mathbf{z}), \tag{3.25}$$

where ${\sigma_n}^2$ is the noise variance.

Smoothness is an assumption underlying a wide range of physical phenomena. However, it was mentioned earlier that careless imposition of the smoothness criterion can result in undesirable, over smoothed solutions. This could happen at points of discontinuities either in the image or in the depth map. Hence it is necessary to take care of discontinuities. We introduce separate line fields for the two processes S

and Z. After incorporating the first order smoothness term as defined in Eq. (3.21) and Eq. (3.22), the posterior energy function to be minimized is now defined as

$$U(\mathbf{s},\mathbf{z}) = \sum_{m=1}^{K} \frac{\| \mathbf{y}_m - DH(\mathbf{s}_m)\mathbf{z} \|^2}{2{\sigma_n}^2} + \sum_{i,j} \{[\mu_s e_{ss} + \gamma_s e_{sp}] + [\mu_z e_{zs} + \gamma_z e_{zp}]\}. \tag{3.26}$$

Parameters μ and γ correspond to the relative weights of the smoothness term and the penalty term necessary to prevent occurrence of spurious discontinuities.

When the energy function is non-convex, there is a possibility of the steepest descent type of algorithms getting trapped in a local minima. Hence, simulated annealing is used to minimize the energy function and to obtain the MAP estimates of the high resolution space-variant blur and the super-resolved image simultaneously. Simulated annealing applies a sampling algorithm such as the Metropolis algorithm or Gibbs sampler, successively at decreasing values of a temperature variable T. In this work, we have chosen a linear cooling schedule, *i.e.*, $T_{(k)} = \delta \ \ T_{(k-1)}$ where δ is typically between 0.9 and 0.99. The parameters for the MRF models are chosen by trial and error and the optimization is done through sampling the configuration spaces Z and S alternately. The details of the optimization process can be succinctly given by the following steps.

```
begin
    1. Initialization:
    Obtain low resolution depth map using complex spectrogram
    [116].
    Set $\mathbf{s}(initial)$ = bilinearly interpolated depth map.
    Set $\mathbf{z}(initial)$ = bilinearly interpolated least blurred observation.
    Choose $T_0, \mu_s, \mu_z, \gamma_s, \gamma_z, \delta, \theta_s, \theta_z, M', M", \sigma_z,$ and $\sigma_s$.
    Set $\mathbf{s}(old) = \mathbf{s}(initial)$
    Set $\mathbf{z}(old) = \mathbf{z}(initial)$
    Set $k = 0$.
    2. Repeat    (annealing loop)
              for $l = 1$ to $M'$, do (Metropolis loop)
              begin
                  for $i = 0$ to $N-1$, $j = 0$ to $N-1$, do
                  begin
```

```
                Get $s_{i,j}(new)$ from Gaussian sampler with
                mean $s_{i,j}(old)$ and variance ${\sigma_s}^2$.
                if $U(\mathbf{z}(old), \mathbf{s}(new)) \leq U(\mathbf{z}(old), \mathbf{s}(old))$ then
                        $\mathbf{s}(old) = \mathbf{s}(new)$,
                else
                if $\exp\left(\frac{U(\mathbf{z}(old),\mathbf{s}(old))-U(\mathbf{z}(old),\mathbf{s}(new))}{T_k}\right) > \text{rand}[0,1]$
                        then
                        $\mathbf{s}(old) = \mathbf{s}(new)$.
                end if
                Get $z_{i,j}(new)$ from Gaussian sampler with
                mean $z_{i,j}(old)$ and variance ${\sigma_z}^2$.
                if $U(\mathbf{z}(new), \mathbf{s}(old)) \leq U(\mathbf{z}(old), \mathbf{s}(old))$ then
                        $\mathbf{z}(old) = \mathbf{z}(new)$,
                else
                if $\exp\left(\frac{U(\mathbf{z}(old),\mathbf{s}(old))-U(\mathbf{z}(new),\mathbf{s}(old))}{T_k}\right) > \text{rand}[0,1]$
                        then
                        $\mathbf{z}(old) = \mathbf{z}(new)$.
                end if
            end
        end
        $k = k + 1$.
        $T_k = \delta T_{k-1}$.
        until $k > M$".
end
```

It is interesting to note the effect of the value of K (the number of observations) in Eq. (3.26) in super-resolving the fields. For an upsampling factor of r, one requires to estimate $2r^2$ parameters (intensity and depth values) per pixel. Hence one would ideally like to have $K \geq 2r^2$. However, this would be tantamount to using only the first (data fitting) term of the equation and it does not exploit the power of model based restoration techniques. Due to the punctuated smoothness terms one can obtain a very good estimate of both the fields even when $K < 2r^2$. Imposition of the penalty for the line detection saves the algorithm from the possibility of excessive smoothing.

3.7 Neighborhood of Posterior Distribution

If the posterior distribution has local dependencies that enable it to be modeled as an MRF, then it is possible to reduce the computational load of the estimation problem. This allows us to update the fields to be estimated locally, as it was suggested in the previous section. The following theorem asserts that such a neighborhood structure does exist.

Theorem 3.1. *(i) For each* $\mathbf{y}_m$ *fixed,* $P[S = \mathbf{s}, Z = \mathbf{z} | Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K]$ *is a Gibbs distribution over* $\{L, \mathcal{N}\}$ *with energy function*

$$U(\mathbf{s},\mathbf{z}) = \sum_{m=1}^{K} \frac{\| \mathbf{y}_m - DH(\mathbf{s}_m)\mathbf{z} \|^2}{2\sigma_n^2} + \sum_{c\in C^s} V_c^s(s) + \sum_{c\in C^z} V_c^z(z)$$

(ii) The posterior neighborhood corresponding to the site (k,l) *is given by*

$$\mathcal{N}_{k,l} = \mathcal{N}_{k,l}^s \cup \mathcal{N}_{k,l}^z \cup_{m=1}^{K} \{(a,b) \downarrow r : (k,l) \downarrow r \in \zeta_{(a,b)\downarrow r}^m$$

for some level of $s(k,l)$ *or* $z(k,l)\}$. *Here* $\mathcal{N}$ *represents the neighborhood system with line fields included, while* $\zeta_{(a,b)\downarrow r}^m$ *is the neighborhood corresponding to the subsampled lattice* $(a,b) \downarrow r$ *in the low resolution observation* $\mathbf{y}_m$.

Proof : (i) The first part of the theorem has been derived according to equation (3.25).

(ii) Due to the space varying nature of the blur, $\zeta_{(a,b)\downarrow r}^m$ would not be translationally invariant and would also be different, in general, from $\mathcal{N}_{k,l}^s$ and $\mathcal{N}_{k,l}^z$, the neighborhoods corresponding to the MRF models of the space variant blur parameter and the intensity process, respectively. The conditional probability of $s(k,l)$ and $z(k,l)$ given all the remaining pixels and the observations $\mathbf{y}_m$, is given by

$$P[S(k,l) = s(k,l), Z(k,l) = z(k,l), 0 \le (k,l) \le N-1 | Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K]$$

$$\begin{aligned} &= P[S(k,l) = s(k,l), Z(k,l) = z(k,l) | S(a,b) = s(a,b), Z(a,b) = z(a,b),\\ &\qquad 0 \le (a,b) \le N-1, (a,b) \ne (k,l); Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K] \times\\ &\qquad P[S(a,b) = s(a,b), Z(a,b) = z(a,b),\\ &\qquad 0 \le (a,b) \le N-1,\ (a,b) \ne (k,l) \mid Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K]. \end{aligned}$$

But

$$\begin{aligned} P[S(k,l) = s(k,l), Z(k,l) = z(k,l) \mid S(a,b) = s(a,b), Z(a,b) = z(a,b),\\ 0 \leq (a,b) \leq N-1,\ (a,b) \neq (k,l); Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K]\\ = \frac{\exp(-U(\mathbf{s},\mathbf{z}))}{\sum_{\text{all config. of } s(k,l),z(k,l)} (\exp(-U(\mathbf{s},\mathbf{z}))}. \end{aligned} \tag{3.27}$$

In the above equation, the components $s(k,l)$ and $z(k,l)$ can take any of the B_s and B_z possible levels, respectively. We define the vectors

$$\bar{\psi}_m = \frac{1}{\sqrt{2}\sigma_n}(\mathbf{y}_m - DH(\mathbf{s}_m)\mathbf{z}), \quad m = 1, \ldots, K.$$

Now the posterior energy function can be written as

$$U(\mathbf{s},\mathbf{z}) = \sum_{c \in \mathcal{C}^s} V_c^s(\mathbf{s}) + \sum_{c \in \mathcal{C}^z} V_c^z(\mathbf{z}) + \sum_{m=1}^{K} \sum_{0 \leq (a,b)\downarrow r < N/r} \bar{\psi}_m^2((a,b) \downarrow r).$$

Let $\Upsilon = \{(a,b) \downarrow r : 0 \leq (a,b) \downarrow r < N/r\}$ and $\Lambda_m = \{(a,b) \downarrow r : (k,l) \downarrow r \notin \zeta^m_{(a,b)\downarrow r}$, for all levels of $s(k,l)$ and $z(k,l)\}$. Next, we decompose $U(\mathbf{s},\mathbf{z})$ as follows :

$$\begin{aligned} U(\mathbf{s},\mathbf{z}) = & \sum_{c \in \mathcal{C}^s : (k,l) \in c} V_c^s(\mathbf{s}) + \sum_{c \in \mathcal{C}^z : (k,l) \in c} V_c^z(\mathbf{z})\\ & + \sum_{m=1}^{K} \sum_{\{\Upsilon - \Lambda_m\}} \bar{\psi}_m^2((a,b) \downarrow r) + \sum_{c \in \mathcal{C}^s : (k,l) \notin c} V_c^s(\mathbf{s})\\ & + \sum_{c \in \mathcal{C}^z : (k,l) \in c} V_c^z(\mathbf{z}) + \sum_{m=1}^{K} \sum_{\{\Lambda_m\}} \bar{\psi}_m^2((a,b) \downarrow r) \end{aligned} \tag{3.28}$$

Substituting the above in equation (3.27) and canceling terms common to the numerator and the denominator, we get

$$\begin{aligned} P[S(k,l) = s(k,l), Z(k,l) = z(k,l) \mid S(a,b) = s(a,b), Z(a,b) = z(a,b),\\ 0 \leq (a,b) \leq N-1,\ (a,b) \neq (k,l); Y_1 = \mathbf{y}_1, \ldots, Y_K = \mathbf{y}_K]\\ = \frac{\exp(-X_1 - X_2 - X_3)}{\sum_{s(k,l),z(k,l)} \exp(-X_1 - X_2 - X_3)}, \end{aligned} \tag{3.29}$$

where

$$X_1 = \sum_{c \in C^s : (k,l) \in c} V_c^s(\mathbf{s}),$$

$$X_2 = \sum_{m=1}^{K} \sum_{\{\Upsilon - \Lambda_m\}} \bar{\psi}_m^2((a,b) \downarrow r),$$

and

$$X_3 = \sum_{c \in C^z : (k,l) \in c} V_c^z(\mathbf{z}).$$

Hence, the posterior neighborhood structure corresponding to site (k,l) is given by

$$\begin{aligned} \mathcal{N}(k,l) &= \mathcal{N}_{k,l}^s \cup \mathcal{N}_{k,l}^z \cup_{m=1}^K \{\Upsilon - \Lambda_m\} \\ &= \mathcal{N}_{k,l}^s \cup \mathcal{N}_{k,l}^z \cup_{m=1}^K \{(a,b) \downarrow r : (k,l) \downarrow r \in \zeta_{(a,b)\downarrow r}^m \\ &\qquad \text{for some level of } s(k,l) \text{ or } z(k,l)\}. \end{aligned} \tag{3.30}$$

The method of simultaneous depth recovery and image restoration described in [118] is a specific instance of the more general scheme presented here in that the estimated depth map and intensity values are at the same resolution as the defocused observations. This fact leads to the following corollary which is proved in [105].

Corollary 1 *(i) For each* $\mathbf{y}_1$ *and* $\mathbf{y}_2$ *fixed,* $P[S = \mathbf{s}, Z = \mathbf{z} | Y_1 = \mathbf{y}_1, Y_2 = \mathbf{y}_2]$ *is a Gibbs distribution over* $\{L, \mathcal{N}\}$ *with energy function*

$$U(\mathbf{s}, \mathbf{z}) = \sum_{m=1}^{2} \frac{\| \mathbf{y}_m - H(\mathbf{s}_m)\mathbf{z} \|^2}{2\sigma_n^2} + \sum_{c \in C^s} V_c^s(\mathbf{s}) + \sum_{c \in C^z} V_c^z(\mathbf{z})$$

(ii) The posterior neighborhood corresponding to the site (k,l) *is given by*

$$\mathcal{N}_{k,l} = \mathcal{N}_{k,l}^s \cup \mathcal{N}_{k,l}^z \cup \{(a,b) : (k,l) \in \zeta_{(a,b)}^{y_1}$$

$$\textit{for some level of } s(k,l) \textit{ or } z(k,l)\} \cup \{(a,b) : (k,l) \in \zeta_{(a,b)}^{y_2}$$

$$\textit{for some level of } s(k,l) \textit{or } z(k,l)\}.$$

Here $\mathcal{N}$ *represents the neighborhood system with line fields included, while* $\zeta_{(a,b)}^{y_m}$ *is the neighborhood corresponding to the site* (a,b) *in the low resolution observation* $\mathbf{y}_m$.

3.8 Experimental Demonstration

We have explained how the defocus cue can be used in super-resolving an image. We now illustrate the efficacy of the method for simultaneous super-resolved blur identification and image reconstruction through several examples of simulation and real data.

We note that since the blur is a function of depth, it suffices to recover the distribution of the blur parameter over the image. We show all our results for an upsampling factor of $r = 2$. In all simulation experiments, only five low resolution observations were considered *i.e.*, $K < 2r^2$. In addition to the space variant blur, each low resolution observation was also corrupted with an additive white Gaussian noise of variance 5.0. The low resolution blur was estimated from any two of the five observations using the complex spectrogram method described in [116] which offers a good initial estimate of the blur with a very little computation. A square window of size 16 × 16 was used for the purpose. A bilinear interpolation of the estimated low resolution blur yields the initial estimate of the high resolution blur. Similarly, the bilinear interpolation of the least blurred image was chosen as the initial estimate of the true focused image. The number of discrete levels for the space variant blur parameter (σ) was taken as $B_s = 128$. We observe that in most cases, the amount of blur is restricted to $\sigma = 5$. We discretize the interval $[0, 5]$ in 128 levels. For the intensity process, $B_z = 256$ levels were used, which is the same as the CCD dynamic range. The parameters involved in the simulated annealing algorithm while minimizing Eq. (3.25) are as follows.

- θ_s and θ_z - thresholds for the line fields corresponding to the blur parameter and the intensity process, respectively.
- T_0 - initial temperature.
- δ - cooling schedule (rate of cooling).
- σ_z and σ_s - standard deviations of the Gaussian sampler for the intensity and the blur process, respectively.

In order to generate the data for the simulation experiments, we adopt the following strategy. First, we consider the image to have been taken with a pin-hole camera implying that we obtain a focused image of the scene. Next we assign an arbitrary depth map to the scene. Since the depth at a point in the scene is a function of the amount of blur at that point, the depth and the blur are deemed to be analogous. Using the space varying blur, we carry out a space varying convolution with the scene map to obtain a defocused image. What we are doing, in essence, is that we are mapping a texture to a particular depth map. Appropriate noise sequences are added to obtain the observations.

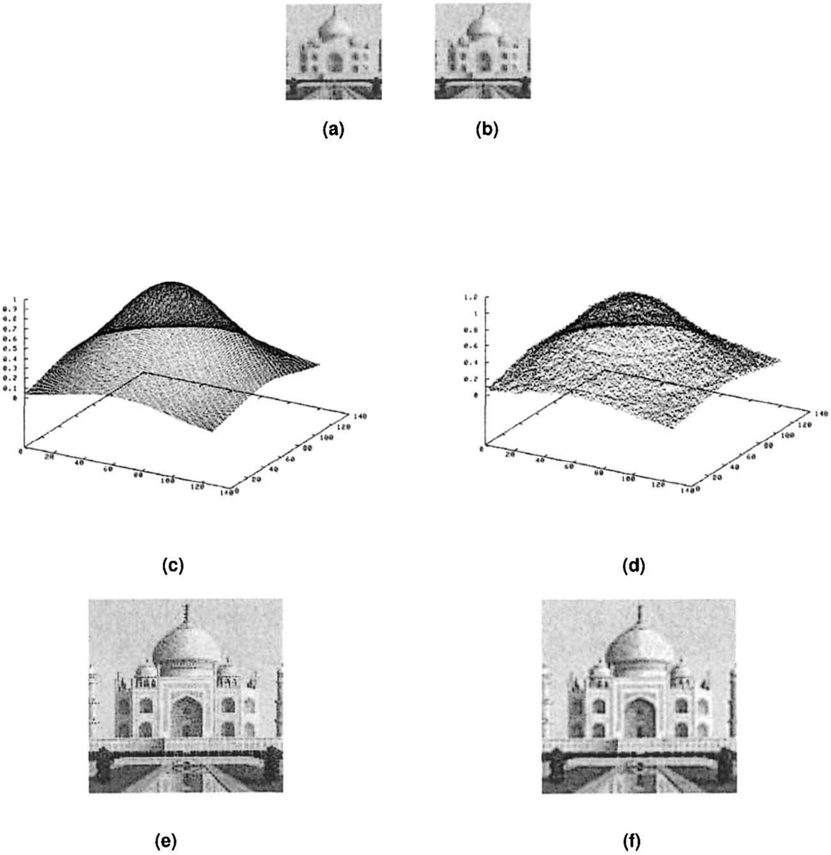


Fig. 3.4. (a, b) Two of the defocused low resolution Taj images. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The original high resolution Taj image. (f) The super-resolved Taj image.

Figures 3.4(a) and 3.4(b) show two of the five low resolution observations of the Taj image. In general we display only the two of the least blurred observations here. The blur parameter in these defocused images are related through $\sigma_{m+1}(i,j) = 0.75\ \sigma_m(i,j)$, $m = 1, \ldots, 4$. We recall that such a linear relationship between the blurs exists when defocused images of a scene are obtained using different values of the camera parameters. The true high resolution blur $\sigma_1(k,l) = a\exp(-\frac{(k-64)^2+(l-64)^2}{2b^2})$, $0 \leq k,l < 128$, is plotted in Figure 3.4(c). In this experiment, the values are $a = 1.0$, and $b = 35.0$. As mentioned earlier, we have chosen a decimation factor of $r = 2$. The original Taj image is blurred using the space-varying Gaussian blurring kernels formed from the blurs given above and then down sampled. White Gaussian noise is now added to the observations. The estimated values of the super-resolved blur parameters using the method described earlier are shown in Figure 3.4(d). We compare the performance in terms of root mean squared error (RMSE) for the blur field. The RMSE between the true entity $\mathbf{f}$ and the estimated one $\hat{\mathbf{f}}$ is defined by the following equation.

$$RMSE = \sqrt{\frac{\sum_{i,j}(f(i,j) - \hat{f}(i,j))^2}{\sum_{i,j}(f(i,j))^2}}. \tag{3.31}$$

Using the above equation, the RMSE in the estimate of the blur was found to be only 0.033. The values of the parameters used in the simulated annealing (SA) algorithm are $\mu_s = 1000.0$, $\gamma_s = 15.0$, $\theta_s = 0.15$, $\sigma_s = 1.2$, $\mu_z = 0.005$, $\gamma_z = 5.0$, $\theta_z = 25.0$, $\sigma_z = 3.0$, $\delta = 0.975$ and $T_0 = 3.0$. It is to be noted here that no attempt has been made in this study to obtain the best parameter set for the optimization purpose. The algorithm has been able to determine the super-resolved blur (depth) field quite accurately. The true high resolution Taj image is shown in Figure 3.4(e), and the super-resolved image is shown in Figure 3.4(f). We observe that the quality of the estimated super-resolved image is also good especially in the region of the main entrance and near the minarets. Note that the technique has worked well even in the case of a fairly non-textured image such as the Taj image.

We next consider a blur profile in the form of a ramp function. The blur varies linearly from a value of 0.02 at the left edge of the image to 0.97 at the right edge. Two of the least blurred low resolution Graveyard images generated using our observation model are shown in Figures 3.5(a) and (b). Once again the blurs are related through $\sigma_{m+1}(i,j) = 0.75\ \sigma_m(i,j)$, $m = 1,\ldots,4$, and the true high resolution blur $\sigma(k,l)$ is plotted in Figure 3.5(c). Since the amount of blur increases from left to right, the right part of the images are severely blurred. As before, the initial estimate of the super-resolved blur is the bilinear interpolation of the low resolution estimate of the blur determined using the complex spectrogram method. The bilin-

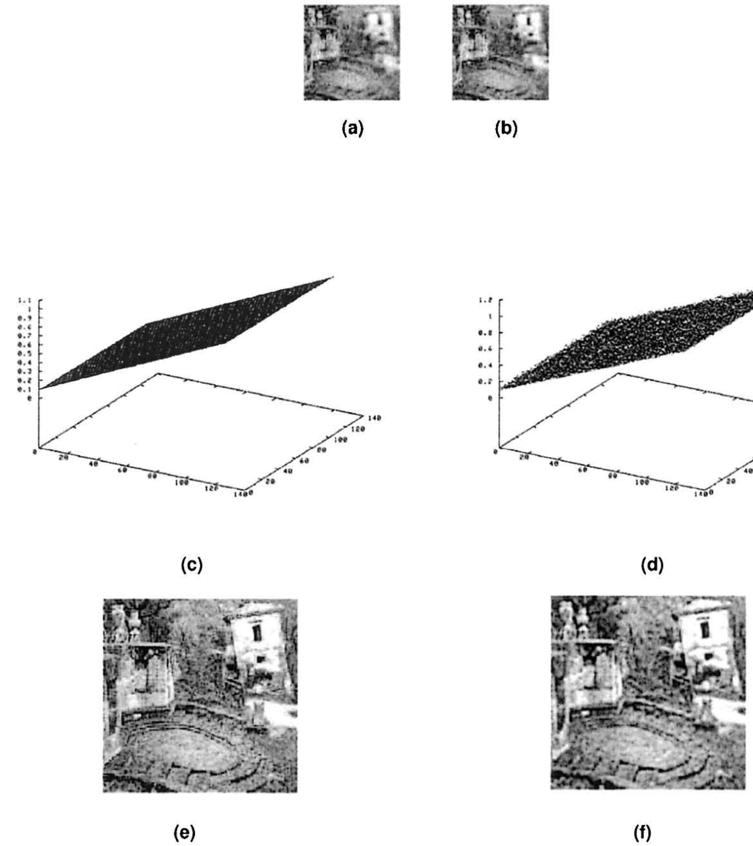


Fig. 3.5. (a, b) Two of the simulated low resolution observations. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The original high resolution Graveyard image. (f) The super-resolved Graveyard image.

early interpolated, least blurred observation is the initial estimate of the super-resolved focused image. The super-resolved blur parameters and the super-resolved Graveyard image are shown in Figure 3.5(d) and Figure 3.5(f), respectively. The RMSE in the estimate for the super-resolved blur is again only 0.019, yielding an average estimation error of about 2%. The parameters of the SA algorithm are kept the same as in the last experiment. The true high resolution Garveyard image is shown in Figure 3.5(e). We observe that the degradations in the observations are eliminated in the super-resolved image. Note that steps in the right end are now clearly visible.

In both the above experiments the field representing the blur process can be assumed to be sufficiently smooth so as to preclude the use of line fields for the blurring process alltogether. In order to see how the method performs in presence of discontinuities in the blur process, we next consider the case where the blurring is constant over a certain contiguous region of the image and then varies linearly over a second region and finally is constant again over the remaining part of the image. Two such blurred observations of the Sail image are shown in Figures 3.6(a) and (b). The true dense depth profile and the true high resolution image are shown in Figures 3.6(c) and (e), respectively. The estimated super-resolved blur parameters are shown in Figure 3.6(d) and the super-resolved image in Figure 3.6(f). The RMSE in the estimation of super-resolved blurs is only 0.020. The super-resolution image recovery technique has performed quite well. The numerals on the sail as well as the thin lines are clearly discernible. Even the left arm of the sailor is visible.

For a higher degree of discontinuity we consider a step profile for the variation in blur/depth in the same Sail image. Two of the defocused sail images resulting from the space varying convolution of the step form of blur variation with the original scene map are shown in Figures 3.7(a) and (b). The true variation of blur is plotted in Figure 3.7(c). The estimated super-resolved blur and image are shown in Figures 3.7 (d) and (e), respectively. Since the blur variation is highly discontinuous, we observed that slightly reduced values of $\mu_s = 500$ and $\gamma_s = 10$ (*i.e.*, less demand for smoothness and lowering the penalty for introducing a discontinuity in the depth field) yield better results. The RMSE in the blur estimates is 0.068 which is slightly on the higher side compared to the previous cases. Still the image reconstruction is very good.

Next we present the results of our technique on low resolution observations of a Text image. The purpose of the experimentation is to subjectively judge the improvement in readability after the super-resolution restoration. Each observation of size 41 × 207 is blurred by a space varying blur which has a similar variation as in the previous example, viz., step variation. Two of the five low resolution images are shown in Figure 3.8(a) and (b). The true high resolution blur parameters and the high resolution intensity map are shown in Figures 3.8(c) and (e), respectively. Due to the step-like variation in the blur profile, we notice the text getting progressively blurred from the left edge to the right edge of the input images. The estimated super-resolved blur pa-

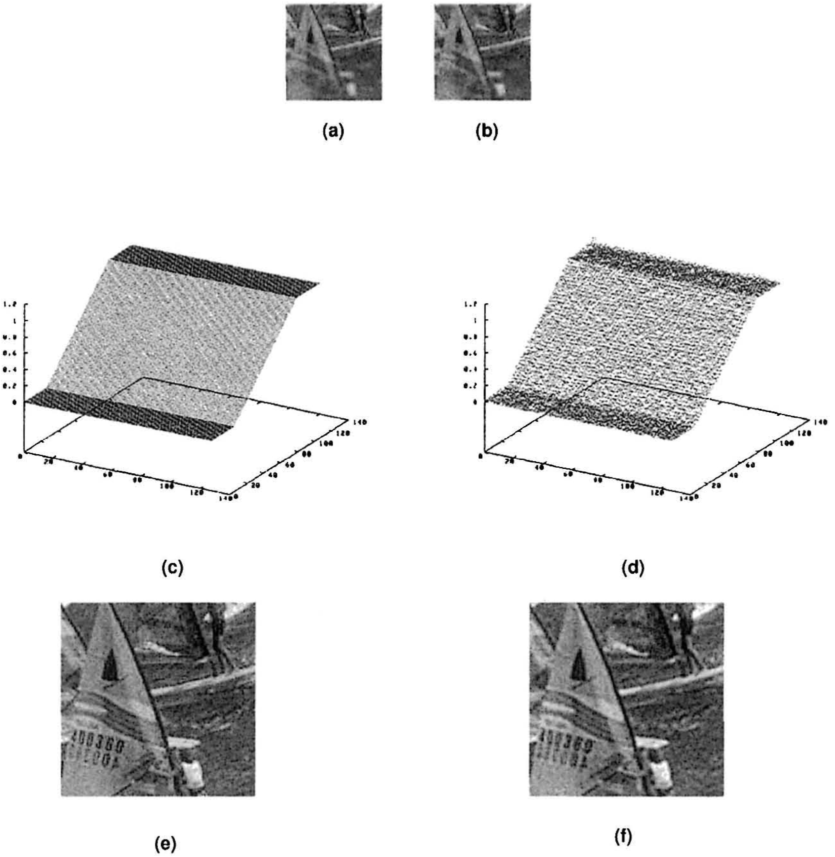


Fig. 3.6. Experimentation with discontinuous blur variation. (a, b) Two of the low resolution Sail images. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The original high resolution Sail image. (f) The super-resolved Sail image.

rameters and the super-resolved text image are shown in Figure 3.8(d) and Figure 3.8(f), respectively. The RMSE for the blur parameters in this case is 0.051. The same parameter set used in the previous experiment is used for the optimization purpose. Since the image field is also very discontinuous, a similar change in the values of μ_z and γ_z tend to yield a partly improved results. The super-resolution blur recovery is very encouraging. The text in the super-resolved image is easily readable. All these experiments substantiate our claim that both these fields can, indeed, be super-resolved.

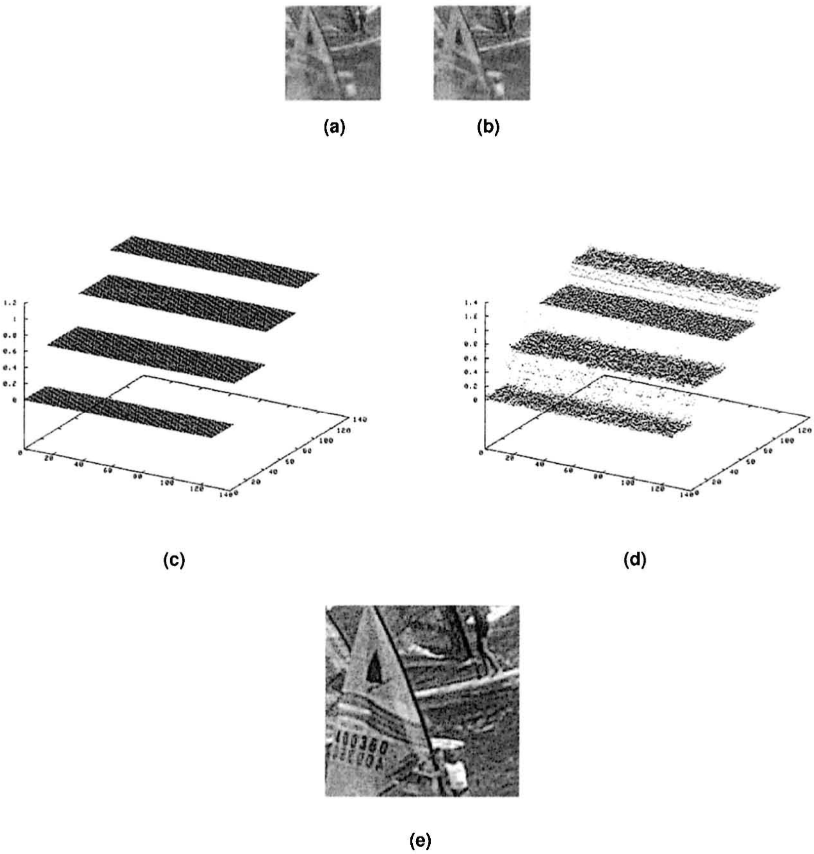
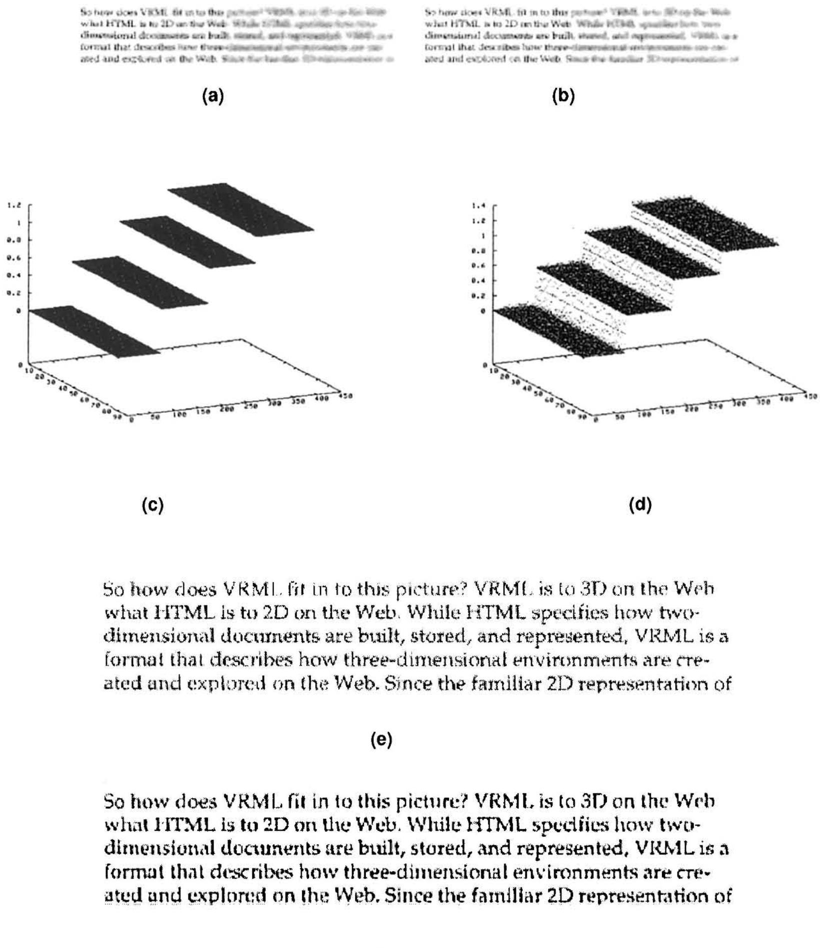


Fig. 3.7. Another experiment with a step-discontinuous blur. (a, b) Two of the low resolution Sail images with step variation in blurs. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The super-resolved Sail image.

We also experimented on the efficacy of the proposed technique by varying the number of available observations K. It was found that with further increase in K, there is barely any improvement in the quality of the restored fields. However, when we reduced the number of observations, there were some degradation in the quality of the output. The improvement tends to saturate for a value of $K = 4$ or 5 when $r = 2$.

The performance of the method was next tested on a real data captured in our laboratory under controlled conditions. The first experimental setup was the "blocks world" where three concrete blocks were



So how does VRML fit in to this picture? VRML is to 3D on the Web what HTML is to 2D on the Web. While HTML specifies how two-dimensional documents are built, stored, and represented, VRML is a format that describes how three-dimensional environments are created and explored on the Web. Since the familiar 2D representation of

So how does VRML fit in to this picture? VRML is to 3D on the Web what HTML is to 2D on the Web. While HTML specifies how two-dimensional documents are built, stored, and represented, VRML is a format that describes how three-dimensional environments are created and explored on the Web. Since the familiar 2D representation of

Fig. 3.8. Experimentation for readability of text image. (a, b) Two of the low resolution Text images. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The original high resolution Text image. (f) The super-resolved Text image.

arranged at different depths, the nearest one at a distance of 73 *cm*, another at 82.7 *cm* and the farthest block at 96.6 *cm*. All the blocks are placed perpendicular to the optical axis of the camera and hence there is no depth variation for a particular face of a block. Newspaper cuttings were pasted on the blocks to provide some texture as well as for the ease of post-operative evaluation. A Pulnix CCD camera fit-

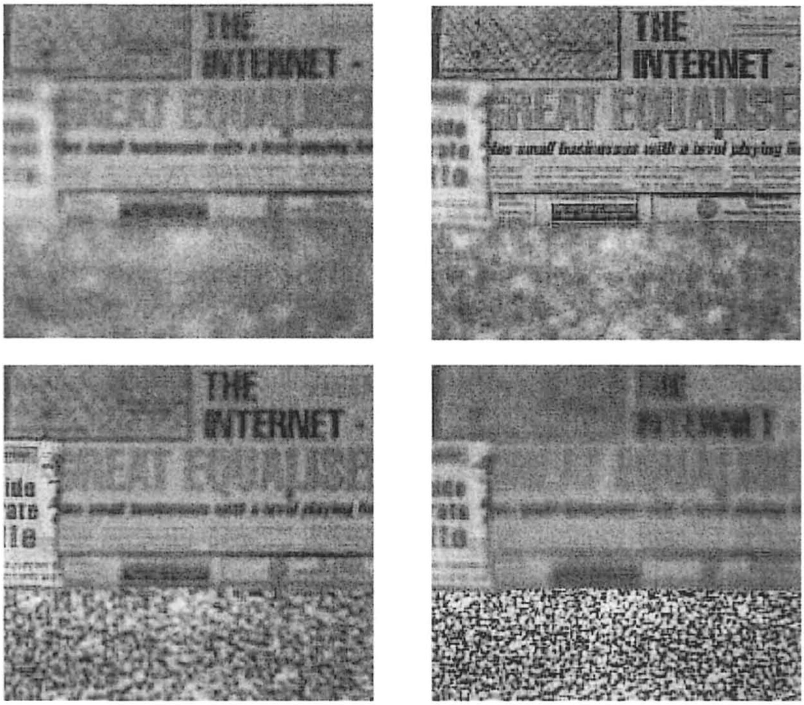


Fig. 3.9. Four low resolution observations of the "blocks world" captured in the laboratory by varying the focus setting of the camera.

ted with a Fujinon HF35A-2 lens of focal length 3.5 *cm* was used to grab the images. The lens aperture was kept at F1.7. The camera was coarsely calibrated using an object at a known depth. Five low resolution images, each of size 240 × 260 were captured. Four of them are shown in Figure 3.9. Depending on the selection of lens-to-image plane distances, one obtains different amount of defocus in different observations. The estimated super-resolved depths are shown in Figure 3.10 and the super-resolved image is shown in Figure 3.11. As we can see, the defocus cue has been able to capture the depth variation quite satisfactorily. The root mean square error in the estimation of depth is 1.768 *cm*, which is equivalent to a ranging error of just 1.96%. Also the super-resolved image has been recovered quite well as is evident from the readability of the text on both the blocks. This is not so in the captured images where the texts on either or both the blocks are always out-of-focus. The random dot pattern pasted on the lower block has also been recovered well.

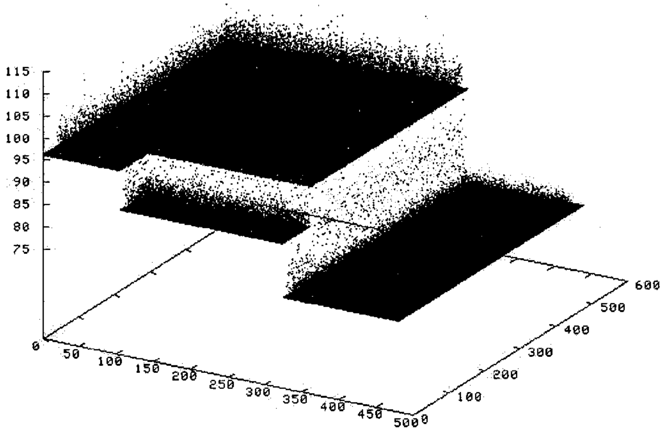


Fig. 3.10. The super-resolved depth estimates in the "blocks world". Here the heights are given in *cm*.

The second experimental setup consisted of a ball resting on a block. The selection of the ball as the scene was motivated primarily to verify the performance of the proposed method when the scene does not have much textural content. The block was at a distance of 117 *cm* from the camera. The point on the ball nearest to the camera was at 121.8 *cm* while the farthest points, viz., the points lying on the occluding boundary of the ball, were at 132.3 *cm* from the camera. Two of the five low resolution images each of size 280 × 280 pixels are shown in Figure 3.12. We have changed the lens-to-image plane distance in our experiments to obtain the differently defocused observations. This introduces a small amount of change in magnification in successive observations. We neglect the effect in our studies. The super-resolved depth is shown in Figure 3.13 in which one out of every four points is plotted in order to avoid clutter. The proposed technique has been able to capture the spherical depth variation well together with the depth of the block on which the ball is resting. The super-resolved Ball image is shown in Figure 3.14. The restored image is also of very good quality as the characters on the spherical surface are clearly visible.

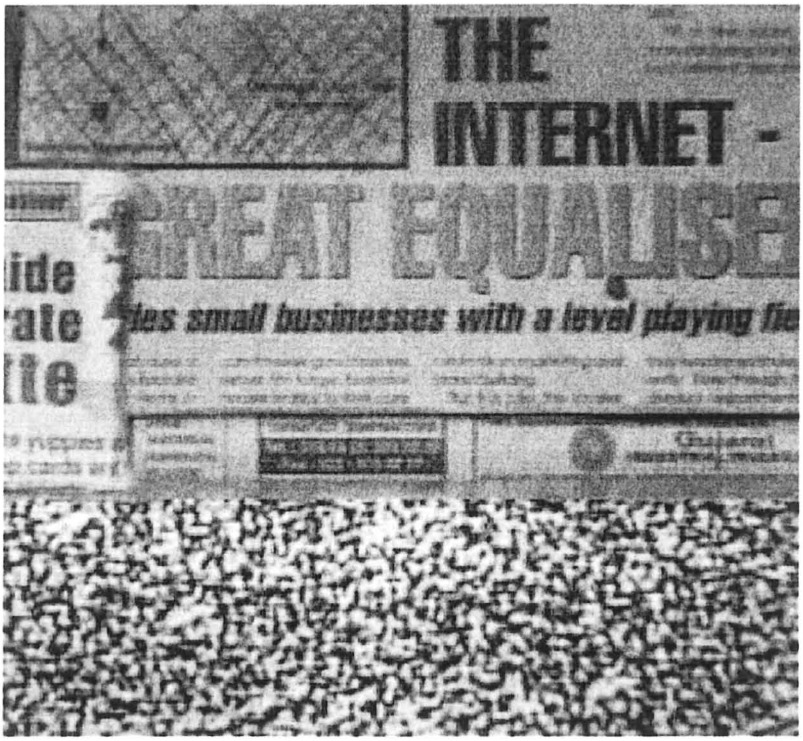


Fig. 3.11. The super-resolved image of the "blocks world".

3.9 Conclusions

We have described a MAP-MRF framework for simultaneously generating the super-resolved depth map and the super-resolved image from low resolution defocused observations. This method avoids the correspondence and warping problems inherent in current super-resolution techniques involving the motion cue in the low resolution observations and uses a more natural depth related defocus as a natural cue in real aperture imaging. Both the super-resolved blur parameter and image are modeled as separate MRFs. It is interesting to note that a large class of problems in computer vision, such as DFD, super-resolution, optical flow, shape from shading, *etc.*, can all be solved in a similar MAP framework. The basic structure of the solution remains the same. The equations of image formation are written in conjunction with appropriate priors and the solution is obtained by optimizing the resultant energy function. The current method is no exception.

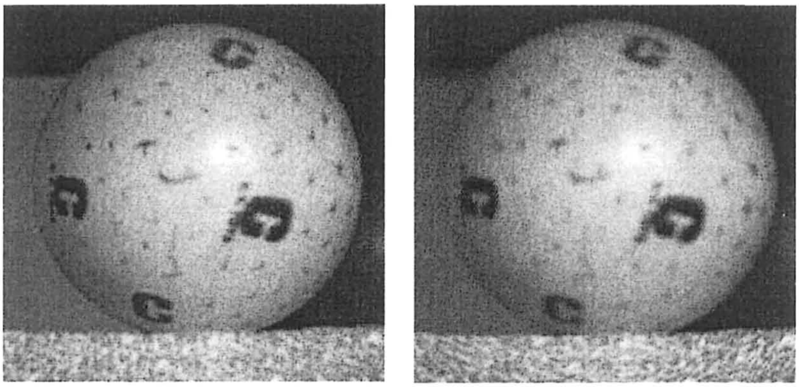


Fig. 3.12. Two of the low resolution Ball images.

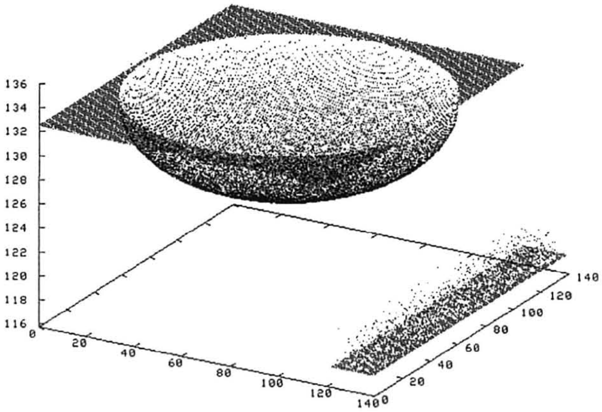


Fig. 3.13. The super-resolved depth map of the Ball image. The height is given in *cm*.

The use of line fields preserves discontinuities in the super-resolved depth and image fields. The super-resolved depth maps have been generated with a very high accuracy. We have chosen the line fields to be binary variables in this study. However, one can use continuous variables as well without much changing the problem formulation. The advantage of using continuous variable line fields lies in having a differ-

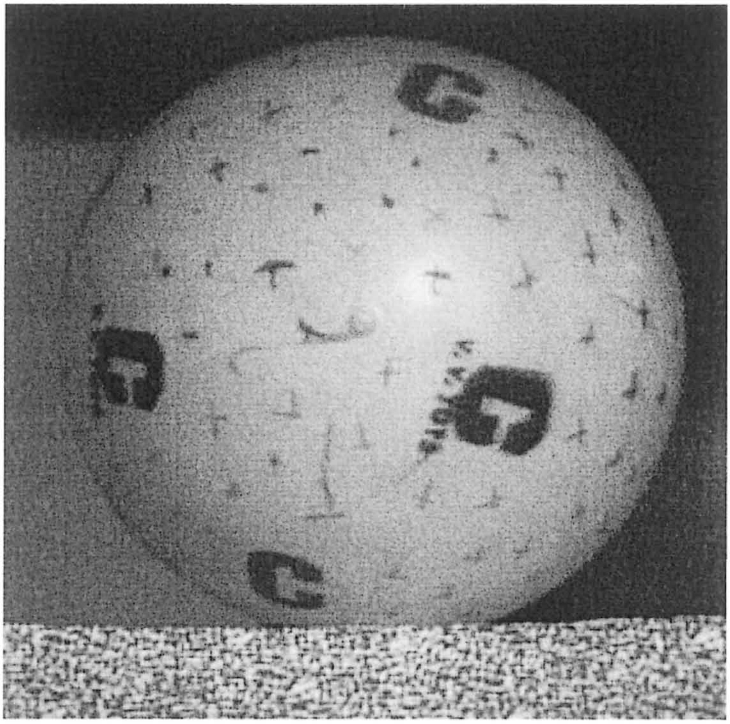


Fig. 3.14. The super-resolved Ball image. Note that one can read the name of the manufacturer (V. V. TOYS).

entiable cost function when a gradient-based optimization method can be used. Hence the restoration process becomes much faster.

The quality of the super-resolved images is also quite good. Through the restoration process we are able to obtain an ideal pin-hole equivalent image (*i.e.*, there is no depth of field) of an arbitrary scene using commercially available real-aperture cameras. However, there are some limitations of using the defocus cue for super-resolution. First of all, it works for a limited range of depth values in the scene. For a typical commercial camera with a $50mm$ lens, the depth variation in the scene should be restricted to only a few meters, with an average depth of $1-2$ meters. Beyond this, either the defocus cue is insignificant or the scene becomes so badly blurred that the corresponding blur parameters cannot be reliably estimated, making the system fail to restore the scene.

When we change the camera parameters like the lens-to-image plane distance and the focal length of the lens, there is an associated change in the magnification. If this magnification is not negligible, we may have to perform a rescaling operation on the observed images before the method can be used. Alternately, one can change the aperture size that does not result in a magnification. However, one now gets a different amount of light intensity at the CCD elements. One may have to adjust the change in image brightness through either adjusting the shutter time or post processing before applying the method.

Further, we have assumed a first order smoothness of the fields to be estimated to serve as the MRF priors. However, one can obtain a better result if the MRF parameters of these fields are known apriori. Ideally one would also like to estimate the MRF parameters simultaneously while super-resolving them. This issue of MRF parameter estimation has been discussed in chapter 8 while using zoom as a cue for image super-resolution.

4

Photometry Based Method

In the previous chapter we demonstrated that the depth related defocus can be used for image super-resolution. We also mentioned the merits and demerits of using the defocus as the cue. One can obtain the structure information also while using the defocus cue. The implicit assumption there is that the shape and the intensity fields are independent. As explained earlier, this assumption may not always hold good due to the area foreshortening. Hence we explore if the 3D shape and texture can be super-resolved as dependent entities. This is the subject matter of the current chapter. We explore a super-resolution technique where the 3D shape preservation is used as a constraint while super-resolving a scene. Given the observations under different illuminant positions, we obtain the super-resolved image and the spatially enhanced scene structure simultaneously. The use of shape cue in the form of photometric measurements, instead of the motion cue, eliminates the need for image registration. We model the high resolution image, the structure and the albedo of the surface as separate Markov random fields and super-resolve them using a suitable regularization scheme. Quite naturally, the proposed method is applicable to indoor scenes where the ambient illumination can be controlled.

Image super-resolution using the motion cue in spite of being inherently a 2D dense feature matching technique, it does not consider the 3D structure of the scene being imaged, albeit such an information is inherently available from the disparity map. Since the structure of an object is embedded in the images in various forms, e.g, texture, shading, *etc.*, it limits the quality of the super-resolved image and its applicability for subsequent use in 3D computer vision problems. This motivates us to explore a *structure preserving* super-resolution techniques. Hence we explore the usefulness of the photometric cue instead of the motion cue for super-resolving a scene. Since there is no relative motion between the camera and the scene, the super-resolution technique based on the differential sampling of the plenoptic function is no longer valid. We decompose the plenoptic function into a number of sub-functions and a generalized upsampling process with the prior based regularization is used. The problem then can be stated as: given a set of observations of a static scene taken with different light source positions, obtain a super-resolved image not only for a particular light source direction but also for an arbitrary illuminant pattern. In addition, obtain the super-resolved (dense) depth map of the scene and the albedo simultaneously. Clearly, by doing this, we get a more accurate description of the depth as well as the reflectance properties of the scene, which eventually leads to a better performance of the vision task at hand. Since such a problem is inherently ill-posed, we need suitable regularization of all these fields. We show in this chapter that the entire problem can again be expressed as a simple problem of regularization, just as it was in the previous chapter, and hence can be solved using existing mathematical tools.

The plenoptic function proposed in [128] is a seven dimensional function which describes the radiance received by the observer along any direction, at any point in the space, at any time and over any range of wavelength. Existing methods are based on a dense sampling of this plenoptic function along the direction or in the space. The time parameter implicitly models the change in illumination and the change of scene. When it is constant, the scene is static and the illumination is fixed. The authors in [129] have proposed a new formulation of the plenoptic function to include the illumination component and call it as plenoptic illumination function. They extract the illumination component from the aggregate time parameter and explicitly specify it in the new formulation. Thus when the light source position is changed, new information is available at each pixel to capture the surface properties of the object. We sample this extended plenoptic function by taking photographs of the same scene with different light source positions and decompose them into sub-functions, perform a high resolution model fitting and obtain the super-resolved image and the structure.

4.1 Past Work

A general survey of work done in the super-resolution area has been carried out in chapter 2. We now refer to some specific work that relate to structure preserving super-resolution. Some researchers have already explored the possibility of super-resolving a scene for its intensity distribution along with the depth map (surface reconstruction). For super-resolution surface reconstruction authors in [130] formulate the problem as that of expectation maximization (EM) and tackle it in a probabilistic framework using MRF modeling. An iterative algorithm is proposed to recover the high resolution albedo and the depth maps. They assume that the low resolution images are formed by projecting a Lambertian surface of varying albedo onto the lens of a distant camera while the surface is being illuminated by a distant light source. Also they consider that a sequence of low resolution images are available with subpixel shifts keeping the light source position fixed. Initializing the super-resolution image to an arbitrary estimate, they first simulate the imaging process to obtain a set of low resolution observations. Comparing these with the observed sequence of low resolution frames they minimize a penalty function iteratively, and update the initial guess until a stopping criteria is met.

As already referred in chapter 3, Cheeseman *et al.* [120] use a Bayesian method for constructing a super-resolved surface model by combining information from a set of images of the given surface.

Shekarforoush *et al.* use MRFs to model the images and obtain a high resolution 3D visual information (albedo and depth) from a sequence of displaced low resolution images [119]. The effect of sampling a scene at a higher rate is acquired by having interframe subpixel displacements. Using a probabilistic interpretation of Papoulis' generalized sampling theorem, an iterative algorithm is developed for 3D reconstruction of a Lambertian surface at a subpixel accuracy in [131]. The generalized sampling theorem gives conditions for reconstructing a φ-bandlimited (Fourier transform $F(w) = 0$ for $|w| \geq \varphi$) function when it is passed through different filters and the filter outputs are sampled. In this context, they assume the low resolution observations to contain the recurring samples of a nonuniform sampling sequence obtained by applying a common input function to a set of linear shift invariant systems. Their reconstruction gives the "emittance" of the surface, which is a combination of the effects of surface albedo, illumination conditions and ground slope for landsat images. In [132] the authors make use of images from different view points and a neighborhood correlation prior with a Gaussian noise model to reconstruct the surface. In all these approaches low resolution image frames have to entail subpixel overlap and have to be registered at a subpixel accuracy which imply a very accurate preprocessing or registration. Errors in registration reflect in the quality of the super-resolved image generated as well as the structure recovered.

For super-resolution applications authors in [96] propose a generalized interpolation method. Here a space containing the original function is decomposed into appropriate subspaces. These subspaces are chosen so that rescaling operation preserves properties of the original function. On combining these rescaled sub-functions, they get back the original space containing the scaled or zoomed function. This method allows an alias-free interpolation of the original image provided the sub-functions satisfy certain bandlimiting conditions. The content in this chapter is based initially on their approach, but is extended much beyond to obtain both the high resolution intensity and the depth map represented by surface gradients and also the high resolution albedo using a suitable regularization approach.

4.2 Generalized Interpolation

Let us assume that we need to interpolate or zoom up a function $f(x)$. Consider the following abstract parametric decomposing of the function

$$f(x) = \phi(a_1(x), a_2(x), \cdots, a_K(x)), \tag{4.1}$$

where $a_i(x)$, $i = 1, 2, \cdots, K$ are different functions of the interpolating variable x and when they are combined by an appropriate K-variate function ϕ, one recovers the original function. Assume that these functions $a_i(x)$, $\forall i$ and ϕ are arbitrary but continuous. We can now interpolate the individual functions $a_i(x)$ and combine them by using Eq. (4.1) to obtain rescaled $f(x)$. In [96] such an interpolation technique has been called as generalized interpolation. It has been shown in the book that if $f(x)$ is not bandlimited but all these sub-functions are bandlimited to a frequency W, and if these sub-functions are upsampled by a factor of r, the corresponding $f(x \uparrow r)$ will be free from aliasing artifacts upto a frequency rW. Here we use the symbol $\uparrow r$ to denote the upsampling by the factor r. Let us illustrate with an example. Consider

$$f(x) = a_1(x)a_2(x) = \sin w_1 x(1 + \; sinw_2x)$$

where the bandwidth of $f(x)$ is $w_1 + w_2$ and it exceeds the individual sub-function's bandwidths w_1 and w_2, respectively. If we sample $f(x)$ at a frequency w_s where $2[\max(w_1, w_2)] < w_s < 2[(w_1 + w_2)]$, then the function $f(x)$ will be aliased. However, the functions $a_1(x)$ and $a_2(x)$ are not aliased. Hence $a_1(x \uparrow 2)a_2(x \uparrow 2)$ will give us an alias-free reconstruction of $f(x \uparrow 2)$

We will now consider how the theory of generalized interpolation can be applied to photometric stereo to upsample the image. Given an ensemble of images captured with different light source positions, we can express the low resolution image $E_l(x, y)$ at a pixel location (x, y) for a Lambertian surface by the irradiance equation

$$E_l(x,y) = \rho_l(x,y)\mathcal{R}(p_l(x,y), q_l(x,y)) = \rho_l(x,y)\hat{n}_l(x,y).\hat{s}, \qquad (4.2)$$

where $\hat{n}_l$ is the low resolution unit surface normal, $\hat{s}$ is the unit vector defining the light source direction and ρ_l is the low resolution albedo of the surface. Here $\mathcal{R}()$ is the reflectance model of the surface. The surface gradient $(p_l,\, q_l)$ is used to specify the unit surface normal given by

$$\hat{n}_l(x,y) = (-p_l(x,y), -q_l(x,y), 1)^T / \sqrt{1 + (p_l(x,y))^2 + (q_l(x,y))^2}. \qquad (4.3)$$

We can recover the surface gradients p_l, q_l and the albedo by using a minimum of three observations for different light source positions provided the three equations due to the measurements are linearly independent. Having obtained the surface normal and the albedo for say, an $M \times N$ image, a suitable interpolation method is applied individually on $p_l(x, y)$, $q_l(x, y)$ and $\rho_l(x, y)$ subspaces to get the spatially magnified (denser) surface normal and albedo spaces of dimension $rM \times rN$, where r represents the magnification factor. Here $p_l(x, y), q_l(x, y)$ and $\rho_l(x, y)$ represent the sub-functions $a_1(x, y)$, $a_2(x, y)$ and $a_3(x, y)$, respectively with $K = 3$ in Eq. (4.1). The interpolated surface normals and the albedo are now used to reconstruct the high resolution intensity image $z(x, y)$ according to equation

$$z(x,y) = \rho_l(x \uparrow r, y \uparrow r)\hat{n}_l(x \uparrow r, y \uparrow r).\hat{s} = \rho(x,y)\hat{n}(x,y).\hat{s}. \qquad (4.4)$$

Note that we use the subscript l for low resolution field. For the corresponding field in high resolution we do not use any subscript. Now in order to obtain ϕ^{-1}, one requires several photometric measurements - a normal practice in super-resolution imaging techniques to fuse information from many observations. However the advantage here is that one does not require to establish the subpixel registration of different observations. On the other hand, this method has the disadvantage that the measurements cannot be combined directly in the intensity domain as the illumination pattern is different for each measurement and one requires a controlled environment. The advantage of this approach to upsampling is that both photometric and structural properties are preserved while super-resolving the scene. This is due to the fact that the interpolation is carried out on the surface normal and on albedo individually. Thus, if both of these fields vary slowly for a given surface such that they may be assumed to be bandlimited, an alias-free super-resolution is possible. The demerit of the proposal is that the method still relies on basic interpolation techniques. The process of super-resolving an image does not imply only an alias-free upsampling of the image lattice. The issues like blur, noise, and model and data inconsistencies must also be addressed.

From Eq. (4.2) and Eq. (4.3) we observe that the image intensity is a nonlinear function of the surface normal given by $p_l(x, y)$ and $q_l(x, y)$, but linear with respect to the albedo function. As illustrated earlier with an example, even if both p_l and q_l are quite bandlimited, the corresponding intensity field $E_l(x, y)$ need not be. Or in other words, for a given $E_l(x, y)$ we expect the functions $p_l(x, y)$ and $q_l(x, y)$ to vary a lot more smoothly than the intensity domain. Hence any interpolation technique is expected to serve well for the upsampling of the surface normal. Unless the albedo varies very sharply, an alias-free upsampling is possible. This is the motivation for developing the contents of this chapter.

4.3 Difficulties with Generalized Interpolation

The generalized interpolation based upsampling technique as discussed in the previous section, although has a sound mathematical footing, fails to achieve very good results due to the following reasons.

1. The interpolation of the surface normal $\hat{n}_l(x, y)$ should be carried out as a vector field and not just $p_l(x, y)$ and $q_l(x, y)$ as separate scalar fields. Hence the interpolated surface $\hat{n}(x, y)$ may not satisfy the integrability constraint, *i.e.*, the second order partial derivatives

of the depth are not independent of the order of differentiation. Any inconsistency may badly affect the high resolution rendering of the intensity field using Eq. (4.4).

2. No spatial constraints in the form of maintaining relationships among neighboring pixels are used while interpolating these subfunctions which are essential in vision problems due to their ill-posed nature. This helps the solution to converge towards the true entity and to provide a numerical stability.

3. For photometric analysis, a particular reflectance model (say, the Lambertian model as used in this study) or the BRDF of the scene is assumed. However, in practice it may differ significantly from the assumed model and this may lead to a significant departure while either estimating the surface normals or recovering the surface intensity.

4. There could be self-occlusion or self shadowed surface patches in the image since the viewing direction and the source direction are different (for example, see near the mouth of the dog or the shadow on the rear side of the shoe in the images given in the results section). Since we are using several measurements, it is possible to obtain the surface normal even in the shaded region as long as the surface patch is visible under at least three different illuminations. The use of image irradiance equation (given in Eq. (4.4)) to render the shaded image removes any effect due to self-shadowing.

5. The presence of sensor noise may also affect the quality of the reconstructed image.

6. It is assumed that all observations are free from blurring. It is possible that due to improper setting of the camera parameters, some or all of these observations are poorly focused.

A good super-resolution algorithm must take care of all these issues while utilizing the photometric cue. In this chapter we present a comprehensive scheme under which all the above shortcomings can be alleviated.

We know that it is difficult to interpolate a vector field where the integrability relationship $p_y(x, y) = q_x(x, y)$ can be ensured everywhere. In the literature on shape from shading, this is solved as a regularized but unconstrained optimization problem. This ensures that $p_y(x, y) \approx q_x(x, y)$, where the subscripts x and y refer to partial derivatives. We follow a similar integrability constraint here.

In order to bring in the contextual or spatial dependency while interpolating a sub-function, we use the Markov random field (MRF) to model it over a uniformly gridded lattice. The MRF provides a convenient and consistent way of modeling context dependent entities such as pixel intensities, depth of the object and other spatially correlated features. In this work, all high resolution interpolating functions, *i.e.*, $p(x, y)$, $q(x, y)$ and $\rho(x, y)$ are modeled as separate MRFs. It should be noted that, as it was mentioned in section 4.1, there has been a large body of literature on super-resolution technique that models the field to be super-resolved as an MRF. The current proposal is different in the sense that each high resolution sub-function is modeled as a separate MRF, allowing us to preserve each physical attribute of a scene such as the albedo and the 3D structure.

Let us now discuss how one can rectify errors due to improper modeling of the reflectance function. We assume the standard Lambertian reflectance model of the surface (without any loss of generality, as any other non-degenerate model would fit our discussion equally well), but any practical surface would hardly ever be a Lambertian one. Thus the estimation of the low resolution albedo (ρ_l) and the surface normal ($\hat{n}_l$) based on photometric cues will be quite different from their true values. One way to constrain the super-resolved ρ and $\hat{n}$ fields to stay close to the corresponding true fields is by re-projecting the reconstructed super-resolved image on the actual low resolution observation and making sure that they match well. The same constraint also helps us to alleviate the problem due to self-shadowing or surface invisibility. Because, if there is self shadowing and the albedo function is not appropriately modified while interpolating, the intensity at the corresponding surface will be very different from the actual observation. We call this as a data consistency constraint.

In this chapter we refrâin from handling the issue of the presence of blur in the observations. This is due to the fact that the blur is usually unknown. We devote the next chapter entirely on how to handle the blur in photometric observations.

The proposed method can now be illustrated with Figure 4.1. We obtain K low resolution observations of a static scene by varying the direction of a point light source. It is assumed that the directions are known failing which they can be estimated using the technique proposed in [133]. We also assume that the reflectance model is approximately known. One can now obtain a least squares estimate of the

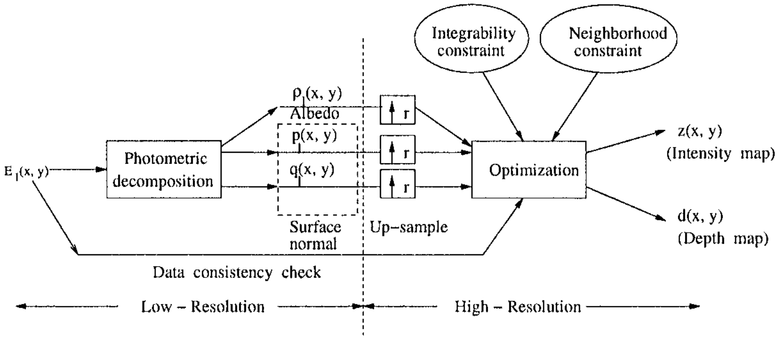


Fig. 4.1. Illustration of the proposed method of simultaneous super-resolution recovery of depth and the intensity maps using the photometric cue.

surface normal

$$\hat{n}_l(x, y) = (-p_l(x, y), -q_l(x, y), 1)/\sqrt{(1 + p_l^2(x, y) + q_l^2(x, y))}$$

and the surface albedo $\rho_l(x, y)$. One can use any interpolation technique such as the bilinear or the spline interpolation to up-sample the lattices over which these sub-functions are defined by a factor of r to obtain an initial estimate of the high resolution structure. Now an optimization technique can be developed that explicitly incorporate the constraints discussed earlier. As a result, one obtains high resolution estimates of both the intensity and the depth maps at a lattice size r times denser than that of the given observations satisfying the required constraints. In the next section we explain in detail the technique for estimating the above high resolution fields simultaneously.

4.4 Super-Resolution Estimation

Regularization is the most investigated approach to solve the ill-posed problems in computer vision, which was originally proposed by Tikhonov [134, 135, 136, 137]. Regularization is a popular method for interpolating sparse data, as well as smoothing the data obtained from noisy measurements. Simply put, regularization looks for an interpolating or approximating function which is both close to the data and also "smooth" in some sense. Formally this function is obtained by minimizing an error functional which is the sum of two terms, one measuring the distance from the data, the other measuring the smoothness of the function. The MRF based regularization approach is quite amenable to the incorporation of information from multiple observations with the smoothness function chosen from the prior knowledge of the fields to be estimated. The prior knowledge here, serves as a contextual constraint used to regularize the solution.

Let **z** represent the lexicographically ordered vector containing pixel intensities from the high resolution image for a particular light source position. Similarly, let **p**, **q** and $\boldsymbol{\rho}$ be the vector representations of the high resolution surface gradients and the albedo. The low resolution image formation model can be expressed as

$$\mathbf{E}_{l_k} = F_{C_k}(D, \mathcal{S}, \boldsymbol{\rho}) + \mathbf{v}_k, \quad k = 1, \ldots, K \tag{4.5}$$

where $\mathbf{E}_{l_k}$ is the $MN \times 1$ lexicographically ordered vector containing pixels from the k^{th} low resolution observation $E_{lk}(i,j)$. Here F_{C_k} indicates that these low resolution observations are a function of

- the decimation matrix D representing the high resolution to low resolution image down sampling process,
- the high resolution structural information $\mathcal{S}$ in the scene representing the surface gradients **p** and **q**, and
- the high resolution reflectance field such as the albedo $\boldsymbol{\rho}$, with C_k denoting the lighting conditions for the k^{th} observation.

In Eq. (4.5) v_k is the i.i.d Gaussian distributed noise vector with variance σ_v^2. Here K denotes the number of low resolution observations. It is quite simple to derive that the solution to a typical high resolution restoration problem can be obtained by minimizing the following cost function

$$\epsilon = \Sigma_{k=1}^{K} ||\mathbf{E}_{l_k} - F_{C_k}(D, \mathcal{S}, \rho)||^2 + ||\mathcal{L}(\mathcal{S})||^2 + ||\mathcal{L}(\boldsymbol{\rho})||^2 \tag{4.6}$$

with respect to the structure and the albedo fields. In the above equation if we drop the decimation matrix D the problem becomes similar to the photometric stereo with smoothness constraint [138]. Here $\mathcal{L}$ stands for a suitable regularization operator. The above cost function does not include constraints on the high resolution field **z** and the data consistency terms. These terms are added in the next subsection.

4.4.1 MRF Model Based Approach

As discussed in section 4.3 the MRF model provides a most general and powerful approach for prior field modeling and is often adopted in solving computer vision problems. The use of MRF models for the priors can allow us a lot more freedom in capturing the neighborhood relationships for the individual functions (fields). We know that the change in intensity of a scene is usually gradual. Also the the depth of a scene and the reflectance function of a scene is gradual. Thus we can consider the high resolution intensity field, the high resolution surface gradients and the estimated high resolution albedo to be smooth fields. We introduce the context dependencies in the estimated high resolution image $\mathbf{z}$, by modeling it as an MRF. An alternative way of explaining this is that we want the restored high resolution intensity map to be smooth one.

Due to equivalence with the Gibbs random field, the prior term in the form of potential energy can be given by $U(\mathbf{z})$. Similarly, the fields $\mathbf{p}$, $\mathbf{q}$ and $\boldsymbol{\rho}$ are also modeled as separate (independent) MRFs and the corresponding priors are $U(\mathbf{p})$, $U(\mathbf{q})$, $U(\boldsymbol{\rho})$, respectively. The exact form of the function $U(.)$ is given in Eq. (4.7). It is well known that in images, pixels with significant changes in intensities carry important information such as edges. The surface gradients representing the variation in depth over the surface exhibit a sudden change if there are depth discontinuities. Albedo also carries important information about the variations in surface reflectance. In order to incorporate provisions for detecting such discontinuities, so that they can be preserved in the reproduced entity, we also use the concept of line fields located on a dual lattice. The use of line fields $\mathbf{l}$ and $\mathbf{v}$ has been explained earlier in the chapter 3. Having chosen them as binary variables 1 and 0 we use the following energy function $U(\mathbf{w})$

$$\begin{aligned}
U(\mathbf{w}) &= \sum_{c \in C} V_c(\mathbf{w}) \\
&= \sum_{i,j} (\mu_w[(w_{i,j} - w_{i,j-1})^2(1 - {v^w}_{i,j}) + (w_{i,j+1} - w_{i,j})^2(1 - {v^w}_{i,j+1}) \\
&+ (w_{i,j} - w_{i-1,j})^2(1 - {l^w}_{i,j}) + (w_{i+1,j} - w_{i,j})^2(1 - {l^w}_{i+1,j})] \\
&+ \gamma_w[{l^w}_{i,j} + {l^w}_{i+1,j} + {v^w}_{i,j} + {v^w}_{i,j+1}]) \\
&= \sum_{i,j} [\mu_w e_{ws} + \gamma_w e_{wp}],
\end{aligned} \tag{4.7}$$

where $\mathbf{w} = \mathbf{z}$, $\mathbf{p}$, $\mathbf{q}$, or $\boldsymbol{\rho}$ as the field we consider. μ_w is the penalty term for departure from the smoothness and γ_w prevents the occurrence of spurious discontinuities in the estimated fields. Any other potential function can also be used if we have some information about the prior model. In absence of any other information about the high resolution fields, we restrain ourselves in using the punctuated first order smoothness model (estimation of these field parameters has been dealt with in chapter 8). Considering the integrability constraint and the data consistency constraint, the overall cost function can be obtained as

$$\begin{aligned}\epsilon = &\|\mathbf{z} - \rho(x \uparrow r, y \uparrow r)\hat{n}_l(x \uparrow r, y \uparrow r).\hat{s}\|^2 \\ &+ \lambda\|(\mathbf{p}_y - \mathbf{q}_x)(1 - \mathbf{l}^p)(1 - \mathbf{v}^q)\|^2 \\ &+ U(\mathbf{z}) + U(\mathbf{p}) + U(\mathbf{q}) + U(\rho) + \alpha\|\mathbf{E}_l - DH\mathbf{z}\|^2. \qquad (4.8)\end{aligned}$$

Here H is the camera point spread function (PSF) for any blur which may be implicitly present while observing the low resolution images. In our study we have considered H as an identity matrix, *i.e.*, the observations are not blurred. If they indeed are blurred, the corresponding PSF must be known or estimated. This issue has been discussed in the next chapter. $E_l(x, y)$ is the observed low resolution image for a particular light source position for which the super-resolution is sought. The matrix D is the decimation matrix, the form of which is given in chapter 3. Here α and λ are the constants used as weighting factors for carrying out the minimization for the observation term and the integrability term, respectively. Comparing this cost function with Eq. (4.6) we observe that Eq. (4.8) is modified to include the constraint term in the high resolution field $\mathbf{z}$ and the data consistency constraint (the last term). This cost function is non-convex due to the inclusion of line fields and is minimized using the simulated or mean field annealing optimization algorithm in order to obtain the global minima. Having obtained the high resolution surface gradients $\mathbf{p}$ and $\mathbf{q}$ we use the following equation (see [138] for details) to obtain the super-resolved (dense) depth map $d(x, y)$

$$\nabla^2 d(x, y) = p_x(x, y) + q_y(x, y), \qquad (4.9)$$

where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator. p_x and q_y represent the derivatives of the estimated high resolution surface gradients along x and y directions, respectively. The depth map $d(x, y)$ can be easily obtained by iteratively solving Eq. (4.9). At this point one may argue that, instead of using the surface gradients $p(x, y)$ and $q(x, y)$, one can use the depth variable $d(x, y)$ itself in the formulation since $p(x, y)$ and $q(x, y)$ are nothing but the partial derivatives of $d(x, y)$. Eq. (4.8) would then simplify into

$$\epsilon = \|\mathbf{z} - \rho(x \uparrow r, y \uparrow r)\hat{n}_l(x \uparrow r, y \uparrow r).\hat{s}\|^2$$

$$+ U(\mathbf{z}) + U(\mathbf{d}) + U(\boldsymbol{\rho}) + \alpha||\mathbf{E}_l - DH\mathbf{z}||^2,$$

where $\hat{n}_l(x \uparrow r, y \uparrow r)$ is now obtained from the high resolution depth field $d(x, y)$. Here the integrability term is no longer needed, and one requires to handle one set of variables less. Although this is also a perfectly valid method, we continue with Eq. (4.8) due to the motivation that generalized interpolation yields better results when the function is decomposed into more number of sub-functions, each changing very smoothly.

Now a few words about the different terms used in Eq. (4.8) follow. The first term corresponds to the requirement that the reconstructed super-resolved image be close to the high resolution image synthesized from the shape at high resolution as per the given reflectance model, ensuring structure preservation while reconstructing the image. We need a term $||(\mathbf{p}_y - \mathbf{q}_x)||^2$ to enforce the integrability condition, *i.e.*, to impose the condition that the reconstructed surface should correspond to a valid physical surface. Since as we are preserving discontinuities in $\mathbf{p}$ and $\mathbf{q}$ by using line fields it is necessary to multiply the integrability constraint term with $(1 - \mathbf{l}^p)(1 - \mathbf{v}^q)$ where $\mathbf{l}^p$ and $\mathbf{v}^q$ are the associated horizontal and vertical line-field terms for the fields $\mathbf{p}$ and $\mathbf{q}$, respectively. This would prevent minimizing the cost due to $U(\mathbf{p})$ and $U(\mathbf{q})$ terms whenever there is a discontinuity in $\mathbf{p}$ or in $\mathbf{q}$, so that the abrupt changes in surface gradients are preserved. The integrability condition should not be enforced at places of surface discontinuities.

The next four terms use the MRF prior with the line fields in order to provide the local context dependencies for the various high resolution fields as defined by $U(.)$ in Eq. (4.7). To enforce a data consistency check, we expect that the low resolution observation $\mathbf{E}_l$ should be very close to the down-sampled version of the super-resolved image $\mathbf{z}$, as given in the last term. Since there are K such observations due to various locations of the light source, $\mathbf{E}_l$ corresponds to that particular light source direction for which the super-resolved display of the scene is being sought. Thus the last term checks the consistency of the estimated high resolution fields against modeling and other errors such as those due to self shadowing. It should be noted that the last term is present when we super-resolve a view which is part of the set of observations. In case we plan to perform a high resolution rendering of view for a virtual light source position, we do not have the corresponding measurement and hence the last term should be removed from the cost function. Or in other words, the data consistency check is possible only when the low resolution observation of a scene for a particular light source position is available.

A comment is now due about the propriety of modeling the intensity field $\mathbf{z}$ as an MRF when the constituent functions $\mathbf{p}$, $\mathbf{q}$ and $\boldsymbol{\rho}$ have already been modeled as separate MRFs. One cannot prove that it is, indeed, so. Strictly speaking $\mathbf{z}$ is not an MRF due to the presence of $\mathbf{p}$ and $\mathbf{q}$ in the denominator in Eq. (4.3). The locality property as discussed in section 3.7 is no longer satisfied by $\mathbf{z}$. Any local perturbation in the fields $\mathbf{p}$ or $\mathbf{q}$ will ideally result in an infinitely extended perturbation in $\mathbf{z}$ in spatial domain. A justification for the current method is that we simply want to use it as a smoothness constraint and hence we choose a simple first order smoothness term as the prior. Alternately one may view it as a purely regularization based approach which can also be solved using calculus of variations.

4.4.2 Variational Approach

As already discussed the MRF model proves to be better in choosing the neighborhood relationship among pixels. Thus by making use of line fields in deriving the cost helps in preserving discontinuities. But the computational complexities shoot up as one cannot use simple optimization methods such as the gradient descent in order to obtain a solution. One way to speed up the computation is to choose a smoothness constraint without using the line fields. Although this makes the solution a bit too smooth it is very much advantageous computationally when compared to the MRF prior with line fields. In order to speed up the computation we demonstrate the use of the variational approach. In the variational approach one looks for extrema of expression that depend on some other functions called functionals rather than fixed parameters. This leads to a set of differential equations rather than ordinary algebraic equations. In general, a problem involving partial differential equations is ill-posed without some additional constraints. Hence we impose smoothness and other necessary constraints so as to make the problem well-posed. Thus we can consider the high resolution intensity field, the high resolution surface gradients and the estimated high resolution albedo all to be smooth fields. In lieu of MRF modeling with binary line fields we use the smoothness of the MRF prior for the fields to be recovered along with the integrability and the data consistency constraints as the regularization term and a faster implementation can be obtained using the iterative solution of the corresponding Euler equation. This also helps one to avoid explicitly assuming the intensity field $\mathbf{z}$ to be an MRF. By dropping the line fields from Eq. (4.8), the error functional (thus, in effect, we are using the same MRF parameterization) can be written as

$$\epsilon = \int\int V(z, p, q, \rho, p_x, p_y, q_x, q_y, \rho_x, \rho_y, z_x, z_y)dxdy. \tag{4.10}$$

The corresponding Euler equations are

$$\begin{aligned} V_z - \frac{\partial}{\partial x}V_{z_x} - \frac{\partial}{\partial y}V_{z_y} &= 0 \\ V_p - \frac{\partial}{\partial x}V_{p_x} - \frac{\partial}{\partial y}V_{p_y} &= 0 \\ V_q - \frac{\partial}{\partial x}V_{q_x} - \frac{\partial}{\partial y}V_{q_y} &= 0 \\ V_\rho - \frac{\partial}{\partial x}V_{\rho_x} - \frac{\partial}{\partial y}V_{\rho_y} &= 0. \end{aligned} \tag{4.11}$$

In our case V can be written as

$$\begin{aligned} V = {} & (z(x,y) - \mathcal{R}(p,q))^2 + \beta_i(p_y - q_x)^2 \\ & + \beta_z(z_x^2 + z_y^2) + \beta_p(p_x^2 + p_y^2) + \beta_q(q_x^2 + q_y^2) + \beta_\rho(\rho_x^2 + \rho_y^2) \\ & + \beta_d(E_l(x,y) - DHz(x,y)))^2, \end{aligned} \tag{4.12}$$

where z, p, q, ρ are the high resolution fields and $(.)_x$ and $(.)_y$ represent their partial derivatives along x and y directions, respectively. While $\mathcal{R}$ represents the reflectance map of the high resolution scene, β_z, β_p, β_q, β_ρ denote the weights for smoothness for each field used in the cost functional. The parameters β_i, β_d denote the weightage for the integrability and data consistency terms, respectively. The difference between the above equation and the Eq. (4.8) is that no attempt is made to preserve the abrupt changes occurring in different fields. The advantage is that the Eq. (4.12) is differentiable and hence avoids the use of computationally taxing optimization algorithms. But the disadvantage is that the solution becomes a bit smooth.

The Euler equations for the minimization of integral of the functional V given in Eq. (4.12) can now be written as

$$\nabla^2 z = \frac{1}{\beta_z}(z(x,y) - \mathcal{R}(p,q)) - \frac{\beta_d}{\beta_z}(E_l(x/2, y/2) - DHz(x,y))$$

$$
\begin{aligned}
\nabla^2 p &= -\frac{1}{\beta_p}(z(x,y) - \mathcal{R}(p,q))\mathcal{R}_p - \frac{\beta_i}{\beta_p}(p_{yy} - q_{xy}) \\
\nabla^2 q &= -\frac{1}{\beta_q}(z(x,y) - \mathcal{R}(p,q)\mathcal{R}_q - \frac{\beta_i}{\beta_q}(q_{xx} - p_{yx}) \\
\nabla^2 \rho &= -\frac{1}{\beta_\rho}(z(x,y) - \mathcal{R}(p,q))\mathcal{R}_\rho.
\end{aligned}
\tag{4.13}
$$

Since all the terms except $E_l(x,y)$ involve the high resolution field, and the updates are carried out on the high resolution lattice (x,y), we use the location $(x/2, y/2)$ to relate to the corresponding pixel in the low resolution observation. We solve the above set of equations using the iterative method by approximating the continuous solution by its discrete version. We used the initial estimates for $\mathbf{z}$, $\mathbf{p}$, $\mathbf{q}$ and $\boldsymbol{\rho}$ from those obtained from the generalized interpolation explained in section 4.2. The super-resolved fields $\mathbf{p}$ and $\mathbf{q}$ are then used to estimate the high resolution depth field $d(x,y)$ of the scene by using the Eq. (4.9).

4.5 Experimentations

We present some of the experimental results to demonstrate the efficacy of this regularization based approach for super-resolution image and surface reconstruction for a static scene and a camera. All the experiments were conducted on real images. The images were captured with different light source positions. A controlled environment was created by capturing the images in a dark room with no ambient light. In this arrangement, the distance between the object and the camera is much larger than the object size, so that we can assume an orthographic projection. The light source is also located at a sufficiently large distance, so that the light source direction can be assumed to be constant for the whole surface.

First we consider an object where the imaged scene gives a smooth intensity variation, but has arbitrary depth variations. Eight different images were captured for varying source directions to recover the low resolution albedo and the surface normal for the object. Figure 4.2 shows the captured low resolution images of the fluffy dog 'Jodu' of size 235×235 pixels each. These pictures were taken for different positions of the light source, for example, Figures 4.2(d, h) represent two images for the light source positions (0.1763, 0.5596, 1) and (−0.8389 −0.7193, 1), respectively. Each observation provides some additional information

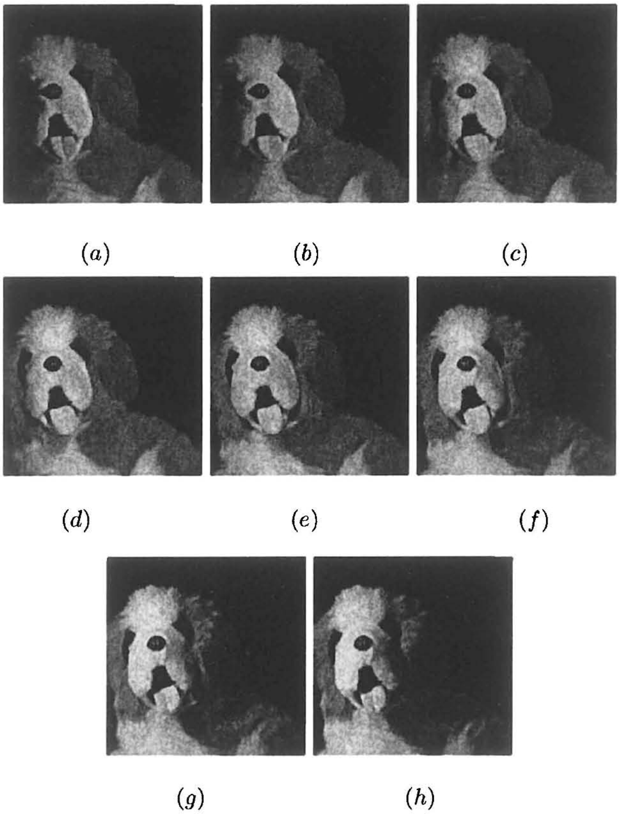
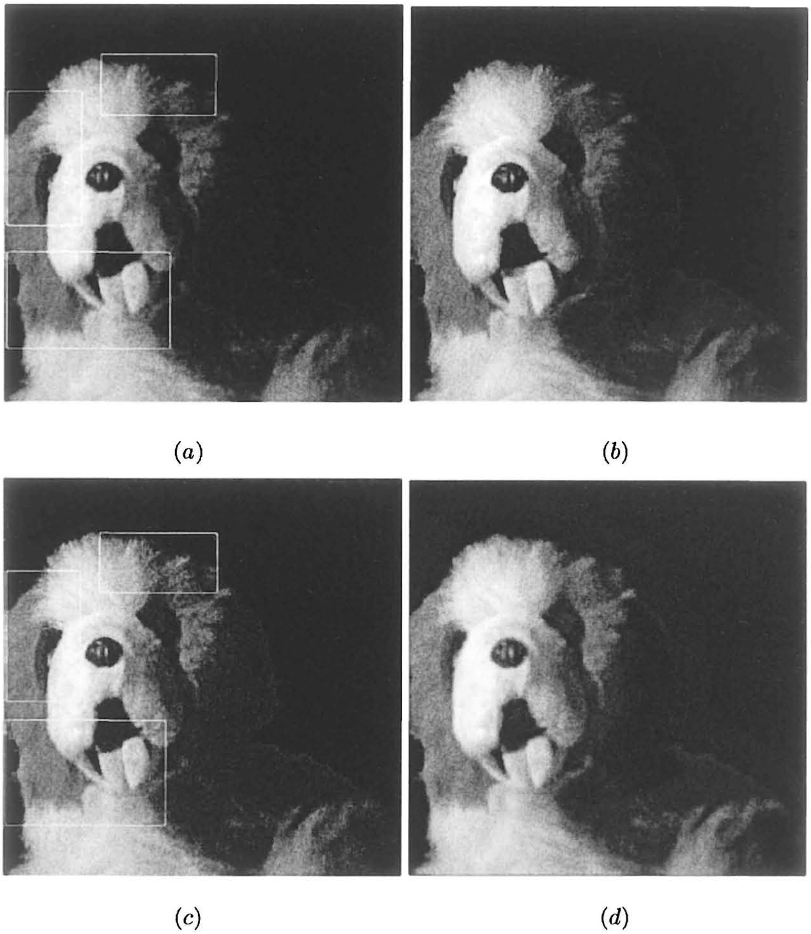


Fig. 4.2. Observed images of a doll 'Jodu' captured with eight different light source positions.

and we want to use this information for super-resolution purposes. The bilinearly interpolated image with a magnification factor of $r = 2$ for the observation in Figure 4.2(h) is shown in Figure 4.3(a). The super-resolved image obtained by interpolating in the p, q, ρ space individually and combined using the image irradiance equation, *i.e.*, by using only the generalized interpolation scheme (as discussed in section 4.2) is shown in Figure 4.3(b). Figure 4.3(c) shows the result obtained with the proposed MRF-based method and the super-resolved image using the variational approach is shown in Figure 4.3(d). We can clearly observe that the estimated super-resolved image obtained with the MRF-based method is more sharp revealing better details, such as the fur on the body of Jodu. The shadow on Jodu's tongue is better estimated as



(a) (b) (c) (d)

Fig. 4.3. (a) Bilinear interpolation of the image Jodu for the source position given in Figure 4.2(h). (b) High-resolution image obtained by using the method of generalized interpolation, super-resolved Jodu image (c) using the MRF-based approach, and (d) using the variational approach.

compared to the high resolution image obtained by the generalized interpolation (refer to Figure 4.3(b)) due to data consistency check. Also, the shadow under the left eye is also lost due to generalized interpolation. The image domain interpolation (refer to Figure 4.3(a)) does not suffer from such a problem, as it is obtained directly from the input image but it is too smooth compared to what can be achieved using the proposed scheme. In order to highlight the sharpness of the picture obtained with the MRF-based method, we draw rectangular boxes at three different regions in the images given in Figures 4.3((a) and (c)). These highlighted areas do show a substantive difference in sharpness in these two figures, justifying the use of the MRF-based technique. We can see that the result obtained with the variational approach is a bit too smooth as compared to using the MRF-based approach with discontinuity preservation. This is quite expected as the cost term for minimization does not include the line fields, which means that the edges are not well preserved. But it may be noted that due to the data consistency term, the shadow on the Jodu's tongue is better preserved in both the cases. It may also be noted that the motivation for the variational approach is its almost insignificant computational requirement compared to the MRF-based approach.

We also demonstrate that we can render high resolution views for arbitrary light source positions. The super-resolved image corresponding to an arbitrary source direction, *i.e.*, one which is not captured as an observation, is shown in Figure 4.4. This corresponds to a source position (0, 0, 1). The estimated super-resolved image is quite correct as the source position, essentially at the same place as the viewer (camera), causes the entire scene to be illuminated. There is no shadow. Thus, if the available source positions are insufficient to illuminate a particular part of an object, the method is suitable to reveal the details there in.

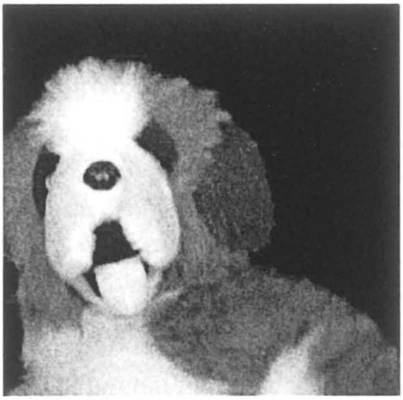


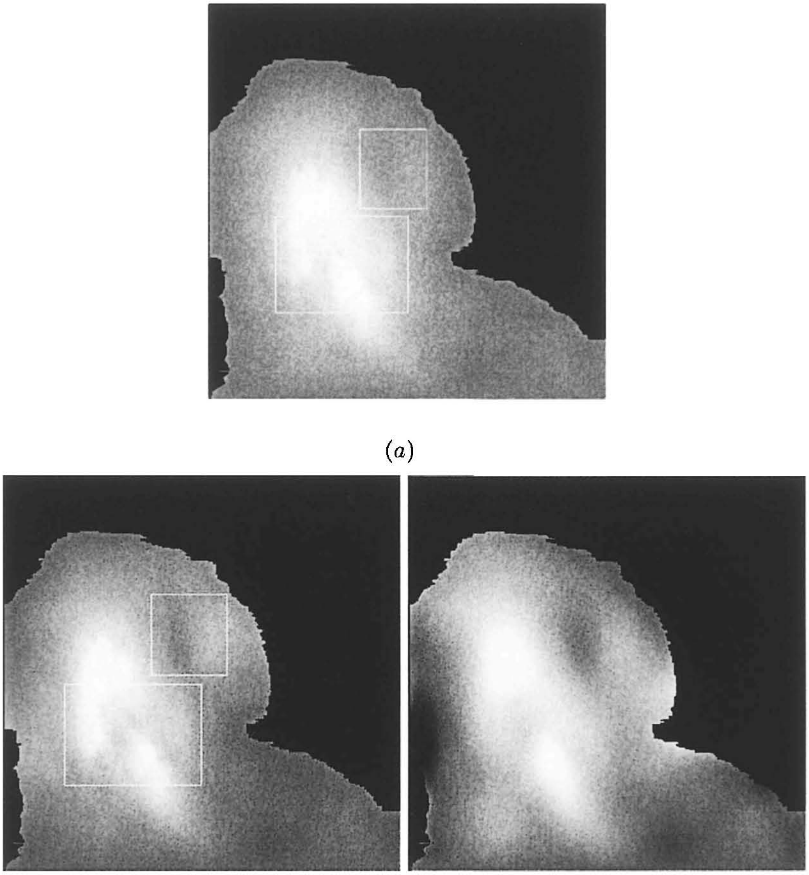
Fig. 4.4. Super-resolved, synthesized view of Jodu corresponding to a virtual light source position (0, 0, 1).

We now demonstrate how good is the structure recovery. Figure 4.5(a) shows the high resolution depth map of Jodu image shown as an intensity variation using the generalized interpolation method. The brighter it is, nearer it is to the camera. The segmentation of the foreground is done using the visibility criterion, *i.e.*, $(\hat{n}.\hat{s})$ should be greater than zero. The depth map estimated using the MRF-based method is shown in Figure 4.5(b) and the same using the calculus of variations is shown in Figure 4.5(c), which compares quite favorably with the depth map shown in Figure 4.5(b) obtained using the MRF-based method. We do notice an improvement in the depth estimation using the methods discussed; a better depth estimate was observed for the protrusion near the tongue and the dip seen near the left eye. These regions are highlighted in Figures 4.5((a) and (b)).

Let us now observe how well the high resolution albedo is recovered. Figure 4.6(a) shows the low resolution albedo for the dog image recovered from the photometric observations and the estimated high resolution albedo using the MRF-based approach is displayed in Figure 4.6(b). Quite expectedly the shadows do not affect the computation of albedo.

We have thus far not mentioned any thing about the choice of parameters in Eq. (4.8) and Eq. (4.12). One can use a cross-validation technique to select the best parameter set. However, we refrain from doing this in this study. These parameters were chosen mostly in an ad-hoc fashion with a little bit fine tuning. As a rule of thumb, we have chosen the parameters in such a way that the relative magnitudes of each component in the cost function are nearly equal. Thus the values of the parameters used for recovering the Jodu image, its surface gradients and the albedo using the MRF-based method are $\mu_z = 0.001$, $\mu_p = \mu_q = \mu_\rho = 100$, $\gamma_z = \gamma_p = \gamma_q = \gamma_\rho = 10$, $\alpha = 10.0$, $\lambda = 3.0$. Similarly the various parameters used in the cost function for the variational approach are $\beta_z = 800$, $\beta_p = \beta_q = \beta_\rho = 5000$, $\beta_d = 300$, $\beta_i = 20$. In order to reduce the computation time we used the output from the generalized interpolation as the initial estimate for both the cases.

We now consider a scene which has a gradual (almost linear) depth variation. We considered sixteen observations for this experiment. Four of them captured by varying the light source direction are shown in Figures 4.7(a-d). Each of these images have a size of 260×260 pixels. Figure 4.8(a) shows the bilinearly interpolated image for the low resolution image shown in Figure 4.7(b) with the source position



(a)

(b)

(c)

Fig. 4.5. Super-resolved depth map (a) using the method of generalized interpolation, (b) using the MRF-based method, and (c) using the variational approach.

(0.4663, 0.3523, 1). The image super-resolved with the generalized interpolation scheme is shown in Figure 4.8(b) and the corresponding one using the MRF-based method is displayed in Figure 4.8(c). We notice again that there is a considerable improvement in the reconstructed image. The stitches on the shoe are more clear with sharp transitions. Also the image in Figure 4.8(b) suffers from the fact that the generalized interpolation cannot handle shadows. Figure 4.8(d) show the super-resolved image using the variational approach for compari-

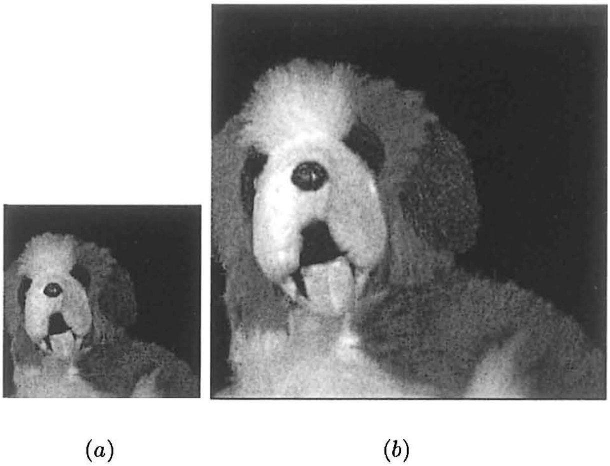


Fig. 4.6. (a) Recovered low resolution albedo of the doll Jodu and (b) the estimated high resolution albedo using the MRF-based approach.

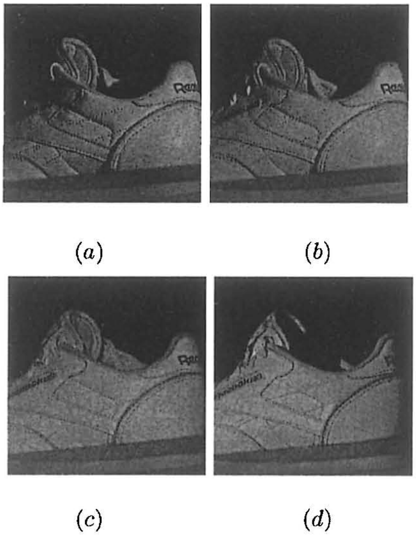


Fig. 4.7. Observed low resolution shoe images captured using four different light source positions.

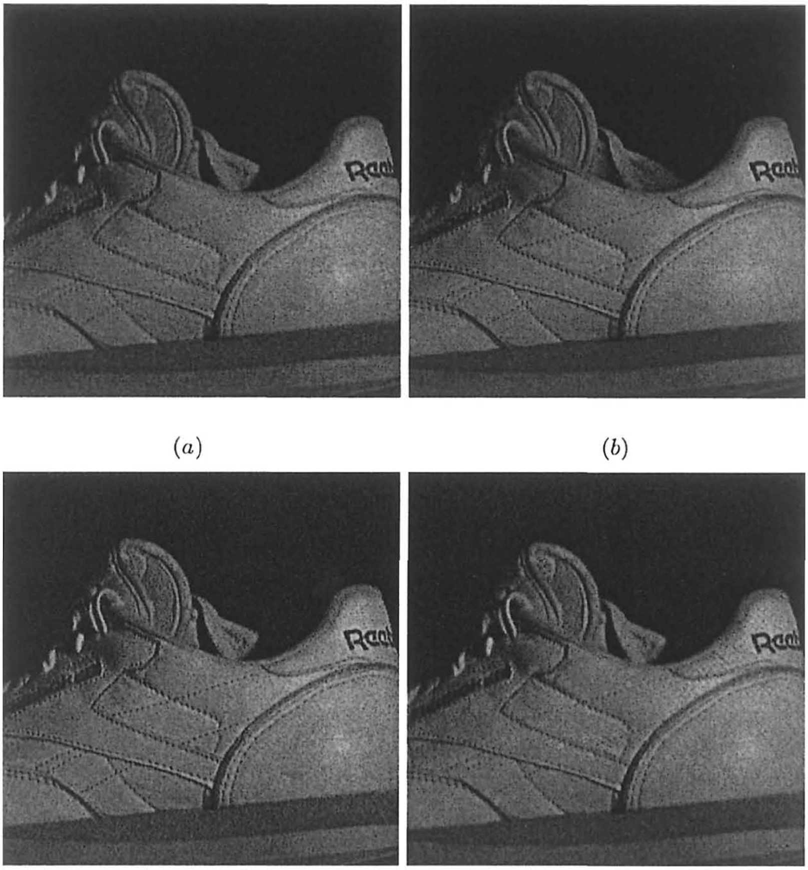


Fig. 4.8. (a) Intensity domain magnification with bilinear interpolation for the low resolution image shown in Figure 4.7(b), super-resolved shoe image for the same source position obtained (b) using the generalized interpolation technique, (c) using the MRF-based method, and (d) using the variational approach.

son. As expected this image is slightly smoother as compared to Figure 4.8(c). Figures 4.9(a, b) show the low resolution albedo recovered using the photometric measurements and the corresponding high resolution albedo using MRF-based approach, respectively. Once again all the details are clearly visible in the high resolution albedo as the shadows do not affect the calculation of albedo. The values of the parameters used for this experiment for recovering the shoe image and its

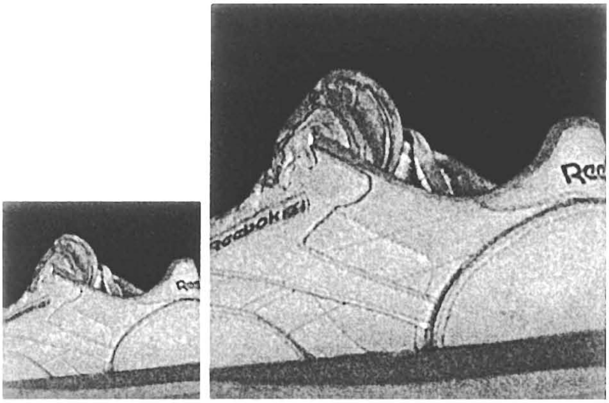


Fig. 4.9. (a) Recovered low resolution albedo of the shoe image and (b) the estimated high resolution albedo using the MRF-based approach.

surface gradients with the albedo are $\mu_z = 0.005$, $\mu_p = \mu_q = \mu_\rho = 2.0$, $\gamma_z = 20.0$, $\gamma_p = \gamma_q = \gamma_\rho = 0.05$, $\alpha = 10$, $\lambda = 5.0$. The parameters were kept the same as in the previous experiment for minimization scheme using the calculus of variations.

In order to test our algorithm for a higher magnification factor we now consider $r = 4$ for the shoe image with the same set of experimental data. We observed that the algorithm works well for higher magnification factor as well, as is evident from the results. The values of the various parameters used for this experiment are same as in the previous experiment. In this case we show the intensity domain super-resolved image for the low resolution image with source position $(-0.8390,\ -0.4168,\ 1)$ shown in Figure 4.7(d). Figure 4.10 shows the bilinearly interpolated shoe image for the corresponding source position. The estimated super-resolved image using the MRF-based method is shown in Figure 4.11. We notice again that the stitches on the shoe are more clear even after such a large magnification. Also the shadow to the rear side, just under the pull-up lace flap, is better preserved. Similarly, observe how nicely the shadow of the loosely coupled shoe lace on the top is preserved in Figure 4.11. Once again we highlight certain areas in the images to bring out the improvement in sharpness achieved using the proposed method. The estimated high resolution depth maps with the MRF-based approach and the generalized interpolation scheme are shown in Figures 4.12 and 4.13. Observe that the intensity from the right side of the shoe gradually decreases which is quite correct as the shoe was placed at an angle with the sensor plane while capturing the image.

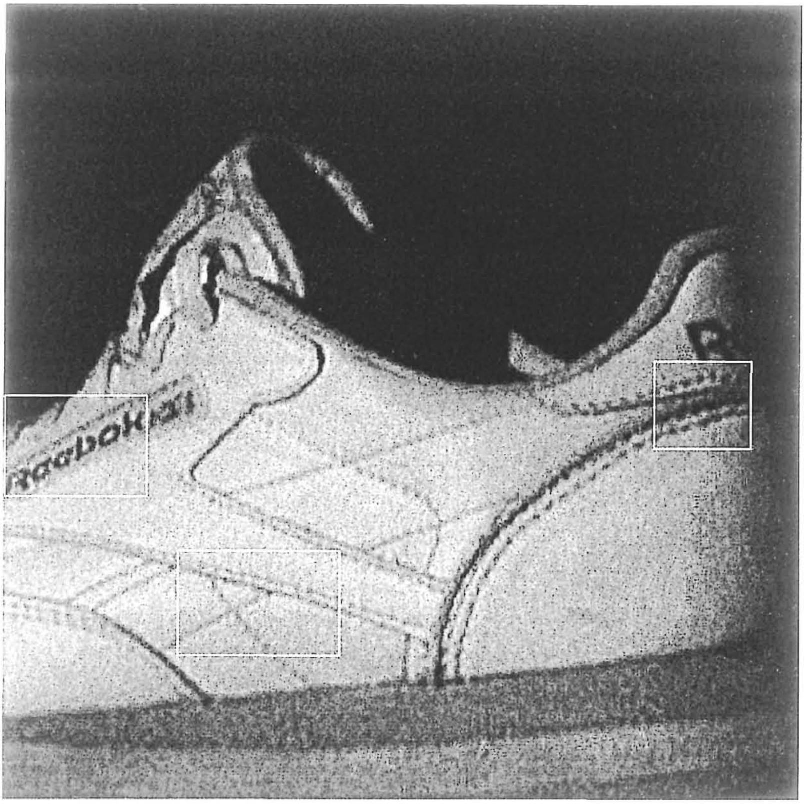


Fig. 4.10. Bilinear interpolation of the shoe image with $r = 4$ for the image shown in Figure 4.7(d)

We have till now not mentioned about the quantitative performance of the approach. We know that it is difficult to give quantitative assessment when the true high resolution images are unavailable for reference as all the captured images are real. But in order to quantify the performance of the proposed approach, all the captured Jodu images were first decimated by a factor of $r = 2$. The decimated low resolution image set is used as the input (observations) and the original undecimated real images are considered as the true reference. The experiments for

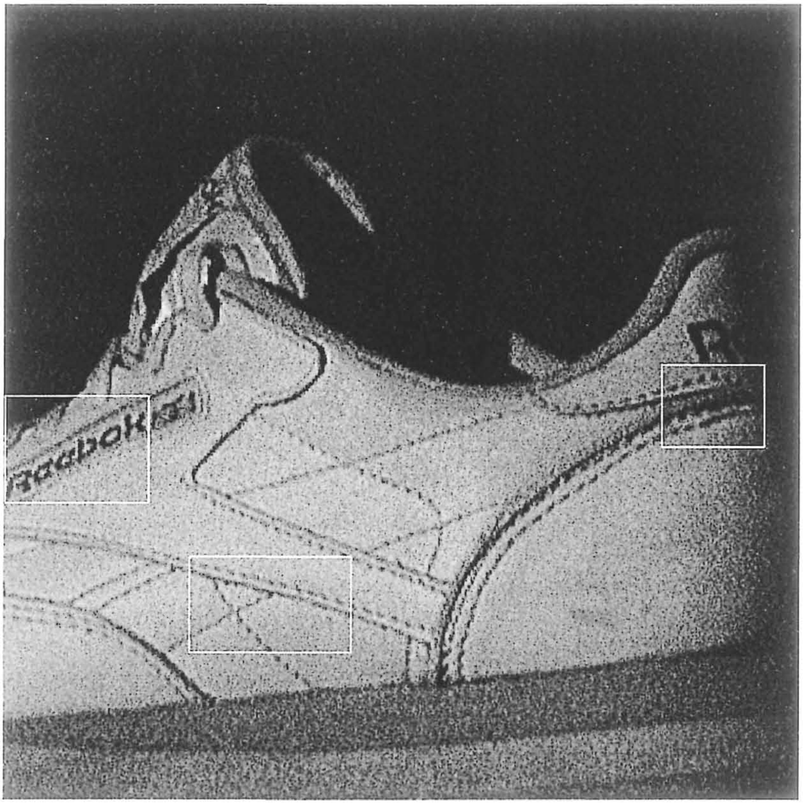


Fig. 4.11. Super-resolved shoe image with $r = 4$ obtained using the proposed MRF-based approach for the image shown in Figure 4.7(d)

super-resolution of intensity field and depth field are conducted on these decimated observations. We compare the mean squared error between the reference images and the recovered high resolution images and use the peak signal to noise ratio (PSNR) as the figure of merit for the restored images. The PSNR in decibels (dB) is computed by using the following equation.

$$PSNR = 10 * \log_{10} \frac{255^2}{\frac{1}{MN}||(f - \hat{f})||^2}, \tag{4.14}$$

where f and $\hat{f}$ are the true and estimated entities, respectively and $M \times N$ is the size of the image. We compare the performance of the MRF-based method with those of the bicubic interpolation and the generalized interpolation in table 4.1. We also compare the performance

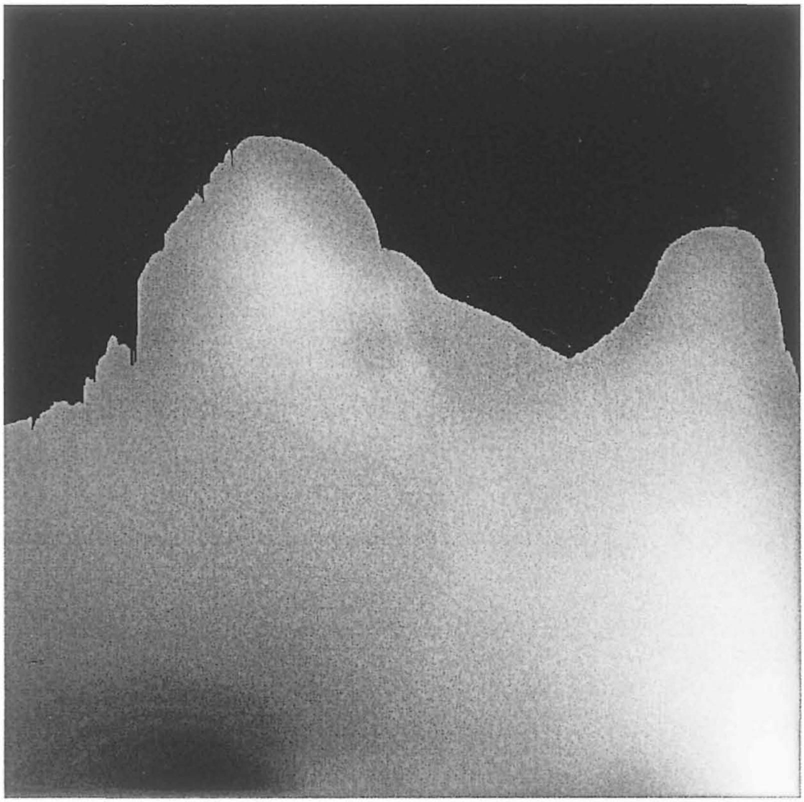


Fig. 4.12. Super-resolved depth map with $r = 4$ using the MRF-based approach

of the scheme in terms of the PSNR of the recovered high resolution depth map for the experimental set up. In order to generate the reference (true) depth map, we use the photometric stereo to compute the surface gradients on the undecimated observations. From the table we see that our approach outperforms the bicubic interpolation and the method of generalized interpolation by about 0.5 to 3.0 dB in PSNR depending on the direction of the light source. We also have a substantial improvement in the depth estimation using the proposed method as the high resolution estimates of the gradient fields are much better due to imposition of smoothness and integrability constraints.

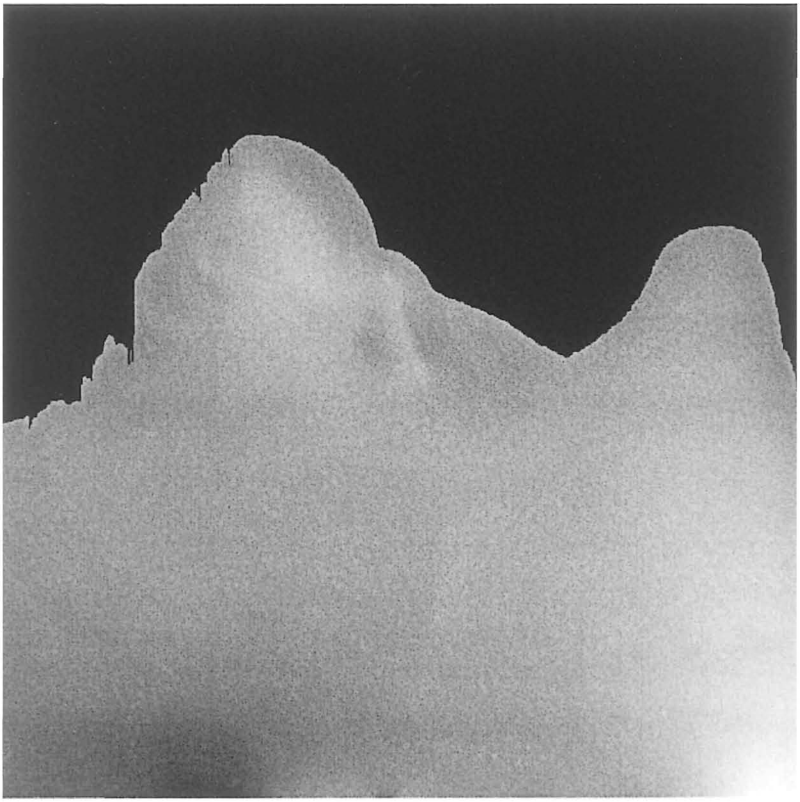


Fig. 4.13. Super-resolved depth map with $r = 4$ using the generalized interpolation scheme.

4.6 Conclusions

In this chapter we have investigated a regularization based approach to simultaneously enhancing the surface gradients and the albedo representing the object shape and the reflectance of the surface and obtain the super-resolved image for the scene. The observations for different light source positions were used to obtain the photometric measurements. We model the super-resolved image and the structure of the object as separate MRF fields. This method avoids the correspondence and warping problems inherent in current super-resolution techniques involving the motion cue in the low resolution observations. Also the method has the advantage that the 3D structure of the scene is better preserved during the super-resolution process. We also showed that the

Image for source position	PSNR in dB Bicubic Interpolation	Generalized Interpolation	MRF Approach
(0.8389, 0.7193, 1)	35.83	30.85	36.16
(0.5773, 0.6363, 1)	36.43	34.69	37.34
(0.3638, 0.5865, 1)	36.17	34.32	37.54
(0.1763, 0.5596, 1)	34.60	31.76	37.62
(-0.1763, -0.5596, 1)	33.36	31.60	35.15
(-0.3638, -0.5865, 1)	33.32	32.49	35.27
(-0.5773, -0.6363, 1)	33.19	32.19	35.49
(-0.8389, -0.7193, 1)	33.80	30.14	36.08
(DEPTH)	17.78	36.16	41.82

Table 4.1. PSNR Comparison for Jodu image for a magnification factor of 2 for different source positions. The last row in the table gives the PSNR comparison for the depth field.

same problem can also be solved under the variational approach when the line fields are eliminated. The corresponding iterative solution is much faster, albeit it yields a little smoother output.

Naturally there are also several limitations of this particular way of super-resolving a scene. We need a controlled environment so that the photometric observations can be obtained. This excludes the method from being applied to super-resolve an outdoor scene. Further, it is assumed that the reflectance model of the scene is known. In many real world applications there will be objects in the scene that are either metallic or have a chemical paint on them. They behave far from being a Lambertian surface. Any specularity on the surface has not been accounted for in this study. Nevertheless, this study provides us with an idea that, in some cases, it may be advisable not to interpolate an image directly. We may be able to perform better (as Table 4.1 suggests) by decomposing it into a number of constituent functions or fields and upsampling them individually. In chapter 7 we shall demonstrate that a principal component analysis (PCA) can be used as an alternative way of performing the generalized interpolation.

5

Blind Restoration of Photometric Observations

In the previous chapter we addressed the problem of image super-resolution using photometric observations. However, we assumed that the observations are all free from blurring. Image super-resolution is a restoration problem where we attempt to undo the aliasing as well as the blurring introduced in the observed images. This chapter addresses the problem of simultaneous estimation of scene structure and restoration of images from blurred photometric measurements. In photometric stereo, structure of an object is determined by using a particular reflectance model (the image irradiance equation) without considering the blurring effect. What we show in this chapter is that, given arbitrarily blurred observations of a static scene captured with a stationary camera under different illuminant directions, we still can obtain the structure represented by the surface gradients and the albedo and also perform a blind image restoration. As before, the surface gradients and the albedo are modeled as separate Markov random fields (MRFs) and a suitable regularization scheme is used to estimate the different fields as well as the blur parameter. The results of the experimentations are illustrated with real as well as synthetic images.

In the existing literature on shape from shading or photometric stereo the researchers have treated the problem of shape estimation without considering the blur introduced by the camera. This is not true when one captures the images with a real aperture camera as the blur could be introduced due to camera jitter or out-of-focus blur. The variations in image intensity due to camera blur affects the estimates of the surface shape. Thus the estimated shape differs from the true shape. This limits the applicability of these techniques in image super-resolution and in 3D computer vision problems.

Among many cues that are used for structure estimation, the photometric stereo (PS) and the shape from shading (SFS) methods recover the shape from the gradual variation in shading. While the PS employs two or more observations under different illuminations or light source positions, the SFS provides the shape estimation from a single observation. In many practical applications, PS has been found to yield better shape estimates. It requires that the entire surface area of the object be illuminated from different light source positions and hence gives better results with more number of observations taken under different illuminant positions. PS estimates the slope of the surface by measuring how the intensity varies with the direction from which they are illuminated.

As discussed in the next section researchers traditionally treat the shape from shading or photometric stereo problem without considering the blur introduced by the camera. These approaches assume a pinhole model that inherently implies that there is no camera blur during observation. However, when one captures the images with a camera, the degradation in the form of blur and noise is often present in these observed image intensities. Some of the low end commercial cameras fail to set the auto-focus properly when the illumination is not bright – typically the case in photometric measurements. Similarly when the illuminant direction is changed for the subsequent shots, the camera tries to re-adjust the focus (although there is no need for it as the object and the camera are both stationary), and focusing error does creep in. It is natural that the variations in image intensity due to camera blur affects the estimates of the surface shape. Thus, the estimated shape differs from the true shape in spite of possibly having the knowledge of the true surface reflectance model. This limits the applicability of these techniques in 3D computer vision problems. This motivates us to restore the image as well, while recovering the structure.

The problem can then be stated as follows: given a set of blurred observations of a static scene taken with different light source positions, obtain the true depth map and the albedo of the surface as well as restore the images for different light source directions, *i.e.*, estimate the true structure and also the images. Since the camera blur is not known, in addition, we need to estimate the point spread function (PSF) of the blur which caused the degradation. We assume a point light source illumination with known source directions and an orthographic projection. The problem can be classified as a joint blind restoration and surface recovery problem. Since such a problem is inherently ill-posed, we need suitable regularization of all the fields to be estimated, *i.e.*, surface gradients as well as albedo. We cast the entire problem as a problem of regularization and hence solve it iteratively.

5.1 Prior Work

Here we do a quick review of the current status of research in shape from shading, which was not covered in chapter 2. Shading is an important cue for human perception of surface shape. Researchers in computer vision have attempted to use the shading information to recover the 3D shape. Horn was one of the first researchers to study this problem by casting it as a solution to a second order partial differential equation [139]. Shape from shading (SFS) problem is typically solved using four different approaches. These approaches include the regularization approach, the propagation approach, the local approach and the linear approach. Ikeuchi and Horn [140] were the first to use the energy minimization technique using a reflectance model and the smoothness constraint. The propagation approach is basically the characteristic strip method proposed by Horn [139]. It starts with a solution along a special direction and obtains the profile of the surface called characteristic curve along this direction. If we are given initial information as a profile along some curve and not just as a point, we can integrate along curves starting at a point on this initial curve and obtain the depth profile for the whole surface. Pentland [141] uses the local approach to recover the shape information by using the image intensity and its first and the second derivatives. He assumes that the surface is locally spherical at each point. Linear approaches proposed by Pentland [142], and Tsai and Shah [143] linearize the reflectance map and solve for the shape.

It is well known that the SFS problem is ill-posed and hence the solution may not be reliable. Also, most of the traditional SFS algorithms assume that the surface has constant albedo values. This assumption restricts the class of recoverable images. The researchers thus attempted to solve the problem of shape recovery by using the photometric stereo (PS) by making use of multiple images to provide additional information for robust shape recovery. Even though some of the details of the local surface characteristics may be lost due to the least squares approach in PS, the global accuracy, in most cases, is better than the SFS. The idea of PS was initially formulated by Woodham [144, 145] and later applied by others [146, 147]. The authors in [146, 147] use multiple images using distant light sources at different directions. They consider both the Lambertian as well as the non-Lambertian reflectance models of the surface.

Authors in [148] propose two robust shape recovery algorithms using photometric stereo images. They combine the finite triangular surface model and the linearized reflectance image formation model to express the image irradiance. The recovery of albedo values for color images using photometric stereo has been considered by Chen *et al.* [149]. They show that the surfaces rendered using the calculated albedo values are more realistic than surfaces rendered using a constant albedo. The authors in [150, 151, 152] use a calibration object of known shape with a constant albedo to obtain the nonlinear mapping between the image irradiance and the surface orientation in the form of a lookup table. A neural network based approach to photometric stereo, for a rotationally symmetric object with a nonuniform reflectance factor is considered in [153]. In [154] shape from photometric stereo with uncalibrated light sources is discussed. The recovery of the surface normal in a scene using the images produced under a general lighting condition which may be due to a combination of point sources, extended sources and diffused lighting, is considered by Basri and Jacobs [155]. They assume that all the light sources are isotropic and distantly located from the object.

In order to improve the performance of the shape recovery method, the SFS algorithm is integrated with the PS in [156]. Here the recovered albedo and the depth from photometric stereo are used in SFS to obtain a better depth estimate. A different method for obtaining the absolute depth from multiple images captured with different light source positions is presented in [157]. It involves solving a set of linear equations and is applicable to a wide range of reflectance models. Another approach to PS where the input images are matched through an optical flow is presented in [158]. The resulting disparity field is then used in a structure-from-motion reconstruction framework which does not require the reflectance map information. Recently, several researchers have applied PS to the analysis, description and discrimination of surface textures [159, 160, 161, 162]. It has also been applied to the problems of machine inspection of paper [163] and identification of machined surfaces [164]. But none of these methods consider the presence of blur in the observations.

We now discuss, in very brief, a few of the research works carried out on image restoration and blind image deconvolution. Image restoration is probably the area that has attracted the maximum amount of research focus over the last three decades, resulting in a huge volume of work by many researchers. It will be an injustice if we claim that we are doing a literature review on image restoration here. We restrict ourselves to reviewing only a few relevant papers. The general approaches for image restoration include both stochastic and deterministic methods. Stochastic approaches assume that the original image is a realization of a random field, usually a Markov random field (MRF). Maximum likelihood (ML), maximum entropy (ME), and maximum *a posteriori* (MAP) approaches are specific types of stochastic methods for restoration. The deterministic approaches frequently use constraints to make the restoration problem tractable. Popular among these methods are projection onto convex sets (POCS) and regularization. For a comprehensive survey of various digital image restoration techniques published prior to 1997, the reader is referred to [165]. A plethora of methods have also been proposed to solve the problem of blind image deconvolution [166, 167, 168]. A variant of the ML estimation, *i.e.*, expectation-maximization (EM) algorithm has been used for blur identification and image restoration in [169, 170]. Some of the POCS-based iterative restoration techniques are discussed in [171, 172, 173, 174, 175]. Recently, Candela *et al.* use local spectral inversion of a linearized total variation model for denoising and deblurring [176]. The model here consists of a system of partial differential equations (PDEs) obtained as a local linearization of a variational problem. Reconstruction of degraded images corrupted with impulsive noise for restoration of digital films is considered in [177]. An unsupervised edge-preserving image restoration and estimation of Gibbs hyperparameters is discussed by Bedini *et al.* [178]. They model the image to be restored as a Markov random field (MRF) and use a mixed annealing algorithm for maximum *a posteriori* (MAP) restoration which is periodically interrupted to compute the ML estimates of the MRF parameters.

As discussed above, the researchers have treated the shape estimation and restoration problems separately. For shape estimation using the shading cue, the blur introduced by the camera is never considered. We demonstrate in this chapter that both the shape estimation and restoration problems can be handled jointly in a unified frame-

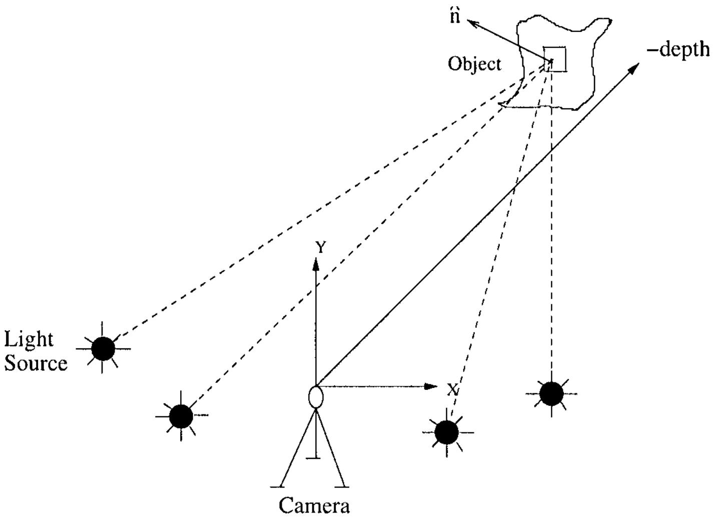


Fig. 5.1. An illustration of an observation system for photometric stereo.

work. In other words we extract the depth values considering the effect of observation blur and restore the image by simultaneously identifying the blur parameter using a suitable regularization approach.

5.2 Observation Model

Consider a scene illuminated with different light source positions. Figure 5.1 shows the setup for capturing the scene using the photometric stereo method. Here we consider that both the object and the camera positions are stationary. We capture the images with a large distance between the object and the camera, thus making a reasonable assumption of orthographic projection and neglect the depth related perspective distortions. The light source is assumed to be a distant point light. Under this assumption the incident light rays can be characterized by unit vectors. Now given an ensemble of images captured with different light source positions, using the theory of photometric stereo we can express the intensity of the image at a point using the image irradiance equation as

$$E(x,y) = \rho(x,y)\mathcal{R}(p(x,y),q(x,y)) = \rho(x,y)\hat{n}(x,y).\hat{s}, \qquad (5.1)$$

where $\hat{n}$ is the unit surface normal at a point on the object surface, $\hat{s}$ is the unit vector defining the light source direction and ρ is the albedo or the surface reflectance of the surface. The surface gradients (p, q) are used to specify the unit surface normal given by

$$\hat{n} = (-p, -q, 1)^T / \sqrt{1 + p^2 + q^2}.$$

$\mathcal{R}$ is called the reflectance function. We concentrate on the Lambertian model in our study, but the method can be expanded to other reflectance models also. As discussed in the previous chapter, the surface gradients (p, q) and the albedo can be recovered using a minimum of three observations provided the three equations due to the measurements are linearly independent. In practice, one often uses more than three observations due to inconsistency of measurement equations and a least squares solution is sought. The solution to Eq. (5.1) using the different measurements gives the true surface gradients and the albedo in the least squares sense only when we do not consider the camera blur. However, due to improper focus setting or camera blur the observations are often blurred. Thus, considering the effect of blur the observed image in Eq. (5.1) can be expressed as

$$g(x, y) = h(x, y) * E(x, y) + \eta(x, y),$$

where $h(x, y)$ represents the two-dimensional point spread function (PSF) of the imaging system, and $\eta(x, y)$ is an additive noise introduced by the system. Considering K light source positions, the observed images can then be expressed as

$$g_m(x, y) = h(x, y) * E_m(x, y) + \eta_m(x, y), \quad m = 1, \cdots, K \tag{5.2}$$

Here $E_m(x, y)$ corresponds to the actual shaded image due to the m^{th} light source position. From Eq. (5.1), $E_m(x, y)$ is a function of ρ, p, q. It may be noted that, since there is no relative movement between the camera and the object, the PSF remains the same for all the observations. This assumption is violated if there is a camera jitter. Note that we assume here that there is no chromatic aberration due to the lens in the observations. If it does exist, it will affect each color channel differently.

If $\mathbf{E}_m$ is the $N^2 \times 1$ lexicographically ordered vector representing the image irradiance for m^{th} light source position, and let $\mathbf{g_m}$ be the corresponding observation vector, then using Eq. (5.2), the observed images can be modeled as

$$\mathbf{g}_m = H(\sigma)\mathbf{E}_m(\rho, p, q) + \boldsymbol{\eta}_m, \quad m = 1, \cdots, K \tag{5.3}$$

where $H(\sigma)$ is a $N^2 \times N^2$ matrix with σ representing the blur parameter. $\mathbf{E}_m$ is the true focused image for the m^{th} light source position and is of size $N^2 \times 1$. Note $\mathbf{g}_m$ is a function of the surface gradients and the albedo as seen from Eq. (5.1). We assume that the blur is due to the camera out-of-focus which can then be modeled by a pillbox blur or by a Gaussian PSF characterized by the spread parameter σ [105]. We also assume that the blur is space invariant. This is tantamount to assuming that the variation in depth in the object is very small compared to its distance from the camera. Hence H becomes a block-Toeplitz matrix representing the linear convolution operator. Here $\boldsymbol{\eta}_m$ is the $N^2 \times 1$ noise vector which is zero mean i.i.d Gaussian. Our problem now is to estimate the blur parameter σ, the albedo $\boldsymbol{\rho}$, the surface gradients $\mathbf{p}$ and $\mathbf{q}$, and also perform blind image restorations given the observations $\mathbf{g}_m$, $m = 1, \cdots, K$. This is an ill-posed inverse problem, requiring a suitable regularization.

5.3 Restoration and Structure Recovery

As we are using a regularization based approach for simultaneous estimation of different parameter fields ($\mathbf{p}$, $\mathbf{q}$, and $\boldsymbol{\rho}$), we need to use suitable priors for the fields to be estimated. As discussed in earlier chapters, the MRF provides a convenient and consistent way of modeling context dependent entities such as image pixels, depth of the object and other spatially correlated features. Once again we model the structure $\mathbf{p}$ and $\mathbf{q}$, and the albedo ρ as separate MRFs.

5.3.1 Structure and Image Recovery with Known Blur

For an easier understanding, we first consider the case where we estimate the albedo, the surface gradients and also restore the image when the blur is known. This will be relaxed in the next subsection.

The proposed method can be illustrated using Figure 5.2. We obtain K observations of a static scene by varying the direction of the point light source. It is assumed that the directions are known. We also assume that the reflectance model is known. One can now obtain a least squares estimate of the surface gradients ($p^{(0)}(x,y)$, $q^{(0)}(x,y)$) and the albedo $\rho^{(0)}(x,y)$ assuming that the observations are free from blur, which are used as the initial estimates. Expectedly these estimates are quite poor due to blurred observations. We introduce the context dependencies in the estimated fields by modeling them as separate MRFs.

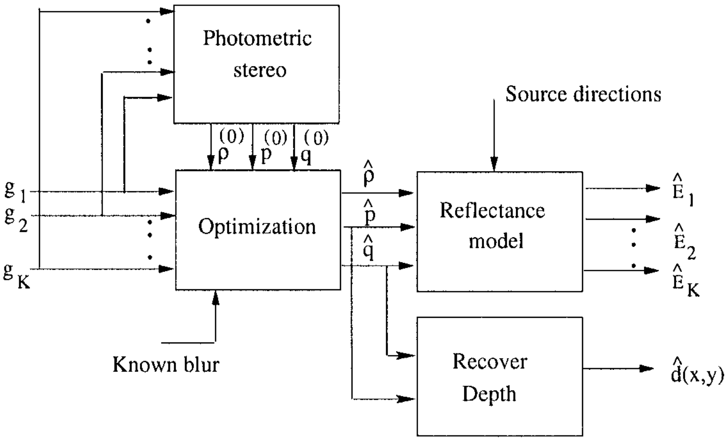


Fig. 5.2. Illustration of how the image and the structure can be recovered when the blur PSF is known.

Thus the corresponding priors are $U(\mathbf{p})$, $U(\mathbf{q})$, and $U(\boldsymbol{\rho})$. We use the following energy function $U(\mathbf{w})$ for each of the fields $U(\mathbf{p})$, $U(\mathbf{q})$, and $U(\boldsymbol{\rho})$.

$$U(\mathbf{w}) = \mu_w \sum_{k=1}^{N-2} \sum_{l=1}^{N-2} [(w_{k,l} - w_{k,l-1})^2 + (w_{k,l} - w_{k-1,l})^2 + (w_{k,l+1} - w_{k,l})^2 + (w_{k+1,l} - w_{k,l})^2], \tag{5.4}$$

where $\mathbf{w} = \mathbf{p}$, $\mathbf{q}$, or $\boldsymbol{\rho}$. Here μ_w is a penalty term for departure from smoothness. Thus considering the brightness constraint term and the smoothness term for regularizing the solution, the final cost function can be expressed as

$$\epsilon = \sum_{m=1}^{K} ||\mathbf{g}_m - H(\sigma)\mathbf{E_m}(\rho, p, q)||^2 + U(\mathbf{p}) + \mathbf{U}(\mathbf{q}) + \mathbf{U}(\boldsymbol{\rho}). \tag{5.5}$$

This cost function is convex and can be minimized with respect to the fields $\mathbf{p}, \mathbf{q}$, and $\boldsymbol{\rho}$ using a gradient descent method. The initial estimates $\mathbf{p}^{(0)}$, $\mathbf{q}^{(0)}$, and $\boldsymbol{\rho}^{(0)}$ are used here to speed up the convergence. Having obtained the estimated fields $\hat{\mathbf{p}}$, $\hat{\mathbf{q}}$, and the albedo $\hat{\boldsymbol{\rho}}$ one can obtain the restored image for a particular light source direction by using Eq. (5.1). Since the surface gradients $\mathbf{p}$ and $\mathbf{q}$ are already estimated, it is straight forward to restore the depth $d(x, y)$ of the scene, which is obtained by solving the Eq. (4.9) iteratively, where p_x and q_y now represent the derivatives of the estimated surface gradients in x and y directions and not the super-resolved fields. It should be noted here that we are not performing direct image deconvolution which is highly ill-posed and leads to numerical instability.

5.3.2 Blur Estimation

We now extend the method to a more realistic situation in which the blur PSF σ in $H(\sigma)$ is also unknown. In order to do that we must estimate the amount by which an image is blurred. When the images are captured with a camera, the blur phenomenon could occur due to various reasons even when the camera is stationary. Thus it is required to estimate the blur while restoring the image and estimating the structure. Considering that the unknown blur is due to the effect of improper focusing, it can be modeled by a Gaussian PSF, when we need to estimate the blur parameter σ (standard deviation) that determines the severity of the blur. We have already discussed in chapter 3 how the blur parameter can be efficiently computed using the method proposed in [109]. This is due to the fact that the blur is parameterized by a single parameter σ.

Since the blur is mostly due to camera defocus, the PSF can be easily parameterized by a single parameter σ. Hence the PSF estimation problem simplifies drastically. Let $g(x,y)$ and $f(x,y)$ be two images with σ being the blur parameter where $g(x,y)$ represents the blurred image while $f(x,y)$ is the true focused image. Then $g(x,y)$ can be expressed in terms of $f(x,y)$ by a simple convolution operation as

$$g(x,y) = h(x,y;\sigma) * f(x,y). \tag{5.6}$$

Using the fact that blur PSF $h(x,y;\sigma)$ is Gaussian, it can be expressed as

$$h(x,y;\sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}}. \tag{5.7}$$

Now taking the the Fourier transform on both sides of Eq. (5.6), and making use of the derivation given earlier in chapter 3, one can show that

$$\sigma^2 = \frac{1}{\mathcal{A}} \int\int_A \frac{-2}{w_x^2 + w_y^2} \log \frac{\hat{G}(w_x, w_y)}{\hat{F}(w_x, w_y)} dw_x dw_y, \tag{5.8}$$

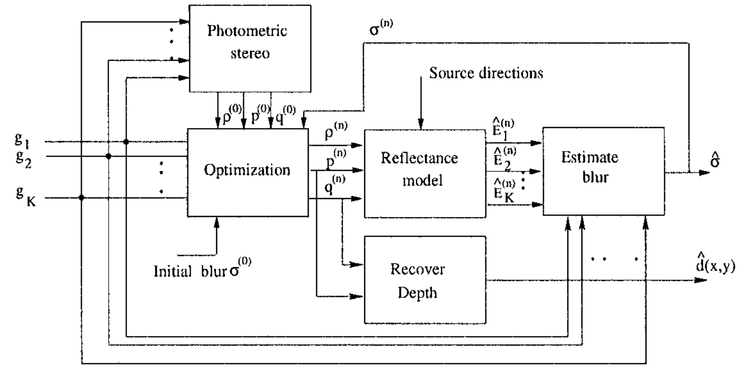


Fig. 5.3. Illustration of the proposed method for image and structure recovery when the blur PSF is unknown. A comparison with Figure 5.2 shows that we have added another block "Estimate blur" to estimate the blur between ($\mathbf{g}_i$, $\hat{\mathbf{E}}_i^{(n+1)}$). Further, the output of this block is fed back to the optimization block to improve the structure estimation.

where A is a small region in the frequency domain and $\mathcal{A}$ is the area of A. Compare this equation with Eq. (3.6) derived in chapter 3. One may notice that the only difference is that we set $\sigma_1 = \sigma$ and $\sigma_2 = 0$.

5.3.3 Structure Recovery and Blind Image Restoration

The blur estimation technique as discussed in the previous section 5.3.2 gives the estimate of blur only when the true focused image $f(x, y)$ and its blurred version $g(x, y)$ are available. But our problem is to estimate the blur given only the blurred observations, since the true focused images for different light source positions, *i.e.*, $E_m(x, y)$ in Eq. (5.2), are unknown. In this section we describe an approach for simultaneously estimating the blur parameter and the structure, given only the blurred photometric observations. As already mentioned, for blind image restoration we use a Gaussian blur which can be parameterized by the standard deviation σ.

An iterative approach for joint structure recovery and blind image restoration can be obtained by suitably modifying the block diagram given in Figure 5.2. The approach is schematically presented in Figure 5.3. Using the photometric stereo we obtain the least squares estimates of the fields $\mathbf{p}$, $\mathbf{q}$ and $\boldsymbol{\rho}$ that serve as the initial estimate as before. The optimization as given in Eq. (5.5) is carried out with these initial estimates using an initial value of the blur parameter $\sigma^{(0)}$. The cost function in Eq. (5.5) is minimized for $\mathbf{p}$, $\mathbf{q}$, $\boldsymbol{\rho}$ keeping $\sigma^{(0)}$ constant. Having estimated $\mathbf{p}$, $\mathbf{q}$, and $\boldsymbol{\rho}$, we obtain a revised estimate of σ as follows. The new estimates of fields $\mathbf{p}^{(n)}$, $\mathbf{q}^{(n)}$, $\boldsymbol{\rho}^{(n)}$ are used in image irradiance Eq. (5.1) along with the source directions to get the estimates of the focused images for different light source positions. We then obtain the new estimate of $\sigma^{(n)}$ by using the Eq. (5.8) holding $\mathbf{p}^{(n)}$, $\mathbf{q}^{(n)}$, $\boldsymbol{\rho}^{(n)}$ constant. Here the blurs (σ) are calculated between the observed images $\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_K$ and the estimated images $\hat{\mathbf{E}}_1^{(n)}$, $\hat{\mathbf{E}}_2^{(n)}$, $\cdots, \hat{\mathbf{E}}_K^{(n)}$ and the average value of the estimated blur parameter σ is used as the updated one. This new value of $\sigma^{(n)}$ is then used again in the optimization (Eq. (5.5)) to update the fields $\mathbf{p}$, $\mathbf{q}$, $\boldsymbol{\rho}$. The blur parameter and the structure (along with the images for different light source directions) are then estimated in an alternative way by keeping the blur parameter constant and updating the structure and vice-versa. The estimation of blur parameter and the different fields are carried out until the convergence is obtained in terms of the update for the parameter $\sigma^{(n)}$. The blur thus obtained is the final estimated one. The corresponding gradient fields are then used to calculate the depth map. It should be mentioned here that the mask size chosen for the PSF should be sufficiently large compared to the value of σ. Typically we use the size to be larger than 6σ. Since σ is not known, we use the PSF kernel size to be 13×13 pixels as the defocus blur during the experimentation is rarely expected to exceed $\sigma = 2$ pixels. Further, one may note that there has been no attempt to perform any deconvolution of the observed data after having estimated the blur parameter. Experimentally we found that such an effort always leads to an inferior performance compared to that of the proposed method. The complete procedure is summarized below in terms of the steps involved.

STEP 1: Obtain initial estimates $\mathbf{p}^{(0)}$, $\mathbf{q}^{(0)}$, $\boldsymbol{\rho}^{(0)}$ using the photometric stereo on blurred data.

STEP 2: Choose an initial blur parameter $\sigma^{(0)}$. Typically $\sigma^{(0)} = 0$.

STEP 3: Set $n = 0$.

STEP 4: Update the albedo and the scene structure

$$\{\mathbf{p}^{(n+1)}, \mathbf{q}^{(n+1)}, \boldsymbol{\rho}^{(n+1)}\} \leftarrow \arg\min_{\mathbf{p},\mathbf{q},\boldsymbol{\rho}} \{U(\mathbf{p}) + \mathbf{U}(\mathbf{q}) + \mathbf{U}(\boldsymbol{\rho})$$

$$+ \sum_{m=1}^{K} \|\mathbf{g}_m - H(\sigma^{(n)})\mathbf{E}_{\mathbf{m}}(\rho, p, q)\|^2\}.$$

STEP 5: Resynthesize focused images $\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2, \cdots, \hat{\mathbf{E}}_K$ for different light source positions using the Eq. (5.1)

$$\hat{E}^{(n+1)}(x,y) = \rho^{(n+1)}(x,y)\mathcal{R}(p^{(n+1)}(x,y), q^{n+1}(x,y)).$$

STEP 6: Estimate the blurs between $(\mathbf{g}_1, \hat{\mathbf{E}}_1^{(n+1)})$, $(\mathbf{g}_2, \hat{\mathbf{E}}_2^{(n+1)})$, $\cdots$, and $(\mathbf{g}_K, \hat{\mathbf{E}}_K^{(n+1)})$ using the Eq. (5.8) after replacing $\hat{F}$ by $\mathcal{F}[\hat{E}_i]$.

$$\sigma_i^{(n+1)} = \sqrt{\frac{1}{\mathcal{A}} \int\int_A \frac{-2}{w_x^2 + w_y^2} \log \frac{\hat{G}_i(w_x, w_y)}{\mathcal{F}[\hat{E}_i](w_x, w_y)} dw_x dw_y},$$

for the i^{th} pair and calculate the average value of blur $\hat{\sigma}^{(n+1)}$ from K such observations. Here $\mathcal{F}[\hat{E}_i]$ represents the Fourier transform of $\hat{E}_i$. Readers are requested to make note of this departure of the symbol.

STEP 7: Set $n = n + 1$ and go to step 4 until the convergence in the estimate of σ is obtained.

STEP 8: Solve for depth using Eq. (4.9)

$$\nabla^2 d(x,y) = p_x(x,y) + q_y(x,y).$$

A comment about the convergence of the proposed technique is now in order. Like many similar optimization algorithms wherein one alternately estimates two sets of parameters by freezing one of the sets and updating the other set of parameters, a global convergence cannot be proved. However, it has been shown in [179] that the computation is quite stable and it converges to a local minima. Since steps *5* and *6* in the above description involve nonlinearities, a good initial estimate may be required for obtaining a quality solution. As per the observation of [179], even an initial estimate of $\sigma = 0$ provide a good starting point. We have carried out extensive experiments under varying initial conditions and different measurement sets and we never experienced any difficulty in convergence. We also experimented on simulated data sets when the observation noise is quite high and the amount of defocus blur is large. Under such taxing circumstances we found the estimate of the blur parameter σ to be a bit underestimated. Barring the above case, the convergence of the blind restoration method has been found to be very satisfactory.

5.4 Demonstrations

5.4.1 Experiments with Known Blur

We now demonstrate the efficacy of the regularization based approach to shape recovery from blurred observations. First we show the experiments on real images for image restoration, depth recovery and albedo estimation when the blur is known. The parameter for the gradient descent algorithm *i.e.*, the step size is chosen as 0.01 for the estimation of all the three fields namely, $\mathbf{p}$, $\mathbf{q}$ and the $\boldsymbol{\rho}$. The same value is used in all experiments in this section. The camera blur were simulated by using a uniform circular blur mask, *i.e.*, the captured images with different light source positions were convolved with the uniform blur mask which are then used as blurred observations. This blur approximates an out-of-focus blur as a pillbox function, and is used in many research simulations [165]. Here the blur is parameterized in terms of the window size and is modeled as a uniform intensity distribution within a circular disc of radius b,

$$h(x,y) = \begin{cases} \frac{1}{\pi b^2}, & \text{if } \sqrt{x^2+y^2} \leq b \\ 0, & \text{otherwise.} \end{cases}$$

First we consider an object where the imaged scene gives a smooth intensity variation, but has arbitrary depth variations. We continue with the images of Jodu shown in chapter 4. Figures 5.4(a) and (b) show two of the eight focused images (before being blurred) with source positions $(-0.8389, -0.7193, 1)$ and $(-0.3639, -0.5865, 1)$, respectively. The blurred images using a mask of size 5×5 *i.e.*, $b = 2$ for the same source positions are shown in Figures 5.5(a) and (b), respectively. The corresponding restored Jodu images using the suggested approach are shown in Figures 5.6(a) and (b), respectively. It can be observed that because of the blurring the edge details are lost (see Figure 5.5). We note that these details are very well recovered using the proposed approach. Observe the shadow and the bisecting line on the tongue in the restored image as depicted in Figures 5.6(a) and (b). For the sake of comparison, we show the restored images through direct image deconvolution (using the MATLAB function "*deconvlucy*" that follows the Lucy-Richardson algorithm) in Figures 5.6(c, d). As expected the result obtained under direct deconvolution is poor when compared to the proposed approach (see Figures 5.6(a, b)). See the nose, tongue and the hind leg regions of the doll and compare the performances.

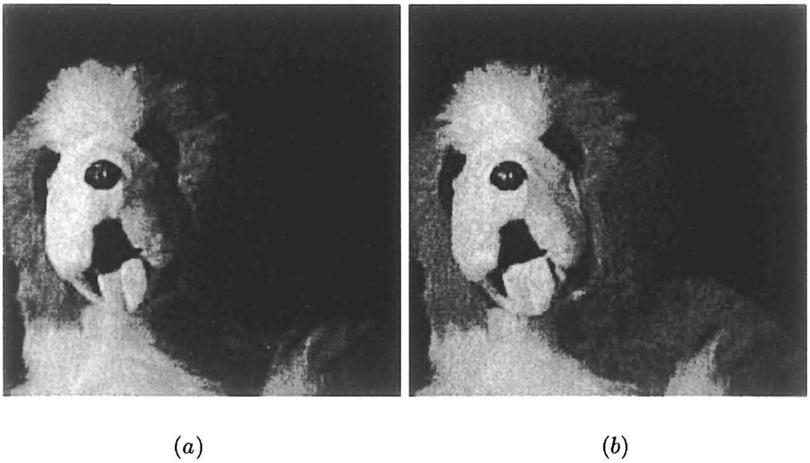


Fig. 5.4. Focused images of Jodu captured with two different light source positions.

Now we consider how well the depth map can be recovered from the blurred observations. Figures 5.7(a) and (c) show the depth map as obtained from the PS using the focused images (not suffering from blur) and the estimated one using the proposed method from blurred observations, respectively, displayed as an intensity variation. The depth values were calculated with the estimated values of **p** and **q** by using the Eq. (4.9) and then scaled between 0 and 255. The brighter it is, nearer it is to the camera. The depth map using the surface gradients obtained from a standard photometric stereo (PS) applied directly on the blurred images is shown in Figure 5.7(b). This result does not account for the presence of blur in the observations. It can be clearly seen that the estimated depth shown in Figure 5.7(c) using the proposed technique is very much similar to the depth map due to focused images shown in Figure 5.7(a). The distortion introduced in the depth map shown in Figure 5.7(b) is very much removed in Figure 5.7(c). The

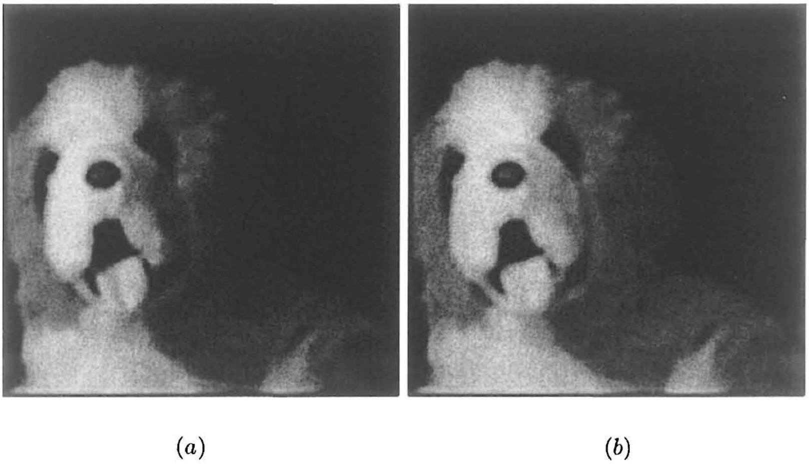


Fig. 5.5. Simulated blurred observations of Jodu using a mask size of 5 × 5 for the images for two different source directions shown in Figure 5.4.

depth distortion near the chest and the mouth region is clearly visible. We observed an improvement in terms of the MSE (mean squared error) as well. The MSE calculated between the true depth map (Figure 5.7(a) that does not have blurring) and the depth map due to blurred Jodu observations (Figure 5.7(b)) was 0.0784, while it reduced to just 0.0005 for the depth map obtained using the proposed approach. Thus, there is a substantial improvement in the recovered depth map.

Next we consider the goodness of the recovered albedo. Figure 5.8(a) shows the recovered albedo using nonblurred, focused observations and Figure 5.8(c) corresponds to the recovered one using the proposed technique. As seen from the figures, the shadows do not affect the computation of the albedo. Figure 5.8(a) represents the true albedo subject to the object surface satisfying the assumption of a Lambertian surface. The result of albedo recovery using the standard PS method applied on the blurred observations without any rectification for the blurring effect is shown in Figure 5.8(b). Comparison of Figures 5.8(b) and (c) clearly indicate that due to the blurring process, the recovered albedo does not give the true reflecting property of the surface when the blurring effect is not compensated. The resulting albedo is very smooth.

In order to test the performance this algorithm for a higher amount of blur, we now consider a blurring mask size of 9 × 9 with $b = 4$. We observed that the algorithm works well even for a higher amount of

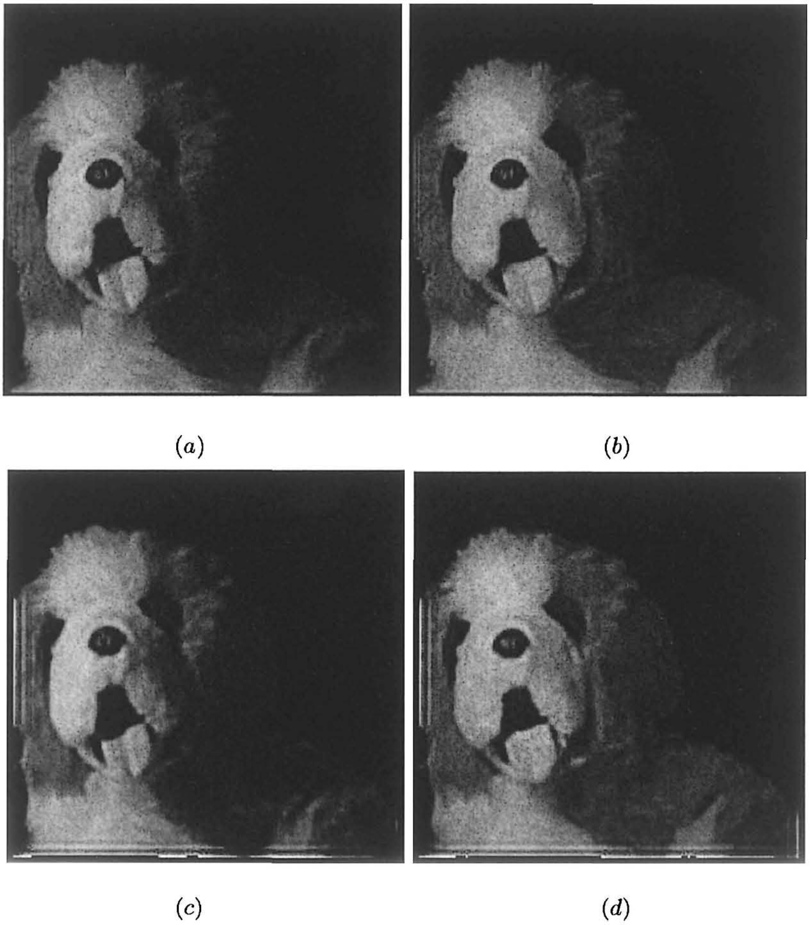
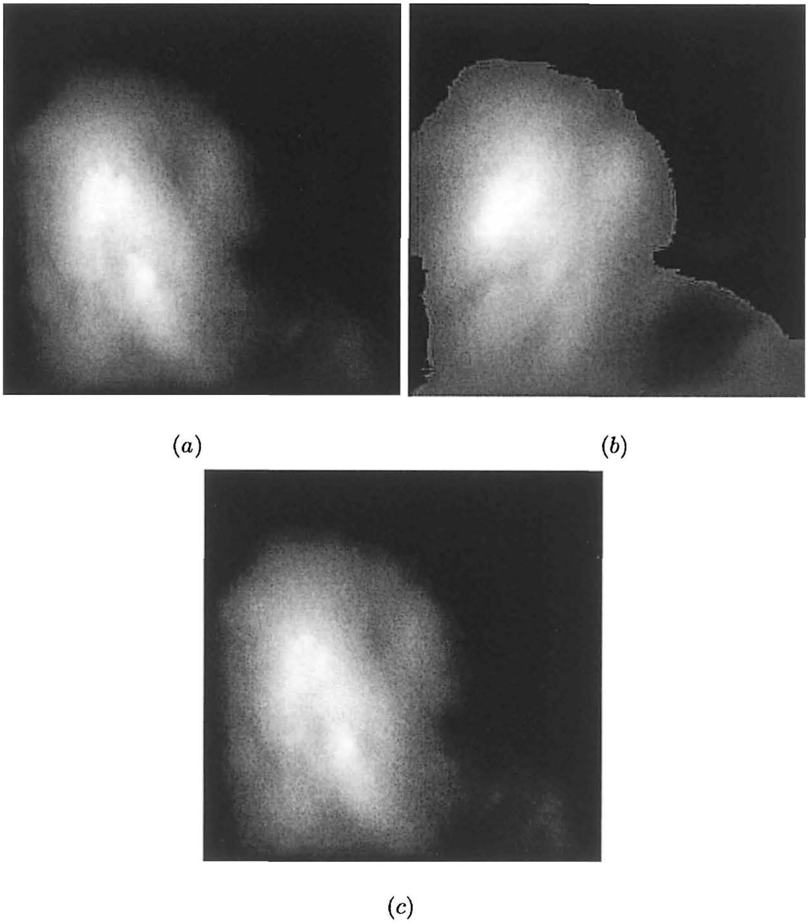


Fig. 5.6. (a, b) Restored Jodu images for the observations given in Figure 5.5 using the proposed approach assuming the blur to be known. (c, d) Restored Jodu images using the image domain deconvolution operation. Compare Figure (a) with Figure (c) and Figure (b) with Figure (d).

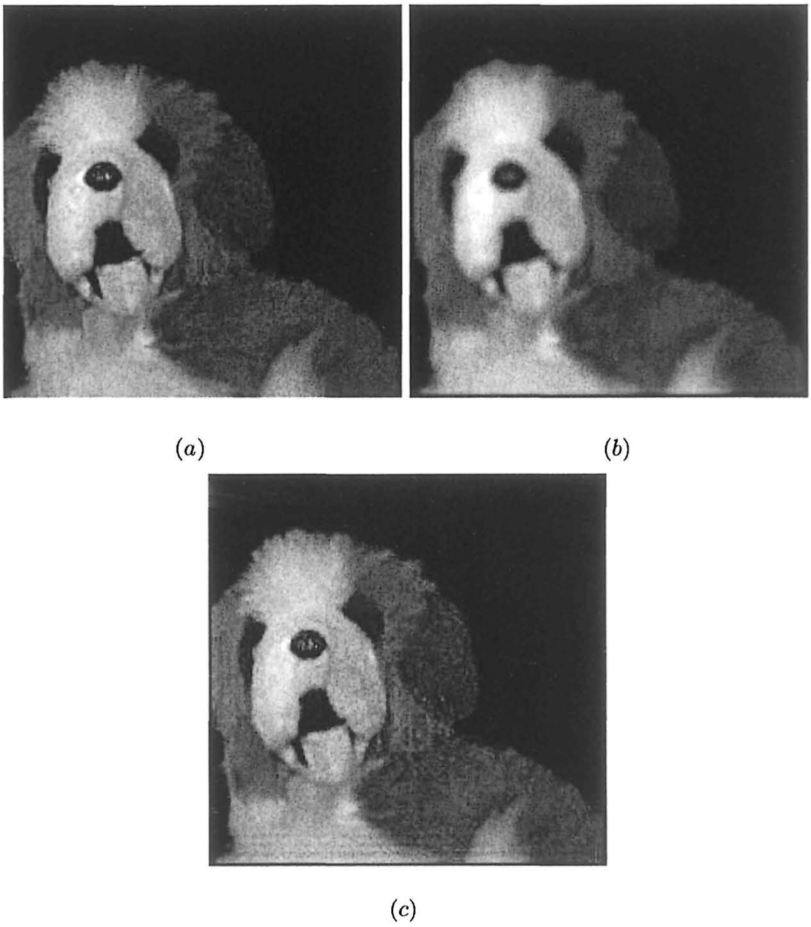
blur, as is evident from the results. For this experiment we considered a region that shows only the face of the doll, and the number of images were kept as eight, same as used in the previous experiment. Figures 5.10(a, b) show the simulated blurred observations corresponding to the focused (pin-hole approximation) images depicted in Figures 5.9(a, b). The restored images for the same using the proposed approach are shown in Figures 5.11(a, b). We notice again that there is a consid-



(a) (b) (c)

Fig. 5.7. (a) True depth map as obtained from PS using the observations not suffering from blurring. (b) Depth map obtained using the standard PS method on blurred observations. (c) Recovered depth map using the proposed technique utilizing the knowledge of PSF.

erable improvement in the reconstructed images. The high frequency details are clearly restored back as is evident from the protruding nose and eye boundaries of the dog. We show the result of image restoration using direct deconvolution in Figure 5.12. Comparing these results with those obtained using the proposed approach in Figures 5.11(a, b), once again we observe that the reconstructions using the proposed approach are much better. When the blur is very severe, the direct image decon-



(*a*) (*b*)

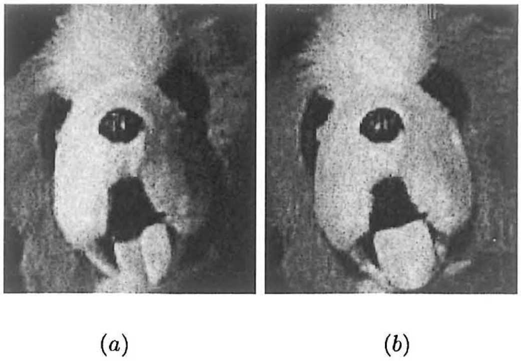
(*c*)

Fig. 5.8. Recovered albedo (a) from the nonblurred or focused images, (b) using the standard PS on blurred observations, and (c) using the proposed technique.

volution techniques offer poor results. However, the indirect method of image restoration yields a much better results.

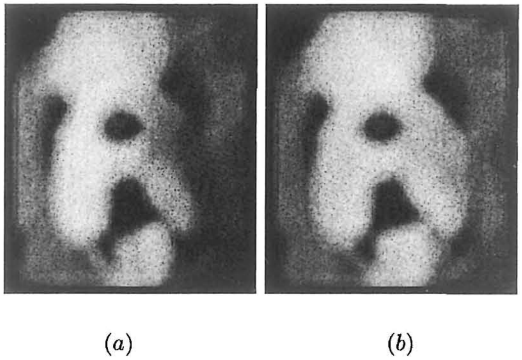
The estimated depth map (Figure 5.13(c)) is quite comparable to the true depth map shown in Figure 5.13(a). The depth map calculated using the standard PS method applied on the blurred observations using the surface gradients recovered from the blurred images looks smooth, lacking the depth variation (see Figure 5.13(b)).

Once again we investigate how well the albedo and the structure are recovered for severely blurred observations. The recovered albedo



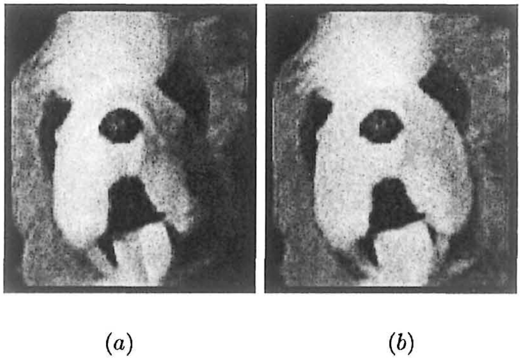
(a) (b)

Fig. 5.9. Focused images of Jodu captured with light source positions (a) (−0.8389, −0.7193, 1), and (b) (−0.1763, −0.5596, 1), respectively.



(a) (b)

Fig. 5.10. Simulated severely blurred observations of Jodu using a mask size of 9 × 9 for the images shown in Figure 5.9.



(a) (b)

Fig. 5.11. Restored Jodu images using the proposed method.

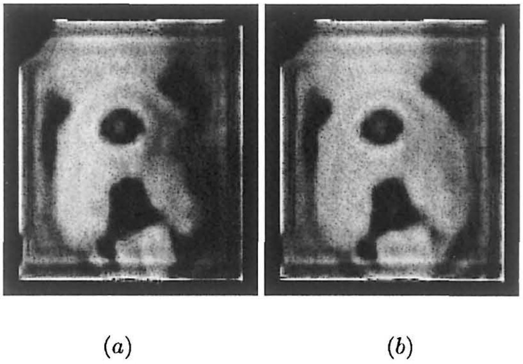


Fig. 5.12. Restored Jodu images in Figure 5.10 using direct image deconvolution.

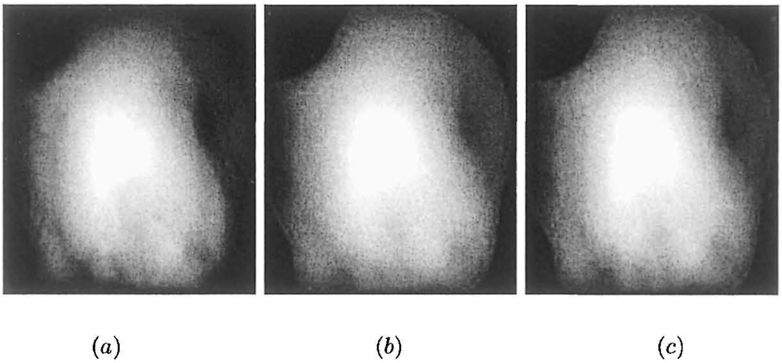


Fig. 5.13. (a) True depth map without considering the blur. (b) Depth map obtained from direct application of PS on blurred observations. (c) Reconstructed depth map using the proposed technique.

map using the proposed technique and the albedo map estimated using the PS on the blurred observations are shown in Figures 5.14(b, a). As can be seen from the figures, the recovered albedo estimated from the blurred observations appears too smooth indicating that the albedo is very poorly estimated when the observations are blurred and are not rectified.

5.4.2 Experiments with Unknown Blur

We now present the results for the more general case where the blur is unknown and need to be estimated along with the estimation of scene structure and the image restoration. Here we consider experiments us-

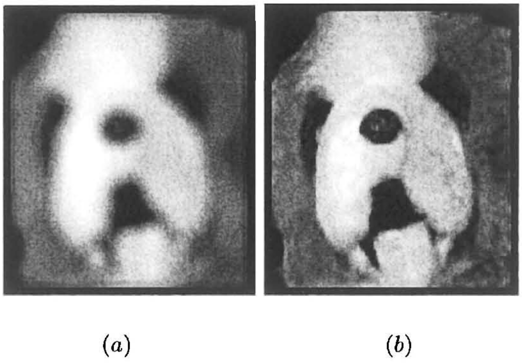
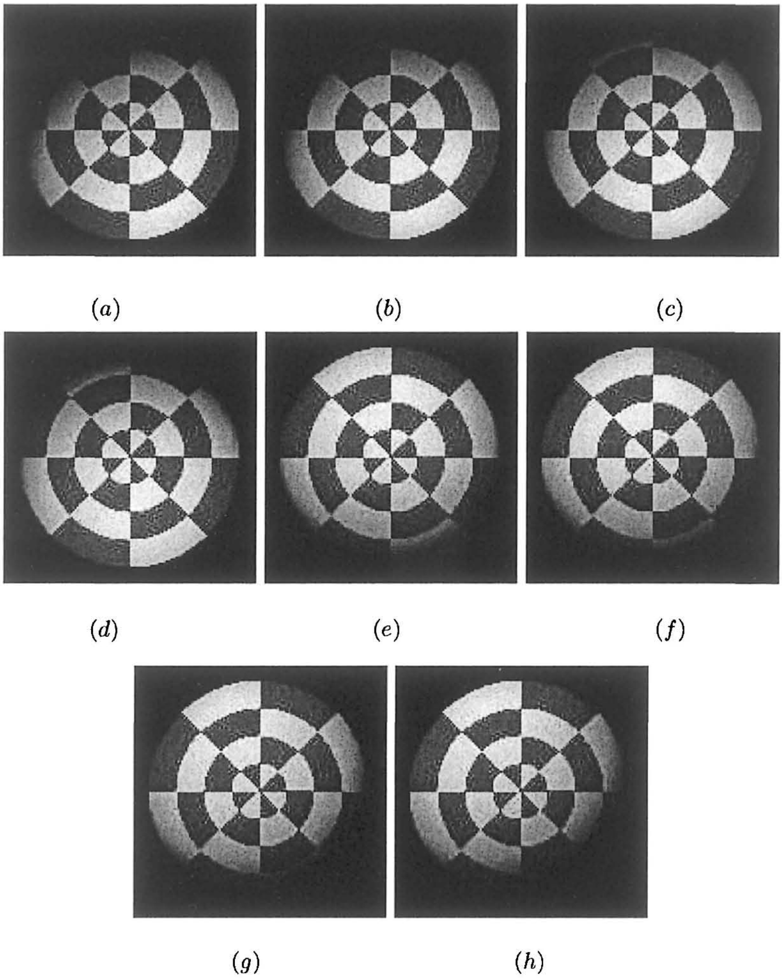


Fig. 5.14. (a) Surface albedo for Jodu computed from blurred data. (b) Estimated albedo using the proposed technique.

ing the real images as well as the synthesized ones. First, we consider an experiment using synthetic images for validation purposes. For this experiment, we generated a set of eight images of a spherical surface for different source positions. An arbitrary texture (albedo) is mapped onto this surface. The corresponding images are shown in Figures 5.15.

The sphere had a checker-board patterned albedo. The obtained images are then blurred by using a Gaussian blur mask of size 7 × 7 with a standard deviation of $\sigma = 1$ and are corrupted by adding a Gaussian noise of zero mean and a standard deviation of 0.01 for a normalized gray value of pixel in the range [0, 1]. The blur is not assumed to be known during the restoration process. We used an initial value of 0.6 for σ for this experiment and a mask size of 13 × 13. The final estimated σ for this experiment is 1.0352 which is very close to the actual defocus blur. Of the eight images generated with different light source positions, we show experimental results on two images with source positions (0.45, 0.80, 1) and (−0.20, −0.60, 1) that correspond to figures (a) and (f) in Figure 5.15. The corresponding simulated blurred and noisy observations are shown in Figure 5.16. Figures 5.17(a, b) show the efficacy of our algorithm for the estimation of true texture from their blurred observations. Compare these images to the original images in Figures 5.15((a) and (f)) and observe that the blind restoration is very good. We see that the boundary curves on each segment in the restored checker-board images are sharper when compared to the blurred observations indicating the restoration of high frequency de-



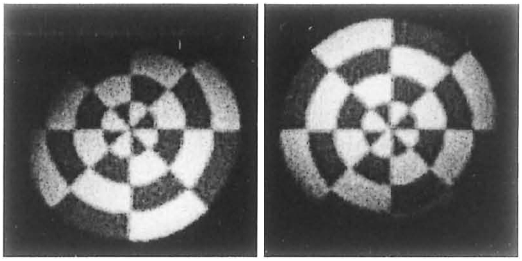
(*a*) (*b*) (*c*)

(*d*) (*e*) (*f*)

(*g*) (*h*)

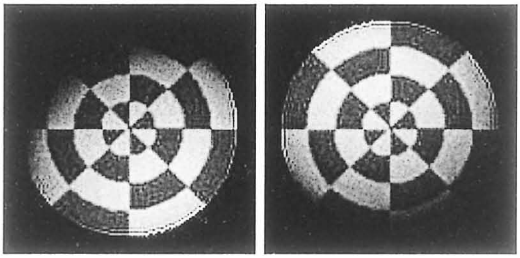
Fig. 5.15. Synthesized images of a sphere for eight different light source positions.

tails. The images restored using the standard image domain blind deconvolution are shown in Figures 5.18(a, b). The estimated depth map (Figure 5.19(b)) is also quite correct as the intensity is highest at the center and decreases as we move away from it which definitely reflects the shape of a hemisphere. The shape distortion seen in the depth map of Figure 5.19(a) recovered from the blurred observations using the standard PS method clearly indicates the loss of depth details.



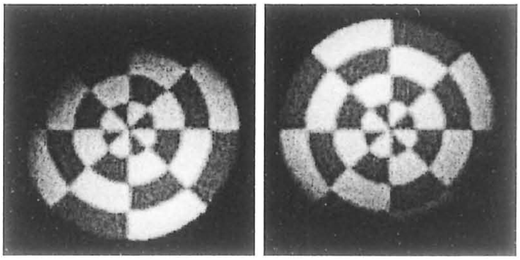
(*a*) (*b*)

Fig. 5.16. Simulated observations using a Gaussian blur $\sigma = 1$ and additive noise corresponding to figures in 5.15(a) and (f).



(*a*) (*b*)

Fig. 5.17. Restored checker-board images using the proposed technique.



(*a*) (*b*)

Fig. 5.18. Restored checker-board images using a standard image domain blind deconvolution.

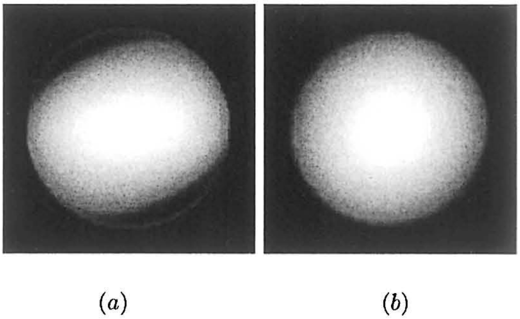
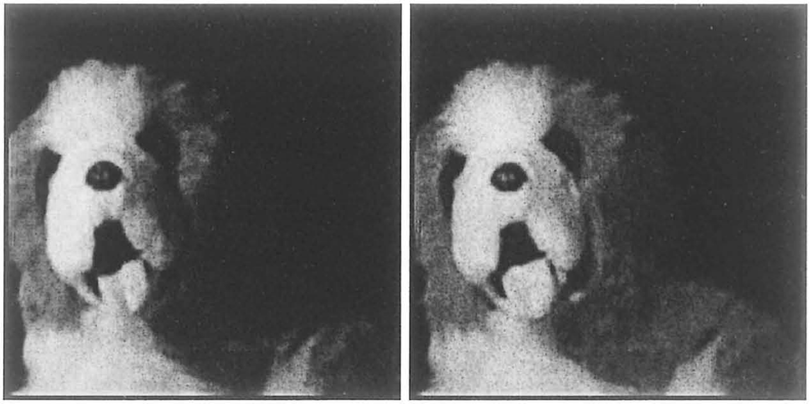


Fig. 5.19. Recovered depth map using (a) standard PS method, and (b) proposed method.

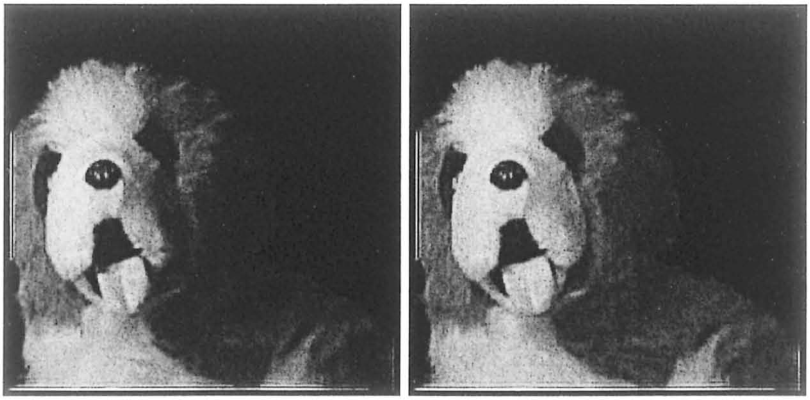
Finally we consider the experiments with a real data set for blind restoration and structure estimation, for which we use the same object Jodu captured with eight different light source positions. However, no attempt was made to bring the object Jodu in focus and hence the observations are slightly blurred. Common in real aperture imaging, the blur due to defocus is modeled by a Gaussian shaped PSF parameterized by the variable σ. The blurred observations are then used to derive the fields $\mathbf{p}^{(0)}$, $\mathbf{q}^{(0)}$ and the $\boldsymbol{\rho}^{(0)}$ which are used as the initial estimates for our algorithm as discussed in section 5.3.3. Again an initial value of $\sigma = 0.6$ and a mask size 13×13 (which is the same as used in the previous experiment) were used in order to estimate the different fields and to restore the images iteratively. After every 100 iterations in gradient descent operation to estimate the surface normals, the new value of σ is calculated and is used again to refine the fields related to structure and albedo, and the images. The algorithm is terminated when no further improvement in the estimate of σ is obtained. The final estimated σ for the given example was found to be 1.0577. The results of the experiment are illustrated with the following figures.

Two of the blurred observations are shown in Figures 5.20(a, b). The restored Jodu images for the same are displayed in Figures 5.21(a, b). As can be seen, the restored images have sharper details. Observe the nose, mouth and tongue regions. The blur which is clearly visible in Figures 5.20(a, b) is well removed in Figures 5.21(a, b). The restored images through the blind image deconvolution (using the MATLAB function "*deconvblind*") are shown in Figures 5.22(a, b). For the MATLAB-program, the PSF mask size and the initial value of σ were kept the same as used in the proposed approach and the number of iterations were again chosen as 100. As we can see from the figures, the restoration is quite poor. The images shown in Figures 5.21 are definitely sharper than those in Figure 5.22.



(*a*) (*b*)

Fig. 5.20. Observed images of Jodu with an arbitrary camera defocus for two different light source positions.



(*a*) (*b*)

Fig. 5.21. Restored Jodu images using the proposed method.

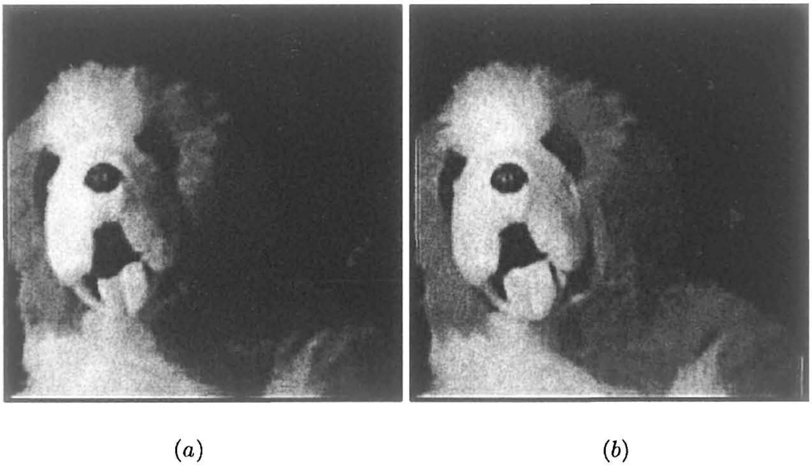


Fig. 5.22. Restored Jodu images using a standard blind deconvolution tool.

We also look at the quality of depth and albedo recovery for these observations. Although it is difficult to visualize much perceptual improvement in the estimated depth map using the proposed technique from that obtained using a standard PS method on blurred observations due to the use of gray levels to encode the depth map (see Figures 5.23(b, a)), there is a definite perceptual difference when the depth is viewed as meshplots. The corresponding plots are displayed in Figure 5.24. As can be seen from the meshplot, the recovered depth map corresponding to the standard PS method is smoother when compared to the plot for the proposed approach. There was also a noticeable change in the estimated **p** and **q** fields using these two methods. Since it is difficult to visualize the **p** and **q** fields for an arbitrary shaped object, we display only the recovered depth maps. The process of smoothing the gradient fields using the expression

$$\nabla^2 d(x,y) = p_x(x,y) + q_y(x,y)$$

to obtain the depth map somehow blurs the difference between the two results. However, we did observe a gain in terms of accuracy in the estimated depth map measured in terms of the MSE (mean squared error). The MSE between the depth map due to focused observations shown in Figure 5.7(a) and the depth map shown in Figure 5.23(a) was found to be 0.0100, whereas it was only 0.0045 considering the depth map obtained using the proposed method. This clearly indicates an improvement in the depth map estimation using the proposed algorithm. Similar conclusions about the efficacy of the proposed scheme can also be drawn from the Figures 5.25(a, b) for the albedo recovery. The recovered albedo is sharper when the proposed method is used.

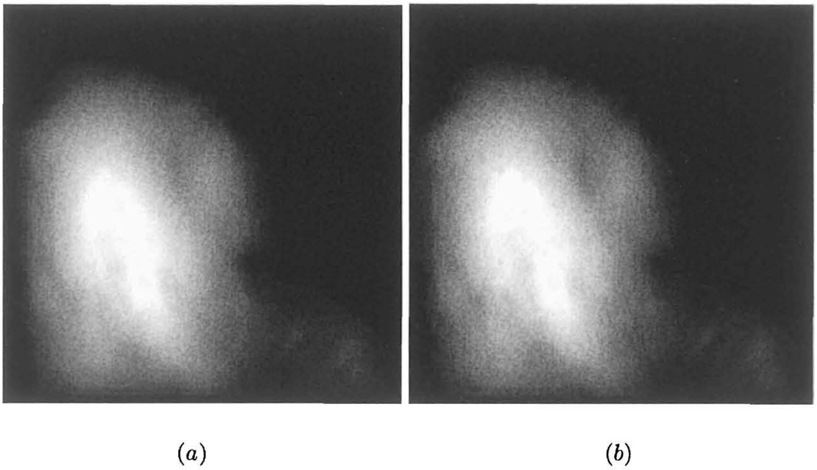


Fig. 5.23. Recovered depth map using (a) standard PS method, and (b) proposed method.

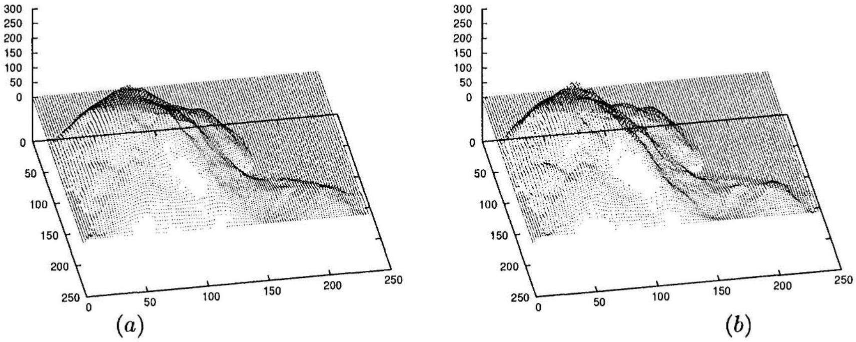


Fig. 5.24. Recovered depth map shown as a mesh plot using (a) standard PS method, and (b) proposed method.

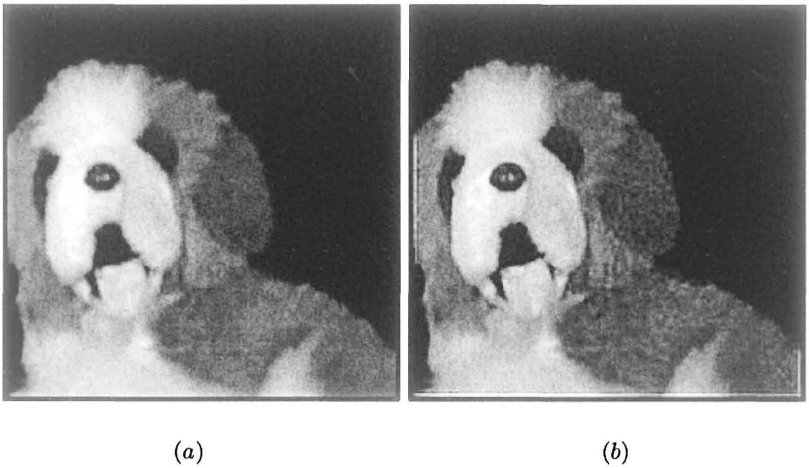


Fig. 5.25. Recovered albedo using (a) standard PS method, and (b) proposed method.

5.5 Conclusions

We have described a method for simultaneous estimation of scene structure and blind image restoration from blurred photometric observations. The structural information is embedded within the observations and, through the unified framework we have described, we were able to recover the restored images as well as the structure. We model the surface gradients and the albedo as separate MRF fields and use a suitable regularization scheme to estimate the different fields and the blur parameter alternately. No problem was faced regarding the convergence of the proposed method during the experimentations. Experiments were carried out on the synthetic as well as the real images to show the effectiveness of our approach.

Although we do not discuss the super-resolution aspect in this chapter one can now easily extend it to the super-resolution problem discussed in chapter 4 where we neglected the presence of possible blur in the observations. The contents of these two chapters can be combined to handle super-resolution reconstruction from photometric observations even in the presence of unknown blur.

6

Use of Learnt Wavelet Coefficients

Until chapter 5 we assumed that a number of observations of the same scene under varied camera or lighting conditions are available for image super-resolution. In particular we investigated the usefulness of defocus and photometric cues for high resolution reconstruction purposes. Now we relax the above requirement of having to have multiple observations. We show that instead of multiple observations of the same scene, if we have a set of high resolution observations of an arbitrary set of objects as exemplars it may suffice as we may still be able to improve the resolution of the given image.

In this chapter we investigate a learning based super-resolution restoration technique by using the wavelet coefficients to define a constraint on the solution. Only a single image is used for super-resolution. An arbitrary set of high resolution images are taken as training images. Wavelet coefficients at finer scales of the unknown high resolution image are learnt from a set of high resolution training images and the learnt image in the wavelet domain is used for further regularization while super-resolving the picture. We use an appropriate smoothness prior with discontinuity preservation in conjunction with the learnt wavelet based prior to estimate the super-resolved image. The smoothness term ensures the spatial correlation among the pixels whereas the learnt wavelet term chooses the best high resolution edges from the training set. Since this amounts to extrapolating the high frequency components, this method does not suffer from oversmoothing effects. The results demonstrate the effectiveness of this approach. The advantages of this method would include avoidance of point correspondences, no need to estimate the blur and there is no need to model the reflectance property of the surface.

6.1 Introduction

In many applications more than one low resolution observations may not be available, but we may have a database of a number of similar or arbitrary images at a higher spatial resolution. For example we may have a snapshot of an object or a person alone. In order to have a better view, we may interpolate the image to double its size. At places where there are not much of variations in gray levels or no edges, any interpolation technique would yield acceptable results. However, any edge in the low resolution image would get blurred when interpolated. Noting the fact that sharpness of edges are very crucial for image representation and perception, we ask the question if the high resolution edges can be learnt from the high resolution training data to replace the corresponding low resolution edges while upsampling, thus avoiding edge smearing. This is the primary motivation for developing the contents of this chapter.

In this chapter we consider having an access to a set of high resolution training images to learn the edge prior. The basic problem we solve in this chapter is as follows. One captures an image using a low resolution camera. We are interested in generating the super-resolved image for the same using a set of available high resolution images of different objects. It is assumed that the high frequency contents to be extrapolated are locally present in the training set. We use a wavelet-based multi-resolution analysis to learn the wavelet coefficients at a given location at the finer scales for the image to be super-resolved. The learnt coefficients are then used as a prior that enforces the condition that the wavelet coefficients at the finer scales of the super-resolved image should be locally close to the best matching coefficients learnt from the training set. In order to preserve the spatial continuity of the restored image, we use a smoothness constraint in conjunction with the learnt prior to obtain the super-resolved image.

Since edges in the image are places where one requires a better clarity, there have also been some efforts in the literature on preserving the edges while interpolating an image. Chiang and Boult [62] use edge models and a local blur estimate to develop an edge-based super-resolution algorithm. In [180] authors propose an image interpolation technique using a wavelet domain approach. They assume that the wavelet coefficients scale up proportionately across the resolution pyramid and use this property to go down the pyramid. Thurnhofer and Mitra [181] have proposed a non-linear interpolation scheme based on a polynomial operator wherein perceptually relevant features (say, edges) are extracted and zoomed separately. Different reconstruction methods to improve the resolution of digital images while zooming have been discussed in [182]. The authors here focus on both the linear and the non-linear methods based on total variation to study the ability of these methods to preserve the directionality of the edges while zooming.

It was mentioned earlier in chapter 2 that researchers have also attempted to solve the problem by using learning based techniques [97, 98, 100, 101]. Here the new information required for predicting the high resolution image is obtained from a set of training images rather than from subpixel shifts among low resolution observations. The method investigated in this chapter can also be classified under the learning based super-resolution schemes. However, here we use a different type of learning where we use a prior term that enforces the condition that the wavelet coefficients of the super-resolved image at the finest scale should be locally close to the best matching wavelets learnt from the high resolution training set. A smoothness constraint is imposed on the restored image to obtain a regularized solution.

6.2 Wavelet Decomposition of an Image

Wavelets are mathematical functions that split up data into different frequency components locally, and then study each component with a resolution matched to its scale. They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities or a local analysis is required. The discrete wavelet transform (DWT) provides us with a sufficient information for analysis and synthesis of a time series data or an image and is easier to implement. The idea here is similar to the continuous wavelet transform (CWT) which is computed by changing the scale of the analysis, shifting the window in time, multiplying it by the data and integrating over time. In the case of DWT, filters of different cut-off frequencies are employed to analyze the sequence at different scales. The input sequence is passed through a series of high pass and low pass filters to analyze the high and low frequency components, respectively. The procedure starts with passing the sequence through a half band ($0 - \pi/2$ radians) digital low pass filter with impulse response $h(n)$, thus removing all the frequencies that are above half of the highest frequency in the sequence. The filtered output is then subsampled by a factor of 2, simply by discarding every other sample since the sequence now has a highest frequency of $\pi/2$ radians instead of π. The low pass filter thus halves the resolution, but leaves the scale unchanged. The subsequent subsampling by a factor of 2, however, changes the scale. Subsequently the low pass signal is passed through another low pass filter whose passband is just half of the previous filter bandwidth. The process is continued several times until a coarse description of the signal is achieved at a desired level. This is illustrated in Figure 6.1

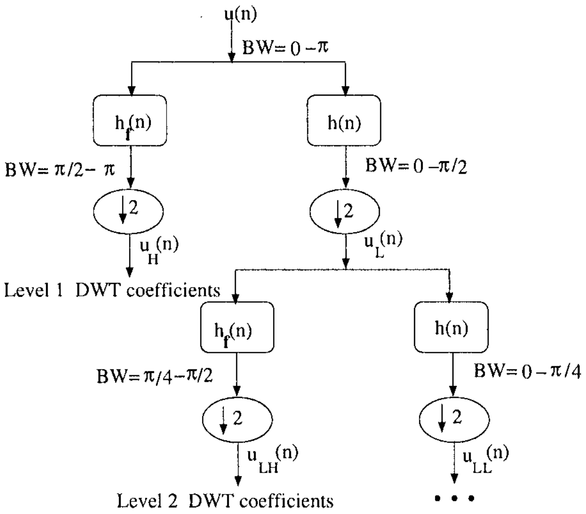


Fig. 6.1. Illustration of subband wavelet decomposition. Here $u(n)$ is the original sequence to be decomposed and $h(n)$ and $h_f(n)$ are low pass and high pass filters, respectively. The bandwidth of the resulting signal is marked as "BW".

The wavelet transform for a 2D sequence is similar to that of 1D decomposition. A 2D wavelet decomposition is first performed (horizontally) on the rows by applying low pass and high pass filters. Then we perform the same operations vertically (on the columns) resulting in four subbands LL, LH, HL, HH. Here L stands for the lower band signal and H stands for the higher band. Needless to say, we assume that the filter kernel is separable so that the wavelet decomposition can be carried out along the rows and columns separately. We repeat the operation with 'LL' as the input image for further decomposition. We illustrate wavelet decomposition of an image in Figure 6.2. An input image is shown in Figure 6.2(a). The corresponding 3-level wavelet decomposition is shown in Figure 6.2(b). Note that in the top left we have the down sampled dc component.

Wavelet analysis of a signal, by itself, is an actively pursued area of research. There are many text books available for a comprehensive discussion on this topic. We refrain from introducing this topic here. We do assume a certain familiarity with this topic by a reader in this chapter. The readers are referred to [183, 184, 185] for further discussion on wavelet decomposition.

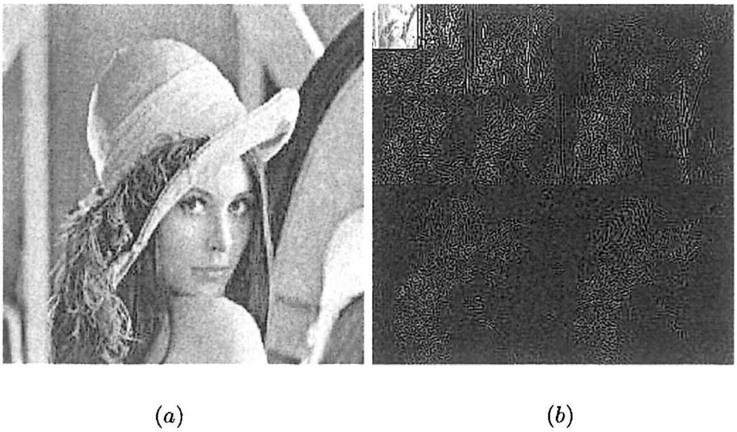


Fig. 6.2. Illustration of wavelet decomposition of an image. (a) Input Lena image, and (b) the corresponding 3-level wavelet decomposition.

6.3 Learning the Wavelet Coefficients

As discussed in the previous section the wavelet decomposition splits the data into high and low frequency components. As seen from Figure 6.1, given a high resolution sequence $u(n)$ having a bandwidth support of $[0 - \pi]$, it can be decomposed into u_L and u_H sequences constituting the low frequency and the high frequency components in the sequence, respectively. Let us consider that u_L (the low resolution sequence) is given and we need to generate the high resolution sequence $u(n)$. In order to do that we need to know the u_H so that when we take the inverse discrete wavelet transform (IDWT) we get back the original sequence $u(n)$. However, for the current problem on super-resolution, we do not have the high frequency components u_H to obtain the high resolution sequence $u(n)$. In the absence of any information on u_H, we plan to estimate the coefficients u_H by learning them from a set of high resolution sequences.

Similarly, when a low resolution image or a 2D signal is considered we need to learn the corresponding unknown high frequency components u_{LH}, u_{HL} and u_{HH}. Since the problem of super-resolution involves handling data at multiple resolutions, and since the wavelets are best suited for a multi-resolution analysis, it motivates us to use a wavelet-based approach for learning the wavelet coefficients at the finer resolution. These wavelet coefficients indicate the high frequency details in an image. For example, in the illustrative Figure 6.2(b) all three quadrants other than the top-left one display the high resolution components or the edges in the image. The top-right quadrant shows the vertical edges, the bottom-left quadrant shows the horizontal edges and the bottom-right the diagonal edges. These edge information are needed to reconstruct the high resolution image given in 6.2(a). If these quadrants are not available, as is the case in low resolution observations, can they be learnt?

The learning of higher band wavelet coefficients is done from a set of high resolution training images. If the high resolution data in a region does not have much high frequency components, the region can easily be obtained from its low resolution observation through a suitable interpolation. However, if a region has edges, the corresponding wavelet coefficients (u_H in Figure 6.1) are quite significant and they cannot be neglected while obtaining the high resolution image. These coefficients must be learnt from a database of training images. We assume that a primitive edge element in the high resolution image is localized to an 8×8 pixel area, and we observe the corresponding edge elements over a 4×4 pixel area in the low resolution image. From the high resolution data base, can we obtain the best 8×8 region by matching it in the wavelet domain with the given 4×4 pixel observation? Note that such a matching should be brightness (dc-shift) independent.

We make use of a two level wavelet decomposition of the given low resolution observation while learning the wavelet coefficients at the finer scale. Figure 6.3 illustrates the block schematic of how the wavelet coefficients at finer scales are learnt from a set of N training images using a two level wavelet decomposition of the low resolution test image. The high resolution training images are decomposed into three levels and

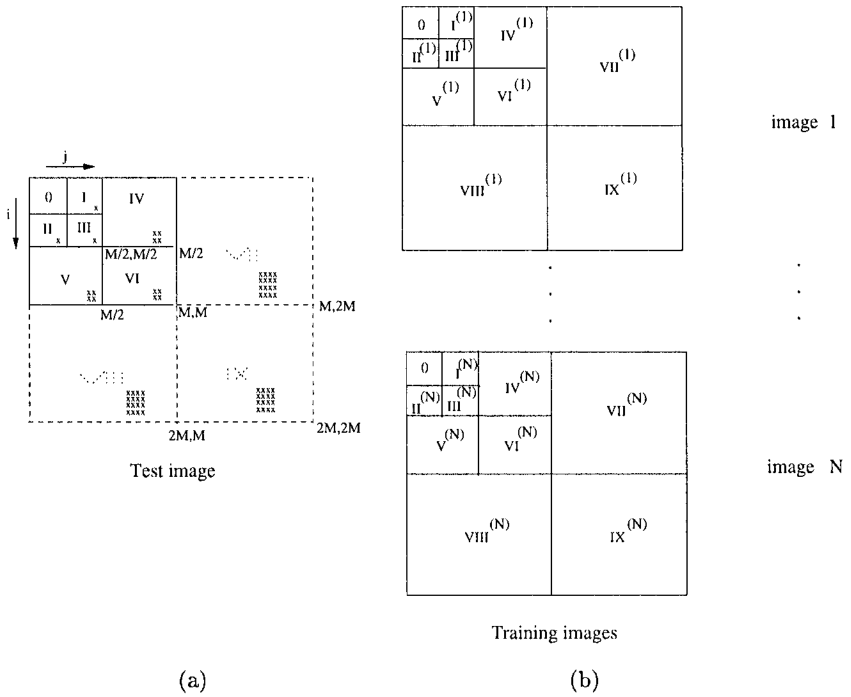


Fig. 6.3. Illustration of learning of wavelet coefficients at a finer scale. (a) A Low resolution image with a two level wavelet decomposition. Wavelet coefficients (marked as x) in subbands shown with the dotted lines are to be estimated for subbands $VII - IX$. (b) High resolution training set in wavelet domain with three level decompositions.

the test image is compared to the training images in the wavelet domain at the coarser two scales. This decomposition is used to extrapolate the missing wavelet coefficients in subbands $VII - IX$ (shown as dotted in Figure 6.3(a)) for the test image. They correspond to the estimated high pass wavelet coefficients at the finest level decomposition of the unknown high resolution image. Here the low resolution image is of size $M \times M$ pixels. Considering an upsampling factor of 2, the high resolution image, now has a size of $2M \times 2M$ pixels. For each coefficient in the subbands $I - III$ and the corresponding 2×2 blocks in the subbands $IV - VI$, we need to extrapolate a block of 4×4 wavelet coefficients in each of the subbands VII, $VIII$ and IX.

In order to learn the wavelet coefficients we exploit the idea from zero tree concept, *i.e.*, in a multi-resolution system, every coefficient at a given scale can be related to a set of coefficients at the next coarser scale of similar orientation [186]. Using this idea we follow the minimum absolute difference (MAD) criterion to estimate the wavelet coefficients. We take the absolute difference locally between the wavelet coefficients in the low resolution image and the corresponding coefficients in each of the high resolution training images. The learning process is as follows. Consider the subbands $0 - VI$ of the low resolution image. Denote the wavelet coefficient at a location (i,j) as $\psi(i,j)$. Consider the range $0 \leq i,j \leq M/4$. The wavelet coefficients $\psi_I(i, j + M/4)$, $\psi_{II}(i + M/4, j)$, $\psi_{III}(i + M/4, j + M/4)$ corresponding to subbands $I - III$ and a 2×2 block consisting of $\{\psi_{IV}(k, l + M/2)\}_{k=i,l=j}^{k=i+1,l=j+1}$, $\{\psi_V(k + M/2, l)\}_{k=i,l=j}^{k=i+1,l=j+1}$, $\{\psi_{VI}(k + M/2, l + M/2)\}_{k=i,l=j}^{k=i+1,l=j+1}$, in each of the subbands $IV - VI$ are then considered to learn a 4×4 wavelet block in each of the subbands $VII - IX$ consisting of unknown coefficients $\{\psi_{VII}(k, l + M)\}_{k=i,l=j}^{k=i+3,l=j+3}$, $\{\psi_{VIII}(k + M, l)\}_{k=i,l=j}^{k=i+3,l=j+3}$, and $\{\psi_{IX}(k + M, l + M)\}_{k=i,l=j}^{k=i+3,l=j+3}$, *i.e.*, we need to estimate or learn these 48 coefficients for every 4×4 region in the low resolution image. In order to illustrate which set of wavelet coefficients we select for learning purposes, we denote them with 'x' marks in Figure 6.3(a). To obtain the wavelet coefficients for the test image at a finer resolution, we consider the wavelet coefficients in subbands $I - VI$ in each of the high resolution training images (see Figure 6.3(b)). We search for the best matching training image at a given location (i,j) that matches to the wavelet coefficients for the test image in the subbands $I - VI$ in the MAD sense and copy the corresponding high resolution wavelet coefficients in subbands $VII - IX$ to those subbands for the test image. In effect, we use the following equation to find the minimum.

$$
\begin{aligned}
\hat{m}(i,j) = \arg\min_m \{ & |\psi_I(i, j + M/4) - \psi_I^{(m)}(i, j + M/4)| \\
& + |\psi_{II}(i + M/4, j) - \psi_{II}^{(m)}(i + M/4, j)| \\
& + |\psi_{III}(i + M/4, j + M/4) - \psi_{III}^{(m)}(i + M/4, j + M/4)| \\
& + \sum_{k=i}^{k=i+3} \sum_{l=j}^{l=j+3} |\psi_{IV}(k, l + M/2) - \psi_{IV}^{(m)}(k, l + M/2)| \\
& + \sum_{k=i}^{k=i+3} \sum_{l=j}^{l=j+3} |\psi_V(k + M/2, l) - \psi_V^{(m)}(k + M/2, l)|
\end{aligned}
$$

$$+\sum_{k=i}^{k=i+3}\sum_{l=j}^{l=j+3}|\psi_{VI}(k+M/2,l+M/2)-\psi_{VI}^{(m)}(k+M/2,l+M/2)|\},\tag{6.1}$$

where $m = 1, 2, \cdots, N$. Here $\psi_J^{(m)}$ denotes the wavelet coefficients for the m^{th} training image at the J^{th} subband. For each (i, j) in subbands $I - III$ of low resolution observation, a best fit 4×4 block of wavelet coefficients in subbands $VII - IX$ from that training image given by $\hat{m}(i, j)$ which gives the minimum are then copied into subbands VII, $VIII$, IX of the observed image. In effect, Eq. (6.1) helps in matching edge primitives at low resolutions. Thus we have,

$$\begin{aligned}\psi_{VII}(i,j) &:= \psi_{VII}^{(\hat{m})}(i,j),\\ \psi_{VIII}(i,j) &:= \psi_{VIII}^{(\hat{m})}(i,j),\\ \psi_{IX}(i,j) &:= \psi_{IX}^{(\hat{m})}(i,j),\end{aligned}$$

for $(i, j) \in$ subbands $(VII - IX)$. Here $\hat{m}$ is the index for the training image which gives the minimum at location (i, j). This is repeated for each coefficient in subbands I, II, III of the low resolution image. Thus for each coefficient in subbands $I - III$, we learn a total of 16 coefficients for each of the subbands $VII - IX$ from the training set.

It may be mentioned here that each 4×4 region in the low resolution image could be learnt from different training images. In case the error (MAD) term in Eq. (6.1) is quite large, it signifies that the 4×4 block does not find a good match in the training data, *i.e.*, an edge primitive does not have its corresponding high resolution representation in the database. In order to avoid such spurious learning, we accept the wavelet coefficients only when the MAD is less than a chosen threshold. The goodness of the learning depends on how extensive and useful is the training data set. In our experiments we use Daub4 wavelet bases for computing the discrete wavelet transform. The issue of which particular wavelet basis best fits the learning scheme has not been investigated in this chapter.

The subband 0 corresponds to the low resolution portion 'LL' (see Figure 6.3(a)) in the wavelet decomposition and since the corresponding 'LL' portions in the training set may have different brightness averages, inclusion of the pixels from 'LL' portion of the low resolution image does not yield a good match of an edge primitive as we want the edges to be brightness independent. Hence, we refrain from using the 'LL' portion of the low resolution image for learning. The complete learning procedure is summarized below in terms of the steps involved.

STEP 1: Perform a two level wavelet decomposition of the low resolution test image of size
$M \times M$ and three level decompositions of all training images each of size $2M \times 2M$.

STEP 2: Consider the wavelet coefficients at locations $(i, j + M/4)$, $(i + M/4, j)$ and $(i + M/4, j + M/4)$ in subbands I, II and III, and the corresponding 2×2 blocks in subbands $IV - VI$ of the low resolution image as well as the high resolution training set.

STEP 3: Obtain the sum of absolute difference between the wavelet coefficients in the low resolution image and the corresponding coefficients for each of the training images. Obtain the best match.

STEP 4: If MAD $<$ threshold, obtain the unknown high resolution wavelet coefficients (4×4 block) from the appropriate training image for each of the subbands $VII - IX$, else set them all zeros.

STEP 5: Repeat steps (2 - 4) for every wavelet coefficient in subbands $I - VI$ of the low resolution image.

A few comments about the learning of the wavelet coefficients are in order now. The high frequency coefficients are estimated using the nearest neighbor criterion from the training images. The process is not adaptive in the sense that no adaptive updating of these coefficients is performed based on previously learned values at a given location or from its neighborhood. The coefficients have to be learned afresh at every location. Furthermore, there is no reinforcement of the learnt coefficients through a posterior analysis. This may yield inferior values of the coefficients, but the advantage is that one does not have to worry about the convergence issues. A similar learning procedure is typically adopted in other learning based techniques in super-resolution. such as in [97, 100].

In this study we select a 4×4 edge primitive in the low resolution image for learning the coefficients. A smaller primitive could provide a better localized result, but more spurious matches negate the advantage. A larger primitive yields even better matches in the coefficient, but the localization is poor and the reconstruction suffers from severe blockiness. Furthermore, the requirement for the training data size goes up drastically when the size of the edge primitive is increased.

6.4 Regularization with Wavelet Prior

With the wavelet coefficients learnt from the high resolution training set as discussed in the previous section, we would like to obtain the super-resolution image for the given low resolution observation. Since we pick up the high frequency components of each 8 × 8 region as per the best fit edge element from different training data independently, there is no guarantee that the corresponding high resolution image would be a good one as it lacks any spatial context dependency. One may occasionally find an unwanted abrupt variation across the 8 × 8 blocks. In order to bring in a spatial coherence during the high resolution reconstruction, we must use a smoothness constraint. Thus the constraints are chosen based on enhancing the edges as well as ensuring the smoothness of the high resolution image. Near the edges in the low resolution image, we learn the wavelet coefficients from the high resolution database to have edge preserving upsampling. Also a smoothness constraint is enforced while upsampling at relatively smooth regions.

We use the wavelet coefficients learnt from the training set to enforce the constraint that the wavelet coefficients of the super-resolved image should be locally close to the best matching wavelet coefficients learnt from the training images in a least squares sense. Let $\mathbf{Z}_w$ be wavelet transform of the high resolution image $z(i,j)$ to be estimated and $\hat{\mathbf{Z}}_w$ be the wavelet transform of the learnt image as discussed in the previous section. Then the learnt prior term can be expressed as

$$C(\mathbf{z}) = ||\mathbf{Z}_w - \hat{\mathbf{Z}}_w||^2. \tag{6.2}$$

Now in order to enforce the smoothness constraint we make use of the fact that the image pixel intensities have a spatial correlation. But this constraint pushes the reconstruction towards a smooth entity. Hence in order to enforce a smoothness in the smooth regions alone while upsampling, we use a discontinuity preserving smoothness prior. Since the high frequency details learnt by using the wavelet-based prior constitute the discontinuities it would ensure undistorted edges in the super-resolved image while smoothing the regions with spatial continuity. In order to incorporate provisions for detecting such discontinuities, so that they can be preserved in the reconstructed image, the binary variables $l_{i,j}$ and $v_{i,j}$ which detect the horizontal and vertical edges, respectively, are used. The use of line fields in the context of MRF modeling has already been explained earlier in chapter 3. It may be noted that we are not modeling the high resolution image as an MRF unlike in previous chapters. The line fields are included to preserve the discontinuities while smoothing the edges. We use the following prior for the smoothness constraint in this study.

$$\begin{aligned} U(\mathbf{z}) = \sum_{i,j} \{ & \mu[(z_{i,j} - z_{i,j-1})^2(1 - v_{i,j}) + (z_{i,j+1} - z_{i,j})^2(1 - v_{i,j+1}) \\ & + (z_{i,j} - z_{i-1,j})^2(1 - l_{i,j}) + (z_{i+1,j} - z_{i,j})^2(1 - l_{i+1,j})] \\ & + \gamma[l_{i,j} + l_{i+1,j} + v_{i,j} + v_{i,j+1}]\}. \end{aligned} \tag{6.3}$$

Here μ is the penalty term for departure from the smoothness. The second term in the above equation enforces a penalty for over-punctuation in the smoothness constraint. In effect we are considering only a first order spatial relationship along with the scope for handling the discontinuities. Thus by making use of the data fitting term, the learning term and the smoothness constraint the final cost function to be minimized for the high resolution image $\mathbf{z}$ can be expressed as

$$\varepsilon = ||\mathbf{y} - D\mathbf{z}||^2 + \beta C(\mathbf{z}) + U(\mathbf{z}), \tag{6.4}$$

where D is the decimation matrix and β is a suitable weight. The above cost function is nonconvex and also consists of terms in both spatial domain (the first and the third term) and in wavelet domain (the second term). Hence it cannot be minimized by using a simple optimization technique such as gradient descent since it involves a differentiation of the cost function. We minimize the cost by using the simulated annealing technique which is expected to lead to a global minima. However, in order to provide a good initial guess and to speed up the computation, the result obtained by using the inverse transform of the learnt wavelet coefficients is used as the initial estimate for $\mathbf{z}$.

We now explain the various terms in Eq. (6.4) with respect to the wavelet-based learning method. The first term relates to the consistency in data fitting. If $\mathbf{z}$ is the actual high resolution image, we observe that $||\mathbf{y} - D\mathbf{z}||^2$ need not be zero as the chosen decimation operator D as defined in Eq. (3.12) need not be close to the wavelet decomposition (LL image in Figure 6.3) of the high resolution image, in general. The above is true only for Haar basis. However, the use of Haar basis introduces a lot more blockiness in the reconstructed image when the third (smoothness) term becomes very large. Alternately one may set all the wavelet coefficients in the finer subbands to be zero prior to taking the inverse wavelet transform. Although this may be similar in idea to the sinc interpolation, the corresponding interpolation results are quite inferior. The choice of Daub4 as the basis function in the study was more on an ad hoc basis, and a proper selection of the basis function would be an interesting topic of research. The selection of various weighting parameters in Eq. (6.4) was based on the idea that each term in the equation should have comparable magnitudes when the algorithm converges to the high resolution image.

6.5 Experimental Illustrations

We demonstrate the efficacy of the proposed technique to super-resolve a low resolution observation using the wavelet coefficients learnt from a high resolution training data set.

First we consider experiments with face images. A number of high resolution images of different objects were downloaded from the Internet arbitrarily to use them as a training set. We considered a high resolution training set of size $N = 200$. The same training data set has been used in all experiments. We show a random sample of these images as thumbnails in Figure 6.4

We notice that we have images of faces, indoor and outdoor scenes all mixed up in training set. In order to obtain a low resolution test image, we consider a high resolution image from the training set and downsample it by a factor of 2. Figure 6.5(a) shows one such low resolution face image of size 64 × 64. Figure 6.5(b) shows the same image upsampled by a factor of 2 using the bicubic interpolation technique. The super-resolved image is shown in Figure 6.5(c). A comparison of the Figures 6.5(b) and (c) shows more clear details in the super-resolved image. The features such as eyes, nose and the mouth appear blurred in the interpolated image shown in Figure 6.5(b), while they are restored well in Figure 6.5(c). Also the eye balls are sharper in the displayed super-resolved image. However, one can see some amount of blockiness at chin boundaries. It has been experimentally found that the best results are obtained with the parameters $\mu = 0.01$, $\gamma = 25$, and the weight for the learning term $\beta = 0.08$. These parameters were selected so that all the components in the cost function (refer to Eq. (6.4)) have comparable contributions. We retain the same values for the parameters in all subsequent experiments.

We show the results of experiments on another face image. The low resolution observation obtained by down sampling the high resolution

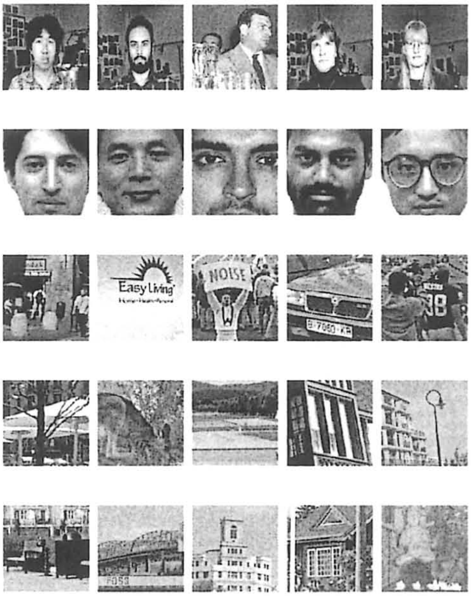


Fig. 6.4. A random sample of the training images used in learning the wavelet coefficients are shown as thumbnails.

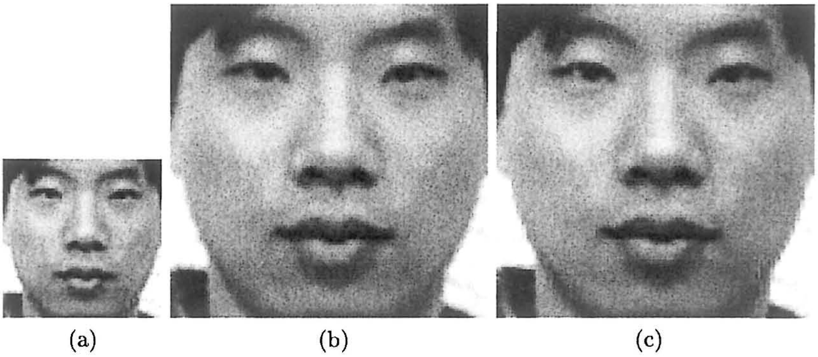
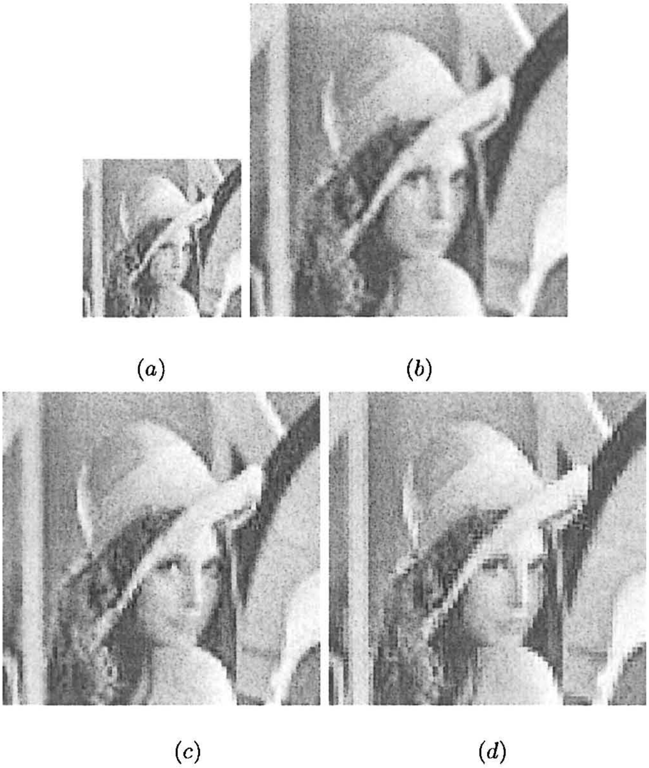


Fig. 6.5. (a) A low resolution observation (face1), (b) bicubic interpolated image, and (c) the super-resolved image using the proposed approach.

Lena image is shown in Figure 6.6(a). The super-resolution result obtained using the proposed approach is displayed in Figure 6.6(c), and Figure 6.6(b) shows the bicubic interpolated image. Once again we see that the high frequency details are better preserved in the super-resolved image. Various surface boundaries are much sharper. The hair strand on the shoulder and the lace on the hat appears more clearly.



(a) (b) (c) (d)

Fig. 6.6. (a) Another low resolution observation (Lena), (b) bicubic interpolated image, (c) the super-resolved image, and (d) the result of simple pixel replication.

The eyes and the nose are also clear. However, we observe some blockiness on the boundary curves of the hat and the slanted structure on the upper right corner of the picture. Since, the edge primitives are chosen over a 8 × 8 block in the wavelet domain, the learnt edges may suffer from blockiness. The smoothness constraint is supposed to take care of such jaggedness. However, the given choice of parameters in the smoothness term fails to undo the jaggedness. An increase in the weight for the smoothness term may render the solution very smooth and this may not be desirable. We could have played with the parameter set in Eqs. (6.3) and (6.4), but the various parameters for recovering the super-resolved image for this experiment were kept the same as used in the previous experiment. But this blockiness is nothing compared to the blockiness one obtains when a simple pixel replication is used.

Comparing this with the simple zero order hold expanded image shown in Figure 6.6(d) (in which every feature in the image appears blocky), we see that the blockiness is quite tolerable here. In order to test the algorithm for an image which has prominent edges, we considered a portion of a building image. The results for the same are shown in Figures 6.7(b-c) with the low resolution observation depicted in Figure 6.7(a). We can clearly see that the discontinuities are better estimated in the super-resolved image shown in Figure 6.7(c), but they appear blurred in the bicubic interpolated image (see Figure 6.7(b)). This substantiates our claim that the learning of wavelet coefficients does help in improving the resolutions.

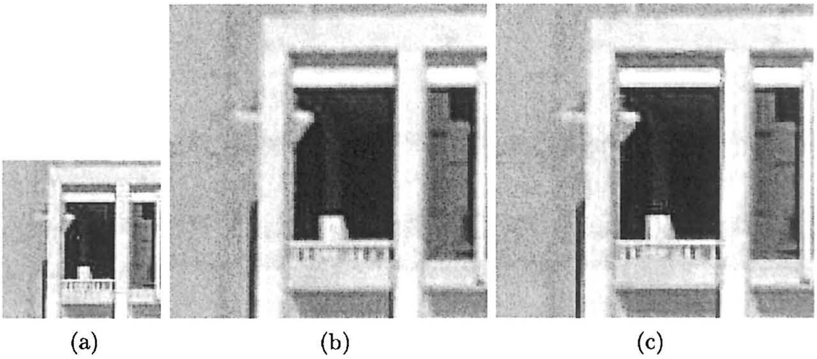


Fig. 6.7. (a) Low resolution observation of a building, (b) upsampled by bicubic interpolation, and (c) the super-resolved image.

We now consider a few experiments on the color image super-resolution. For these experiments we first convert the low resolution color image into $Y - C_b - C_r$ format. The learning of the wavelet coefficients is then done using the Y (luminance) plane only. The recovered high resolution image in the luminance plane after optimization is then combined with the bicubic interpolated version of the data in low resolution $C_b - C_r$ planes in order to obtain the super-resolved color image. The idea is quite similar to the way a macroblock is represented by $4:1:1$ DCT blocks in the $Y - C_b - C_r$ domain while using an MPEG coder. The training images used were kept the same as in the previous experiments on gray scale images. One may note here that learning of the wavelet coefficients for the Y, C_b, and C_r planes can also be done separately in order to obtain the super-resolution on each of the low

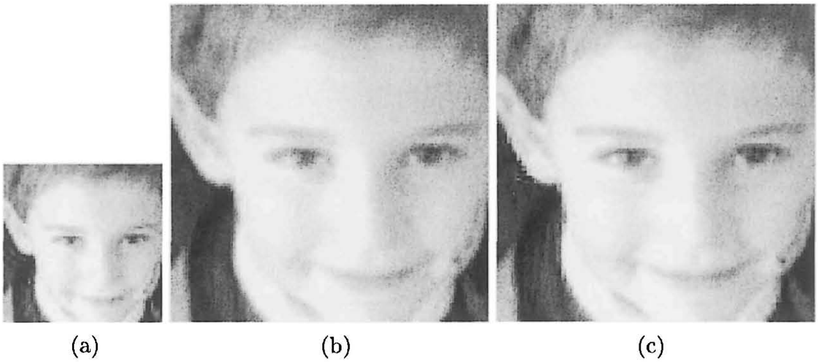


Fig. 6.8. (a) A low resolution observation, (b) upsampling using the bicubic interpolation, and (c) the super-resolved image using the proposed approach.

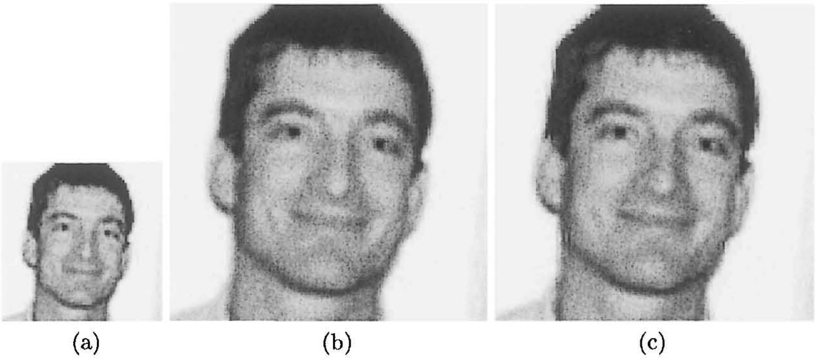


Fig. 6.9. (a) Another low resolution face image, (b) upsampling using the bicubic interpolation, and (c) the super-resolved image using the proposed approach.

resolution images. However, we refrain from doing it in this chapter as any possible error in learning in any of the color planes may introduce chromatic distortions and the human vision appears to be very sensitive to that.

We show results of two experiments conducted on the color face images. However, the results are shown here in gray tone. Figure 6.8(c) shows the result of the proposed approach on a low resolution observation shown in Figure 6.8(a). Compare this with the bicubic interpolated image shown in Figure 6.8(b). We observe that the super-resolved image appears sharper. Few areas of interest where such an enhancement can be observed are the mark on the left chin, eye balls and the hair. Some amount of blockiness can be observed near the ear. The results for another low resolution face image are displayed in Figures 6.9(a-c). Similar conclusions can again be drawn from this experiment. Observe the eye balls, eye brows, frontal hair, chin and the nose shown in Figure 6.9(c) which appear sharper when compared to the bicubic interpolated image given in Figure 6.9(b). However, some blockiness does appear along the silhouette. Thus we may conclude that the approach works well for color images as well.

In order to convey the comparative edge over the conventional interpolation techniques, we show the PSNR during interpolation for the gray scale images. Table 6.1 shows the comparison of the proposed method with the standard bilinear interpolation and the bicubic interpolation. In order to be able to compute the PSNR, we started with a high resolution image and the decimated version of that was used as the low resolution observation. We can observe that in addition to the perceptual betterment in all observed images there is also a gain in PSNR for the wavelet-based approach. This illustrates the usefulness of the wavelet-based learning scheme in super-resolving the images.

Method	face1	Lena	building
Bilinear	30.87	26.84	25.23
Bicubic	31.54	27.57	26.27
Proposed	32.74	28.05	26.97

Table 6.1. Comparison of PSNR achieved under different schemes.

6.6 Conclusions

We have described a method for super-resolution restoration of images using a wavelet-based learning technique. The wavelet coefficients at finer scales, learnt from a set of several high resolution training images, are used as a constraint along with an appropriate smoothness prior to estimate the super-resolved image. The learning term selects the best high resolution edges from the training set given a low resolution observation, while the discontinuity preserving smoothness term ensures a proper spatial correlation among pixel intensities. The results obtained for both gray scale and color images show perceptual as well as quantifiable improvements over conventional interpolation techniques. The method is useful when multiple observations of a scene are not available and one must make the best use of a single observation to enhance its resolution.

An inherent drawback of this learning method is that the learning process is very much resolution dependent. If we want to super-resolve a $2m$/pixel satellite image by a factor of $r = 2$ the training data must be of $1m$/pixel resolution. If one wants to perform super-resolution on a $2.5m$ image, none of the images in existing database could be used for training. For a commercial camera, if we change the zoom factor, it requires that a completely different set of training images be provided. Even if the zoom provides the details of a scene at a higher resolution, the wavelet-based approach fails to make use of this information. This provides us with the motivation to develop a different scheme for image super-resolution where the incremental information about a scene through camera zooming can be handled efficiently. This will be discussed in chapter 8.

Another major difficulty with wavelet-based learning lies in the fact that the wavelet decomposition kernel is separable. Although this provides computational advantages, we expect to catch only the horizontal and vertical edges properly. Hence we do not have difficulties in learning horizontal and vertical edges, but we do have some problem in learning edges oriented along arbitrary directions. This leads to blockiness in the learnt edges. A better way to handle this is to use directionally selective wavelet decomposition to learn the oriented edges. A recent development in oriented multi-resolution analysis include the concepts curvelet, contourlet, edgelet, ridgelet, *etc.* [187, 188, 189, 190, 191]. It will be nice to explore the usefulness of these concepts in learning the oriented edge better.

We observe that the edge primitives have been defined locally and then they are learnt from the database images. Thus the learning is local and independent and hence it required the imposition of a smoothness constraint. Is it then possible to define the learning process globally? This would then eliminate the problem of blockiness in the reconstructed image. However, the global learning would imply that the input image should be somehow globally similar to the training images. We explore this issue in the next chapter.

7

Extension of Generalized Interpolation to Eigen Decomposition

In chapter 4 we demonstrated that any function may be decomposed into a set of sub-functions and each of these sub-functions may be suitably interpolated to upsample the given function. We decomposed the image intensity function in terms of structural and albedo components and showed that a better high resolution representation of the image can be obtained. Motivated by the above, we ask the question if any other alternative form of decomposition is possible. Such a decomposition should have some advantages over traditional image domain interpolation.

In this chapter we show that the same concept of generalized interpolation developed with respect to decomposing a function in terms of a number of subfunctions can also be applied to a finite dimensional vector space. In particular we explore an eigen-image based high resolution reconstruction technique. Eigen-images of a database of several similar training images are obtained and the given low resolution image is projected onto the eigen-images to compute the projection coefficients. The eigen-images are then interpolated using a suitable interpolation method and approximated to the nearest orthonormal bases. The high resolution image is reconstructed using these interpolated basis functions. This method is applicable to images of a particular class of object and results are demonstrated for both face and fingerprint images. This method offers a significant advantage when the input image is blurred and noisy. Thus, a blind restoration is possible using the eigen decomposed observation. We also explore the case when a training database of high resolution images are available.

7.1 Introduction

In chapter 4 we needed several low resolution observations of the scene with a moving point light source to decompose the given intensity image into structural and albedo components. Here we do away with the requirement of having to have several low resolution observations of the scene to achieve generalized interpolation. Thus, in effect, we have a single low resolution, noisy and blurred observation which needs to be super-resolved. In chapter 6 we discussed a method where a single observation was super-resolved by learning the high resolution wavelet coefficients representing the edges in the image from a given set of arbitrary but high resolution training image. However, the method could not handle an input image with arbitrary blurring as the method was unable to decide the scale at which the edge primitives should be searched from the training data. In this chapter, we relax this constraint, *i.e.*, the low resolution input image may have an arbitrary amount of blurring. Further, we do not impose any restriction on the form of blur point spread function (PSF). For example, we assumed the blur to be parameterizable using a single variable σ (read a Gaussian shaped blur) in chapter 3. Thus, we look at the general case of having an arbitrary blurred and noisy input image. However, we do put a constraint on the available training image database. We require that all training images must conform to the same class of objects, like a face or a fingerprint.

In many biometric databases, a large number of images of similar contents, shape and size are available. For example, in investigative criminology one has available face and fingerprint databases. These are often taken at controlled environment and can be registered easily. The question we ask is that if one encounters a poor quality input image, can it be enhanced using the knowledge of the properties of the database images? Thus, the basic problem that we solve in this chapter is as follows. Given a low resolution input image belonging to a particular class (face, fingerprint, *etc.*) and a database of several good quality images of the same class, obtain a high resolution output. We perform a principal component analysis (PCA) on the image database and an appropriate interpolation is carried out on the eigen-images, using which the high resolution image is reconstructed. We show that this method is particularly useful when the input image is noisy and partly blurred so that the other existing learning-based methods do not provide a good solution.

In chapter 6, we have explained the concept of learning based super-resolution reconstruction. The method discussed in this chapter can also be classified under that. Previously we learnt the high resolution edge primitives from the training data set for the low resolution edges in the observation. Thus, the learning was local and hence, we opted for the wavelet based representation of the image due to its nice localization property. However, the PCA-based method to be discussed in this chapter utilizes a global learning. Imagine what would happen if the wavelet bases are replaced by the Fourier bases. The edges can no longer be learnt locally. Certain aspects of the input image now has to be learnt globally. This imposes the constraint that all the training images should be globally similar, *i.e.*, they should represent a similar class of objects or signals. This calls for the use of eigenvectors in discrete space domain instead of the Fourier bases for an efficient utilization of the signal statistics.

7.2 PCA-Based Generalized Interpolation

We have discussed the concept of generalized interpolation while using the photometric cue in chapter 4. This is done by decomposing the image into appropriate subspaces, carrying out interpolation in individual subspaces and subsequently transforming the interpolated values back to the image domain. The given function was decomposed into subspaces consisting of the structure of the object represented by the surface gradient and the albedo as given by $a_1(x,y) = p(x,y), a_2(x,y) = q(x,y)$ and $a_3(x,y) = \rho(x,y)$ where p and q are the surface gradients and ρ is the albedo, respectively. In this chapter we use the same parametric decomposition given by Eq. (4.1), *i.e.*,

$$f(x,y) = \phi(a_1(x,y), a_2(x,y), \ldots, a_K(x,y)). \tag{7.1}$$

However, we decompose the given function into an eigen-space containing the principal components $\bar{a}_i, \;\; i = 1, \cdots, K$ and use a linear function ϕ, *i.e.*,

$$\phi(a_1(x,y), a_2(x,y), \ldots, a_K(x,y)) = \sum_{i=1}^{K} w_i \bar{a}_i. \tag{7.2}$$

Note that here $\bar{a}_i$'s are orthogonal to each other and they are derived from the database of training images. Similarly, w_i represents the projection of the given image on the i^{th} basis vector (eigen-image). We may now apply a suitable interpolation in the eigen-space and then combine them to get the super-resolved image. Thus the method is a special case of generalized interpolation. Since the eigen-images are not dependent on the input image $f(x, y)$, and they are computed from the database of training images, all these interpolated basis vectors can be pre-computed and stored. Hence the method, if it does at all provide a good image reconstruction, will be a very fast one.

7.2.1 Eigen-Image Decomposition

An image can be reconstructed from eigen-images in the PCA representation as described in [192]. Eigen-image decomposition for a class of similar objects is currently the most popular and actively pursued area of research in pattern recognition. The terms like eigen-face, eigen-shape, eigen-palm, eigen-iris, eigen-ear, eigen-nose, *etc.*, are increasingly being used in the literature and product brochures to specify the domain of recognition. The concept derives its origin from the task of finding principal components of a signal from an ensemble of its observations. We refrain from discussing this in the monograph for reasons of brevity.

The basic procedure for computing the eigen-space is as follows: We have a dataset of N similar training images, represented by the matrix $\Lambda = [\boldsymbol{\mathcal{F}}_1, \boldsymbol{\mathcal{F}}_2, \ldots, \boldsymbol{\mathcal{F}}_N]$, where $\boldsymbol{\mathcal{F}}_i$ is the i^{th} training image. Note that the training image of size $M \times M$ is converted to a vector $\boldsymbol{\mathcal{F}}$ of size M^2 through a raster scan conversion. Thus the matrix Λ has the dimension $M^2 \times N$. In PCA, a set of top K eigenvectors $E = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_K]$, also called eigen-images, of dimension $M^2 \times K$ are computed from the covariance matrix,

$$\Sigma = \sum_{i=1}^{N} (\boldsymbol{\mathcal{F}}_i - \mathbf{m}_{\mathcal{F}})(\boldsymbol{\mathcal{F}}_i - \mathbf{m}_{\mathcal{F}})^T. \tag{7.3}$$

where $\mathbf{m}_{\mathcal{F}}$ is the average image intensity defined by

$$\mathbf{m}_{\mathcal{F}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\mathcal{F}}_i. \tag{7.4}$$

Note that the size of the training database is much smaller than the dimension of the image, *i.e.*, $N << M^2$. Hence Σ in Eq. (7.3) is rank deficient. Further, one does not store all eigen-images for the non-zero eigenvalues. We retain only top K (where $K < N$) eigen-images based on the magnitude of the eigenvalues. Since $K << M^2$, given the eigen-images one cannot reconstruct an image exactly. However, since all images are similar in nature (like face or ear images), only a small value of K suffices to reconstruct an image to good enough details.

For a given low resolution image $\mathbf{f}_l$, a weight vector can be computed by projecting it onto eigen-images using

$$\mathbf{w} = E^T(\mathbf{f}_l - \mathbf{m}_{\mathcal{F}}). \tag{7.5}$$

An approximate reconstruction of $\mathbf{f}_l$ can be obtained from the top K eigen-images,

$$\hat{\mathbf{f}} = E\mathbf{w} + \mathbf{m}_{\mathcal{F}}. \tag{7.6}$$

Since K is typically much smaller than the size of the image vector, the image representation through the eigen-image expansion is not complete. Hence $\hat{\mathbf{f}}$ is an approximation of $\mathbf{f}_l$ and the quality of approximation depends on its nearness to the class of images in the database.

7.2.2 Eigen-Image Interpolation

Now we wish to form a set of high resolution eigen-images using which we can construct the high resolution output corresponding to the given low resolution input image. In order to do this all the K low resolution eigenvectors E and the mean vector $\mathbf{m}_{\mathcal{F}}$ are upsampled using the bicubic interpolation. Any other suitable interpolation scheme can also be used. But we restrict to bicubic interpolation in this study. The interpolated mean vector is given by $\mathbf{m}_z$ and the upsampled set of eigenvectors are given by $E_h = [\mathbf{e}_{1h}, \mathbf{e}_{2h}, \cdots, \mathbf{e}_{Kh}]$, *i.e.*,

$$\mathbf{m}_z = \mathbf{m}_{\mathcal{F}}(\uparrow r)$$

and

$$E_h = [\mathbf{e}_1(\uparrow r), \mathbf{e}_2(\uparrow r), \cdots, \mathbf{e}_K(\uparrow r)],$$

where the symbol $\uparrow r$ represents upsampling by a factor of r. One may use an appropriate upsampling factor such as $r = 2, 3, 4$, *etc.* The new set of interpolated eigenvectors need not be orthonormal. They are then transformed into the nearest set of orthonormal vectors using the Gram-Schmidt orthogonalization procedure. Since all these vectors are of unit norm, the weights (eigenvalues associated with the corresponding eigen-images) must be multiplied by the upsampling factor r

($i.e.$, $\mathbf{w}_h = r\mathbf{w}$) to preserve the average brightness of the interpolated pictures. The high resolution image is now reconstructed using

$$\begin{aligned}\hat{\mathbf{z}} &= E_h \mathbf{w}_h + \mathbf{m}_z \\ &= r \sum_{i=1}^{K} w_i \mathbf{e}_i(\uparrow r) + \mathbf{m}_z,\end{aligned} \tag{7.7}$$

where w_i is the projection of the input image on the i^{th} eigen-image. Compare this to Eq. (7.2) and observe that Eq. (7.7) is nothing but the generalized interpolation. There are two primary differences with the way the generalized interpolation is carried out compared to what we presented in chapter 4.

1. One does not require several observations to decompose the original image into constituent eigen-images unlike in the previous case. The decomposition is based on the statistics learnt from the training images.
2. The decomposition of the original image $\mathbf{f}_l$ into eigen-images $\mathbf{e}_i$ is linear.

The reader may recall that the motivation for using the generalized interpolation lay in the fact that the original function $\mathbf{f}_l$ may not be band limited but the constituent function a_i's may very well be. However, the above argument is no longer valid for PCA-based decomposition. If all $\mathbf{e}_i$'s in Eq. (7.7) are bandlimited, so is $\mathbf{f}_l$ due to linearity. Hence the PCA-based upampling cannot eliminate aliasing from the given image $\mathbf{f}_l$. Then what is the motivation for PCA-based upsampling? Does it provide any benefit in terms of having a lower interpolation error while upsampling? Let us look at this issue in more details.

According to Lagrange's theorem, if a function $f(x)$ possesses the $(n+1)^{th}$ derivative $f^{(n+1)}(x)$ at all points in an interval containing the point x_0, the remainder $R_n(x)$ is representable in the form

$$R_n(x) = f^{(n+1)}(\xi)\frac{(x - x_0)^{(n+1)}}{(n+1)!}$$

for every point x in this interval where ξ is a number lying between x_0 and x. Using this, it can be easily shown that for a n^{th} order polynomial approximation of the original (unknown) function $\tilde{f}$ at a point δx away from the nearest grid point, the approximation error is bounded by [193]

$$|f - \tilde{f}| < \frac{|\delta x|^{n+1}}{(n+1)!} \max_x |(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})^{n+1} f(x)|. \tag{7.8}$$

For a thin plate fitting spline over a square grid of size h, the maximum error is bounded by [194]

$$|f - \tilde{f}| \leq ch\sqrt{|\log h|}\ \|\mathcal{L}f\|, \tag{7.9}$$

where c is a positive number given by $[(32\pi)^{-1}(3\log 2)]^{\frac{1}{2}}$ and $\mathcal{L}$ stands for the corresponding regularization term.

Let us now consider the following abstract parametric decomposition of the function $\tilde{f}(x)$.

$$\tilde{f}(x) = \phi(a_1(x), a_2(x), \ldots, a_K(x)), \tag{7.10}$$

where $a_i(x),\ i = 1, 2, \ldots, K$ are different functions of the interpolating variable x and when they are combined by an appropriate K-variate function ϕ, one recovers the original function. We can now interpolate the individual functions $a_i(x)$ and combine them using Eq. (7.10) to obtain a rescaled $\tilde{f}(x)$.

The interpolation error at a point x can be written as

$$\begin{aligned} |f_g - \tilde{f}| &= |\phi(a_1(x) + \epsilon_1, \ldots, a_K(x) + \epsilon_K) \\ &\quad -\phi(a_1(x), a_2(x), \ldots, a_K(x))| \\ &\approx |\epsilon_1 \frac{\partial \phi}{\partial a_1} + \cdots + \epsilon_K \frac{\partial \phi}{\partial a_K}| \end{aligned} \tag{7.11}$$

where f_g represents the result of generalized interpolation. Here ϵ_i, $i = 1, 2, \ldots, K$ are the interpolation error at the same point x for the associated interpolant $a_i(x)$. In order to get a feel for the behavior of the error function for the PCA-based upsampling method, we consider ϕ to be a linear function, *i.e.*,

$$\phi(a_1(x), a_2(x), \ldots, a_K(x)) = \sum_{i=1}^{K} \alpha_i a_i(x),\ \ \alpha_i \geq 0\ \forall i. \tag{7.12}$$

From Eq. (7.12), the interpolation error using a n^{th} order polynomial at a point δ away from a grid point x is given by

$$|f_g(x) - \tilde{f}(x)| \leq \sum_{i=1}^{K} \alpha_i |\epsilon_i|,$$

i.e.,

$$|f_g(x) - \tilde{f}(x)| < \frac{|\delta|^{(n+1)}}{(n+1)!} \sum_{i=1}^{K} \alpha_i \max_x |(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})^{n+1} a_i(x)|. \qquad (7.13)$$

On the other hand, if one performs an n^{th} order polynomial interpolation at the same location on the scattered data $f(x_i)$ itself, the corresponding error bound is

$$|f_g(x) - \tilde{f}(x)| < \frac{|\delta|^{(n+1)}}{(n+1)!} \max_x |(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})^{n+1} f(x)|. \qquad (7.14)$$

We need to determine whether we gain anything by individually interpolating the constituent functions of ϕ instead of interpolating the function $f(x)$ directly? In order to prove that there is, indeed, some gain, one should compare Eq. (7.13) and (7.14) and must prove that

$$\sum_i \alpha_i \max_x |\frac{\partial^{n+1} a_i(x)}{\partial x^{n+1}}| \leq \max_x |\frac{\partial^{n+1} f(x)}{\partial x^{n+1}}| \qquad (7.15)$$

Similarly, for a thin plate spline interpolation, it can be shown that if one were to achieve a lower approximation error using the parametrically decomposed generalized method, we must have

$$\sum_{i=1}^{K} \alpha_i ||\mathcal{L} a_i(x)|| \leq ||\mathcal{L} f(x)||. \qquad (7.16)$$

Unfortunately, all the above relationships are not valid when ϕ is a linear function of polynomials. Thus, a direct interpolation of the function $f(x)$ seems to be an equally good option instead of the indirect one.

7.3 Usefulness of PCA

In the last section we noticed that the PCA-based upsampling method neither provides an alias-free reconstruction nor achieves a lower interpolation error. In a typical image super-resolution problem, one is required to restore the input image from its noisy, blurred and aliased observations. Quite naturally, the PCA-based method cannot handle aliasing in the observation. We refrain from discussing this issue further in this chapter. But does it help in removing sensor noise and the image blur?

Let us first assume that the observation f is free from blur, but is quite noisy. Since the eigen-image representation (see Eq. (7.6)) is an incomplete representation and since the noise present in the input image is expected to be totally uncorrelated to all the available basis vectors, the reconstruction process reduces the noise drastically. Due to the incompleteness of the basis vectors, the reconstructed image may be partly distorted. But it is the incompleteness of the eigen decomposition that removes the noise from the data. Since the input image conforms to the given class of objects, only a few eigen-images are required to reconstruct the image without much error. It is this property of the PCA that allows us to filter out the noise.

Let us now consider the case when the input image is blurred, *i.e.*,

$$\mathbf{f}_b = \mathbf{h} * \mathbf{f}$$

where $\mathbf{f}$ is the true image and $\mathbf{h}$ is the blur PSF. For simplicity, let us assume that the blur kernel is of finite impulse response (FIR) in nature, when

$$\mathbf{f}_b = \sum_i \alpha_i f(x+i) \overset{\Delta}{=} \sum_i \alpha_i \mathbf{f}_i, \tag{7.17}$$

with $\sum \alpha_i = 1$ (mean preserving blurring) and $\alpha_i \geq 0 \;\; \forall \; i$. Without loss of generality, we may assume the true image $\mathbf{f}$ to be zero mean. Using Eq. (7.5) we can compute the projections on the eigen-images

$$\mathbf{w}_b = E^T \mathbf{f}_b = \sum_i \alpha_i (E^T \mathbf{f}_i) \overset{\Delta}{=} \alpha_0 \mathbf{w} + \sum_{i=1} \alpha_i \mathbf{w}_i, \tag{7.18}$$

where $\mathbf{w}$ is, as before, the projection coefficient vector for the true (non-blurred) image, and $\mathbf{w}_i$ corresponds to the projection coefficients for the shifted image $\mathbf{f}_i$. Since the eigen-images represented by E correspond to the principal components, and since an image is typically correlated over its neighbors, it is expected that $\mathbf{w}_i \simeq \mathbf{w} \; \forall i$. The reconstructed image $\hat{\mathbf{f}}_b$ is given by

$$\hat{\mathbf{f}}_b = E\mathbf{w}_b = \alpha_0 E\mathbf{w} + \sum_{i=1} \alpha_i E\mathbf{w}_i. \tag{7.19}$$

The error in reconstruction due to the blurred observation with respect to the previously obtained (see Eq. (7.6)) image $\hat{\mathbf{f}}$ is given by

$$\hat{\mathbf{f}} - \hat{\mathbf{f}}_b = (1 - \alpha_0) E\mathbf{w} - \sum_{i=1} \alpha_i E\mathbf{w}_i.$$

Using the fact that $\sum \alpha_i = 1$, we get

$$|\hat{\mathbf{f}} - \hat{\mathbf{f}}_b| = |\sum_{i=1} \alpha_i E\mathbf{w} - \sum_{i=1} \alpha_i E\mathbf{w}_i| = |\sum_{i=1} \alpha_i E(\mathbf{w} - \mathbf{w}_i)|. \quad (7.20)$$

Since E form an orthonormal basis set,

$$|\hat{\mathbf{f}} - \hat{\mathbf{f}}_b| \leq \sum_{i=1} \alpha_i |(\mathbf{w} - \mathbf{w}_i)|. \quad (7.21)$$

As mentioned earlier $|\mathbf{w} - \mathbf{w}_i|$ is very small and similarly $\alpha_i < 1$. Hence $|\hat{\mathbf{f}} - \hat{\mathbf{f}}_b|$ is quite negligible, *i.e.*, the reconstructed image is quite good even if the observation was blurred. Or in other words, if the input image is blurred, it will still have significant correlation with the corresponding eigen-images of the *ideal* image. Since the eigen-images have been computed using the good quality training images, the reconstruction process is expected to remove the blur present in the data. Needless to say, if the input image is badly blurred, the associated eigen expansion may be very different from that of the ideal image, when the reconstruction will be quite poor. Direct interpolation of the input image does not solve any of the above two problems of blurring and noise perturbation, justifying the claim that the PCA-based restoration does help.

7.4 Description of Algorithm

The PCA-based restoration algorithm is summarized below in terms of the steps involved.

STEP 1: Perform the PCA decomposition on the low resolution image database to get k eigen-images represented by the matrix E and also obtain the mean image $m_{\mathcal{F}}$.

STEP 2: Project the given low resolution image f_l onto the eigen- images to get the eigen-image coefficients $\mathbf{w}$.

STEP 3: Interpolate the eigen-images E and the mean image $\mathbf{m}_{\mathcal{F}}$ to get the corresponding high resolution eigen-image matrix E_h and the high resolution mean image $\mathbf{m}_z$.

STEP 4: Approximate the high resolution eigen-images to the nearest orthonormal bases. These are precomputed and stored while obtaining the principal components.

STEP 5: Obtain the super-resolved image using Eq. (7.7).

It may be noted that only steps 2 and 5 need to be computed for a given input image for restoration. Hence the method is very fast. Since no high pass filter is used for deblurring, it does not boost the noise. However, the method may fail if the blurring is very severe or if the input images are not properly registered with those of the database images.

7.5 Use of High Resolution Database

We have assumed thus far that the training image database is at the same spatial resolution as the input image. The use of upsampling of the eigen-images does not recover the high frequency details. However, we observed that the PCA-based method is able to undo the image blurring and to remove noise. Blurring in a signal is closely related to its scale at which the signal is viewed. We expect a good correlation between the eigen-images at different scale. If the correlation structure remains quite unchanged over the scale, we may be able to move the upsampling process at the output end in step-3 in section 7.4 to the input side before step-1 itself. Let us see what this achieves for us.

We do now have a number of high resolution training images of a particular object class. The principal components of these training data are obtained. A low resolution (say, a decimation factor of r), blurred and noisy image is first upsampled by a factor of r using any interpolation technique. The upsampled input image is now projected onto the eigen-images, and the high resolution restored image is obtained using equation Eq. (7.6). Since the training images are all of high resolution, the input image is, indeed, super-resolved in the sense that it is now able to recover the high frequency details.

The performance of this super-resolution scheme depends on how good (or correlated) the training images are with respect to the input image. Hence the method is applicable to an image of a specific class of object such as fingerprint or face images.

7.6 Experimental Results

We now demonstrate the performance of the PCA-based upsampling method. We show results for both the cases, *i.e.*, the training images are at the same resolution as the test image, and when the test image is at a lower resolution. Experiments were conducted on both face and fingerprint images. For face images the database consisted of 105 good quality images (in the sense that there is no blur in the training data) of size 82 × 96 pixels. All the images were of frontal face and no preprocessing was done on them. A high resolution image is blurred using a 3 × 3 Gaussian kernel with standard deviation 0.5, and added with zero mean Gaussian noise of different standard deviations (σ) to form the input image. For the second case, the input image is decimated by a factor of r to serve as the input image to be super-resolved. The same database is used to serve as the high resolution training data.

Figure 7.1 shows the first 10 eigen-images computed from the database of 105 face images. The eigen-images were then upsampled by a factor of r (say $r = 2, 3, 4$, *etc.*) and stored for subsequent usage. In Figure 7.2 the noisy input image with $\sigma = 0.1$ (the gray values for the images considered in this chapter have been normalized in the range $[0, 1]$) and the corresponding bicubic interpolated image and the super-resolved images for zoom factors of 2 and 4 are shown. It can be observed that the super-resolved image is almost noise free and more clear than the bicubic interpolated image which is highly noisy. This is quite expected as the bicubic interpolated image takes the given noisy image itself as the input and hence it can remove neither blur nor the noise. For the PCA-based approach, the lips, the eye-brows and the hairlines appear quite clearly.

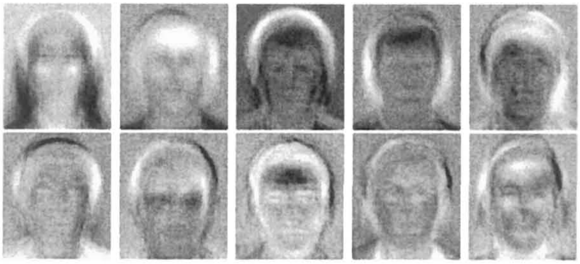
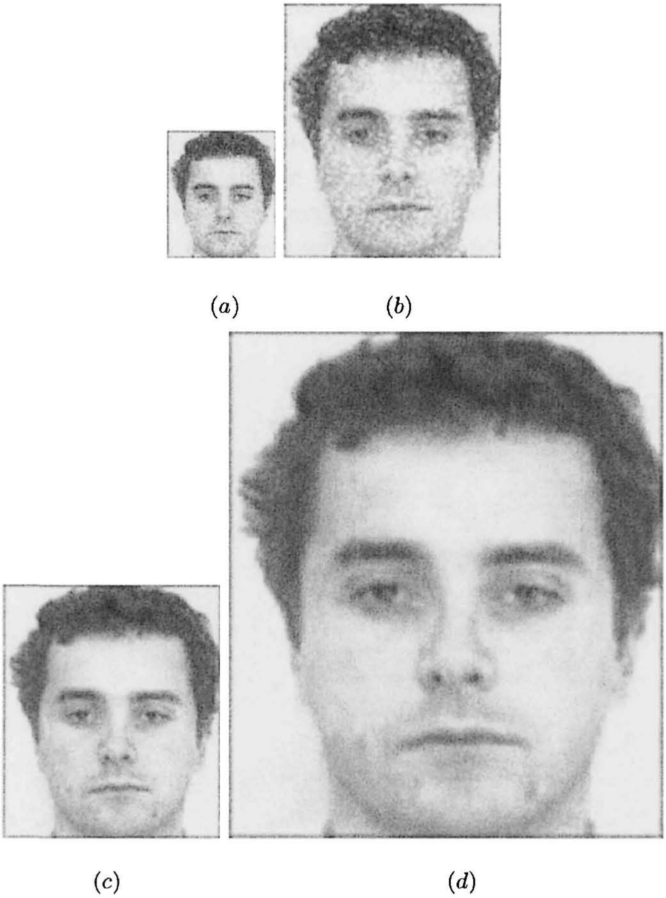


Fig. 7.1. First ten eigen-images obtained from the training data set.

We now experiment on what happens if the noise level is increased. In Figure 7.3, even though the given observation is much more noisy ($\sigma = 0.5$), the super-resolved image is of far better quality compared to the bicubic interpolated image which is very noisy for obvious reasons. The quality of reconstruction is now inferior to what we obtained in



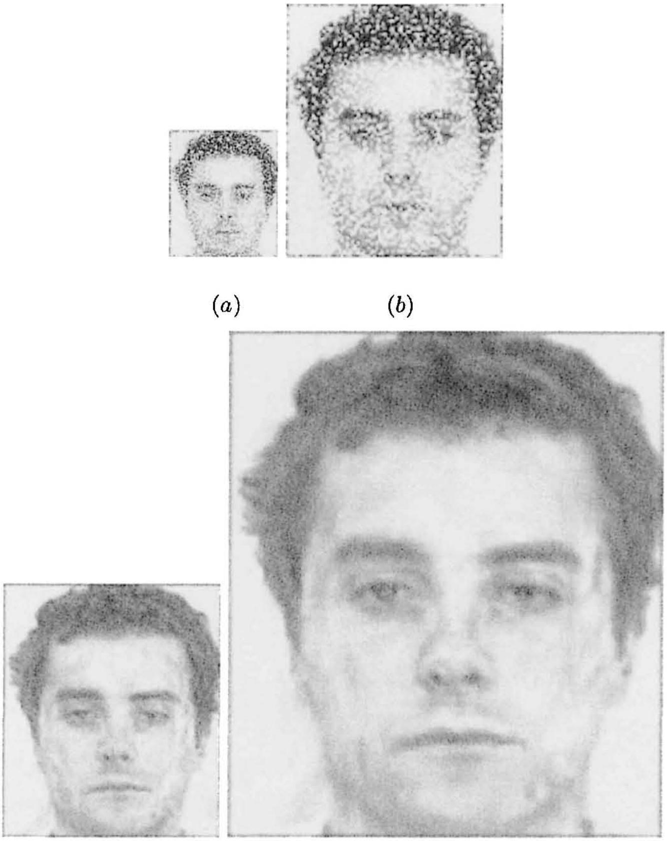
(a) (b)

(c) (d)

Fig. 7.2. (a) A low resolution noisy observation ($\sigma = 0.1$), (b) bicubic interpolated image with $r = 2$, PCA-based restoration with (c) $r = 2$ and (d) $r = 4$.

Figure 7.2(c). Some artifacts are now visible on the left cheek. The performance is quantified in terms of the PSNR tabulated in table 7.1 where the PSNR for the bicubic interpolated image and super-resolved image for a zoom factor $r = 4$ and for different values of noise level are shown. As mentioned in section 7.3 it is observed that when the noise level σ is very large, the reconstructed image deviates from the original face image.

We now investigate the performance of the PCA-based method when the input image is severely blurred. In Figure 7.4(a) an input image



(*a*) (*b*)

(*c*) (*d*)

Fig. 7.3. (a) A very noisy observation ($\sigma = 0.5$). (b) Result of bicubic interpolation with $r = 2$. PCA-based reconstruction for (c) $r = 2$, and (d) $r = 4$.

which is blurred with a 7×7 Gaussian mask with a standard deviation of 2 is shown. The details on the face is quite lost in the input. As expected, the output due to bicubic interpolation is heavily blurred, but the super-resolved image is almost free from blur. The details on the face are now quite restored in Figure 7.4(c). However, we observe a bit of artifacts on the face. This demonstrates that as long as there is a good correlation of the input image with the eigen-images, a good reconstruction is, indeed, possible. Thus the key aspect about the PCA-based method is its capability to recover a good quality image when the input image is blurred and noisy.

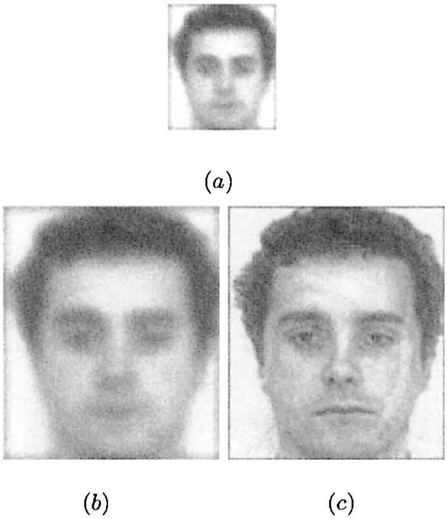
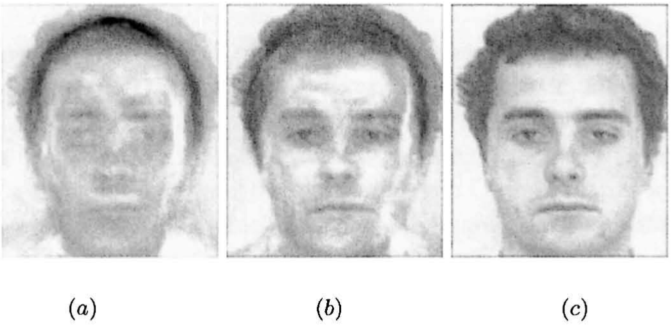


Fig. 7.4. (a) A highly blurred low resolution observation, (b) bicubic interpolated image, with $r = 2$, and (c) the reconstructed image with $r = 2$.

Now we experiment on how many eigen-images are required for a good reconstruction. Figure 7.5 shows the reconstructed image obtained using 10, 20 and 50 eigen-images. Here the figure given in 7.2(a) served as the input image. It is observed that using the top 50 eigen-images a good quality output can be reconstructed. Compare this to the result given in Figure 7.2(c) which was obtained using 100 eigen-images. They are nearly indistinguishable in quality.

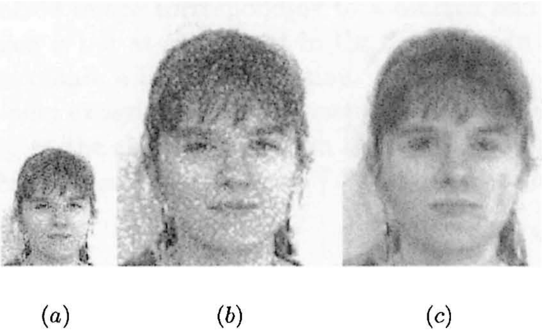
In all the above experiments the low resolution input image was a part of the database which consisted of 75 male faces and 35 female faces. The database had the picture of the same person but at a different orientation. Figure 7.6 shows the bicubic interpolated image and the super-resolved image corresponding to a blurred and noisy input face image which is not at all present in the database. In this case also we are able to obtain a better restoration.

In the next experiment we demonstrate that if the input image does not belong to the class of objects in the database, one cannot do any meaningful reconstruction. Figure 7.7 shows the reconstructed image



(a) (b) (c)

Fig. 7.5. PCA-based reconstructions using different numbers of eigen-images. (a) $K = 10$, (b)$K = 20$, and (c)$K = 50$.



(a) (b) (c)

Fig. 7.6. Restoration of an input image not present in the database. (a) Noisy observation, (b) bicubic interpolated image, and (c) result of PCA-based restoration for $r = 2$.

for some arbitrary input image using the face image database and 100 eigen-images. Here the output is not at all related to the input, which indicates clearly that the PCA-based method is applicable only for a specific class of images.

We now show results of experiments on a different database. Figure 7.8 shows the poor quality input, bicubic interpolated result and the super-resolved images for zoom factors of $r = 2$ and 4 for a fingerprint image. The results are shown for a noise level of $\sigma = 0.1$. In this experiment the low resolution database consisted of 150 fingerprint images of size 32 × 32 pixels, and the top 100 eigen-images were used for reconstruction. It can be observed that the super-resolved image is more clear and noise free compared to the bicubic interpolated image. We also compare the performance in terms of the PSNR measure and

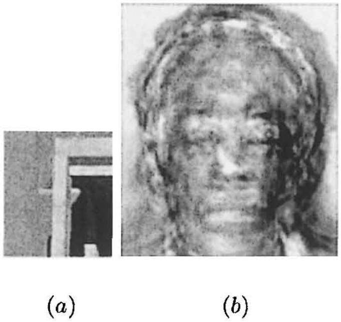


Fig. 7.7. Illustration of PCA-based restoration for an arbitrary input image very different from the given class of face images. (a) The input image, and (b) restored image!

the corresponding values are given in Table 7.1 for different values of noise variance. We observe a substantial improvement in PSNR for the PCA-based approach.

Image	Method	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.5$
Face	Bicubic	22.93	20.27	16.78
	Proposed	24.23	22.88	19.79
Fingerprint	Bicubic	20.88	17.76	13.72
	Proposed	21.36	20.01	16.50

Table 7.1. Comparison of PSNRs for a zoom factor of $r = 4$ for different levels of noise.

We now investigate the performance of the method when the upsampler at the output end is replaced by an upsampler at the input end (section 7.5). Thus the input image is at a lower resolution, but the training database is at a higher resolution. For convenience, we use the same training database, but the input image is decimated by a suitable factor r to serve as the low resolution observation. This observed image is then appropriately interpolated before applying the PCA-based restoration.

In Figure 7.9(a), we show a low resolution observation. We add a Gaussian white noise with $\sigma = 0.1$ to simulate the presence of noise in the data. Figure 7.9(b) shows the result of bicubic interpolation. Quite naturally it is poor due to the presence of noise. In Figure 7.9(c) we show the result of PCA-based high resolution restoration for the

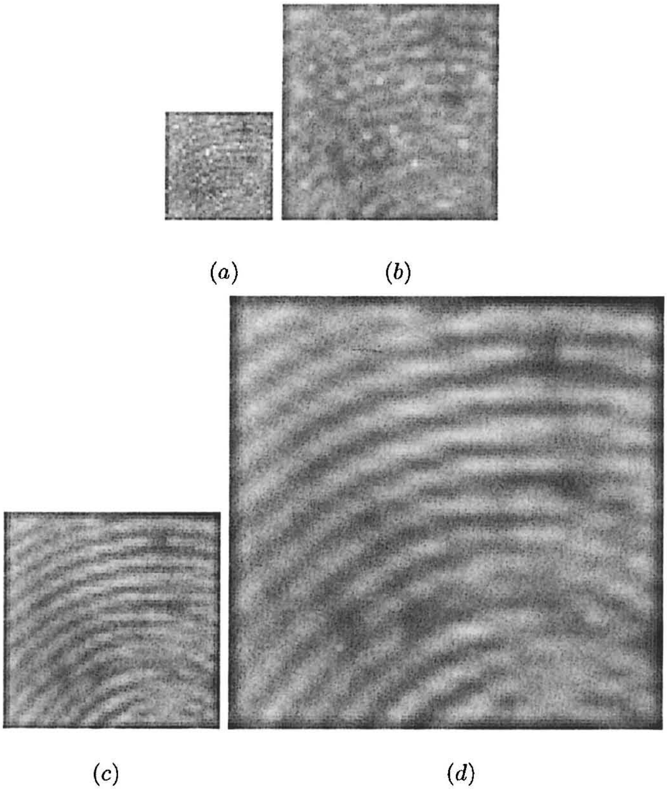


Fig. 7.8. Illustration of results for a different object class. (a) A poor quality fingerprint image. Results of (b) bicubic interpolation, and PCA-based interpolation for (c) $r = 2$, and (d) $r = 4$.

upsampling factor of $r = 2$. The effect of noise is almost removed and the quality of reconstruction is very good.

In Figure 7.10(a), the same input as shown in Figure 7.9(a) is further corrupted with additive noise. We use $\sigma = 0.5$ and the corresponding input image is of very poor quality. Hence we do not accept to view the image content in Figure 7.10(b) when the image is bicubically interpolated. The result of high resolution PCA-based restoration for $r = 2$ is shown in Figure 7.10(c). Although the corresponding reconstruction is inferior compared to what is given in Figure 7.9(c), the face is still identifiable. There appears to be significant distortion near the lips.

Now we show some results of experimentation for an upsampling factor of $r = 4$. The input image having a good amount of noise cor-

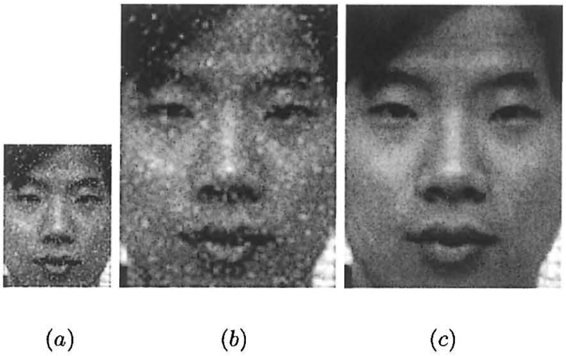


Fig. 7.9. (a) A low resolution noisy observation ($\sigma = 0.1$), (b) bicubic interpolated image, and (c) PCA-based restoration with $r = 2$ using a high resolution training data set.

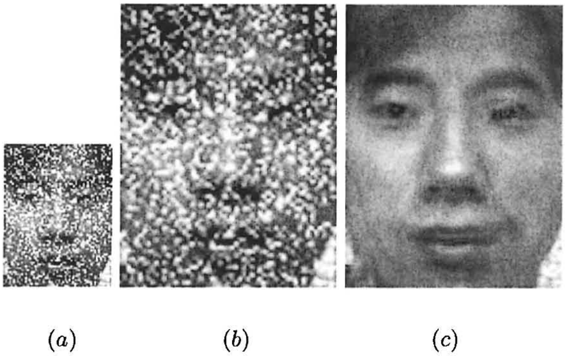


Fig. 7.10. (a) A low resolution, extremely noisy observation ($\sigma = 0.5$), (b) bicubic interpolated image, and (c) the high resolution PCA-based restoration with $r = 2$.

ruption ($\sigma = 0.1$) is shown in Figure 7.11(a). The corresponding bicubic interpolated image with $r = 4$ is shown in Figure 7.11(b). The result of PCA-based reconstruction is shown in Figure 7.11(c). Compare this with the corresponding result for $r = 2$ given in Figure 7.9(c). We notice that reconstruction is still very good even for an upsampling factor of $r = 4$. The face is clearly recognizable and the effects of noise are no longer visible. Likewise in the case of upsampling by a factor of $r = 2$, we now experiment with the case when the input is very noisy. The noise level is increased to $\sigma = 0.5$. There is hardly anything visible either in the input image or in the bicubic interpolated image shown in Figures 7.12(a, b). The result of PCA-based reconstruction is shown in Figure 7.12(c). We now notice a considerable distortion in the reconstructed image compared to what we obtained for $r = 2$ in Figure 7.10(c). The nose and the lips are poorly distorted. The strands of hair over the right eye is now missing! Although the reconstructed image does look like a face image, it is hardly of any consolation as the training data consists of face images only. Probably a face recognition engine would fail to recognize the reconstructed image. Hence we may conclude that if the noise level in the input image is very high and the magnification factor r is also large, the PCA-based reconstruction method will fail.

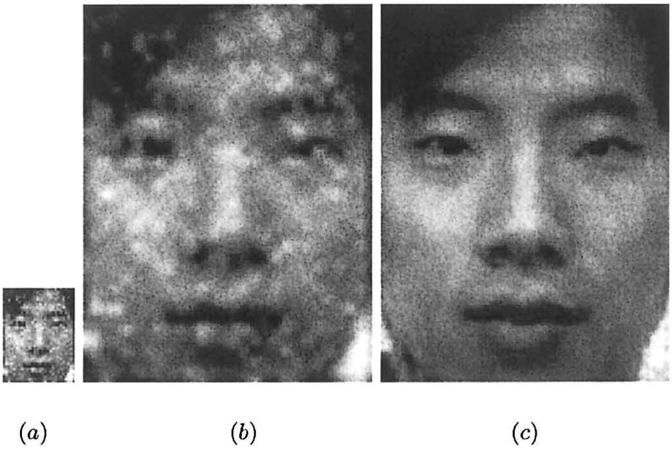


Fig. 7.11. (a) A low resolution noisy observation ($\sigma = 0.1$), (b) bicubic interpolated image, and (c) high resolution PCA-based restoration with $r = 4$.

In the next experiment, we do not corrupt the image with random noise. But the low resolution image was convolved with a 7 × 7 pixels Gaussian mask with $\sigma = 1$ to simulate a blurred observation (see 7.13(a)). The image, when upsampled by a factor of $r = 4$ using bicubic interpolation, shows a large amount of blur in Figure 7.13(b). In Figure 7.13(c) we show the result of corresponding PCA-based reconstruction. There is definitely some distortion in the reconstructed image near the lips and near the left eye. However, the face is still quite recognizable. All these experiments substantiate the claim that a high resolution database can, indeed, be used for super-resolving a low resolution blurred and noisy observation.

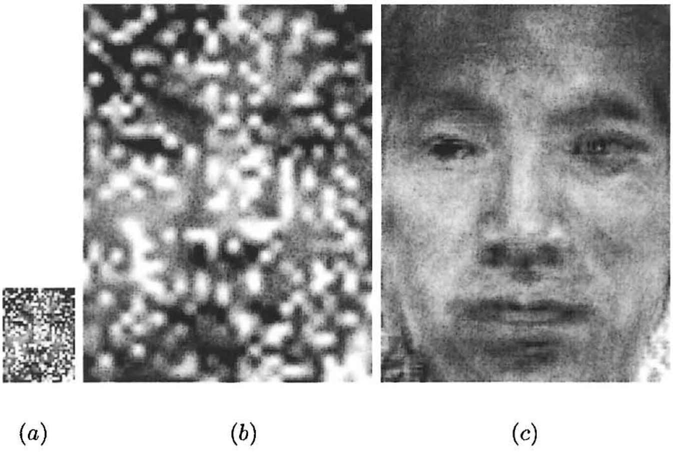


Fig. 7.12. (a) A low resolution very noisy observation ($\sigma = 0.5$), (b) bicubic interpolated image, and (c) high resolution PCA-based restoration with $r = 4$.

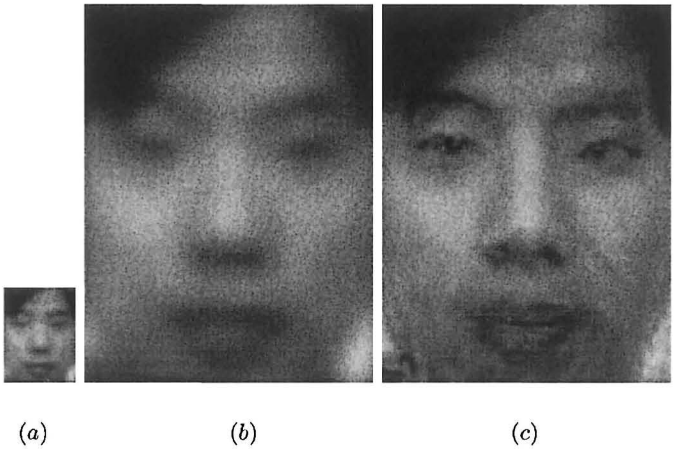


Fig. 7.13. (a) A low resolution image blurred with a Gaussian mask with a standard deviation of 1.0 serves as an observation. (b) Bicubic interpolated image, and (c) result of a high resolution PCA-based restoration with $r = 4$.

In the previous set of experiments shown in Figures 7.9-7.12, a variant of the input image was a part of the training database. We now show the results when the input image was not a part of the high resolution training database. Figures 7.14(a, b) show the noise corrupted low resolution observation and its bicubic interpolation, respectively. Figure 7.14(c) shows the results of high resolution PCA-based reconstruction

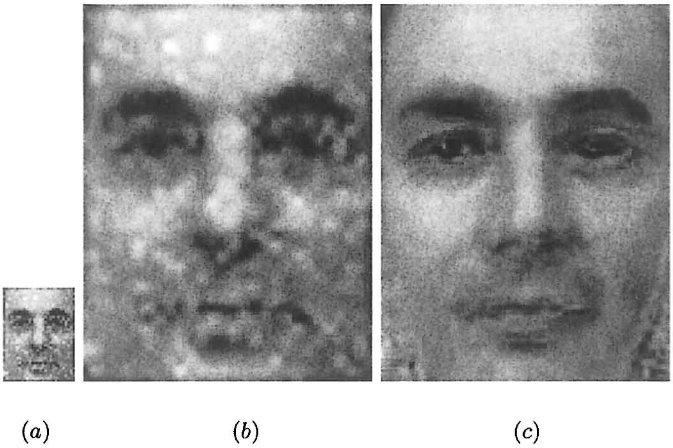


Fig. 7.14. (a) A low resolution image, without any corresponding high resolution image in the database is corrupted by noise with $\sigma = 0.1$. (b) Bicubic interpolated image, and (c) high resolution PCA-based restoration with $r = 4$.

for the upsampling factor of $r = 4$. The quality of reconstruction does appear to be quite good.

In the last experiment on face images we combine the use of a high resolution database for PCA with the generalized interpolation through the eigen-images. Here the input image is at a lower resolution than the training images. The output image is also at a higher resolution than the training images. This is obtained by using the generalized interpolation of the corresponding high resolution eigen-faces. In effect, we have upsamplers $r = r_1$ and $r = r_2$ at both the input and the output ends, respectively.

In Figure 7.15(a) a low resolution observation of size 41×48 pixels is shown. The database training images were of dimension 82×96 pixels. Figure 7.15(b) shows the bicubic interpolated output for a zoom factor of $r_1 r_2 = 8$. The low resolution input is first bicubic interpolated by a factor of $r_1 = 2$ and then super-resolved by a factor of $r_2 = 4$ using the proposed approach and the corresponding result is shown in 7.15(c). As expected, the super-resolved image is less blurred than the bicubic result.

Before we end this section on experimental results, we show some more results on the usage of a high resolution training data for fingerprint images. The input images Figures 7.16(a) and 7.17(a) show two low resolution observations at two different levels in the resolution

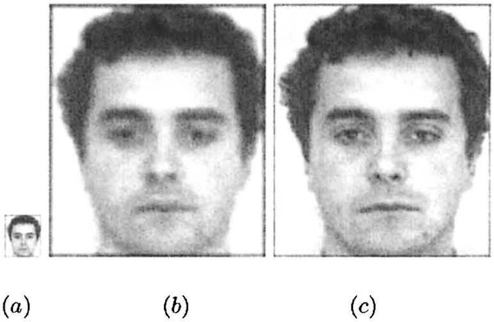


Fig. 7.15. (a) A low resolution observation of different image size 41 × 48 pixels, (b) bicubic interpolated image with $r = 8$ and, (c) PCA-based reconstruction.

pyramid. The images have been corrupted with additive white Gaussian noise with $\sigma = 0.1$. In Figures 7.16(b and c), we compare the performance of the bicubic interpolation with that of the PCA-based method for an upsampling factor of $r = 2$. Figures 7.17(b and c) show the same results for $r = 4$. We can clearly observe an improvement in the picture quality when the PCA-based reconstruction is used.

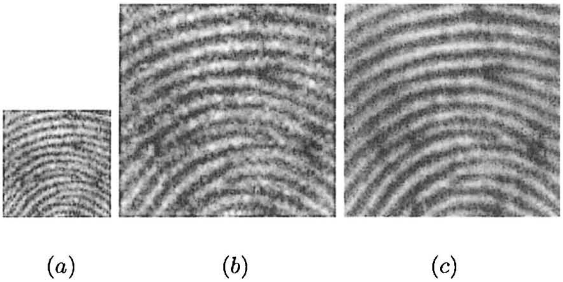


Fig. 7.16. (a) A low resolution noisy fingerprint observation ($\sigma = 0.1$), (b) bicubic interpolated image, and (c) high resolution PCA-based restoration with $r = 2$.

7.7 Conclusions

We have described a method for super-resolution restoration of images of a particular class of object using a PCA-based generalized interpolation technique. The low resolution eigen-images obtained from PCA decomposition are interpolated and transformed into an orthonormal

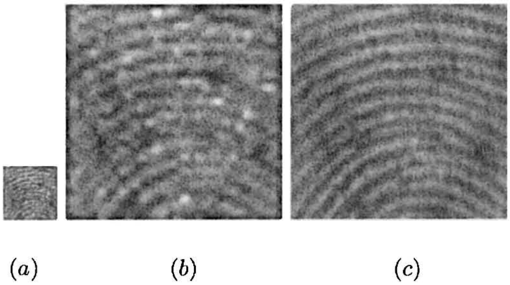


Fig. 7.17. (a) A low resolution very noisy fingerprint observation ($\sigma = 0.1$), (b) bicubic interpolated image, and (c) high resolution PCA-based restoration with $r = 4$.

basis to reconstruct the super-resolved image. The results obtained for both face and fingerprint images show far better perceptual as well as quantifiable improvements over conventional interpolation techniques. The proposed method is useful when multiple observations of the input are not available and one must make the best use of a poor quality single observation to enhance its resolution.

We have also shown the usefulness of having a high resolution training dataset instead of a low resolution dataset, when there is no need to perform a generalized interpolation of the eigen-images. However, the reconstruction process breaks down if the low resolution observations are quite noisy and one requires an upsampling factor of 2 or 3. The use of the observation at the same resolution as the training data and then the use of a subsequent generalized interpolation appears to be a more robust technique.

The proposed method cannot be classified under a general purpose super-resolution technique as the scope of applicability is very much restricted to images of a specific class of objects. For example, we cannot use it for an outdoor scene. However, we envisage that the method may be found quite suitable for biometric authentification or recognition purposes.

8

Use of Zoom Cue

In chapter 6 on wavelet-based learning for image super-resolution it was mentioned that the concept falls apart whenever one tries to adjust the zoom setting of the camera. Any change in zoom changes the effective resolution of the scene being imaged and hence the existing set of training images become redundant. However, we do acquire new and additional information about the scene while zooming. For a continuously zooming camera, we do get an enhanced resolution. However, the field of view being gradually smaller and smaller, one can see only a part of the entire scene. This is a typical dilemma faced by people working in the area of remote sensing with satellite imagery. By going into higher resolutions, say $1m$ resolution, we loose much of the area coverage that one could have had with a low resolution camera, say $5.8m$ resolution. Spatial details and the area coverage are both important.

In order to have a high spatial resolution and wider area coverage, a possible solution is to use image mosaicing [195]. However, one requires to have a complete set of high resolution observations of the entire landmass to construct the mosaic. Quite often, we may not have access to high resolution observations for all parts of the scene. We may have access with varying resolutions at different points in the scene. Is there any way we can reconstruct a high resolution description of the entire scene from observations at varying resolutions? This has been the primary motivation behind developing the contents of this chapter.

8.1 Introduction

When one captures an image with different zoom settings the amount of aliasing differs with zooming. This is because, with different zoom settings, the least zoomed entire area of the scene is represented by a very limited number of pixels, *i.e.*, it is sampled with a very low sampling rate and the most zoomed scene with a higher sampling frequency. Therefore, the larger the scene, the lower will be the resolution with more aliasing effect. By varying the zoom level, one observes the scene at different levels of aliasing. Thus one can use zoom as a cue for generating high resolution images at the lesser zoomed area of a scene. As mentioned earlier, this would help us in representing a larger landmass with a high spatial details for remote sensing applications. An interesting application of super-resolution using zoom as a cue is also in remote sensing for fusion of remotely sensed images where it is often required to construct both high spectral and high spatial resolution multi-spectral (MS) images using the high spatial resolution panchromatic (PAN) image of the same geographical area. Due to the technological limitations the multi-spectral images are generally acquired with larger instantaneous field of view than the panchromatic image leading to a low spatial resolution for these images. However, the fusion of MS and PAN images can lead to a high spatial resolution for MS images and thus results in a better analysis of remotely sensed images in terms of understanding of the observed terrain. Many researchers have tackled the fusion problem [196, 197, 198, 199, 200], but these are mostly based on point statistics and require accurate registration among PAN and MS images. The contents developed in this chapter could be of help in multiresolution fusion. Another interesting application is in developing a high quality immersive viewing system where the zoom amount can be changed while making the virtual-walk through without compromising in the spatial resolution.

The objective of super-resolution imaging is to undo the distortions introduced in an image due to undersampling, loss of high frequency information due to sensor blur or out-of-focus optical blurring. In this chapter, we discuss a technique for super-resolution imaging of a scene from observations at different zoom levels, where there is loss of information due to undersampling during the zooming process. Given a sequence of images with different zoom factors of a static scene, our problem is to obtain a picture of the entire scene at a resolution corresponding to the most zoomed image in the scene. We not only obtain the super-resolved image for known integer zoom factors, but also for unknown arbitrary zoom factors. After a lapse of two chapters in between, we again model the super-resolution image as a Markov random field (MRF) and a maximum *a posteriori* (MAP) estimation method is used to derive a cost function which is then solved to recover the high resolution field. The entire observation is assumed to conform to the same MRF, but is viewed at the different resolution pyramid. Since there is no relative motion between the scene and the camera, we do away with the correspondence problem.

As discussed in the chapter on literature survey, researchers traditionally use the motion cue to super-resolve the image. However, the methods based on the motion cue cannot handle observations at varying levels of spatial resolution. It assumes that all the frames are captured at the same spatial resolution. Previous research work with zoom as a cue to solve computer vision problems include determination of depth [12, 13, 14], minimization of view degeneracies [15], and zoom tracking [16]. Lavest and his co-authors [12, 13] develop a depth reconstruction method for a static object and camera using the thick lens model. They conclude that simpler pinhole model can be used (instead of a more accurate thick lens model) only if the effective change of focal point during zooming is considered, enabling 3D information to be inferred by triangulation. Ma and Olsen [14] develop two depth from zooming methods for a pinhole camera model applicable to static objects using both optical flow and feature matching. They conclude that feature matching in a zoom sequence is more accurate and reliable, because it is less sensitive to noise than the optical flow analysis by presenting results for synthetic models. In [15] Wilkes *et al.* use zoom to reduce the probability of view degeneracies. Degenerate views occupy a significant fraction of the viewing sphere surrounding an object. Furthermore, these view degeneracies cannot be detected from a single viewpoint. Wilkies *et al.* choose a focal length that reduces the probability of view degeneracies, improving the performance of systems designed to recognize objects from a single, arbitrary view point. Zoom tracking refers to the problem of continuous adjustment of camera focal length in order to keep a constant sized image of an object moving along the camera's optical axis. Two methods for performing zoom tracking presented in [16] are based on the optical flow and the use of depth information from an autofocus camera's range sensor. The authors show that zoom tracking can be used to reconstruct the depth map of the tracked object.

We demonstrate in this chapter that even the super-resolution problem can be solved using zoom as an effective cue by using a simple MAP-MRF formulation. The basic problem that we address can then be defined as follows: One continuously zooms into a scene while capturing its images. The most zoomed-in observation has the highest spatial resolution. We are interested in generating an image of the entire scene (as observed by the most wide angle or the least zoomed view) at the same resolution as the most zoomed-in observation.

We use observations at arbitrary levels of resolution (scale) and these scale factors are estimated while super-resolving the entire scene. One may observe that the approach generates a super-resolved image of the entire scene although only a part of the observed scene has multiple observations. In effect what we do is as follows. If the wide angle view corresponds to a field of view of α^o, and the most zoomed view corresponds to a field of view of β^o (where $\alpha > \beta$), we generate a picture of the α^o field of view at a spatial resolution comparable to β^o field of view by learning certain image statistics from the most zoomed view. The details of the method are presented in this chapter.

8.2 Low Resolution Image Formation Model

In chapter 3 we have discussed the low resolution image formation model. We continue to use the same model in this chapter. However, because of the zooming process, it needs some elaboration. The zooming based super-resolution problem is cast again in a restoration framework. There are K observed images $\{Y_i\}_{i=1}^{K}$ each captured with different zoom settings and of size $M_1 \times M_2$ pixels each. Figure 8.1 illustrates the block schematic of how the low resolution observations of a scene at different zoom settings are related to the high resolution image. Here we consider that the most zoomed observed image of the scene Y_K ($K = 3$ in the figure) has the highest resolution.

A zoom lens camera system has complex optical properties and thus it is difficult to model it. As Lavest *et al.* [13] point out, the pinhole model is inadequate for a zoom lens, and a thick-lens model has to be used; however, the pinhole model can be used if the object is virtually shifted along the optical axis by the distance equal to the distance between the primary and secondary principal planes of the zoom lens. Since we capture the images with a large distance between the object and the camera and if the depth variation in the scene is not very significant compared to its distance from the lens, it is reasonable to assume that the paraxial shift about the optical axis as the zoom varies is negligible. Thus, we can make a reasonable assumption of a pinhole

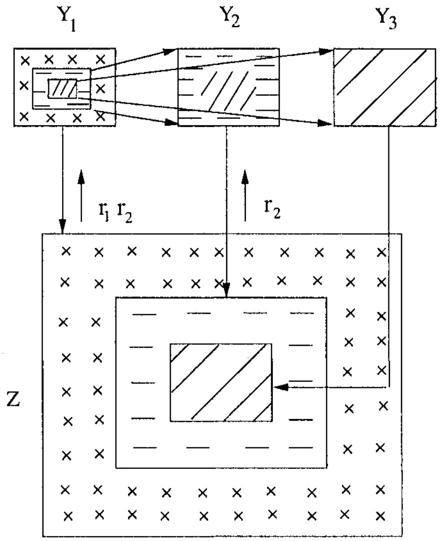


Fig. 8.1. Illustration of observations at different zoom levels: Y_1 corresponds to the least zoomed and Y_3 to the most zoomed images. Here Z is the high resolution image of the same scene. The image Y_3 does not need upsampling while filling up the central region in Z. But Y_1 and Y_2 need appropriate amount of upsampling.

model and neglect the depth related perspective distortion due to the thicklens behavior. We are also assuming that there is no rotation about the optical axis between the observed images taken at different zooms. However we do allow lateral shift of the optical center as explained in section 8.3.1.

Since different zoom settings give rise to different resolutions, the least zoomed scene corresponding to entire scene needs to be upsampled to the size of $(r_1 r_2 \cdots r_{K-1}) \times (M_1 \times M_2) = (N_1 \times N_2)$ pixels, where $r_1, r_2, \cdots, r_{K-1}$ are the zoom factors between observed images of the scene $Y_1 Y_2, Y_2 Y_3, \cdots, Y_{(K-1)} Y_K$, respectively. Given Y_K, the remaining $(K-1)$ observed images are then modeled as decimated and noisy versions of this single high resolution image of the appropriate region in the scene. With this the most zoomed observed image will have no decimation. If $\mathbf{y}_m$ is the $M_1 M_2 \times 1$ lexicographically ordered vector containing pixels from differently zoomed images Y_m, the observed images can be modeled as (refer to Figure 8.2 for illustration)

$$\mathbf{y}_m = D_m R_m \mathbf{z}_{\alpha_m} + \mathbf{n}_m, \qquad m = 1, \cdots, K \qquad (8.1)$$

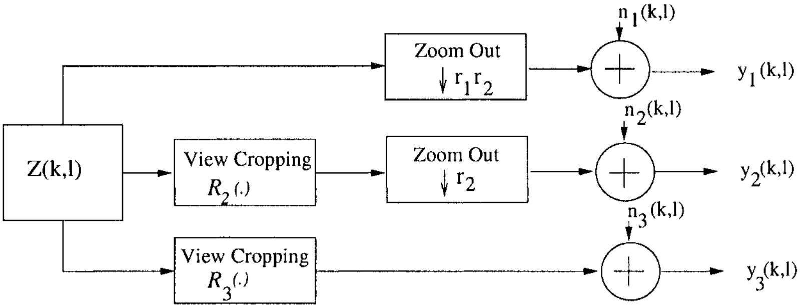


Fig. 8.2. Low resolution image formation model is illustrated for three different zoom levels. View cropping block just crops relevant part of the high resolution image Z as the field of view shrinks with zooming along with a possible lateral shift.

where $z_{\alpha_m}(x,y) = z(x - \alpha_{m_x}, y - \alpha_{m_y})$ with $\alpha_m = (\alpha_{m_x}, \alpha_{m_y})$ representing the lateral shift of the optical center due to zooming by the lens system for the m^{th} observation. The matrix D is the decimation matrix which takes care of aliasing present while zooming. The subscript m in D denotes that the amount of decimation depends on the amount of zoom for the m^{th} observation. For an integer zoom factor of r, decimation matrix D has the form as given in Eq. (3.12) and is repeated here for recapitulation (note that a re-ordering of the elements of $\mathbf{z}$ is needed to get D in this form).

$$D = \frac{1}{r^2}\begin{bmatrix} 11\ldots1 & & & 0 \\ & 11\ldots1 & & \\ & & \ddots & \\ 0 & & & 11\ldots1 \end{bmatrix}.$$

However, we do not restrict ourselves to integer zoom factors alone as any practical implementation using an optical zoom mechanism would involve an arbitrary value of r. Here R_m is a cropping operator on $\mathbf{z}$ needed to handle the shrinkage of view angle during the zooming process. The cropping operator is similar to a characteristic function that crops out a $\lfloor r_1 r_2 \cdots r_{m-1} N_1 \rfloor \times \lfloor r_1 r_2 \cdots r_{m-1} N_2 \rfloor$ pixel area from the high resolution image $\mathbf{z}$ at an appropriate position for the m^{th} observation. In case there is no lateral shift while zooming along the optical axis, R_m would involve cropping from the center. In Eq. (8.1), K is the number of observations, $\mathbf{n}_m$ is the $M_1 M_2 \times 1$ noise vector. We assume the sensor noise to be zero mean Gaussian i.i.d, and hence the multivariate noise probability density is given by

$$P(\mathbf{n}_m) = \frac{1}{(2\pi{\sigma_n}^2)^{\frac{M_1M_2}{2}}}\exp\left\{-\frac{1}{2{\sigma_n}^2}\mathbf{n}_m^T\mathbf{n}_m\right\}, \tag{8.2}$$

where ${\sigma_n}^2$ denotes the variance of noise process. Our problem now reduces to estimating $\mathbf{z}$ given $\mathbf{y}_m$'s, which is clearly an ill-posed, inverse problem requiring suitable regularization.

8.3 MAP Estimation

Since the restoration of the high resolution image from its differently zoomed observations is an ill-posed problem, we model the high resolution data $\mathbf{z}$ as an MRF. Earlier in chapters 3-5 we have used precisely the same model for regularization purposes. We define the corresponding energy function $V(\mathbf{z})$ exactly the way it has been defined in Eq. (3.18)

In order to use the maximum *a posteriori* (MAP) estimation technique to obtain the high resolution image $\mathbf{z}$ given the ensemble of images at different resolutions we need to obtain

$$\hat{\mathbf{z}} = \arg\ \max_{\mathbf{z}} P(\mathbf{z} \mid \mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K). \tag{8.3}$$

From Bayes' rule this can be written as

$$\hat{\mathbf{z}} = \arg\ \max_{\mathbf{z}} \frac{P(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K \mid \mathbf{z})P(\mathbf{z})}{P(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K)}, \tag{8.4}$$

where $P(\mathbf{z})$ represents the prior probability for the super-resolved image. Since the denominator is not a function of $\mathbf{z}$, it can be considered as a constant while maximizing, and hence Eq. (8.4) can be written as

$$\hat{\mathbf{z}} = \arg\ \max_{\mathbf{z}} P(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K \mid \mathbf{z})P(\mathbf{z}). \tag{8.5}$$

Taking the log of the posterior probability,

$$\hat{\mathbf{z}} = \arg\ \max_{\mathbf{z}} \left[\log P(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K \mid \mathbf{z}) + \log P(\mathbf{z})\right]. \tag{8.6}$$

Now since $\mathbf{n}_m$ are independent, we have

$$P(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K \mid \mathbf{z}) = P(\mathbf{y}_1 \mid \mathbf{z})P(\mathbf{y}_2 \mid \mathbf{z}) \cdots P(\mathbf{y}_K \mid \mathbf{z}),$$

and hence

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} \left[\sum_{m=1}^{K} \log P(\mathbf{y}_m \mid \mathbf{z}) + \log P(\mathbf{z}) \right] . \tag{8.7}$$

Using equations (8.1) and (8.2) in $P(\mathbf{y}_m \mid \mathbf{z})$ we obtain

$$P(\mathbf{y}_m \mid \mathbf{z}) = \left[\frac{1}{(2\pi{\sigma_n}^2)^{\frac{M_1 M_2}{2}}} \exp\left\{ -\frac{\|\mathbf{y}_m - D_m R_m \mathbf{z}_{\alpha_m}\|^2}{2{\sigma_n}^2} \right\} \right] . \tag{8.8}$$

Now we use the MRF prior for the high resolution image $\mathbf{z}$. Thus using Eq. (3.17) and substituting in Eq. (8.7) the final cost function is obtained as

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left[\sum_{m=1}^{K} \frac{\|\mathbf{y}_m - D_m R_m \mathbf{z}_{\alpha_m}\|^2}{2{\sigma_n}^2} + U(\mathbf{z}) \right] . \tag{8.9}$$

Here $U(\mathbf{z}) = \sum_{\mathbf{c}\in\mathcal{C}^{\mathbf{z}}} \mathbf{V}^{\mathbf{z}}_{\mathbf{c}}(\mathbf{z})$ is the energy function associated with the random field $\mathbf{z}$ and $V_c^z(\mathbf{z})$ the potential function associated with a given clique. The above cost function is convex since the binary line fields are not included and is minimized using the gradient descent technique. The initial estimate $\mathbf{z}^{(0)}$ is obtained as follows. Pixels in zero-order hold (or bilinear interpolation) of the least zoomed observed image corresponding to the entire scene is replaced successively at appropriate places with zero-order hold (or bilinear interpolation) of the other observed images with increasing zoom factors. Finally the most zoomed observed image with highest resolution is copied at the center after accounting for the lateral shift (α_x, α_y) with no interpolation (see Figure 8.1 for illustration).

In order to preserve discontinuities we modify the cost for prior probability term as discussed in section 3.4 and use the energy function given in Eq. (3.21) that incorporates the line field. The corresponding prior term then becomes

$$U(\mathbf{z}) = \sum_{i,j} \left[\mu e_{zs} + \gamma e_{zp} \right] , \tag{8.10}$$

where μ, γ, e_{zs}, e_{zp} have the same meaning as explained in section 3.4 in chapter 3. On inclusion of binary line fields in the cost function, the gradient descent technique cannot be used since it involves a differentiation of the cost function. Hence, we minimize the cost by using simulated annealing (SA) which leads to a global minima. However, in order to provide a good initial guess and to speed up the computation, the result obtained using the gradient descent method is used as the initial estimate for simulated annealing. Although the simulated annealing optimization results in global minima, it is computationally taxing. In order to reduce the computation Bilbro *et al.* [201, 202] developed the mean field annealing (MFA) technique that approximates the SA, but is based on the deterministic relaxation as opposed to stochastic relaxation of SA. It converts the problem of minimization on $\mathbf{z}$ into the problem of minimization on its mean field. This can be done by using a standard gradient descent algorithm and thus allows a faster convergence. The computational time is greatly reduced upon using mean field annealing which leads to a near optimal solution.

8.3.1 Zoom Estimation

We now extend the method to a more realistic situation in which the successive observations vary by an unknown rational valued zoom factor. Further, considering an actual lens system for the imaging process, the numerical image center can no longer be assumed to be fixed. The zoom factor between the successive observations needs to be estimated during the process of forming an initial guess and while solving Eq. (8.9) for the proposed super-resolution algorithm. Thus D_m and α_m in Eq. (8.1) are unknown and they should be estimated from the observations themselves. We however assume that there is no rotation about the optical axis between the successive observations though we do allow a small amount of lateral shift in the optical axis. The image centers move as lens parameters such as focus or zoom are varied [203, 204]. Naturally, the accuracy of the image center estimation is an important factor first in obtaining the initial guess and then for minimizing Eq. (8.9) for super resolution purposes. Generally, the rotation of a lens system will cause a rotational drift in the position of the optical axis, while sliding action of a lens group in the process of zooming will cause a translation motion of the image center [203]. These rotational and translational shifts in the position of the optical axis cause a corresponding shift in the camera's field of view and the optical center. In variable focal length zoom lenses, the focal length is changed by moving groups of lens elements relative to one another. Typically this is done by using a translational type of mechanism on one or more internal groups. These arguments validate our assumption that there is no rotation of the optical axis in the zooming process and at the same time it stresses the necessity of accounting for the lateral shift in the image centers of the input observations obtained at different zoom settings.

We estimate the relative zoom and shift parameters between two observations by minimizing the mean squared difference between an appropriate portion of the digitally zoomed image of the wide angle view and the narrower view observation. This is illustrated in Figure 8.3. The method searches for the appropriate zoom factor and the lateral shift α that minimizes the distance. We do this by hierarchically searching for the best match by first upsampling the wide angle observation (block 'A' in Figure 8.3) and then searching for the shift that corresponds to a local minima of the cost function. The lower and upper bounds for the upsampling process need to be appropriately defined. Naturally, the efficiency of the algorithm is limited by the closeness of the bounds to the actual solution and the number of steps in which the search space is discretized. The search can be greatly enhanced by first searching for a rough estimate of the zoom factor at a coarse level and slowly approaching the exact zoom factor by redefining the lower and upper bounds as well as the finely discretized step size about the best match at the coarser level.

Let us illustrate this with an example. We initiate the search at a coarse level for discrete zoom factors (say, 1.4 to 2.3 in steps of 0.1). At this point, we need to note that the digital zooming of an image by a rational zoom factor $r = \frac{m}{n}$ (m and n are integers) is obtained by upsampling the lattice by m and then downsampling it by a factor n. We have carried out this using MATLAB routines. We then redefine the increment and the bounds that correspond to the two of the nearest estimates of the zoom factor (say, 1.6 to 1.7 in steps of 0.01). This procedure is continued till the zoom factor is estimated to a desired level of accuracy. Naturally, a greater accuracy in the zoom estimation would result in a refined initial guess. However, it has been observed that an accuracy upto two digits in the decimal place of the zoom factor would be sufficient for accurately aligning the two observations. Further, the efficiency of the algorithm is greatly enhanced by constraining the search for the lateral shift in the image center P of the wide angle view image to a small neighborhood of the image center P' of the narrower view image as the lateral shift in the optical axis in the zooming process is usually very small (about 2 to 3 pixels).

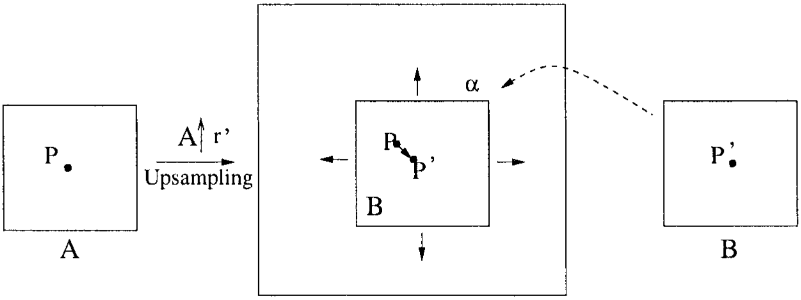


Fig. 8.3. Illustration of zoom and alignment estimation. 'A' is the wide angle view and 'B' is the narrower angle view. A is upsampled by a factor of $r^{'}$ and B is matched to the upsampled A. Find the best value of $r^{'}$ and the shift α that offer the best match.

8.4 Experimental Results

Let us now see how well we can super-resolve the low resolution image from observations at different zooms through examples on real data. We start by presenting the results corresponding to the known integer zoom factors and slowly lead to a more realistic situation in which the zoom factors also need to be estimated along with the lateral shifts in the optical center.

8.4.1 Use of Known Zoom Factors

For this experiment, we have considered three low resolution observations with the zoom factor of $r = 2$ between images Y_1Y_2 and $r = 4$ between Y_1Y_3. Thus the zoom factor between Y_2Y_3 is also $r = 2$. Figures 8.4 (a-c) show the input (observed) images of a house Y_1 ,Y_2 ,Y_3 each of size 72×96 with a zoom factor of $r = 2$ between images (a) and (b) and also a factor of $r = 2$ between (b) and (c).

The automatic gain control (AGC) in the camera automatically sets the camera gain in accordance with the amount of light in the pictured area and the level of zooming. Since we are capturing regions with different zoom settings, the AGC of the camera yields different average brightness for differently zoomed observations. Hence in order to compensate for the AGC effect, we use the mean correction to maintain the average brightness of the captured images approximately the same. This is done for the observation Y_2 by subtracting its mean from each pixel and adding the mean due to its corresponding portion in Y_1 (refer

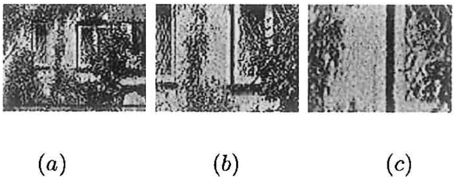
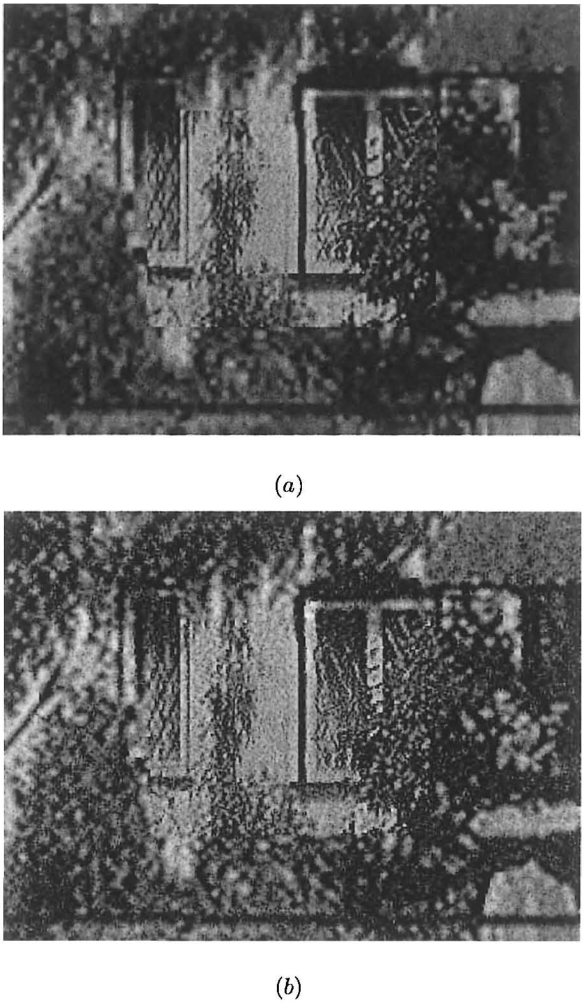


Fig. 8.4. Observed images of a house captured with three different zoom settings.

to Figure 8.1). Similarly for the observation Y_3 we subtract its mean and add the mean of its portion in Y_1. We have used mean corrected images in all our experiments.

Figure 8.5(a) shows the zoomed image of the house of size 288 × 384 pixels obtained by bilinear interpolation of the least zoomed observed image Y_1 of size 72 × 96 pixels with an 144 × 192 pixel sized bilinearly interpolated Y_2 image replacing that part of the interpolated least zoomed observed image, and the 72 × 96 sized most zoomed observed image replacing those corresponding pixels in interpolated Y_1. The corresponding super-resolved image is shown in Figure 8.5(b). Comparison of the figures shows more clear details in the super-resolved image. The seam is clearly visible in Figure 8.5(a). Also the branches in the plants are more clearly distinguishable in the super-resolved image. It may be mentioned that the reconstruction at the peripheral region of the image is expected to be inferior to that at the central region as a very little information is available for the purpose of super-resolution. This confirms to the observation made in [205] that the restoration error increases with an increase in the amount of blurring. It has been experimentally found that the best results are obtained with the parameters $\mu = 0.0095$, $\gamma = 150$, $\theta = 25$, the decrement factor for temperature in the annealing schedule $\delta = 0.999$, and the initial temperature $T_0 = 3.15$ for the optimization purposes.

The previous example was highly undersampled. Next we consider a case where the observed images shown in Figure 8.6(a-c) have a smooth intensity variation. Figure 8.7(a) shows the successive bilinearly interpolated girl image 'Nidhi' and 8.7(b) displays the corresponding super-resolved image. Again notice the seam and also the blockiness on the edges of hair-band in Figure 8.7(a). Note that the central part of both the images in Figure 8.7 are identical. The available highest resolution observation has been copied at these places. We expect to



(a)

(b)

Fig. 8.5. (a) Zoomed house image formed by successive bilinear expansion. (b) The super-resolved house image.

see improvement as we move away from the center. We do notice some improvement near the lips and the right eye of Nidhi in Figure 8.7(b).

The super-resolved image corresponding to the entire scene Y_1 consists of super-resolved image due to Y_2 which in itself has the actual observation Y_3. In this case Y_1, Y_2 and Y_3 correspond to Figure 8.6(a), (b), and (c) respectively. For this experiment, we found discontinuities

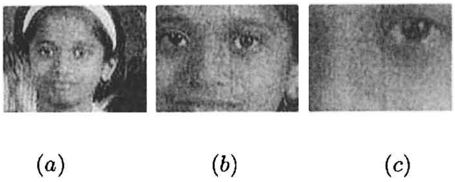
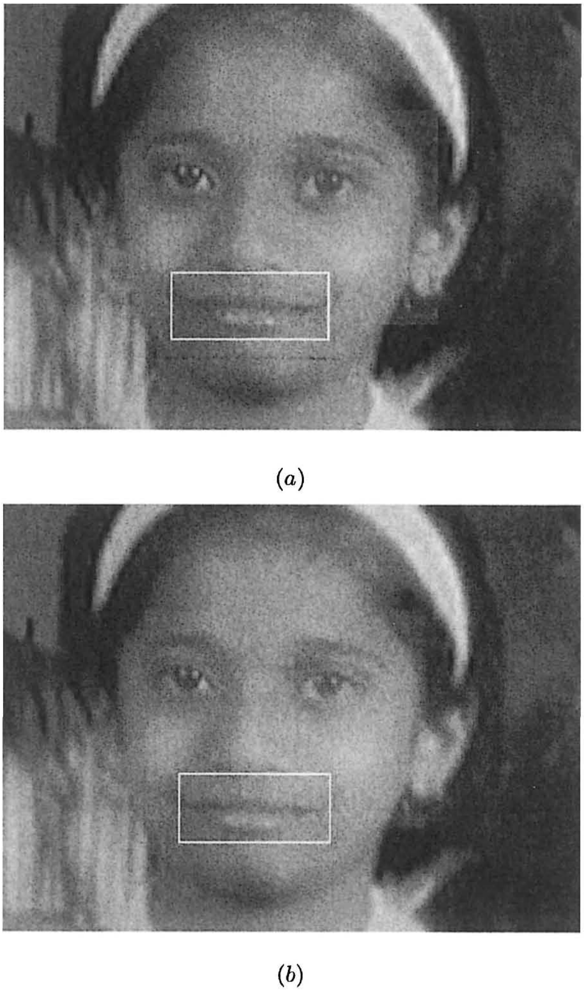


Fig. 8.6. Observed images of the girl Nidhi captured with three different zoom settings.

were better preserved by considering three different threshold values θ for the three super-resolution regions, and the parameters used were $\theta_{Y_1} = 70$ for the super-resolution region of Y_1 only, $\theta_{Y_2} = 10$ for the super-resolution region of Y_2 only, and $\theta_{Y_3} = 5$, for the super-resolved region Y_3. Here $\theta_1 = \theta_2 \overset{\Delta}{=} \theta$ defines the threshold for detecting the presence of an edge as defined in Eqs. (3.19 and 3.20) in chapter 3. The weightage for smoothness term μ was chosen as 0.08 and the constant γ was selected as 10. The justification for selecting different thresholds for different regions is that one has relatively less information about the peripheral regions, and hence the restoration tends to be smoother at these regions. The choice of lower threshold values tries to prevent oversmoothing. The values of δ and T_o were kept the same as it was in the previous experiment.

We also experiment to find out what would happen if the line fields are dropped so that a gradient descent method can be used resulting in a much faster computation. The super-resolved Nidhi image using the gradient descent optimization scheme is shown in Figure 8.8. A step size of 0.005 was chosen for this experiment. As expected the super-resolved image looks smooth. However, it does not affect too much as the imaged object does not have too much of high frequency details.

We show results of experimentation for another scene having arbitrary textures (with significant amount of high frequency content). Figures 8.9(a-c) show the observations taken at different zooms. The scene is highly undersampled and severely aliased. The results obtained for this case are shown in Figures 8.10(a) and 8.10(b). It can be clearly seen that the estimated super-resolved image appears sharper. The trees in the background definitely do. Also, the seam present in the interpolated image 8.10(a) is absent in the super-resolved image. The bush in front of the house also appears well in 8.10(b). However, do note that there is a localization error in both the images while combin-



(a)

(b)

Fig. 8.7. (a) Zoomed Nidhi image using successive bilinear expansion, (b) super-resolved Nidhi image.

ing the three views. This is due to the fact that the relative zoom and the alignment were assumed to be known in this experiment. Clearly, the image alignment was not correct and it will be shown in subsequent section that the zoom and alignment estimation takes care of the above problem.

The initial estimates for the high resolution image in all the experiments discussed in this subsection were chosen to be the output

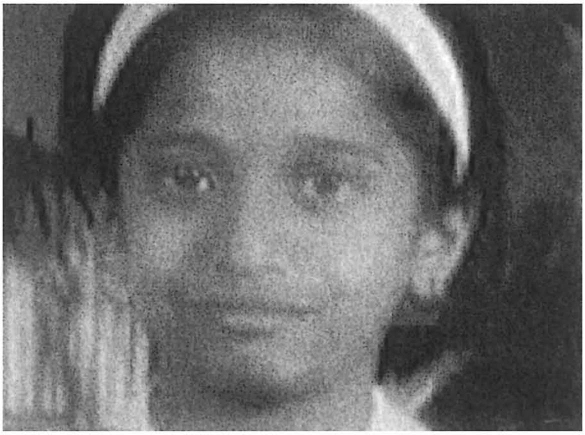
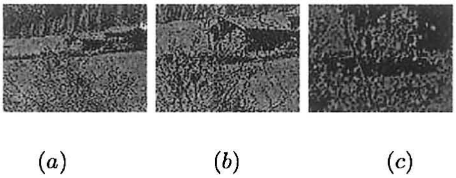


Fig. 8.8. Super-resolved Nidhi image using the gradient descent optimization.

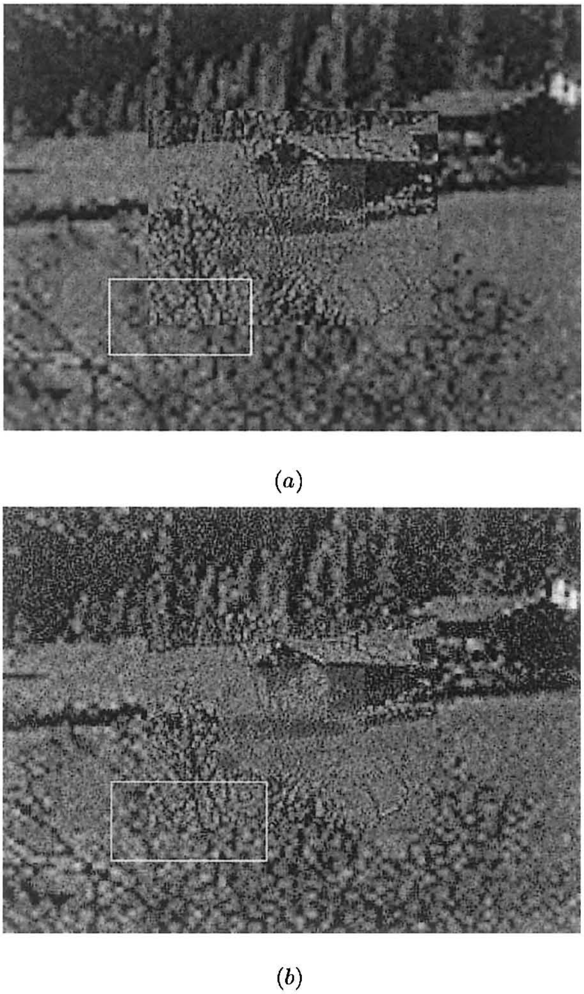
obtained by using the gradient descent technique to speed up the computation. Comparison of these results with the results of successive interpolation demonstrates the usefulness of the proposed approach. However, the restoration tends to be a bit smooth near the periphery. This phenomena is expected as we have used only three observations and the peripheral region has been upsampled by a factor of 4 × 4. The effect of over smoothing is quite visible in the house image where the high frequency region corresponding to the plants appears to be smoothened. However, the picture of Nidhi (Figure 8.7(b)) when super-resolved, does not show this effect prominently. The result does appear to be visually more pleasant. The simulated annealing optimization algorithm used here is slow and in order to decrease the computation time, we have also implemented the mean-field annealing optimization (MFA). The recovered super-resolved Nidhi and house images using MFA are shown in Figures 8.11(a) and (b) respectively. These results compare quite favorably with the results given in Figures 8.7(b) and 8.5(b) despite a substantial reduction in computation time.

All the above results have been obtained with input observations at known, integer zoom factors. It is also assumed that the lateral shift in the center of the image during the zoom process is known. Now we relax these assumptions.



(a) (b) (c)

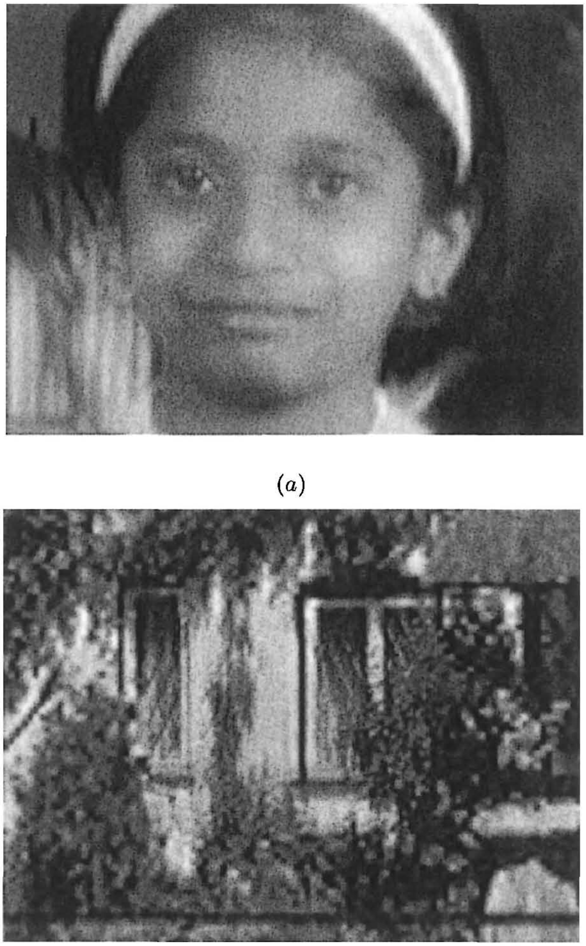
Fig. 8.9. Observed images of a scene captured again with three different zoom settings.



(a)

(b)

Fig. 8.10. (a) Zoomed scene image formed by successive bilinear expansion, (b) super-resolved scene image.



(a)

(b)

Fig. 8.11. Super-resolved (a) Nidhi, and (b) house image using mean field optimization.

8.4.2 Experiments with Unknown Zoom

Now we present the results of the more general case in which the zoom factors also need to be estimated. We do this using the method described in the section 8.3.1. Figures 8.12 (a-c) show the input images Y_1, Y_2, Y_3 each of size 128×128 pixels. We again account for the variation in the average intensity due to the automatic gain control feature

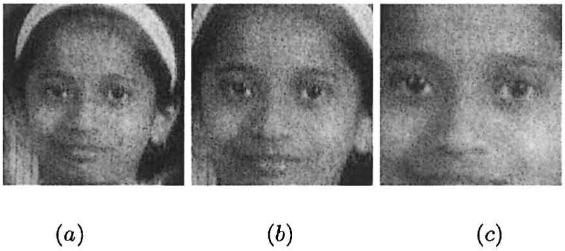


Fig. 8.12. Observed images of Nidhi captured with three different unknown zoom settings.

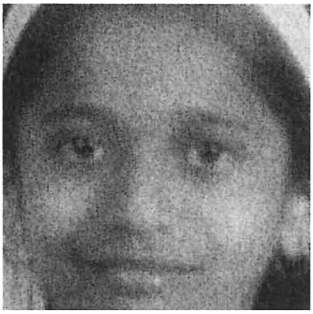


Fig. 8.13. Image obtained by aligning images Y_2 and Y_3 given in Figure 8.12.

of the camera using the mean correction in order to maintain the same average brightness of the captured images. The method described in section 8.3.1 asks for an appropriate interpolation technique (digital zooming) for the process of aligning the images. It has been observed that a nearest neighbor interpolation technique would perform quite effectively for accurately estimating the zoom factor. However, the estimate of the lateral shift obtained in the process may result in an error upto a few pixels due to the inherent repetitive nature of the nearest neighbor interpolation algorithm. The use of a bilinear or bicubic interpolation technique in the proposed alignment algorithm would estimate accurately not only the zoom factor but also the lateral shift that the image center undergoes in the process. Considering the accuracy of the results and the computational efficiency, the bilinear interpolation technique is found to be quite appropriate for zoom estimation.

Using a bilinear interpolation in the proposed algorithm, a zoom factor of $r = 1.72$ and a lateral shift of $\alpha = (3, -2)$ pixels were estimated between the observations 8.12 ((b) and (c)). These two observations are aligned by first interpolating (b) by the factor obtained and by replacing the appropriate section of this interpolated image by the observation (c). The image shown in the Figure 8.13 is obtained by aligning (c) on (b) using the estimated zoom factor and the lateral shift in the image center. We observe that the wide angle view (image (b)) has been aligned quite accurately with the zoomed observation (c). There is no apparent distortion in Figure 8.13 due to the merging process as it was evident in Figure 8.10. We follow the same procedure and align the image obtained in the above process and the other observation (a). A zoom factor of 2.14 and a lateral shift of $(3, -2)$ pixels were estimated in this case. The aligned image using bilinear interpolation is shown in Figure (8.14(a)). We again observe that the input images have been accurately aligned.

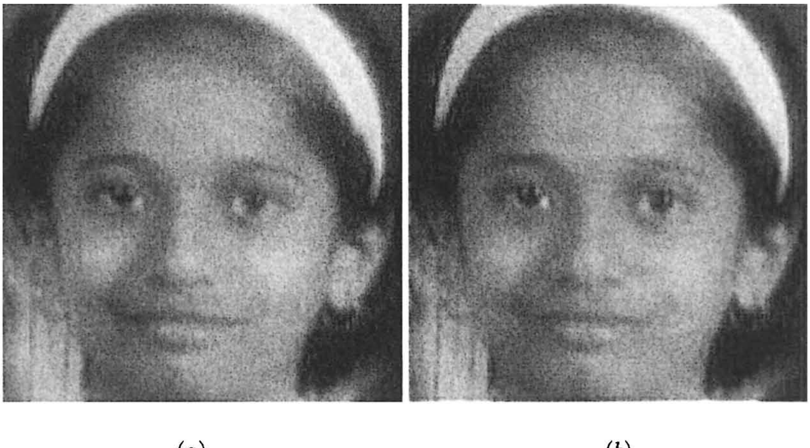


Fig. 8.14. (a) Zoomed Nidhi image formed by using successive bilinear expansion, (b) super-resolved image using the gradient descent method when the zoom factors are not known.

The image as obtained above is used as an initial guess of the high resolution intensity field Z. Starting from this initial estimate, we try to reduce the cost given in Eq. (8.9) using the gradient descent algorithm and obtain the super-resolved image. We use the MATLAB routine '*imresize*' for decimation process. As before, the choice of decimation in the gradient descent algorithm has to be made considering the computational burden and the quality of the output. A bicubic decimation would lead us to a better quality of the result and a nearest neighborhood algorithm would result in a blockiness effect in the image. It has been observed that bilinear interpolation would be the best choice for decimation considering both the computational time and the quality of the result obtained. The step size for the gradient descent algorithm needs to be adjusted appropriately to obtain a good quality output. It has been experimentally found that the best results are obtained with a step size of 0.01 for the case of Nidhi image. The resultant super-resolved image is shown in the Figure 8.14(b). We do observe a resolution enhancement in this case. A few areas of interest in the image would be the region around the eyes and nostrils in the image where the zoom-based algorithm outperforms the traditional bilinear interpolation technique result of which is given in Figure 8.14(a). Also note that since the performance of the gradient descent algorithm was quite satisfactory, no further attempt is made to improve the accuracy through the use of line fields.

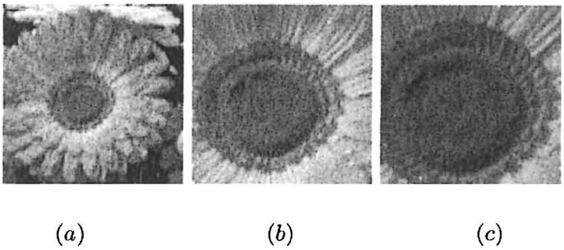
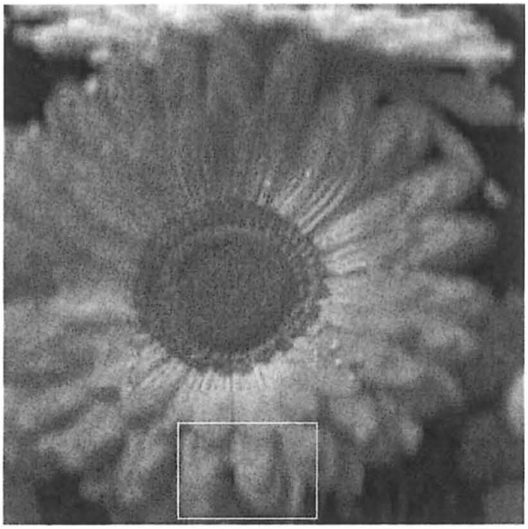
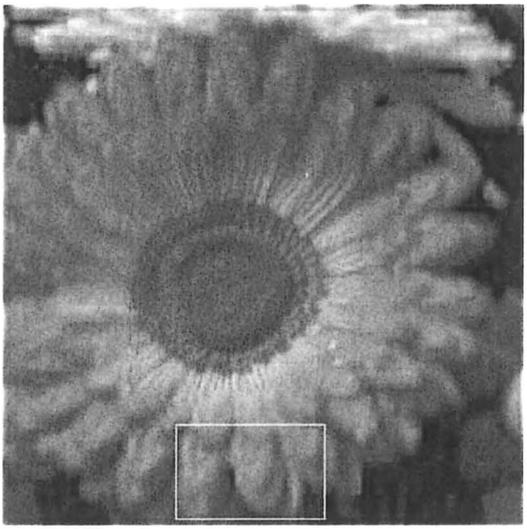


Fig. 8.15. Observed images of a flower captured with three different (unknown) zoom settings.

We now present the results of this method for scenes having a significant amount of high frequency content. Figures 8.15 (a-c) show the observations taken at different zooms. Since the zoom levels were unknown, they were first estimated using the hierarchical cross-correlation technique across the scale, and were found to be $r_1 = 1.33$ between the observations (a) and (b) and $r_1 r_2 = 2.89$ between the observations (a) and (c). A lateral shift of $(3, -2)$ and $(6, -10)$ pixels in the optical centers, respectively, for the above two cases, were detected. We can observe that the super-resolved image (see Figure 8.16(b)) appears sharper as compared to the one obtained by using the bilinear expansion (see Figure 8.16(a)). A few areas of interest where such an



(*a*)



(*b*)

Fig. 8.16. (a) Zoomed flower image formed by using successive bilinear expansion, (b) super-resolved image using the gradient descent method.

enhancement can be observed would be the portion of the image containing a group of petals towards the periphery. One such region has been highlighted by a rectangular box in the Figure. These results once again demonstrate the efficacy of the super resolution algorithm for a wide range of data sets even for the case of fractional zoom. A step size of 0.01 is used in the gradient descent algorithm. We again note as in the previous case, since the performance of the gradient descent was quite satisfactory, no further attempt is made to improve the accuracy through the use of discontinuity preserving line fields.

8.5 Learning of Priors from Zoomed Observations

In the previous section we have selected the MRF model parameters in an ad-hoc fashion. This increases the computational burden, as one needs to adjust the parameters on trial and error basis. This is typically a problem faced in such type of MAP-MRF formulation. What are the correct MRF parameters? We observe that we have a part of the super-resolved image already available in the form of most the zoomed observation. It will be ideal to recover the parameters from this most zoomed observation and use it while super-resolving the entire scene. This is exactly what we discuss in the rest of this chapter.

The super-resolution restoration problem is known to be ill-posed. Thus obtaining a desirable solution requires a reasonable assumption about the nature of the true image. Once the prior model for the true image is chosen, the solution obtained depends on the correctness of model parameters. A proper choice of model parameters leads to a better solution and alleviates the problems due to ill-posedness. Till now the prior has been modeled as an MRF and the model parameters were selected on an adhoc basis for minimizing the cost function by adjusting the parameters on a trial and error basis until a better solution is obtained. A more practical and challenging situation would be one in which these model parameters are learnt from the given observations themselves. We now concentrate on the problem of learning the priors for super-resolution imaging of a scene from observations at different camera zooms. We model the high resolution image as a *homogeneous* Markov random field. Through the most zoomed observation, we get to view a part of the high resolution field. Hence we learn the corresponding field parameters for the model from this high resolution observation assuming a homogeneity of the scene and this prior is later used to super-resolve the rest of the scene captured at a lower resolution.

As discussed in chapter 2, a few researchers have also attempted to solve the super-resolution problem by using the learning based approaches using a set of training images [97, 98, 99, 100, 101, 102]. The advantage of learning based methods is that they provide a very natural way of obtaining the characteristics of the high resolution image . An ideal learning based method should make it possible to learn very complex scenes. The problem of learning is very tricky, since it should preserve the features of the original high resolution image from a set of training images. Also, it is not obvious what features of the training set should be learnt at the high resolution. By choosing a proper feature set from training images, the quality of the results obtained can be improved. Here we use a different type of cue for parameter learning, where we make use of the given observations themselves to learn the parameters and not the training set.

The estimates of the MRF parameters are obtained using a maximum pseudo-likelihood (MPL) estimator in order to reduce the computations avoiding the computation of partition function. Although we use the MAP-MRF approach for super-resolution, our work is fundamentally different from those of [50, 120] in the sense that we learn the field parameters on the fly while the previous works assume them to be known. Further, all previous methods use observations at the same resolution.

Since we are learning the MRF parameters from the given data, let us briefly review the status of research in this area. In [206] authors use Metropolis-Hastings algorithm to estimate the MRF parameters. Lakshmanan and Derin [207] have developed an iterative algorithm for MAP segmentation using an ML estimate of the MRF parameters. Nadabar and Jain [208] estimate the MRF line process parameters using geometric CAD models of the objects in the scene. Potamianos and Goutsias [209],[210] propose the estimation of partition function by approximating the Gibbs random field (GRF) by a mutually compatible Gibbs random field (MC-GRF) through the use of Monte-Carlo simulations. Their work concentrates on binary valued, second order Gibbs random fields.

One of the primary application of MRF modeling is in textured image analysis and synthesis. Zhu *et al.* [211] use the maximum entropy principle to derive a probability density function for the ensemble of images with the same texture appearance. This density function has a form of Gibbs distribution and the estimated GRF parameters are used for texture synthesis and analysis. They extend their work in [212] and describe a stepwise algorithm for selection of filter banks used to extract the features for texture synthesis purposes. Zhu and Liu [213] propose a method for fast learning of Gibbsian fields using a maximum satellite likelihood estimator which makes use of a set of pre-computed Gibbs models called "satellites" to approximate the likelihood function.

8.6 Estimation of MRF Prior

We assume that the high resolution image $\mathbf{z}$ is represented by an MRF. Thus, we have

$$P(Z = \mathbf{z}) = \frac{1}{Z_p} e^{-U(\mathbf{z},\Theta)}, \tag{8.11}$$

where $\mathbf{z}$ is a realization of Z, and Z_p is the partition function. Here Θ represents the parameter set that defines the MRF model and $U(\mathbf{z}, \Theta)$ is the energy function given by

$$U(\mathbf{z}, \Theta) = \sum_{c \in C} V_c(\mathbf{z}, \Theta). \tag{8.12}$$

$V_c(\mathbf{z}, \Theta)$ denotes the potential function associated with a clique c and C is the set of all cliques. Unlike in previous chapters, we explicitly use the parameterization Θ in the potential function $V_c(\mathbf{z}, \Theta)$. The clique c consists of either a single pixel or a group of pixels belonging to a particular neighborhood system. In this study, we consider either the symmetric first order neighborhood consisting of the four nearest neighbors of each pixel or the second order neighborhood consisting of the eight nearest neighbors of each pixel. In particular, we use the following two and four types of cliques shown in Figure 8.17. This is a simpler version of the cliques shown earlier in Figure 3.3. In the figure, β_i is the parameter specified for clique c_i. The Gibbs energy prior for $\mathbf{z}$ can now be written as

$$P(\mathbf{z}, \Theta) = \frac{1}{Z_p} \exp\left\{-U(\mathbf{z}, \Theta)\right\}. \tag{8.13}$$

Depending on whether we choose a first order or a second order neighborhood to model the field, the overall energy function for the image can be given as

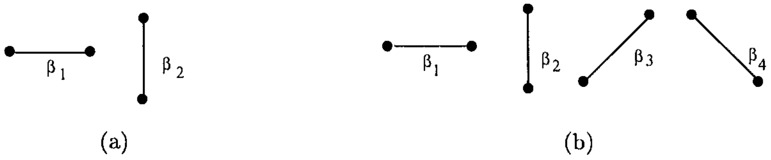


Fig. 8.17. Cliques used in modeling the image. (a) First order, and (b) second order neighborhood.

$$
\begin{aligned}
U(\mathbf{z},\Theta) = \sum_{k=1}^{N-2}\sum_{l=1}^{N-2} \{ & \beta_1[(z_{k,l}-z_{k,l+1})^2+(z_{k,l}-z_{k,l-1})^2] \\
& + \beta_2[(z_{k,l}-z_{k-1,l})^2+(z_{k,l}-z_{k+1,l})^2]\}
\end{aligned}
$$

for two parameters $[\beta_1, \beta_2]$, or

$$
\begin{aligned}
U(\mathbf{z},\Theta) = \sum_{k=1}^{N-2}\sum_{l=1}^{N-2} \{ & \beta_1[(z_{k,l}-z_{k,l+1})^2+(z_{k,l}-z_{k,l-1})^2] \\
& + \beta_2[(z_{k,l}-z_{k-1,l})^2+(z_{k,l}-z_{k+1,l})^2] \\
& + \beta_3[(z_{k,l}-z_{k-1,l+1})^2+(z_{k,l}-z_{k+1,l-1})^2] \\
& + \beta_4[(z_{k,l}-z_{k-1,l-1})^2+(z_{k,l}-z_{k+1,l+1})^2]\}
\end{aligned}
$$

for four parameters $[\beta_1, \beta_2, \beta_3, \beta_4]$. Thus $\Theta = [\beta_1, \beta_2]^T$ represents the parameter vector for the first order and $\Theta = [\beta_1, \beta_2, \beta_3, \beta_4]^T$ for the second order neighborhoods, respectively. We use these particular energy functions in our studies in order to regularize the solution using the estimated prior.

We realize that in order to enforce the prior information while estimating the high resolution image $\mathbf{z}$, we must know the values of the field parameters Θ. Thus the parameters must be learnt from the given observations themselves. However, we notice that a major part of the scene is available only at a low resolution. The parameters of the MRF cannot be learnt from these low resolution observations as the field property is not preserved across the scale or the resolution pyramid [127]. There is only one observation Y_K where a part of the scene is available at the high resolution. Hence, we use the observation Y_K to estimate the field parameters. The inherent assumption is that the entire scene is statistically homogeneous and it does not matter which part of the scene is used to learn the model parameters.

The estimation of the model parameters is, however, a non-trivial task. As already discussed, a large body of literature exists on how to estimate the MRF parameters. Most of these methods are computationally very expensive. We adopt a relatively faster but an approximate learning algorithm, known as the maximum pseudo-likelihood (MPL) estimator [206] to estimate the model parameters. The estimation procedure is briefly explained here.

The parameter estimation formulation for the prior model is based on the following ML optimality criterion

$$\hat{\Theta} = \arg \max_{\Theta} P(Z = \mathbf{z}|\Theta). \tag{8.14}$$

The probability in Eq. (8.14) can be expressed as

$$P(Z = \mathbf{z}|\Theta) = \frac{\exp\left[-U(\mathbf{z}|\Theta)\right]}{\sum_{\zeta} \exp\left[-U(\zeta, \Theta)\right]}. \tag{8.15}$$

In Eq. (8.15) summation is over all possible realizations of Z. From a computational point of view, handling Eq. (8.15) is practically not possible. Hence to overcome the computational difficulty and to make the parameter estimation problem tractable, we approximate Eq. (8.15) using the pseudolikelihood function (see [214]).

$$\hat{P}(Z = \mathbf{z}|\Theta) \triangleq \prod_{k,l} P(Z_{k,l} = z_{k,l}|Z_{m,n} = z_{m,n}, \Theta), \tag{8.16}$$

where $(m, n) \in \eta(k, l)$ form the given neighborhood model (the first order or the second order neighborhood as chosen in this study). Further it can be shown that Eq. (8.16) can be written as

$$\hat{P}(Z = \mathbf{z}|\Theta) \triangleq \prod_{k,l} \left[\frac{\exp\left\{-\sum_{c \in C} V_c(z_{k,l}, \Theta)\right\}}{\sum_{z_{k,l} \in B_z} \left\{\exp\left[-\sum_{c \in C} V_c(z_{k,l}, \Theta)\right]\right\}} \right], \tag{8.17}$$

where B_z is the set of intensity levels used. Considering the fact that the field $\mathbf{z}$ is not available for learning, and that only Y_K is available, the parameter estimation problem can be recast as

$$\hat{\Theta} = \arg \max_{\Theta} \hat{P}(R_K \mathbf{z}_{\alpha_K} = \mathbf{y}_K|\Theta). \tag{8.18}$$

We maximize the log likelihood of the above probability by using Metropolis-Hastings algorithm as discussed in [206] and obtain the parameters.

8.7 Development of an Alternative Prior

Learning of MRF model parameters allows one to obtain the parameters depending on the choice of clique potentials. We have considered here the clique potential as a function of a finite difference approximation of the first order derivative at each pixel location. Thus the learned MRF parameters specify the weightage for smoothness of the super-resolved image. Although the MRF model for prior constitutes a popular statistical model and captures the contextual dependencies very well, the computational complexities for learning these models are high as one needs to compute the partition function in order to estimate the true parameters. The computational burden can be reduced by using a scheme such as the maximum pseudo-likelihood as used in our studies. But to obtain the global minima we still need to use a stochastic relaxation technique, which is computationally very demanding. Also the pseudolikelihood is not a true likelihood except for the trivial case of null neighborhood.This motivates us to use a different but a suitable prior. We can consider the linear dependency of a pixel in a super-resolved image to its neighbors and represent the same by using a simultaneous autoregressive (SAR) model and use this SAR model as the prior. Although this becomes a weaker prior the computation is drastically reduced.

Let $z(\chi)$ be the gray level value of a pixel at site $\chi = (i, j)$ in an $N \times N$ lattice, where $(i, j) = 1, 2, \cdots N$. The SAR model for $z(\chi)$ can then be expressed as [215]

$$z(\chi) = \sum_{\vartheta \in \mathcal{N}_\chi} \Theta(\vartheta) z(\chi + \vartheta) + \sqrt{\varsigma} n(\chi), \tag{8.19}$$

where $\mathcal{N}_\chi$ is the set of neighbors of pixel at χ. $\Theta(\vartheta)$, $\vartheta \in \mathcal{N}_\chi$ and ς are unknown parameters and $n(.)$ is an independent and identically distributed (i.i.d) Gaussian noise sequence with zero mean and variance unity. Here ς is the strength of the white noise sequence which, when passed through a system having an autoregressive model with parameters $\mathbf{\Theta}$, produces the desired sequence $\mathbf{z}$.

The use of linear autoregressive models is very popular in digital signal processing. Much details can be learnt from a text book on signal processing (for example, [216]). We mention only a few past work on 2D signal processing.

We suggested in the previous section the use of a homogeneous MRF to model the high resolution field for learning purposes. However, the accurate learning of MRF parameters is a computationally tedious job. The computation can be drastically reduced if the model is restricted to a linear one such as a SAR [215], although the corresponding prior becomes weaker due to the restriction imposed on it. We circumvent this weakness by learning a larger parameter set by considering a larger neighborhood size. The ML estimates of the SAR model parameters are obtained using the iterative estimation scheme as the loglikelihood function is nonquadratic.

Now we discuss in brief about the use of SAR models in image processing. Kashyap and Chellappa [215] estimate the unknown parameters for the SAR and the conditional Markov (CM) models and also discuss the decision rule for the choice of neighbors using synthetic patterns. Authors in [217] use a multiresolution simultaneous autoregressive model for the texture classification and the segmentation. They derive a rotation invariant SAR model for the texture classification. Multispectral SAR and MRF models for modeling of color images and the procedure for parameter estimation are considered in [218]. As discussed in [97], the richness of the real world images would be difficult to capture analytically. This motivates us to use a learning based approach, where the parameters of the super-resolved image can be learnt from the most zoomed observation and hence can be used to estimate the super-resolution image for the least zoomed entire scene. This method of learning should work well as the parameters are learnt from the super-resolved image itself, as a part of it is available as the most zoomed image. However, the belief is that the same parameters can be used while super-resolving the entire scene.

The number of pixels used in a neighborhood system increases with a larger neighborhood structure. This helps in capturing the statistical dependency of a pixel on its neighbors in a better way. Thus we make use of a larger neighborhood for learning the parameters. While using a fifth order neighborhood we require a total of 24 parameters $\Theta(i,j)$ as shown in Figure 8.18. In order to reduce the computations while estimating these parameters we use a symmetric SAR model where $\Theta(\vartheta) = \Theta(-\vartheta)$. It may be mentioned here that we do not discuss the choice of appropriate order for the neighborhood system and the choice of number of parameters for optimum results.

One of the characteristics of an image data is the statistical dependence of the gray level at a lattice point on those of its neighbors. This statistical dependency is now characterized by using a SAR model

(−2,−2)	(−2,−1)	(−2, 0)	(−2, 1)	(−2, 2)
(−1,−2)	(−1,−1)	(−1, 0)	(−1, 1)	(−1, 2)
(0, −2)	(0, −1)	(0, 0)	(0, 1)	(0, 2)
(1, −2)	(1, −1)	(1, 0)	(1, 1)	(1, 2)
(2, −2)	(2, −1)′	(2, 0)	(2, 1)	(2, 2)

Fig. 8.18. The fifth order neighborhood for a pixel at location (0, 0).

where the gray level at a location is expressed as a *linear* combination of the neighborhood gray levels and an additive noise. Thus we can use a SAR model as a prior instead of an MRF prior where the computational burden is much less. We estimate the SAR model parameters by considering the image as a finite lattice model and using the iterative scheme as given in [215]. We model the most zoomed image as a SAR model and obtain the least squares estimate to initialize the parameters. These initial estimates are then used in the iterative algorithm to obtain the final parameters.

8.8 High Resolution Restoration

8.8.1 Restoration using MRF Prior

Having learnt the MRF model parameters, we now try to super-resolve the entire scene. In order to do that we use the MAP estimator to restore the high resolution field $\mathbf{z}$. Given the ensemble of images at different resolutions the MAP estimate of $\mathbf{z}$ is given by

$$\hat{\mathbf{z}} = \arg \max_{\mathbf{z}} P(\mathbf{z} \mid \mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K, \Theta). \tag{8.20}$$

Note that we include the learnt parameters Θ in the MAP estimator. This is the primary difference between this equation and Eq. (8.3). The scene to be recovered has been modeled as a MRF. Thus using the data fitting term and the prior term it can be easily shown that the final cost function is obtained as (similar to Eq. (8.9))

$$\epsilon = \left[\lambda \sum_{m=1}^{K} \|\mathbf{y}_m - D_m R_m \mathbf{z}_{\alpha_m}\|^2 + \sum_{c \in \mathcal{C}} V_c(\mathbf{z}, \Theta) \right]. \tag{8.21}$$

where λ is a regularization parameter. Since the model parameter set $\boldsymbol{\Theta}$ has already been estimated, a solution to the above equation is, indeed, possible. The above cost functions is convex and is minimized using the gradient descent technique. The initial estimate $\mathbf{z}^{(0)}$ is obtained by using the successive bilinear expansion as already discussed.

8.8.2 Restoration using SAR Prior

Instead of using the MRF prior, we may use the SAR prior for a faster computation. With the SAR parameters estimated as discussed in section 8.7 we would like to arrive at a cost function which has to be minimized to super-resolve the observations. As before, we use the regularization based approach which is quite amenable to the incorporation of information from multiple observations with the regularization function chosen from the prior knowledge of SAR model. Now we use a simple linear dependency of a pixel value on its neighbors as a constraint using the SAR model for the image to be recovered. Using a data fitting term and a prior term one can easily derive the corresponding cost function to be minimized as

$$\epsilon = \lambda \sum_{m=1}^{K} \|\mathbf{y}_m - D_m R_m \mathbf{z}_{\alpha_m}\|^2 + \sum_{\chi} \left(z(\chi) - \sum_{\vartheta \in \mathcal{N}_\chi} \Theta(\vartheta) z(\chi + \vartheta) \right)^2 . \tag{8.22}$$

Here λ is a regularization parameter which is now proportional to $\frac{\sigma_n^2}{\varsigma}$ where ς is the error variance for the SAR model (see Eq. (8.19)). This cost function is also minimized using the gradient descent with initial estimate as $\mathbf{z}^{(0)}$ as discussed in restoration using MRF prior.

8.9 Experiments with Learnt Prior

We demonstrate the usefulness of the proposed technique to recover the super-resolved image from observations at different zooms through learning of model parameters.

Initially we experimented on simulated data. A number of images were chosen from the Brodatz's album. We observe an image at three levels of zoom $r_1 = r_2 = 2$. Figures 8.19(a-c) show one such set of observations, where Figure 8.19(a) shows the entire image at a very low resolution, (b) shows one-fourth of the region at double the resolution, and (c) shows only a small part of Figure 8.19(a) at the highest resolution.

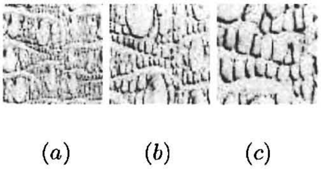


Fig. 8.19. (a-c) Observed images (D10) of a texture captured with three different zoom settings ($r_1 = 2$ and $r_2 = 2$).

We use a first order MRF to model the intensity process in Figure 8.19(c). The estimated values of the parameters were $\beta_1 = 6.9$ and $\beta_2 = 28.8$. These parameters were estimated using the Metropolis-Hastings algorithm by choosing the initial values of the parameters as unity. We observed the convergence of the parameter estimation algorithm for most of the cases within 1000 iterations, although there were convergence difficulties for just a few of the images we considered. For example, the algorithm did not converge for the Nidhi image shown in Figure 8.6(c). Using the learnt parameter set, we now super-resolve the entire scene in Figure 8.19(a) to obtain the Figure 8.20(b). Compare the result to that obtained using a simple bilinear zooming operation given in Figure 8.20(a). We notice that both the images are quite blurred near the periphery, as expected. However, the interpolated image is too blurred to infer anything about the texture. For the super-resolved image, the restoration upto a zoom factor $r = 2$ is quite good. For a zoom factor of $r = 4$, one needs to reconstruct 16 pixels for each observed pixel near the periphery, which is clearly a difficult task. A degradation in the reconstruction is, thus, quite expected even in the estimated high resolution image. We then use the SAR model as an alternative prior for super-resolution. We used a symmetric fifth order neighborhood for SAR modeling. The learnt parameters from the most zoomed observation (Figure 8.19(c)) are used to enforce the dependency of each pixel on its neighbors in the entire scene to be super-resolved by using the prior. For most of the images convergence of the SAR parameter estimation algorithm was obtained within 10 iterations and no convergence problem was ever faced. The super-resolved image using the estimated parameters is shown in Figure 8.20(c). We can clearly see that the super-resolved image is sharper with better details than those obtained either with the bilinear interpolation or the super-resolved image using the MRF prior shown in Figures 8.20(a) and 8.20(b), respectively. In order to highlight the improvement in resolution achieved,

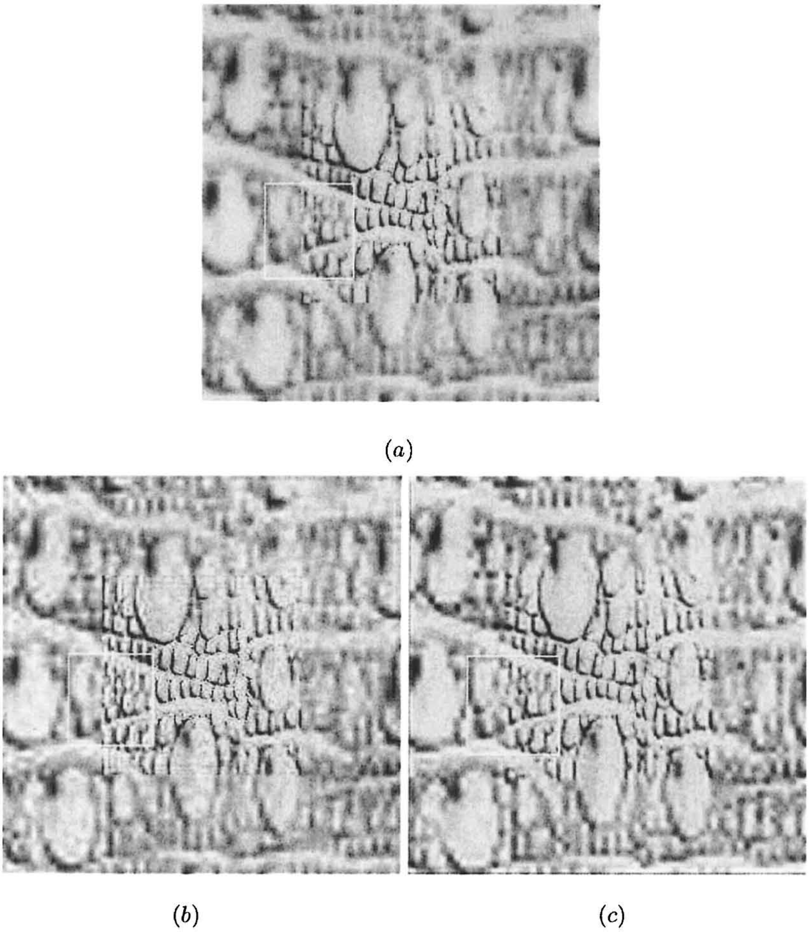


Fig. 8.20. (a)Zoomed texture image formed by successive bilinear expansion. Super-resolved texture image using the learnt (b) MRF prior, and (c) the SAR model.

a rectangular region is marked in all the three results. Note that the edges come out more sharply in the MRF based method compared to the bilinearly interpolated image. However, these edges are even better recovered using the SAR prior. The reason for the better restoration using the SAR approach is that we are using a larger neighborhood with more number of parameters for the model representation. This is able to capture the prior better than the MRF model as we are constrained to use a very few cliques during the MRF modeling for reasons of computational difficulties in learning these model parameters.

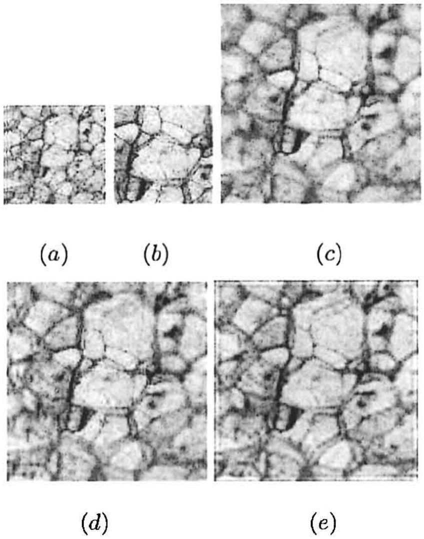


Fig. 8.21. (a, b) Observed images (D112) of another texture captured with two different zoom settings ($r = 2$), (c) Zoomed texture image formed by successive bilinear expansion. Super-resolved image for a zoom factor of $r = 2$ using (d) MRF prior, and (e) SAR model parameters.

In order to demonstrate the performance of these algorithms for a zoom factor of 2, we now consider two simulated observations with $r = 2$ shown in Figure 8.21(a, b). A first order MRF model was used to capture the texture in Figure 8.21(b) and the estimated MRF parameters were $\beta_1 = 29.96$ and $\beta_2 = 38.19$. The bilinearly zoomed image is shown in Figure 8.21(c). The super-resolved image obtained using the MRF based prior and the SAR prior are given in Figures 8.21(d, e), respectively. As can be seen the high frequency details are restored well in the super-resolved images. The bilinearly interpolated image (see Figure 8.21(c)) definitely appears blurred compared to the restored images using the proposed approach (see Figures 8.21(d, e)). The result obtained using the SAR prior is better than that of the MRF prior due to the choice of a larger neighborhood.

Results for another set of observed textures, shown in Figures 8.22(a-c) are given in Figures 8.23(b) and 8.23(c). The zoomed image

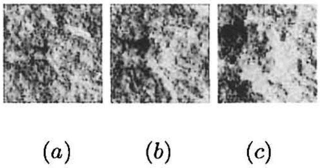


Fig. 8.22. (a-c) Observed images (D2) of yet another texture captured with three different zoom settings.

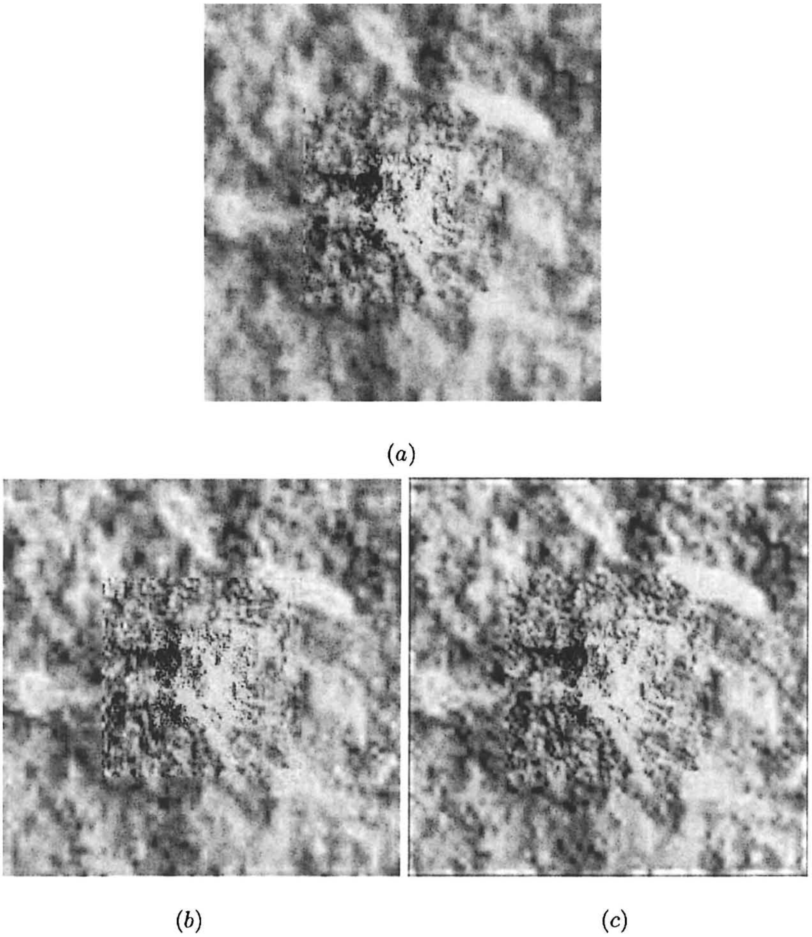


Fig. 8.23. (a) Zooming by successive bilinear expansion. Super-resolution restoration of images given in Figure 8.22 using (b) learnt MRF prior, and (c) the SAR model.

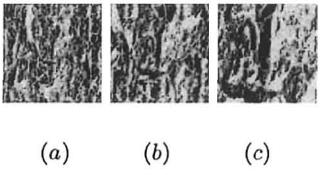


Fig. 8.24. (a-c) Observed texture (D12) at three different zoom settings.

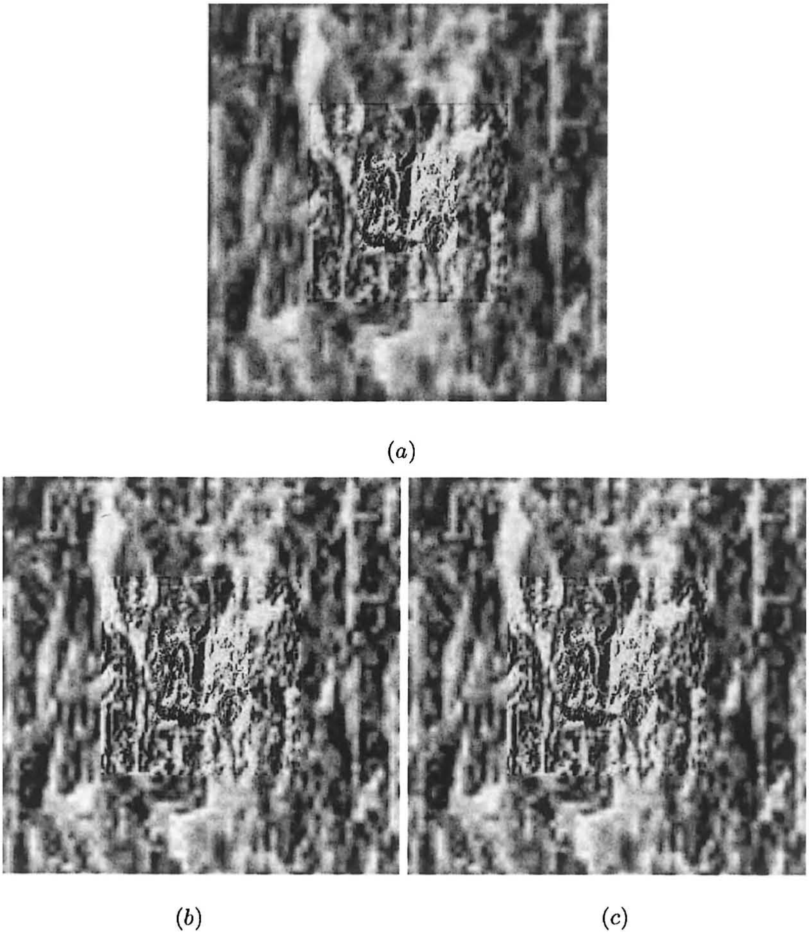


Fig. 8.25. (a) Bilinearly zoomed texture image. (b) The super-resolution restoration using first order MRF prior. (c) Restoration using a second order neighborhood structure.

using the standard bilinear interpolation is shown in Figure 8.23(a). The super-resolved images are definitely sharper than the zoomed image. Although the edges at the outer region are not as sharp as it is in the center, they are a lot more discernible than those in the interpolated image.

We also tested the algorithm for MRF based prior using four parameters instead of just two cliques. Result of the same for a set of observed textures, given in Figures 8.24(a-c), is given in Figure 8.25. Once again, a comparison with the corresponding zoomed image in Figure 8.25(a) brings out a similar conclusion that upto a zoom factor $r = 2$ the results of the proposed super-resolution scheme is very good, but beyond that the quality of restoration starts degrading. This again conforms to the observation that the restoration error increases with an increase in the amount of blurring [205]. However the mean squared error (MSE) comparison for the proposed approach and the successive bilinear interpolated image when measured with respect to the original image showed a significant decrease in all of the above experiments as given in Table 8.1. Further, a comparison between the super-resolved images presented in Figure 8.25(c) and Figure 8.25(b) where the prior term uses a second order neighborhood shows that there is no significant perceptual improvement with an additional order introduced in the prior term. Our experience suggests that the improvement is very gradual as the order of the MRF parameterization is increased. Ideally one requires a large number of cliques to learn the prior. However, the computation goes up drastically while learning the scene prior. Hence we refrain from using a neighborhood structure beyond the second order. One does not have a similar difficulty while using a larger neighborhood structure in the SAR model based approach.

In order to quantify the the improvement in spatial resolution using the proposed approaches, we compute the mean squared error (MSE) of the reconstructed image with respect to the original high resolution image. The result is summarized in Table 8.1 for all the above four simulation experiments for two different levels of zooming, namely $r = 2$ and $r = 4$. From the table we observe that the use of MRF prior helps us in reducing the MSE by at least 30% as compared to the bilinear interpolation. The use of SAR prior helps us to further reduce the MSE by another $5 - 25\%$. This justifies the use of learnt priors in super-resolving the image.

Table 8.1. Comparison of MSE for Bilinear interpolation (BI), MRF Approach and SAR Approach.

Image	$r = 2$ BI	 MRF Approach	 SAR Approach	$r = 4$ BI	 MRF Approach	 SAR Approach
D10	0.0153	0.0097	0.0081	0.0362	0.0264	0.0249
D112	0.0097	0.0065	0.0062	0.0290	0.0185	0.0175
D2	0.0159	0.0106	0.0083	0.0355	0.0239	0.0209
D12	0.0488	0.0333	0.0206	0.1357	0.0759	0.0741

We now present some results of experimentation on real data. Unlike in the case of simulation experiments, the assumption of the homogeneity is not strictly valid for the real data. However in the absence of availability of any other usable priors, we continue to make use of this assumption and show that we still obtain a reasonably good super-resolution reconstruction. First we consider a real image which has a texture similar to the simulated texture. This corresponds to the picture of a bedsheet in a hostel room. Figures 8.26(a-c) show the observations at three different levels of camera zoom. However, the zoom levels were carefully chosen such that the relative zoom factors between two successive observations are again $r = 2$. Since we are capturing the scene with different zoom setting, we used mean correction to compensate for the AGC effect as already described in section 8.4.1. Figure 8.27(a) shows the zoomed image and the super-resolved images are shown in Figure 8.27(b) and Figure 8.27(c), respectively. Comparison of the figures show a more clear details in the super-resolved image using the SAR prior (see Figure 8.27(c)) with a slight improvement in the super-resolved image using the MRF prior. The blur which is clearly visible in Figure 8.27(a) indicating the loss of high frequency details is removed in Figure 8.27(c). The MRF parameters for this experiment were estimated to be $\beta_1 = 33.77$, $\beta_2 = 60.19$.

Now we consider an example where the scene has an arbitrary texture. We repeat the experiment on the data shown in Figure 8.4. We already know what the corresponding super-resolved image (see Figure 8.5)(b) is. This was obtained by manually fine tuning the MRF parameters. We want to verify if the learning of the MRF parameters takes us to a very similar result or not. The corresponding super-resolved image is shown in Figure 8.28(a) and Figure 8.28(b), respectively. Comparison of these figures with those given earlier in Figure 8.5 show that

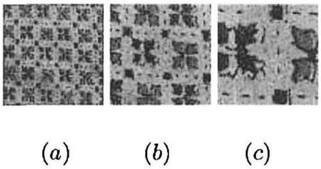


Fig. 8.26. (a-c) Observed images of a bedsheet captured with three different camera zoom settings.

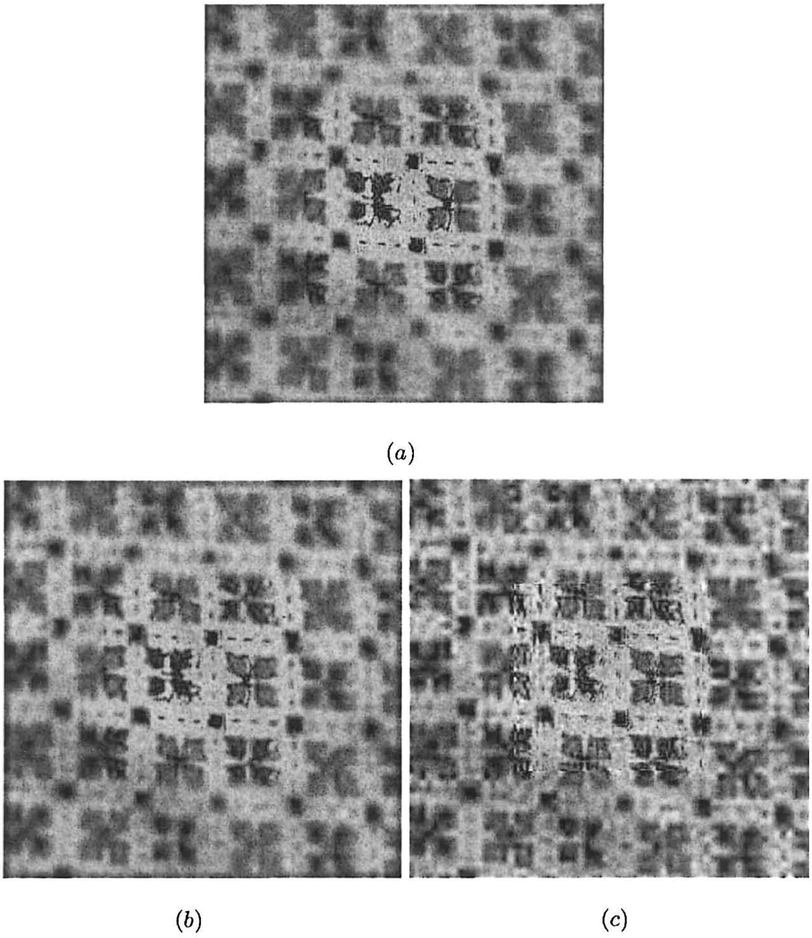
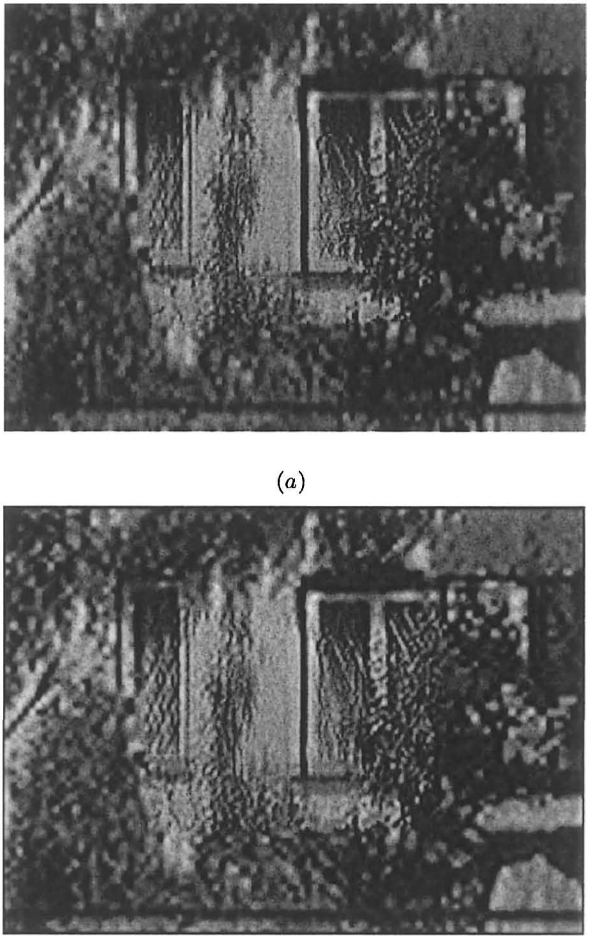


Fig. 8.27. (a) Bilinearly zoomed bedsheet image. (b) Super-resolved bedsheet image using the MRF prior, (c) super-resolved using the SAR prior.



(a)

(b)

Fig. 8.28. Super-resolved house image for observations shown in Figure 8.4. Use of learnt (a) MRF prior, (b) SAR prior.

we are, indeed, able to learn the texture present in the scene. The MRF parameters for this experiment were estimated to be $\beta_1 = 9.1$, $\beta_2 = 155.3$. Again we note that we have assumed the image texture to be homogeneous over the entire scene. The above assumption is, however, not strictly valid for the current example, and hence the quantitative improvement in the super-resolution images is not very significant.

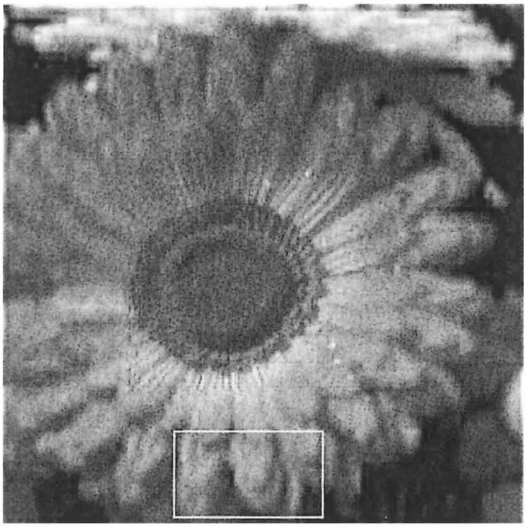


Fig. 8.29. Super-resolved flower image for the observations given in Figure 8.15 using learnt MRF prior.

Nonetheless, we were able to obtain an improved result using this technique.

Next we consider an example of real data acquisition when the zoom levels are totally arbitrary. We continue with the observations shown in earlier in Figure 8.15 where the zoom factors are unknown. The first order MRF model parameters were estimated to be $\beta_1 = 337.3$, $\beta_2 = 463.4$ from Figure 8.15(c). The experimental result of the super-resolution restoration using the MRF prior is given in Figure 8.29. Compare this to the result obtained in Figure 8.16(b). The restoration using the learnt prior is better than what we achieved earlier when these parameters were chosen by a trial and error basis. The extremities of the petals are now much sharper. Even the blobs at the bottom right corner appear more clearly in Figure 8.29 compared to the result in Figure 8.16(b).

8.10 Conclusions

We have discussed in detail a technique to recover the super-resolution intensity field from a sequence of zoomed observations. The resolution of the entire scene is obtained at the resolution of the most zoomed observed image which consists of only a portion of the actual scene. Initially the super-resolved image is modeled as an MRF, and a MAP estimate is used to derive the cost function to be minimized. First we used a convex energy function by selecting the finite difference approximation of the first order derivative of the intensity pixel at each location for the prior term. This made it possible for us to use the gradient descent algorithm for the minimization of the cost. Next, the cost function was modified to include the line fields in order to preserve the discontinuities. As a consequence the cost was not differentiable and we used simulated annealing in order to obtain the global minima. In order to reduce the computational burden, MFA was used to optimize the solution. We demonstrated that it is, indeed, possible to obtain a high resolution image of the scene using zoom as a cue. It is also quite clear that the quality of reconstruction depends on how far a point is from the optical axis - the quality degrades as one moves away from the axis since a very little information is available for super-resolving this region.

We also realize that a part of the scene is available at the highest level of resolution while zooming. This motivated us to learn the MRF parameters directly from the partial data and we use it as the learnt prior while super-resolving the entire scene using the MAP estimator. Although the homogeneity (stationarity) assumption is not strictly valid for the real images, we have attained reasonably good results even for images captured using a real camera. Thus we demonstrate that it is possible to obtain a high resolution image of a scene using zoom as a cue by learning the parameters from the given observations only.

Although MRF is the most general prior model as there is flexibility in choosing the clique potential, the computational burden involved in learning these parameters is high as one needs to consider a much larger neighborhood to capture a more accurate local dependency. Also the calculation of the partition function is a difficult task. Thus SAR seems to be a better prior, though it is a weaker model, as the computational complexity is less as compared to the MRF parameter estimation. We make amends for this by increasing the neighborhood size while using the SAR model. It is always desirable to learn the parameters from the given data so that the computational burden can be reduced and the actual parameters reflecting the characteristics of the high resolution image can be used while super-resolving an image.

9 Looking Ahead

9.1 Summary of Concepts Developed

High resolution images are often desired in most of the imaging applications. However the images captured using a commercially available camera may not offer the required spatial resolution. Super-resolution refers to a technique by which a high resolution image is generated from a sequence of low resolution observations. The idea behind super-resolution is to use the additional information contained in each of the observed low resolution images so as to obtain a high resolution image. One has to look for intensities of missing pixels in the unknown high resolution image by using the pixel intensities of the aliased and blurred observations.

Most of the literature on super-resolution imaging discusses methods that involve taking images of the same scene with subpixel shifts among them. The first task in motion-based super-resolution techniques involve registration. Thus the difficulty with these techniques is the correspondence problem. Obtaining registration with a subpixel accuracy is extremely difficult. Also, the motion-based techniques assume that all the observations are at the same resolution. These difficulties associated with motion-based techniques have been eliminated in this monograph by proposing motion-free super-resolution methods. In addition, most of the super-resolution techniques do not exploit the structural information present in the images. We suggest the use of this important information while improving the quality of the super-resolved image.

First we explored the use of relative blur among low resolution observations in super-resolving a scene. We realize that any real aperture camera offers a finite depth of field, introducing depth related defocus blur in the observations. By suitably changing the camera parameters, we can take multiple observations. We have explained how the observations can be restored at a high resolution and how the dense depth in the scene can be recovered simultaneously. The high resolution intensity and the depth fields have been modeled as separate Markov random fields. The MRF priors were used for regularization purposes while obtaining the MAP estimate. The experimental results have been very encouraging.

We have also explored the possibility of using the the photometric cue where the 3D structure preservation is used as a constraint while super-resolving the scene. The structure or depth of an object is embedded in images in various forms, e.g., texture, shading, *etc.* We make use of the photometric measurements to recover the dense depth map of the scene as well as the intensity map. Thus we expand the scope of super-resolution to include high resolution depth information in a scene, together with recovering the super-resolved intensity values. When we capture images of the same scene by varying the positions of the light source, new information is available at each pixel to capture the surface properties. We make use of this new information for super-resolution. The concept of generalized interpolation in conjunction with appropriate regularization has been used in this study. The high resolution intensity field, the surface gradients and the albedo are all modeled as separate MRFs to provide the regularizing priors.

Experiments using real images captured under controlled lighting conditions offer quite acceptable results. We further demonstrate that the photometric observations need not be free from blur for super-resolution purposes. We do allow blurred photometric observations in our study. An iterative alternate blur and structure recovery algorithm has been suggested for blind restoration.

The super-resolution from photometric observations makes use of the generalized interpolation. We demonstrate that the generalized interpolation could be used as a basic frame work for image upsampling for other cases as well. We consider the principal component analysis of a set of training images in a data base. We explain how an image can be upsampled by individually interpolating each of the eigenvectors. There are certain difficulties using the PCA based interpolation as the interpolated eigenvectors are no longer orthogonal to each other. The method is intrinsically an interpolation technique and it cannot improve the resolution of an aliased image. However, it is quite useful in image restoration if the input image is quite noisy and blurred. Further, we demonstrate that if one does not have an access to a high resolution training dataset, the quality of reconstruction could be further improved. Although the approach is computationally very efficient, the applicability is quite limited to a very specific type of observations such as the fingerprint or face images.

We have also investigated the problem of super-resolving an arbitrary image from a single observation. A set of high resolution images in a database serves as the training pattern. The basic idea is to enhance the sharpness of the edge elements while interpolating the image. The high resolution edge primitive learnt from the training images, that best matches the low resolution edge locally is copied using the wavelet decomposition of the image. The edge primitives have been defined over a 4×4 pixel grid in the low resolution image. Each patch in the low resolution image is learnt from different training images. This may result in some blockiness in the reconstructed image. In order to remove the blockiness we suggest the use of a smoothness prior during the image reconstruction process. The performance of this approach has been evaluated through a number of experiments.

The last topic discussed in the monograph includes a super-resolution scheme using zoom as a cue. In this scheme images were captured with different but arbitrary zoom factors using a commercially available camera. We enhance the resolution of the least zoomed entire area of the scene comparable to that of the most zoomed one. We obtain the super-resolution for both known and unknown arbitrary zoom factors. The zoom factors between different observations were estimated by using a hierarchical cross-correlation technique. We model the super-resolved image to be estimated as a Markov random field (MRF) and a maximum *a posteriori* estimation method is used to recover the high resolution image. The entire set of observations conform to the same MRF but viewed at different resolution pyramid. We observe that a part of the scene is available at the highest resolution. Hence we make use of the high resolution observation itself to learn the MRF parameters for super-resolving the least zoomed entire area of the scene. Inherently we assume that the entire scene corresponds to the same MRF and that the field parameters can be learnt from any part of the scene. Since the estimation of true MRF parameters is a difficult task and the computational complexity is high we also explore solving the problem by using the simultaneous autoregressive (SAR) prior model. Several experiments have been carried out using both the simulated as well as the real images to illustrate the efficacy of these super-resolution schemes.

9.2 Future Research Issues

The research in the area of super-resolution has started about twenty years ago. There has been a large body of research in the area of motion-based super-resolution and the current status in this area has been reviewed in chapter 2. However, the research in the area of motion-free super-resolution is still at its nascent state. The primary emphasis in the book has been to demonstrate that motion-free super-resolution is, indeed, possible. A few different methods for high resolution image reconstruction have been investigated in this monograph. While doing that, a few specific problems and their solutions have been reported here. However, there are still a number of issues which need to be investigated.

In this concluding section, we discuss the future directions in which the research could progress so that both the theoretical foundations as well as practical applications of super-resolution attain a sound footing.

- The inverse procedure in super-resolution reconstruction requires a large computational load as most of the super-resolution algorithms are based on iterative techniques. Super-resolution in real time is, thus, the need of the hour. To apply the super-resolution algorithms to practical real time situations, it is important to develop efficient algorithms that reduce the computational cost. Although, some researchers have attempted to solve this problem, there are several assumptions being made for the image formation process and the structure of the scene in order to reduce the computations. All proposals studied in this monograph, except the one on PCA based reconstruction, are currently unsuitable for even near real-time implementation.
- Learning based super-resolution is quite a new area of research. We have developed a learning based super-resolution technique using zoom as a cue by modeling the prior as either a SAR model or as an MRF. It would be of interest to consider the proper choice of neighborhood and the number of parameters for optimal restoration of the high resolution field. Also the choice of prior itself is an interesting research area. A prior based on the statistical properties such as the histograms of the output of different filters as used

in [212] for texture synthesis instead of the usual smoothness constraint may serve as a better choice. Since the super-resolution is an ill-posed problem, a good choice of prior would always be useful for the restoration.

- For the wavelet based reconstruction method, it is assumed that the low resolution observation is free from blurring. This allows us to restrict the search for the best match of the edge primitive from the training images at the same scale. In case the input image is blurred, then one may have to go down the resolution pyramid of the training images to locate the best match. This appears to be a non-trivial problem as the scale is unknown and it may not exactly match the given levels of scales of the training image after the wavelet decomposition. It will be interesting to extend the method to deal with scale changes.
- The super-resolution using the zoom cue has been shown to be effective. But a zoom lens camera system has complex optical properties and hence it is difficult to model the same. It would be interesting to consider a more realistic thick lens model instead of the pinhole model used in this study, considering the effects of both photometric and geometric distortions, thus extracting the depth field simultaneously.
- While discussing the image and depth super-resolution reconstruction technique using the photometric cue, we assumed that the positions of the light source are known. It would be desirable to estimate the source directions also by using the captured images and use it to estimate the super-resolved intensity and the depth map simultaneously. For example, consider the following problem. A light source moves along an arbitrary trajectory and one obtains a video of a static object with a low resolution stationary camera. Can we track the moving light source, recover the structure and super-resolve scene simultaneously?
- In chapter 5 while simultaneously estimating the blur and recovering the structure we have assumed the blur to be constant (shift invariant). However, in real aperture images, the blur is often shift varying due to depth variation in the scene. Thus the blur and the shading cue become interdependent. Is it then possible to combine the contents of chapter 3 with those of chapter 5 to develop a unified frame work under which both the shading and the defocus cues are jointly used to achieve an improved super-resolution?

- The maximum entropy (ME) principle is applicable to any problem of inference with a well defined hypothesis space and incomplete data. Super-resolution based on ME is an interesting research area. For each pixel we need to estimate the intensities of 4 pixels for an upsampling factor of $r = 2$. We may take all these pixel values to be random variables and may attempt to maximize the entropy of the estimated field satisfying appropriate constraints.
- Most of the current super-resolution techniques in the literature are studied for gray level images. It is necessary to extend the current super-resolution algorithms to real world color imaging systems. Although some work has been carried out in color image super-resolution, a more careful reconstruction method which reflects the characteristics of the color is needed. Since the color space is not a linear one, handling color space as separate R, G, B or Y, C_b, C_r planes and super-resolving them individually do not work well. The properties of the color space must be considered while restoring the colored image.

References

1. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Pearson Education, Inc, 2002.
2. E. L. Hall, *Image Processing and Recognition*, Academic Press, New York, 1979.
3. R. J. Schalkoff, *Digital Image Processing and Computer Vision*, John Wiley, New York, 1989.
4. H. Stark and P. Oskui, "High-Resolution Image Recovery from Image-Plane Arrays Using Convex Projections," *J. Optical Society of America*, vol. 6, no. 11, pp. 1715–1726, Nov. 1989.
5. K. Aizawa, T. Komatsu, and T. Saito, "A Scheme for Acquiring Very High Rerolution Images Using Multiple Cameras," in *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing*, San Francisco, CA, 1992, pp. 289–292.
6. T. Komatsu, K. Aizawa, T. Igarshi, and T. Saito, "Signal Processing Based Method for Acquiring Very High Resolution Image with Multiple Cameras and its Theoretical Analysis," in *Proc. IEE-I*, 1993, pp. 19–25.
7. D. Rajan, *Some New Approaches to Generation of Super-Resolution Images*, Ph.D. thesis, School of Biomedical Engineering, Indian Institute of Technology - Bombay, India, 2001.
8. N. Nguyen and P. Milanfar, "A Wavelet-Based Interpolation Restoration Method for Super-resolution," *Circuits, Systems and Signal Processing, Special issue on advanced signal and image reconstruction*, vol. 19, no. 4, pp. 321–338, August 2000.
9. D. Rajan and S. Chaudhuri, "Generation of Super-Resolution Images from Blurred Observations using an MRF Model," *J. Mathematical Imaging and Vision*, vol. 16, pp. 5–15, 2002.
10. M. Elad and A. Feuer, "Restoration of a Single Superresolution Image from Several Blurred, Noisy and Undersampled Measured Images," *IEEE Trans. on Image Processing*, vol. 6, no. 12, pp. 1646–1658, December 1997.
11. D. Rajan and S. Chaudhuri, "Generalized Interpolation and its Applications in Super-Resolution imaging," *Image and Vision Computing*, vol. 19, pp. 957–969, November 2001.
12. C. Delherm, J. M. Lavest, M. Dhome, and J. T. Lapreste, "Dense Reconstruction by Zooming," in *Fourth Europeon Conf. on Computer Vision*, Cambridge, England, April 1996, pp. 427–438.

13. J. M. Lavest, G. Rives, and M. Dhome, "Three Dimensional Reconstruction by Zooming," *IEEE Trans. on Robotics and Automation*, vol. 9, no. 2, pp. 196–207, April 1993.
14. J. Ma and S. I. Olsen, "Depth from Zooming," *Journal of the American Optical Society*, vol. 7, no. 10, pp. 1883–1890, October 1990.
15. D. Wilkes, S. J Dickinson, and J. K. Tsotsos, "A Quantitative Analysis of View Degeneracy and its use for Active Focal length Control," in *Proc. IEEE Int. Conf. on Computer Vision*, Cambridge, Massachusetts, 1995, pp. 938–944.
16. J. A. Fayman, O. Sudarsky, and E. Rivlin, "Zoom Tracking and its Applications," Tech. Rep. TR CIS9717, Technion-Israel Institute of Technology, December 1997.
17. M. V. Joshi and S. Chaudhuri, "Super-Resolution Imaging : Use of Zoom as a Cue," in *Indian Conf. on Computer Vision Graphics and Image Processing*, Ahmedabad, 2002.
18. D. Rajan, S. Chaudhuri, and M. V. Joshi, "Multi-Objective Super-Resolution Technique : Concept and Examples," *IEEE Signal Processing Magazine, Special issue on Super-Resolution Image Reconstruction*, vol. 20, no. 3, pp. 49–61, May 2003.
19. D. Rajan and S. Chaudhuri, "Simultaneous Estimation of Super-Resolved Scene and Depth map from Low Resolution Defocused Observations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1102–1117, September 2003.
20. M. V. Joshi and S. Chaudhuri, "A Learning-Based Method for Image Super-Resolution from Zoomed Observations," in *Proc. fifth Int. Conf. on Advances in Pattern Recognition*, Kolkata, India, 2003, pp. 179–182.
21. M.V. Joshi and S. Chaudhuri, "Photometric Stereo Under Blurred Observations," in *Proc. of IEEE Intl Conf. on Pattern Recognition*, Cambridge, England, August, 2004.
22. C. V. Jiji, M. V. Joshi, and S. Chaudhuri, "Single Frame Image Super-Resolution using Learnt Wavelet Coefficients," *International Journal of Imaging Science and Technology*, vol. 14, no. 3, pp. 105–112, September 2004.
23. M. V. Joshi and S. Chaudhuri, "Zoom Based Super-Resolution Through SAR Model Fitting," in *Proc. of IEEE Intl Conf. on Image Processing*, Singapore, October, 2004.
24. C. V. Jiji and S. Chaudhuri, "PCA based Generalized Interpolation for Image Super-Resolution," in *Proc. of Indian Conf. on Vision, Graphics and Image Processing*, Kolkata, India, December, 2004.
25. M. V. Joshi, S. Chaudhuri, and R. Panuganti, "Super-Resolution Imaging: Use of Zoom as a Cue," *Image and Vision Computing*, vol. 22, no. 14, pp. 1185–1196, December 2004.
26. M. V. Joshi and S. Chaudhuri, "Joint Blind Restoration and Surface Recovery in Photometric Stereo," *Journal of Optical Society of America-A (Accepted for publication).*
27. M. V. Joshi, S. Chaudhuri, and R. Panuganti, "A Learning-Based Method for Image Super-Resolution from Zoomed Observations," *IEEE Trans. Systems Man and Cybernetics, Part B, Special Issue on Learning in Computer Vision and Pattern Recognition (Accepted for publication).*
28. S. Chaudhuri, *(Ed.), Super-Resolution Imaging*, Kluwer Academic Publisher, Boston, 2001.

29. S. C. Park, M. K. Park, and M. G. Kang, "Super-Resolution Image Reconstruction: A Technical Review," *IEEE Signal Processing Magazine, Special Issue of Super-resolution Image Reconstruction*, vol. 20, no. 3, pp. 21–36, May 2003.
30. R. Y. Tsai and T. S. Huang, "Multiframe Image Restoration and Registration," in *Advances in Computer Vision and Image Processsing*, pp. 317–339. JAI Press Inc., 1984.
31. S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive Reconstruction of High Resolution Image From Noisy Undersampled Multiframes," *IEEE Trans. on Accoustics, Speech and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, June 1990.
32. S. P. Kim and W. Y. Su, "Recursive High Resolution Reconstruction of Blurred Multiframe Images," *IEEE Trans. on Image Processing*, vol. 2, no. 4, pp. 534–539, October 1993.
33. S. H. Rhee and M. G. Kang, "Discrete Cosine Transform based Regularized High Resolution Image Reconstruction Algorithm," *Optical Engineering*, vol. 38, no. 8, pp. 1348–1356, August 1999.
34. C. Srinivas and M. D. Srinath, "A Stochastic Model based Approach for Simultaneous Restoration of Multiple Mis-registered Images," *Proc, SPIE*, vol. 1360, pp. 1416–1427, 1990.
35. A. Papoulis, "Generalized Sampling Theorem," *IEEE Trans. on Circuits and Systems*, vol. 24, pp. 652–654, November 1977.
36. J. L. Brown, "Multi-Channel Sampling of Low Pass Signals," *IEEE Trans. on Circuits and Systems*, vol. CAS-28, no. 2, pp. 101–106, February 1981.
37. H. Ur and D. Gross, "Improved Resolution from Sub-pixel Shifted Pictures," *CVGIP:Graphical Models and Image Processing*, vol. 54, pp. 181–186, March 1992.
38. N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive Implementation of Total Least Squares Algorithm for Image Reconstruction from Noisy, Undersampled Multiframes," in *In Proc. of IEEE Conference on Acoustics, Speech and Signal Processing*, Minneapolis, MN, 1993, pp. 269–272.
39. M. Irani and S. Peleg, "Improving Resolution by Image Registration," *CVGIP:Graphical Models and Image Processing*, vol. 53, pp. 231–239, March 1991.
40. S. Peleg, D. Keren, and L. Schweitzer, "Improving Image Resolution using Sub-pixel Motion," *Pattern Recognition. Letters.*, vol. 5, pp. 223–226, March 1987.
41. M. Irani and S. Peleg, "Motion Analysis for Image Enhancement : Resolution, Occlusion, and Transparency," *Journal of Visual Communication and Image Representation*, vol. 4, pp. 324–335, December 1993.
42. A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High Resolution Image Reconstruction from Lower-Resolution Image Sequences and Space-Varying Image restoration," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, San Francisco,USA, 1992, pp. 169–172.
43. M. K. Ng, N. K. Bose, and J. Koo, "Constrained Total Least Squares Computation for High Resolution Image Reconstruction with Multisensors," *International Journal of Imaging Systems and Technology*, vol. 12, pp. 35–42, 2002.

44. M. K. Ng and N. K. Bose, "Analysis of Displacement Errors in High Resolution Image Reconstruction with Multisensors," *IEEE Trans. on Circuits and Systems 1: Fundamental Theory and Applications*, vol. 49, no. 6, pp. 806–813, June 2002.
45. N. K. Bose, S. Letrattanapanich, and J. Koo, "Advances in Super-resolution Using L Curve," in *Proc. Int. Symp. Circuits and Systems*, Sydney, 2001, pp. 433–436.
46. P. C. Hansen and D. Prost O'Leary, "The Use of the L-Curve in Regularization of Discrete Ill-posed Problems," *SIAM J. Sci. Comput*, vol. 14, no. 6, pp. 1487–1503, November 1993.
47. R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP Registration and High- Resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE Trans. on Image Processing*, vol. 6, no. 12, pp. 1621–1633, December 1997.
48. R. C. Hardie, K. J. Barnard, J. G. Bognar, E. E. Armstrong, and E. A. Watson, "Joint High Resolution Image Reconstruction from a Sequence of Rotated and Translated Frames and its Application to an Infrared Imaging System," *Optical Engineering*, vol. 37, pp. 247–260, January 1998.
49. N. A. Woods, N. A. Galatsanos, and A. K. Katsaggelos, "EM-based Simultaneous Registration, Restoration, and Interpolation of Super-resolved Images," in *Proc. IEEE Int. Conf. on Image Processing*, Barcelona, Spain, 2003.
50. R. R. Schultz and R. L. Stevenson, "A Bayesian Approach to Image Expansion for Improved Definition," *IEEE Trans. on Image Processing*, vol. 3, no. 3, pp. 233–242, May 1994.
51. N. Nguyen, P. Milanfar, and G. Golub, "Efficient Generalized Cross-Validation with Applications to Parametric Image Restoration and Resolution Enhancement," *IEEE Trans. on Image Processing*, vol. 10, no. 9, pp. 1299–1308, September 2001.
52. N. Nguyen, P. Milanfar, and G. Golub, "A Computationally Efficient Super-resolution Reconstruction Algorithm," *IEEE Trans. on Image Processing*, vol. 10, no. 4, pp. 573–583, April 2001.
53. M. Elad and A. Feuer, "Super-Resolution Restoration of an Image Sequence : Adaptive Filtering Approach," *IEEE Trans. on Image Processing*, vol. 8, no. 3, pp. 387–395, March 1999.
54. M. Elad and Y. Hel-Or, "A Fast Super-resolution Reconstuction Algorithm for Pure Translation Motion and Common Space-Invariant Blur," *IEEE Trans. on Image Processing*, vol. 10, no. 8, pp. 1187–1193, August 2001.
55. P. Jorge and S. G. Ferreira, "Two Fast Extrapolation/Super-Resolution Algorithms," in *Proc. IEEE Int. Conf. on Image Processing*, Vancouver, British Columbia, 2000, pp. 343–346.
56. J. H. Shin, J. S. Yoon, J. K. Paik, and M. abidi, "Fast Super-Resolution for Image Sequences using Motion Adaptive Relaxation Parameters," in *Proc. IEEE Int. Conf. on Image Processing*, Kobe, Japan, 1999, pp. 676–680.
57. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and Robust Super-resolution," in *Proc. IEEE Int. Conf. on Image Processing*, Barcelona, Spain, 2003.
58. A. Lorette, H. Shekarforoush, and J. Zerubia, "Super-Resolution with Adaptive Regularization," in *Proc. IEEE Int. Conf. on Image Processing*, Santa Barbara, California, 1997, pp. 169–172.

59. P. Charbonnier, *Reconstruction d'image: regularisation avec prise en compte des discontinuities*, Ph.D. thesis, UNSA, September 1994.
60. T. Hebert, , and R. Leahy, "A Generalized EM algorithm for 3D Bayesian Reconstruction from Poisson Data Using Gibb's Prior," *IEEE Trans. on Medical Imaging*, vol. MI-8, pp. 194–202, June 1989.
61. D. Geman and G. Reynolds, "Constrained Restoration and the Recovery of Discontinuities," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, March 1992.
62. M. C. Chiang and T. E. Boult, "Local Blur Estimation and Super-Resolution," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Puerto Rico, USA, 1997, pp. 821–826.
63. M. C. Chiang and T. E. Boult, "Efficient Super-Resolution via Image Warping," *Image and Vision Computing*, vol. 18, pp. 761–771, December 2000.
64. M. C. Chiang and T. E. Boult, "The Integrating Resampler and Efficient Image Warping," in *Proc. of the DARPA Image Understanding Workshop*, 1996, pp. 843–849.
65. A. Zomet, A. Rav-Acha, and S. Peleg, "Roboust Super-Resolution," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Kauai, HI, USA, 2001, pp. 645–650.
66. S. Zhao, H. Han, and S. Peng, "Wavelet-Domain HMT-Based Image Super-resolution," in *Proc. IEEE Int. Conf. on Image Processing*, Barcelona, Spain, 2003.
67. S. C. Park, M. G. Kang, C. A. Segall, and A. K. Katsaggelos, "Spatially Adaptive High Resolution Image Reconstruction of Low Resolution DCT based Compressed Images," in *Proc. IEEE Int. Conf. on Image Processing*, Rochester, New York, September, 2002.
68. C. A. Segall, R. Molina, A. K. Katsaggelos, and J. Mateos, "Reconctructon of High Resolution Image Frames from a Sequence of Low Resolution and Compressed Observations," in *Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing*, Florida, May, 2002, pp. 1701–1704.
69. Z. Lin and H. Y. Shum, "Fundamental Limits of Reconstruction-Based Super-Resolution Algorithms Under Local Translation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 83–97, January 2004.
70. D. R. Gerswe and M. A. Plonus, "Superresolved Image Reconstruction of Images taken through the Turbulant Atmosphere," *Journal of the Optical Society of America*, vol. 15, no. 10, pp. 2620–2628, October 1998.
71. Q. Yang and B. Parvin, "High Resolution Reconstruction of Sparse Data from Dense Low Resolution Spatio-Temporal Data," *IEEE. Trans. on Image Processing*, vol. 12, no. 6, pp. 671–677, June 2003.
72. B. K. Gunturk, , A. U. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, "Eigenface Based super-resolution for Face Recognition," in *Proc. IEEE Int. Conf. on Image Processing*, Rochester,New York, 2002, pp. 845–848.
73. B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, "Eigenface-Domain Super-resolution for Face Recognition," *IEEE Trans. on Image Processing*, vol. 12, no. 5, pp. 597–606, May 2003.
74. D. Capel and A. Zisserman, "Super-Resolution Enhancement of Text Image Sequences," in *Proc. IEEE Int. Conf. on Pattern Recognition*, Barcelona, Spain, 2000, pp. 600–605.

75. U. Bhosle, S. Gavali, and S. Chaudhuri, "Super-Mosaicing: Mosaicing at High Resolution," in *Proc. Tenth National Conf. on Communications*, Indian Institute of Science, Bangalore, 2004, pp. 106–110.
76. A. Zomet and S. Peleg, "Efficient Super-resolution and Applications to Mosaics," in *Proc. IEEE Int. Conf. on Pattern Recognition*, Barcelona, Spain, 2000, pp. 579–583.
77. D. Capel and A. Zisserman, "Automated Mosaicing with Super-resolution Zoom," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Santa Barbara, 1998, pp. 885–891.
78. A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution Video Reconstruction with Arbitrary Sampling Lattices and Nonzero Aperture Time," *IEEE Trans. on Image Processing*, vol. 6, no. 8, pp. 1064–1076, August 1997.
79. P. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust Object based High-resolution Image Reconstruction from Low Resolution Video," *IEEE Trans. on Image Processing*, vol. 6, no. 10, pp. 1446–1451, October 1997.
80. A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Robust Methods for High-Quality Stills from Interlaced Video in the Presence of Dominant Motion," *IEEE Trans. on Circuits and systems For Video Technology*, vol. 7, no. 2, pp. 328–342, April 1997.
81. R. R. Schultz and R. L. Stevenson, "Extraction of High-Resolution Frames from Video Sequences," *IEEE Trans. on Image Processing*, vol. 5, no. 6, pp. 996–1011, June 1996.
82. N. R. Shah and A. Zakhor, "Resolution Enhancement of Color Video Sequences," *IEEE Trans. on Image Processing*, vol. 8, no. 6, pp. 879–885, June 1999.
83. M. C. Hong, M. G. Kang, and A. K. Katsaggelos, "A Regularized Multichannel Restoration Approach for Globally Optimal High Resolution Video Sequence," in *Proc. SPIE Conf. on Visual Communications and Image Processing*, San Jose, CA, USA, 1997, pp. 1306–1316.
84. B. C. Tom and A. K. Katsaggelos, "Resolution Enhancement of Monochrome and Color Video Using Motion Compensation," *IEEE Trans. on Image Processing*, vol. 10, no. 2, pp. 279–287, February 2001.
85. Y. Altunbask, A. J. Patti, and R. M. Mersereau, "Super-Resolution Still and Video Reconstruction From MPEG- Coded Video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 4, pp. 217–226, april 2002.
86. C. A. Segall, R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian High-Resolution Reconstruction of Low-Resolution Compressed video," in *Proc. IEEE Int. Conf Image Processing*, Thessaloniki, Greece, 2001, pp. 25–28.
87. S. Chaudhuri and D. Taur, "High Resolution Slow Motion Sequencing," *IEEE Signal Processing Magazine (Special Issue on Visual Surveillance)*, Jan 2005.
88. H. Shekarforoush and R. Chellappa, "Data-driven Multichannel Super-Resolution with Applications to Video Sequences.," *Journal of the Optical Society of America A*, vol. 16, no. 3, pp. 481–492, 1999.
89. S. Borman and R. L. Stevenson, "Simultaneous Multi-frame MAP Super-Resolution Video Enhancement using Spatio-temporal Priors," in *Proc. IEEE Int. Conf. on Image Processing*, Kobe, Japan, October 1999, pp. 469–473.
90. Z. Wang and F. Qi, "Super-Resolution with Model Uncertainties," in *Proc. IEEE Int. Conf. on Image Processing*, Rochester, New York, 2002, pp. 853–856.

91. B. K. Gunturk, Y. Altunbasak, and R. Mersereau, "Bayesian Resolution-Enhancement Framework for Transform-Coded Video," in *Proc. IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, 2001, pp. 41–44.
92. E. Shechtman, Y. Caspi, and M. Irani, "Increasing Space-Time Resolution in Video," in *European Conference on Computer Vision*, Copenhagen, 2002, pp. 753–769.
93. D. Rajan and S. Chaudhuri, "Generation of Super-resolution Images from Blurred Observations using an MRF Model," *J. Mathematical Imaging and Vision*, vol. 16, pp. 5–15, 2002.
94. A. N. Rajagopalan and V. P. Kiran, "Motion-free Super-Resolution and the Role of Relative Blur," *Journal of the Optical Society of America A*, vol. 20, no. 11, pp. 2022–2032, November 2003.
95. I. Zakharov, D. Dovanar, and Y. Lebedinsky, "Super-Resolution Image Restoration from Several Blurred Images formed in Various Conditions," in *Proc. IEEE Int. Conf. on Image Processing*, Barcelona, Spain, 2003.
96. D. Rajan and S. Chaudhuri, "Generalized Interpolation for Super-Resolution," in *Super-Resolution Imaging*, S. Chaudhuri, Ed., pp. 45–72. Kluwer Academic Publisher,Boston, 2001.
97. W. T. Freeman, T.R.Jones, and E. C.Pasztor, "Example-Based Super-Resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, March/april 2002.
98. D. Capel and A. Zisserman, "Super-Resolution from Multiple Views Using Learnt Image Models," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Kauai, HI, USA, 2001, pp. II:627–634.
99. L. C. Pickup, S. J. Roberts, and A. Zisserman, "A Sampled Texture Prior for Image Super-Resolution," in *In Proc. of Advances in Neural Information Processing Systems*, 2003.
100. S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, September 2002.
101. A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image Analogies," in *Proc. ACM - SIGGRAPH*, 2001, pp. 327–340.
102. F. M. Candocia and J. C. Principe, "Super-Resolution of Images based on Local Correlations," *IEEE Trans. on Neural Networks*, vol. 10, no. 2, pp. 372–380, March 1999.
103. C. M. Bishop, A. Blake, and B. Marthi, "Super-Resolution Enhancement of Video," in *Int. Conf. on Artificial Intelligence and Statistics*, Key West, Florida, 2003.
104. A. P. Pentland, "Depth of Scene from Depth of Field," in *Proc. of Image Understanding Workshop*, Palo Alto, USA, 1982, pp. 253–259.
105. S. Chaudhuri and A. N. Rajagopalan, *Depth From Defocus: A Real Aperture Imaging Approach*, Springer-Verlog, NewYork, 1999.
106. M. Subbarao, "Efficient Depth Recovery Through Inverse Optics," in *Machine Vision for Inspection and Measurement*, H. Freeman, Ed. Academic Press, 1989.
107. A. P. Pentland, "A New Sense for Depth of Field," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 9, no. 4, pp. 523–531, July 1987.
108. M. Born and E. Wolf, *Principles of Optics*, Pergamon, London, 1965.

109. M. Subbarao, "Parallel Depth Recovery by Changing Camera Parameters," in *Proc. International Conference on Computer Vision*, Florida, USA, 1988, pp. 149–155.
110. K. V. Prasad, R. J. Mammone, and J. Yogeshwar, "3-D Image Restoration using Constrained Optimization Techniques," *Optical Engineering*, vol. 29, no. 4, pp. 277–288, April 1990.
111. W. N. Klarquist, W. S. Geisler, and A. C. Bovik, "Maximum-Likelihood Depth-from-Defocus for Active Vision," in *Proc. of IEEE Intl. Conf. on Intelligent Robots and Systems*, Pittsburgh, PA, 1995.
112. M. Gökstorp, "Computing Depth from Out-of-Focus Blur Using a Local Frequency Representation," in *Proc. of IEEE Intl. Conf. on Pattern Recognition*, Jerusalem, Israel, 1994, pp. 153–158.
113. M. Watanabe and S. K. Nayar, "Minimal operator set for passive DFD," in *Proc. of IEEE Intl. Conf.on Computer Vision and Pattern Recognition*, San Francisco, CA, 1996, pp. 153–158.
114. Y. Y. Schechner and N. Kiryati, "Depth from Defocus vs Stereo: How Different Really are They?," Tech. Rep. EE PUB No. 1155, Technion - Israel Institute of Technology, May 1998.
115. S. K. Nayar, M. Watanabe, and M. Noguchi, "Real-time Focus Range Sensor," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 12, pp. 1186–1198, Dec. 1996.
116. A. N. Rajagopalan and S. Chaudhuri, "Space-variant approaches to recovery of depth from defocused images," *Computer Vision and Image Understanding*, vol. 68, no. 3, pp. 309–329, Dec. 1997.
117. A. N. Rajagopalan and S. Chaudhuri, "A Variational Approach to Recovering Depth from Defocused Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1158–1165, Oct. 1997.
118. A. N. Rajagopalan and S. Chaudhuri, "An MRF Model Based Approach to Simultaneous Recovery of Depth and Restoration from Defocused Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 577–589, July 1999.
119. H. Shekarforoush, M. Bertod, J. Zerubia, and M. Werman, "Sub-pixel Bayesian Estimation of Albedo and Height," *International Journal of Computer Vision*, vol. 19, no. 3, pp. 289–300, 1996.
120. P. Cheeseman, B. Kanefsky, R. Hanson, and J. Stutz, "Super-Resolved Surface Reconstruction from Multiple Images," Tech. Rep. FIA-94-12, NASA Ames Research center, Mofett Field, CA, December 1994.
121. E. Ising, "Beitag Sur Theorie Des Ferromegnetisms," *Zeitschrift Physik*, vol. 31, pp. 253–258, 1925.
122. R. C. Dubes and A. K. Jain, "Random Field Models in Image Analysis," *Journal of Applied Statistics*, vol. 16, no. 2, pp. 131–164, 1989.
123. J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of Royal Statistical Society, Series B*, vol. 36, pp. 192–236, 1974.
124. S. Geman and D. Geman, "Stochastic Relaxation, Gibbs distribution and the Bayesian restoration of image," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
125. J. Subrahmonia, Y.P Hung, and D. B. Cooper, "Model Based Segmentation and Estimation of 3D Surfaces from Two or More Intensity Images Using Markov Random Fields," in *Proc. of IEEE Intl. Conf. on Pattern Recognition*, Atlantic Citys, USA, 1990, pp. 390–397.

126. D. B. Cooper, J. Subrahmonia, Y.P. Hung, and B. Cernuschi-Frias, "The Use of Markov Random Fields in Estimating and Recognizing Objects in 3D Space," in *Markov Random Fields : Theory and applications*, R. Chellappa and A. Jain, Eds., pp. 335–368. Academic Press, 1993.
127. S. Lakshmanan and H. Derin, "Gaussian Markov Random Fields at Multiple Resolutions," in *Markov Random Fields, Theory and Application*, R. Chellappa and A. K. Jain, Eds., pp. 131–157. Academic Press,Inc, 1993.
128. E. H. Adelson and J. R. Bergen, "The Plenoptic Function and the Elements of Early Vision," in *Computational Models of Visual Processing*, M. Landy and J. A. Movshon, Eds., pp. 3–20. Cambridge,MA:MIT Press, 1991.
129. T. Wong, C. Fu, P. Heng, and C. Leung, "The Plenoptic Illumination Function," *IEEE Trans. on Multimedia*, vol. 4, no. 3, pp. 361–371, September 2002.
130. M. Berthod, H. Shekarforoush, M. Verman, and J. Zerubia, "Reconstruction of High Resolution 3D Visual Information," Tech. Rep. RR-2142, INRIA, November 1993.
131. H. Shekarforoush, M. Berthod, and J. Zerubia, "3D Super-Resolution Using Generalised Sampling Expansion," in *Proc. IEEE Int. Conf. on Image Processing*, Washington DC, 1995, pp. 300–303.
132. Y. P. Hung and D. B. Cooper, "Maximum a Posteriori Probability 3D Surface Reconstruction Using Multiple Intensity Images Directly," *SPIE*, vol. 1260, pp. 36–48, 1990.
133. S. G. Deshpande and S. Chaudhuri, "Recursive Estimation of Illuminant Motion from Flow Field and Simultaneous Recovery of Shape," *Computer Vision and Image Undersatanding*, vol. 72, no. 1, pp. 10–20, October 1998.
134. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-posed Problems*, W. H. Winston, Washington D. C., 1977.
135. A. N. Tikhonov, "Solution of Incorrectly Formulated Problems and the Regularization method," in *Soviet Math.*, Dokl., 1963, pp. 1035–1038.
136. M. Bertero, "Regularization Methods for Linear Inverse Problems," in *Inverse Problems*, C. G. Talenti, Ed. Springer Verlog, Berlin, 1986.
137. V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, Berlin, 1984.
138. B. K. P Horn, *Robot Vision*, MIT Press, 1986.
139. B. K. P. Horn, *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*, Ph.D. thesis, Massachussetts Inst. of Technology, 1970.
140. K. Ikeuchi and B. K. P. Horn, "Numerical Shape from Shading and Occluding Boundaries," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 141–184, 1981.
141. A. P. Pentland, "Local Shading Analysis," *IEEE Trans. Pattern analysis and Machine Intelligence*, vol. 6, no. 2, pp. 170–187, 1984.
142. A. P. Pentland, "Shape Information from Shading: A Theory about Human Perception," in *Proc. IEEE Intl. Conf. on Computer Vision*, Florida, USA, 1988, pp. 404–413.
143. P. S. Tsai and M. Shah, "Shape from Shading Using Linear Approximation," *Image and Vision Computing*, vol. 12, no. 8, pp. 487–498, 1994.
144. R. J. Woodham, "Reflectance Map Techniques for Analyzing Surface Defects in Metal Castings," Tech. Rep. 457, MIT AI Lab, June 1978.
145. R. J. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *Optical Engineering*, vol. 19, no. 1, pp. 139–144, 1980.

146. W. M. Silver, "Determining Shape and Reflectance Using Multiple Images," M.S. thesis, Department of Electrical and Computer Science, MIT, Cambridge, 1980.
147. K. Ikeuchi, "Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method," *IEEE Trans. on Pattern analysis and Machine Intelligence*, vol. 3, no. 6, pp. 661–669, 1981.
148. K. M. Lee and C. C. Jay Kuo, "Shape Reconstruction from Photometric Stereo," in *Proc. International Conf. on Computer Vision and Pattern Recognition*, Champaign, Illinois, USA, 1992, pp. 479–484.
149. C. Y. Chen, R. Klette, and R. Kakarala, "Albedo Recovery Using a Photometric Stereo Approach," in *Proc. International Conference on Pattern Recognition*, Quebec, Canada, 2002, pp. 700–703.
150. Y. Iwahori, R. J. Woodham, M. Ozaki, H. Tanaka, and N. Ishi, "Neural Network based Photometric Stereo with a Nearby Rotational Moving Light Source," *IEICE Transactions on Information and Systems*, vol. E-80-D, no. 9, pp. 948–957, 1997.
151. Y. Iwahori, R. J. Woodhan, and A. Bagheri, "Principal Component Analysis and Neural Network Implementataion of Photometric Stereo," in *IEEE Workshop on Physics based Modeling in Computer Vision*, Boston, MA, 1995, pp. 117–125.
152. R. J. Woodham, "Gradient and Curvature from the Photometric Stereo Method including Local Confidence Estimation," *Journal of the Optical Society of America-A*, vol. 11, no. 11, pp. 3050–3068, 1994.
153. Y. Iwahori, R. J. Woodham, Y. Watanabe, and A. Iwata, "Self Calibration and Neural Network Implementation of Photometric Stereo," in *Proc. International Conf. on Pattern Recognition*, Quebec, Canada, 2002, pp. 359–362.
154. O. Drbohlav and R. Sara, "Unambiguous Determination of Shape from Photometric Stereo with Unknown Light Sources," in *Proc. International Conf. on Computer Vision*, Vancouver, BC, Canada, 2001, pp. 581–586.
155. R. Basri and D. Jacobs, "Photometric Stereo with General Unknown Lighting," in *Proc. International Conf. on Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001, pp. 374–381.
156. U. Sakarya and I. Erkmen, "An Improved Method of Photometric Stereo Using Local Shape from Shading," *Image and Vision Computing*, vol. 21, no. 11, pp. 941–954, 2003.
157. J. J. Clark, "Active Photometric Stereo," in *Proc. International Conf. on Computer Vision and Pattern Recognition*, Champaign, Illinois, USA, 1992, pp. 29–34.
158. J. R. A. Torreao, "A New Approach to Photometric Stereo," *Pattern Recognition Letters*, vol. 20, no. 5, pp. 535–540, 1999.
159. G. McGunnigle and M. J. Chantler, "Rotation Invariant Classification of Rough Surfaces," *IEE Proc. Vis. Image Signal Process*, vol. 146, 1999.
160. G. McGunnigle and M. J. Chantler, "Rough Surface Classification Using Point Statistics from Photometric Stereo," *Pattern Recognition Letters*, vol. 21, no. 6-7, pp. 593–604, 2000.
161. G. McGunnigle and M. J. Chantler, "Modelling Deposition of Surface Texture," *Electron. Letters*, vol. 37, pp. 749–750, 2001.
162. M. L. Smith, T. Hill, and G. Smith, "Gradient Space analysis of Surface Defects Using a Photometric Stereo Derived Bump Map," *Image and Vision Computing*, vol. 17, no. 3-4, pp. 321–332, 1999.

163. P. Hansson and P. Johansson, "Topography and Reflectance Analysis of Paper Surfaces Using Photmetric Stereo Method," *Optical Engineering*, vol. 39, pp. 2555–2561, 2000.

164. G. McGunnigle and M. J. Chantler, "Segmentation of Machined Surfaces," in *Irish Machine Vision and Image Processing Conference*, 2001, pp. 200–207.

165. M. R. Banham and A. K. Katsaggelos, "Digital Image Restoration," *IEEE Signal Processing Magazine*, vol. 14, no. 2, pp. 24–41, 1997.

166. D. Kundur and D. Hatzinakos, "Blind Image Deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, 1996.

167. Y. You and M. Kaveh, "A Regularization Approach to Joint Blind Identification and Image Restoration," *IEEE Trans. on Image Processing*, vol. 5, pp. 416–428, March 1996.

168. R. L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*, Kluwer Academic Publishers, 1991.

169. R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Identification and Restoration of Noisy Blurred Images Using the Expectation Maximization Algorithm," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 38, no. 7, pp. 1180–1191, July 1990.

170. K. T. Lay and A. K. Katsaggelos, "Image Identification and Restoration based on the Expectation Maximization Algorithm," *Optical Engineering*, vol. 29, pp. 436–445, May 1990.

171. M. K. Ozkan, A. M. Tekalp, and M. I. Sezan, "POCS based Restoration of Space Varying Blurred Images," *IEEE Trans. Image Processing*, vol. 3, pp. 450–454, July 1995.

172. M. I. Sezan and A. M. Tekalp, "Iterative Image Restoration with Ringing Supression Using POCS," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Tokyo, 1988, pp. 1300–1303.

173. M. I. Sezan and H. J. Trussell, "Prototype Image Constraints for Set Theoretic Image Restoration," *IEEE trans. Acoustics, Speech and Signal Processing*, vol. 39, no. 10, pp. 2275–2285, October 1991.

174. H. J. Trussell and M. R. Civanlar, "The Feasible Solution in Signal Restoration," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 32, no. 2, pp. 201–212, April 1984.

175. Y. Yang, N. P. Galatsanos, and A. K. Katsaggelos, "Projection-based Spatially Adaptive Reconstruction of Block-Transform Compressed Images," *IEEE Trans. on Image Processing*, vol. 4, no. 7, pp. 896–908, July 1995.

176. V. F. Candela, A. Marquina, and S. Serna, "A Local Spectral Inversion of a Linearized TV Model for Denoising and Deblurring," *IEEE Trans. on Image Processing*, vol. 12, no. 7, pp. 808–816, 2003.

177. B. Besserer L. Joyeux, S. Boukir and O. Buisson, "Reconstruction of Degraded Image Sequences. Application to Film Restoration," *Image and Vision Computing*, vol. 19, no. 8, pp. 503–516, 2001.

178. L. Bedini, A. Tonazzini, and S. Minutoli, "Unsupervised Edge-Preserving Image Restoration Via a Saddle Point Approximation," *Image and Vision Computing*, vol. 17, no. 11, pp. 779–793, 1999.

179. T. F. Chan and C. K. Wong, "Convergence of the Alternating Minimization Algorithm for Blind Deconvolution," Tech. Rep. 19, UCLA Computational and Applied Mathematics, June CAM99.

180. N. Kaulgud, J. Karlekar, and U. B. Desai, "Compressed Domain Video Zooming : Use of Motion Vectors," in *Proc. Eighth National Conf. on Communications*, Mumbai, India, Jan. 2002, pp. 120–124.
181. S. Thurnhofer and S. K. Mitra, "Edge-Preserving Image Zooming," *Optical Engineering*, vol. 35, no. 7, pp. 1862–1870, 1996.
182. F. Malagouyres and F. Guichard, "Edge Direction Preserving Image Zooming: A Mathematical and Numerical Analysis," *SIAM Journal on Numerical Analysis*, vol. 39, no. 1, pp. 1–37, 2001.
183. C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, New Jersy, 1998.
184. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pennsylvania, 1992.
185. A. Jensen and A. Cour-Harbo, *Ripples in Mathematics: The Discrete Wavelet Transform*, Springer-Verlag, 2001.
186. H. M. Shapiro, "Embedded Image Coding," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, December 1993.
187. E. J. Candès and D. L. Donoho, "Curvelets - a Surpisingly Effective Nonadaptive Representation for Objects with Edges," in *Curve and Surface Fitting*, A. Cohen, C. Rabut, and L. L. Schumaker, Eds. Saint-Malo: Vanderbilt University Press, 1999.
188. M. N. Do, *Directional Multiresolution Image Representations*, Ph.D. thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland, 2001.
189. D. L. Donoho, "Wedgelets: Nearly Minimax Estimation of Edges," *Ann. Statist*, vol. 27, no. 3, pp. 859–897, 1999.
190. A. Cohen and B. Matei, "Compact Representation of Images by Edge Adapted Multiscale Transforms," in *IEEE Int. Conf. on Image Processing, Invited paper, Greece*, October, 2001.
191. E. J. Candès and D. L. Donoho, "Ridgelets: A Key to Higher Dimensional Intermittency," *Phil. Trans. R. Soc. Lond. A.*, vol. 357, pp. 2495–2509, 1999.
192. M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
193. G. F. Carey, *Computational Grids: Generation Adaptation and Solution Strategies*, Taylor and Francis, 1997.
194. M. J. D. Powell, "The Uniform Convergence of Thin Plate Spline Interpolation in Two Dimensions," *Numerishce Mathematik*, vol. 68, pp. 107–128, 1994.
195. P. Dani and S. Chaudhuri, "Automated Assembling of Images: Image Montage Preparation," *Pattern Recognition*, vol. 28, no. 3, pp. 431–445, March 1995.
196. Hong Jing and Liudi Liu, "Color Space Conversion Methods' Applications to the Image Fusion," *Optical Technology*, , no. 4, pp. 44–48, 1997.
197. P. S. Jr. Chavez, S. C. Sides, and J. A. Anderson, "Comparison of Three Different Methods to Merge Multiresolution and Multispectral Data: Landsat TM and SPOT Panchromatic," *Photogrammetric Engineering and Remote Sensing*, vol. 57, no. 3, pp. 295–303, 1991.
198. M. Ehlers, "Multisensor Image Fusion Techniques in Remote Sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 46, pp. 19–30, 1991.
199. Eryong Yu and Runsheng Wang, "Fusion and Enhancement of the Multispectral Image with Wavelet Transform," *Computer Engineering and Science*, vol. 23, no. 1, pp. 47–50, 2001.

200. B. Aiazzi, L. alparone, S. Baronti, and A. Garzelli, "Context Driven Fusion of High Spatial and Spectral resolution Images based on Oversampled Multiresolution Analysis," *IEEE trans. on Geoscience and remote sensing*, vol. 40, no. 10, pp. 2300–2312, October 2002.
201. G. L. Bilbro, W. E. Snyder, D. E. Van den Vout, and T. K. Millarand M. W. White, "Optimization by Mean Field Annealing," in *Advances in Neural Information Processing*, D. S. Touretzky, Ed., pp. 91–98. Morgan Kaufmann, 1989.
202. G. L. Bilbro and W. E. Snyder, "Range Image Restoration Using Mean Field Annealing," in *Advances in Neural Information Processing*, D. S. Touretzky, Ed., pp. 594–601. Morgan Kaufmann, 1989.
203. R. G. Wilson and S. A. Shafer, "What is the Center of the Image?," Tech. Rep. Technical Report CMU-CS-93-122, Carnegie Mellon University, 1993.
204. R. G. Wilson, *Modeling and Calibration of Automated Zoom Lenses*, Ph.D. thesis, Carnegie Mellon University, January 1994.
205. A. N. Rajagopalan and S. Chaudhuri, "Performance Analysis of Maximum Likelihood Estimator for Recovery of Depth from Defocused Images and Optimal Selection of Camera Parameters," *International Journal of Computer Vision*, vol. 30, no. 3, pp. 175–190, December 1998.
206. Y. Yu and Q. Cheng, "MRF Parameter Estimation by an Accelerated Method," *Pattern Recognition Letters*, vol. 24, pp. 1251–1259, 2003.
207. S. Lakshamanan and H. Derin, "Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields Using Simulated Annealing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 8, pp. 799–813, August 1989.
208. S. G. Nadabar and A. K. Jain, "Parameter Estimation in MRF Line Process Models," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 1992, pp. 528–533.
209. G. Potamianos and J. Goutsias, "Partition Function Estimation of Gibbs Random Field Images Using Monte Carlo Simulations," *IEEE Trans. on Information Theory*, vol. 39, no. 4, pp. 1322–1331, July 1993.
210. G. Potamianos and J. Goutsias, "Stochastic Approximation Algorithms for Partition Function Estimation of Gibbs Random Fields," *IEEE Trans. on Information Theory*, vol. 43, no. 6, pp. 1948–1965, November 1997.
211. S. C. Zhu, Y. N. Wu, and D. Mumford, "Minimax Entropy Principle and Its Application to Texture Modeling," *Neural Computation*, vol. 9, no. 8, pp. 1627–1660, 1997.
212. S. C. Zhu, Y. N. Wu, and D. Mumford, "Filters, Random Fields And Maximum Entropy," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 1–20, March/April 1998.
213. S. C. Zhu and X. Liu, "Learning in Gibbsian Fields: How Accurate and How Fast Can It Be?," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 1001–1006, July 2002.
214. P. K. Nanda, K. Sunil Kumar, S. Ghokale, and U. B. Desai, "A Multiresolution Approach to Color Image Restoration and Parameter Estimation Using Homotopy Continuation Method," in *Proc. IEEE Int. Conf. on Image Processing*, Washington, D.C., USA, 1995, pp. 2045–2048.
215. R. Kashyap and R. Chellappa, "Estimation and Choice of Neighbors in Spatial-Interaction Models of Images," *IEEE trans. on Information Theory*, vol. 29, no. 1, pp. 60–72, January 1983.

216. J. G. Proakis and D. G. Manolakis, *Digital Signal Processing Principles, Algorithms, and Applications*, Prentice Hall, New Jersey, USA, 1995.
217. J. Mao and A. K. Jain, "Texture Classification and Segmentation using Multiresolution Simultaneous Autoregressive Models," *Pattern Recognition*, vol. 25, no. 2, pp. 173–188, 1992.
218. J. Bennett and A. Khotanzad, "Multispectral Random Field Models for Synthesis and Analysis of Color Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 327–332, March 1998.

Index