SINGLE-FRAME IMAGE SUPER-RESOLUTION

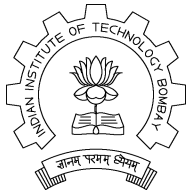
Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

Jiji C.V.

Roll No. : 03407302



DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY - BOMBAY

2007

Approval Sheet

Thesis entitled "SINGLE-FRAME IMAGE SUPER-RESOLUTION" by **Jiji C. V.** is approved for the degree of **DOCTOR OF PHILOSOPHY**.

Examiners

_____

_____

_____

Supervisor

_____

_____

Chairman

_____

Date:_____

INDIAN INSTITUTE OF TECHNOLOGY,

BOMBAY,

INDIA

CERTIFICATE OF COURSE WORK

This is to certify that Mr. Jiji C. V. was admitted to the candidacy of the Ph.D. Degree on July 21, 2003 after successfully completing all the courses required for the Ph.D. Degree programme. The details of the course work done are given below.

Sr. No.	**Course Name**	**Credits**
1.	Image Processing	6
2.	Statistical Signal Analysis	6
3.	Applied Linear Algebra in Electrical Engineering	6
4.	Wavelets	Audit
6.	Computational Vision	6
7.	Credit Seminar	4
8.	Communication and Presentation Skills	

I.I.T Bombay

Date:

Dy. Registrar (Academic)

Abstract

Super-resolution refers to the task of estimating a high resolution image from one or more low resolution observations. Super-resolution techniques handle the issues of alias removal, de-blurring and de-noising while interpolating the low resolution inputs. Most of the research on super-resolution is for multi-frame super-resolution where the nonredundant information among multiple sub-pixel shifted low resolution observations are used to produce a high resolution image. Recently, there have been some works on single frame image super-resolution where only one low resolution observation is used to construct the super-resolved image. These techniques are based on statistical and machine learning approaches where the missing high frequency details in the low resolution image is learned from an appropriate image database. In this thesis we propose some new approaches for recovering a high resolution image from a single low resolution observation making use of a training set containing several high resolution images.

First we propose a single frame image super resolution method using a learned wavelet prior on the high resolution image. We learn the wavelet coefficients at finer scales of the unknown high resolution image locally from a set of high resolution training images and the learned image in the wavelet domain is used for further regularization while super-resolving the picture. We use an appropriate discontinuity preserving smoothness prior in addition to the wavelet based constraint to estimate the super-resolved image. Thus, we minimize a cost function containing the data consistency term, the smoothness constraint and the learned prior in the wavelet domain. The smoothness term ensures the spatial correlation among the pixels whereas the learning term chooses the best edges from the training set while super-resolving

the input low resolution image.

Using wavelets, though we expect to catch the horizontal and vertical edges, edges oriented along arbitrary directions are not handled properly. To overcome this difficulty we propose a learning technique using the contourlet transform which is capable of catching the smoothness along contour naturally. The contourlet coefficients at finer scales of the unknown high resolution image are learned locally from a set of high resolution training images, the inverse contourlet transform of which recovers the super-resolved image. This method is very fast as it is not required to solve the super-resolution problem under a regularizations framework unlike the wavelet-based approach.

We further modify the contourlet-based learning algorithm under a supervised learning approach and making use of the concept of locally linear embedding (LLE), a dimensionality reduction technique. Here the generation of a high resolution patch does not depend on only one of the nearest neighbors in the training set. Instead, it depends simultaneously on multiple nearest neighbors in a way similar to LLE for manifold learning. Here we make use of the low resolution and the corresponding high resolution training image pairs to learn the oriented edges for high resolution embedding.

As the above methods cannot handle the blur and noise present in the observation, we describe a principal component analysis (PCA) based global method which faithfully reproduces a super-resolved image from a blurred and noisy low resolution input. Here we propose two approaches to enhance the low resolution input. In the first method, a few significant eigenimages of a database of several similar low resolution images are computed and the given low resolution image is projected onto the eigenimages to compute the projection coefficients. The high resolution image is reconstructed using these coefficients and the high resolution eigenimages obtained by performing generalized interpolation on the low resolution eigenimages. In the second approach, we use a high resolution database and compute the principal eigenimages. The only difference here is that the upsampling process is shifted to the input side rather than performing generalized interpolation of the eigenimages.

To extract the advantages of both the local and global methods, we also propose a hybrid approach for super-resolution by integrating the global learning using PCA and the local learning approach using wavelets or contourlets. The former extracts the best statistical features from a database of similar class if images while the latter ensures the best edges from the database. We solve the problem under a Bayesian framework incorporating the global and local priors, in addition to the data consistency term. We apply the method particularly for face hallucination where we synthesize the missing high frequency details of the input low resolution face image.

As all the above methods are unable to handle the aliasing effects in the low resolution input, finally we propose an alias-free upsampling scheme. We generate the unknown high frequency components of the given partially aliased low resolution image by minimizing the total variation of the interpolant subject to the constraint that part of unaliased spectral components in the low resolution observation are known precisely and under the assumption of sparsity in the data. This provides a mathematical basis for exact reproduction of high frequency components with probability approaching one, from their aliased observation. Unlike the other methods, here we do not use a training set to produce the interpolated image.

Contents

List of Figures

List of Tables

List of Symbols

f	:	low resolution input image
$\mathbf{f}$	:	Lexicographically ordered vector of low resolution image
z	:	super-resolved image
$\mathbf{z}$	:	Lexicographically ordered vector of super-resolved image
A	:	Blurring matrix
q	:	zoom factor
D	:	Decimation matrix
$\mathbf{n}$	:	noise vector
K	:	Number of training images
$\chi()$	:	Wavelet decomposition
$\Gamma()$	:	Contourlet decomposition
Σ	:	Covariance matrix
$\boldsymbol{\mathcal{F}}_i$	:	i^{th} training image
$\mathbf{m}_{\mathcal{F}}$	:	Average image intensity
$\mathbf{w}$	:	weight vector
E	:	matrix of eigen vectors
$\mathbf{h}$	:	blur point spread function
$\mathbf{f}_b$	:	blurred image vector

Chapter 1

Introduction

Today, one deals with digital images on a daily basis. Digital image processing enables one to extract the pictorial details and useful information from a given low quality image. Image processing ideas were used way back in 1920 for the transmission of pictures through cables. Many of the techniques of digital image processing, were developed in the 1960s at the Jet Propulsion Laboratory, MIT, Bell Labs, etc. with application to satellite imagery, wire photo standards conversion and medical imaging. The field if image processing has grown considerably during the last few decades, with the developments in digital computer technology in terms of improvement in size, speed and cost effectiveness. Today, the technology has reached such a level that one can purchase a compact image processing system off the shelf to capture and manipulate an image the way one wants.

The basic functional elements of an image processing system are (i) image acquisition module such as a video camera, a scanner, a frame grabber (ii) processing and storage module such as a general purpose computer and its memory and (iii) a display module e.g., monitor, printer, etc. The digital image processing field has several applications in microscopy, medical imaging, astronomy, remote sensing, and law enforcement. For all these applications it is always desirable to have an image providing more detailed information. In other words, we always prefer an image with high resolution (HR). For example, with high resolution imaging one can obtain a better classification of regions in a multi spectral image, a more pleasing view in

a high definition television (HDTV), an accurate recognition in a face recognition scheme and a precise localization of a tumor in a medical image, etc. The resolution of an image is dependent on the resolution of the image acquisition device. Thus, to acquire a high resolution image one will need complex sensors which may be very costly and may not be an affordable solution for all situations.

1.1 Definition of Image Resolution

One of the most important technical terms to understand in image manipulation is the word resolution which judges the quality of various image acquisition and processing devices. In its simplest form, *image resolution* is defined as the smallest 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 measures the response of the system to different spatial frequencies. In digital image processing, the term resolution is classified into three different types.

- *Spatial or Pixel Resolution*: An image is composed of several picture elements called pixels. Spatial resolution or pixel resolution refers to the spacing of pixels in an image and is measured with the set of two positive integer numbers, where the first number is the number of pixel columns (width) and the second is the number of pixel rows (height), for example as 640 by 480. Higher spatial resolution allows a clear perception of sharp details and subtle color transitions in an image. If an image having high levels of details is not represented by a spatially dense set of pixels, the image is said to suffer from aliasing artifacts giving rise to a blocky effect. For an output device such as a printer the spatial resolution is expressed in dots per inch (dpi) or meter per pixel in absolute term.

- *Brightness Resolution*: The brightness resolution is also referred to as the gray scale or color range of an individual pixel. The brightness or color value of each pixel is defined by one bit or a group of bits. The more the number of bits used per pixel in an image, the higher the brightness resolution and the larger

the image file. The brightness resolution of monochrome images is usually 256 levels implying that a pixel value is represented by 8 bits. For color images, at least 24 bits are used to represent one brightness level, i.e., 8 bits per color plane(red, green, and blue). This is more commonly referred to as dynamic range or amplitude quantization.

- *Temporal Resolution*: This refers to the frame rate or the number of frames captured per second. Higher frame rates result in less smearing due to movements in the scene. The lower limit on the temporal resolution is directly proportional to the expected motion during two subsequent frames. The typical frame rate for a pleasing view is about 25 frames per second or above.

- *Spectral Resolution*: This refers to the frequency or spectral resolving power of a sensor and is defined as the smallest resolvable wavelength difference by the sensor.

In this report, the term resolution always refers to spatial or pixel resolution.

1.2 Why Resolution Enhancement?

Since the 1970s, charge-coupled device(CCD) and CMOS image sensors have been widely used to capture digital images. The number of sensor elements decide the spatial resolution of the camera. Higher the number, more is the resolution. A camera with less number of sensor elements produces a low resolution (LR) image, giving a blocky effect. Although these sensors are suitable for most imaging applications, the current resolution level and consumer price will not satisfy the future demand. For example, people want an inexpensive HR digital camera/camcorder and scientists often need a very HR level close to that of an analog 35 mm film that has very little visible artifacts when an image is magnified. Thus, finding a way to enhance the current resolution level is essential.

The most direct solution to increase spatial resolution is to reduce the pixel size (i.e., increase the number of pixels per unit area) by sensor manufacturing techniques.

As the pixel size decreases, however, the amount of light available also decreases. It generates shot noise [1] that degrades the quality of the image captured. To reduce the pixel size without suffering the effects of shot noise, therefore, there exists the limitation of the pixel size reduction, and the optimally limited pixel size is estimated at about $40\mu m^2$ for a $0.35\mu m$ CMOS process. The current image sensor technology has almost reached this level. Another approach for enhancing the spatial resolution is to increase the chip size,which leads to an increase in capacitance [2]. Since large capacitance makes it difficult to speed up a charge transfer rate, this approach is not considered effective. The high cost for high precision optics and image sensors is also an important concern in many commercial applications regarding HR imaging. Therefore, a new cost effective approach toward increasing spatial resolution is required to overcome these limitations of the sensors and lens manufacturing technology.

A promising approach is to use post acquisition signal processing techniques to enhance the resolution. These techniques being post processing methods applied on the already acquired low resolution images, they offer flexibility as well as cost benefit since theres is no additional hardware cost involved. However, the increased computational cost may be the burden that the user has to bear.

1.3 What is Super-Resolution?

The image captured using a low resolution camera suffers from three main drawbacks namely aliasing, blurring and the presence of noise. Aliasing occurs due to the loss of high frequency details because of lower spatial sampling. This causes loss of important informations such as edges and textures in the image. Also degradation occurs due to the sensor point spread function (PSF), and optical blurring due to camera motion or out of focus. *Super-resolution* (SR) refers to the process of producing a high spatial resolution image from one or more low resolution images, thereby recovering the missing high frequency details and removing the degradations that arise during the image capturing process. In effect, the super-resolution process extrapolates the high frequency components and minimizes aliasing and blurring.

As explained in the previous section, one way to obtain high resolution images is by reducing the pixel size and thereby increasing the pixel density. But an increase in pixel density causes shot noise and also the sensor cost increases with increase in pixel density. So we have to resort to some signal processing approaches to construct the high resolution image making use of the available low resolution images. A simple approach is to use interpolation techniques to increase the size of the given image. But conventional interpolation methods such as nearest neighbor, bilinear and bicubic interpolation generate images which are severely blurred. This blurring effect is related to the loss of details, which are directly related to the high frequencies in the image. As these methods do not solve any of the three issues in super-resolution they are not considered as SR techniques.

A natural solution to SR reconstruction is to make use of multiple sub-pixel shifted low resolution observations of the same scene and extract the non redundant information among them. This is termed as multi-frame super-resolution in the literature. This technique involves exact registration among the different frames followed by interpolation and finally restoration to remove blur and noise. Most of the literature available on SR is based on multi-frame super-resolution. Recently, there have been some works on single frame image super-resolution where only one low resolution observation is used to construct the super-resolved image. These techniques are based on statistical and machine learning approaches where the missing high frequency details in the low resolution image are learned from an appropriate image database. SR methods under this category are also known as learning based super-resolution techniques which is the subject matter of this thesis.

It may also be noted that there are other forms of super-resolution techniques like optical super-resolution [3] where the high frequency contents beyond the diffraction limit are extracted, and subwavelength imaging which achieve a resolution beyond the diffraction limit (spurred by recent developments in *superlens* made from a material having a negative index of diffraction) [4]. Such issues are not dealt with in this thesis. The focus of this thesis is in the use if digital post processing techniques to address issues such as alias removal, de-blurring and de-noising.

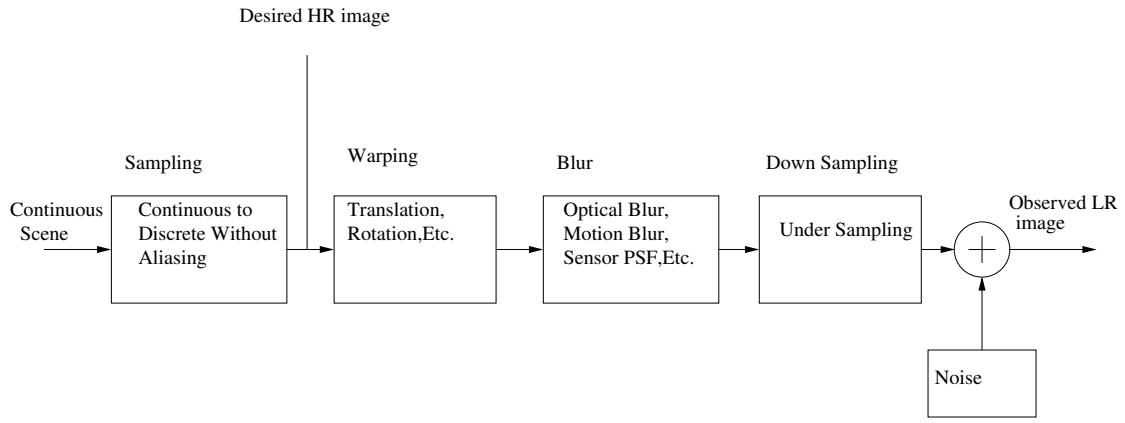


Figure 1.1: Observation model relating LR images to HR image.

1.4 Observation Model for Super-Resolution Reconstruction

The first step to analyze the SR image reconstruction problem is to formulate an observation model that relates the original HR image to the observed LR images. The success of any SR reconstruction method is closely dependent on the correctness of the LR image formation model. The most common model is based on observations which are shifted, blurred and aliased versions of the high resolution image. A general observation model for low resolution observations relating to the high resolution image is shown in 1.1. The desired high resolution image is obtained by sampling at or above the Nyquist rate from a continuous scene which is assumed to be band limited. It is assumed that the HR image remains constant during the acquisition of the multiple LR images, except for any motion and degradation allowed by the model. The relative motion between the camera and the scene may cause blurring and low resolution scanning is equivalent to subsampling or decimation. Therefore, the observed LR images result from warping, blurring, and sub-sampling operations performed on the HR image followed by noise addition.

A simplified low resolution image formation model is illustrated in Figure 1.2. The LR image formed using this model will be aliased as well as blurred and noisy. It is to be noted that without the blur, the LR observation will suffer from aliasing severely.

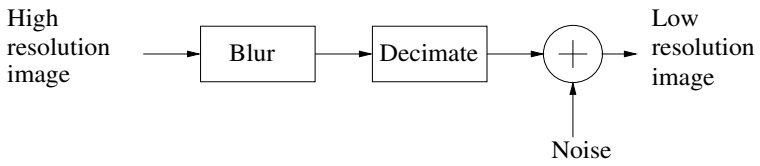


Figure 1.2: Simplified low resolution image formation model.

However, the HR restoration of the original image becomes increasingly difficult with the increase in amount of blurring [5].

Thus the problem we investigate in this thesis is as follows. Given a single instance of the low resolution observation, obtain an estimate of the high resolution image. Ideally the super-resolution process should reduce the effects of noise and blur while recovering an alias-free upsampled version of the given LR image. As it is difficult to deal with all these three issues simultaneously, different forms will be assumed for the blurring kernel and noise in subsequent chapters. Since the decimation process is not invertible, an arbitrary set of high resolution images is used as training data to help recover the unknown high resolution image.

1.5 Contributions of the Thesis

Super-Resolution is the problem of generating a high resolution image from one or more low resolution images. While most of the methods proposed for SR is based on multiple low resolution images of the same scene, the focus of this thesis is in generating a high resolution image from a single low resolution image. We call this problem as single-frame image super-resolution. For this we make use of a set of good quality training images and learn the best features from them. Such single image super-resolution problem arises in a number of real world applications. A common application occurs in digital photography when one wants to increase the resolution of an image while resizing it using image processing software like Adobe Photoshop ®. Another application can be found in web pages with images. To save storage space and communication bandwidth, it would be desirable if the low resolution image is downloaded and enlarged on the user's machine with an appropriate super-resolution

algorithm. To take care of several such applications we are proposing some novel single frame image super-resolution methods in this thesis. For this purpose we make use of a database of several good quality training images. It may be noted that, by the term training we really mean exemplar images from where we try to learn the best features while super-resolving the given low resolution input. In super-resolution one always aims at generating the missing high frequency components. To this end, we also propose a method to quantify that the generated details are really the high frequency details missing in the given low resolution observation. The contributions of the thesis, in summary, are:

- First we propose a single frame, learning based image super resolution method using a learned wavelet prior on the high resolution image. We learn the wavelet coefficients at finer scales of the unknown high resolution image locally from a set of high resolution training images and the learned 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 addition to the wavelet based constraint to estimate the super-resolved image. Thus we minimize a cost function containing the data consistency term, the smoothness constraint and the learned prior in the wavelet domain. The smoothness term ensures the spatial correlation among the pixels whereas the learning term chooses the best HR edges from the training set. Because this amounts to extrapolating the high frequency components, the proposed method does not suffer from over smoothing effects.

- One of the major difficulties 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 gives rise to certain artifacts in the reconstructed image and in order to get a good quality super-resolved image we were forced to use an ap-

propriate discontinuity preserving smoothness constraint under a regularization framework. Thus we ensure spatial correlation among pixels using the smoothness constraint, as well as obtain the best matching edges from the training set using wavelet learning. This required a stochastic optimization technique to obtain the solution which made the reconstruction process very slow.

A better way to handle the above situation is to use directionally selective wavelet decomposition to learn the oriented edges where the reconstruction problem need not be solved under a regularization framework, resulting in a much faster solution. This motivated us to use the contourlet transform [6] which is capable of catching the smoothness along contour naturally. They offer a much richer set of directions and shapes, and thus, they are more effective in capturing smooth contours and geometric structure in the image. The contourlet coefficients at finer scales of the unknown high resolution image are learned locally from a set of high resolution training images, the inverse contourlet transform of which recovers the super-resolved image. In effect, we learn the high resolution representation of an oriented edge primitive from the training data. Our experiments show that the proposed approach outperforms standard interpolation techniques as well as a standard wavelet-based learning both visually and in terms of the peak signal to noise ratio (PSNR) values, especially for images with arbitrarily oriented edges.

- We further modify the contourlet-based learning algorithm under a supervised learning approach and making use of the concept of locally linear embedding (LLE), a dimensionality reduction technique. Here the generation of a high resolution patch does not depend only on one of the nearest neighbors in the training set. Instead, it depends simultaneously on multiple nearest neighbors in a way similar to LLE for manifold learning. Here we make use of the low resolution and the corresponding high resolution training image pairs to learn the oriented edges to estimate the high resolution embedding. The advantage of this lies in capturing the effect of the decimation process explicitly by the

contourlet transform through the matched pair of LR and HR training data. In the previous methods it has been implicitly assumed that the multiresolution decomposition using either wavelet or contourlet has the same effect as that of the decimation process, which is often not true. For a given LR test image patch (in contourlet domain) we find k nearest neighbors from the LR training set by performing an extensive search. Now the test patch is reconstructed from a linear combination of the k neighbors where the reconstruction weights are computed so as to minimize the reconstruction error. The linear combination of the corresponding k HR patches and the same reconstruction weights generate the required super-resolved patch in the contourlet domain. This is repeated for all the patches in the test image.

- All the above methods are based on learning the features locally and are good at enhancing the edges during super-resolution, but are poor in handling noisy data. Further, they are unable to undo the effect of aliasing. But in certain applications one need to super-resolve an image which is severely blurred and noisy. For example, in investigative criminology one has available face and fingerprint databases which are often taken at controlled environment. 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? To solve this problem we propose a global learning approach using principal component analysis (PCA), unlike the above methods where the learning was local. In this global approach, we learn the statistics of a good quality image database and use them to enhance the given LR input. As the learning is global it imposes the constraint that all the training images should be globally similar, *i.e.* they should represent a similar class of objects.

  Here we propose two approaches to enhance the low resolution input. In the first method we use the idea of generalized interpolation. A few significant eigenimages of a database of several similar low resolution images are computed and the given low resolution image is projected onto the eigenimages to compute

the projection coefficients. The high resolution image is reconstructed using these coefficients and the HR eigenimages obtained by performing generalized interpolation on the LR eigenimages. In the second approach, we use a high resolution database and compute the principal eigenimages. The only difference here is that the upsampling process is shifted to the input side rather than performing generalized interpolation of the eigenimages. These methods are applicable to images of a particular class, for example, face and fingerprint images. The algorithm offers a significant advantage when the input image is blurred and noisy. 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. However, this global upsampling method does not provide an alias-free reconstruction. Further, the method has limited applicability, mostly in enhancing face or fingerprint images.

- We realize that the global high resolution reconstruction using generalized interpolation fails to do exact super-resolution, though it removes the blur and noise present in the input low resolution image. But it very well learns the global characteristics from the image database. On the other hand, the local learning methods using wavelets or contourlets obtain the missing high frequency details by locally learning the best edges from the image database. To extract the advantages of both these methods, we now propose a hybrid approach for super-resolution by integrating the global learning using PCA and the local learning approach using wavelets or contourlets. The former extracts the best statistical features from a database of similar class of images while the latter ensures the best edges from the database. We solve the problem under a Bayesian framework incorporating the global and local priors, in addition to the data consistency term. We apply the method particularly for face hallucination where we synthesize the missing high frequency details of the input low resolution face image.

- The final contribution of the thesis is in developing an alias-free interpolation

scheme. All the local and global methods discussed so far are able to provide a super-resolved image, but they are unable to handle the aliasing effects. Even in the local approaches, though the edges are handled properly, the aliasing present in the LR image is not reduced much during the upsampling process. Here we study the possibility of removing aliasing in a scene by designing an alias-free upsampling scheme. Quite naturally this is impossible unless additional information about the HR data is available. Hence we assume that the HR data is piece-wise constant. We generate the unknown high frequency components of the given partially aliased (low resolution) image by minimizing the total variation of the interpolant subject to the constraint that part of unaliased spectral components in the low resolution observation are known precisely and under the assumption of sparsity in the data. This provides a mathematical basis for exact reproduction of high frequency components with probability approaching one, from their aliased observation.

In all the SR methods proposed in literature, the quality of the super-resolved image is measured either by means of visual inspection or using a PSNR check. It can be easily shown that the PSNR measure is heavily biased towards the lower part of the spectrum due to the fact that most of the energy is contained in this region. Hence the PSNR may not be a good measure to evaluate the performance of an SR scheme. Even though, several different measures to evaluate the quality of visually pleasing images are proposed in the literature, the issue that the reconstructed components are really the missing high frequency components has not really been investigated so far. Our focus in this work is a study in this direction. We study the issue of alias-free interpolation at a more fundamental level and restrict ourselves again to using a single observation without using any training database unlike the other methods.

We have carried out extensive experiments to validate the usefulness of the different methods. Comparison between the different methods are also provided.

1.6 Organization of the Thesis

We address the problem of single-frame image super-resolution in this thesis. We develop several methods of single-frame image super-resolution by learning the best features from an appropriate image database. We also develop a method which mathematically establishes the correctness of the generated additional high frequency components in super-resolution by restricting ourselves into certain specific assumptions. The thesis is organized as follows.

In chapter 2, we review the current literature on super-resolution for both multi-frame and single-frame methods.

Single image super-resolution using learned wavelet coefficients is discussed in chapter 3. We show that the wavelet coefficients learned at finer scales can be used as an appropriate prior while solving the super-resolution problem under a regularization frame work.

In order to avoid solving the above problem under a regularization scheme to speed up the super-resolution process, we propose a contourlet transform-based learning method for super-resolution in chapter 4. Here we exploit the capability of the contourlet transform to capture the smoothness along contours while learning the oriented edges.

As taking the best match for a test patch from a database does not always provide a better solution, and since the process of decimation may not be captured well by the wavelet or contourlet coefficients implicitly we further consider a locally linear embedding method making use of several neighbors under a supervised learning frame work. This is the subject matter of chapter 5.

In chapter 6 we describe a PCA based global learning approach for super-resolving images of a particular class like face and fingerprint images. This method specifically solves the de-blurring and de-noising issues in super-resolution but is incapable of handling aliasing present in the data. In chapter 7 we propose a hybrid approach integrating the local and global approaches and apply the method for face hallucination.

In chapter 8 we study the exact super-resolvability of a scene from a single observation by designing a partial alias-free interpolation scheme. This provides a mathematical basis for exact reproduction of high frequency components with probability approaching one, from their aliased observation.

The thesis concludes in chapter 9 where we summarize our work and also discuss some issues for further research in the area of super-resolution.

Chapter 2

Literature Review

Most of the literature available on super-resolution is for multi-frame super-resolution where several low resolution observations of the same scene are used to super-resolve a picture. A majority of them are based on the motion cue where the non redundant information among sub-pixel shifted low resolution observations are extracted to produce a single high resolution image. But the goodness of these methods depend directly on how accurately one can register the low resolution observations. In order to tackle this problem, recently there have been some work in the field of motion-free super-resolution where cues other than motion like blur, defocus, zoom, photometry, etc., are used in super-resolving the given low resolution input. These methods do not require observations with relative motion among them. Another category of super-resolution is called single image (frame) super-resolution where only one low resolution input is used to produce the high resolution image. Most of the papers available in this field are based on learning the appropriate features from an image database to recover the missing high frequency details for super-resolution. These techniques are also called learning-based super-resolution techniques. In this section we review the literature on super-resolution reconstruction techniques for both multi-frame and single-frame cases.

2.1 Multi-frame Super-Resolution

The idea of image super-resolution was first introduced by Tsai and Huang [7] in 1984, motivated by the need for improved resolution images from Landsat data. Landsat acquires images of the same areas of the earth in the course of its orbits, thus producing a sequence of similar, but not identical images. Observed images are modeled as under-sampled versions of an unchanging scene undergoing global translational motion. Here the authors propose a frequency domain formulation based on the shift and aliasing properties of the continuous and discrete Fourier transforms for the reconstruction of a band-limited image from a set of under-sampled, and therefore aliased, observation images. The shift and aliasing properties are used to formulate a system of equations which relate the aliased discrete Fourier transform (DFT) coefficients of the observed images to samples of the continuous Fourier transform of the unknown original scene. The system of equations are then solved to recover the image in the Fourier domain and the original image is recovered using inverse DFT. Solution of the equations requires that each observation contribute *independent* equations, which places restrictions on the inter-frame motion that contributes useful data. Here the observations were assumed to be noise free. An extension of this algorithm for noisy data was provided by Kim *et al.*[8], resulting in a weighted least squares algorithm for computing the high resolution estimate. They consider the same blur and noise characteristics for all the low resolution observations. Kim and Su in [9] consider different amount of blur for each low resolution image and used the Tikhonov regularization to obtain the solution of an inconsistent set of linear equations.

Theoretical simplicity is a major advantage of the frequency domain approaches discussed above. That is, the relationship between LR images and the HR image is clearly demonstrated in the frequency domain. The frequency domain method is also convenient for parallel implementation capable of reducing hardware complexity. However, the observation model is restricted to only global translational motion and linear space invariant (LSI) blur. Also, since the data is uncorrelated in the frequency domain, it is also difficult to apply any spatial domain *a priori* knowledge for regu-

larization. Researchers have also explored the use of other transforms like discrete cosine transform (DCT) to achieve super-resolution [10].

The most intuitive method for SR reconstruction is based on a nonuniform interpolation approach. This is a three stage process starting with the estimation of relative motion among successive frames followed by nonuniform interpolation to produce a high resolution image. The final step is a restoration process to remove blur and noise. Ur and Gross [11] performed a nonuniform interpolation of an ensemble of spatially shifted LR images by utilizing the generalized multichannel sampling theorem of Papoulis [12] and Brown[13] to obtain an improved resolution picture. The relative shifts are assumed to be known precisely. A recursive total least squares method for SR reconstruction to reduce the effects of registration error is discussed in [14]. In [15] the authors developed a technique for real-time infrared image registration and SR reconstruction utilizing a gradient-based registration algorithm for estimating the shifts between the acquired frames and presented a weighted nearest neighbor interpolation approach. In [16] Nguyen and Milanfar proposed an efficient wavelet-based SR reconstruction algorithm. They exploit the interlacing structure of the sampling grid in SR and derive a computationally efficient wavelet-based interpolation for interlaced two-dimensional (2-D) data. All these methods are restricted either to a global translational displacement between the measured images or an LSI blur, and a homogeneous additive noise.

Irani and Peleg [17, 18] formulated the iterative back-projection (IBP) SR reconstruction approach that is similar to the back projection method used in tomography. In this approach, the HR image is estimated by back projecting the error between simulated LR images via imaging blur and the observed LR images. This process is repeated iteratively to minimize the energy of the error. A set theoretic approach to the super-resolution restoration problem was suggested in [19]. The main result in this method is to define convex sets which represent tight constraints on solution as well as use the amplitude constraint to improve the results. A projection on to convex sets (POCS) formulation of the SR reconstruction was first suggested by Stark and Oskoui [1]. They use arbitrary smooth motion, linear space variant blur,

and a non-homogeneous additive noise. Ng *et al.* develop a regularized, constrained total least squares solution to obtain a high-resolution image in [20]. They consider the presence of perturbation errors of displacements around the ideal sub-pixel locations in addition to noisy observations. The effect of the displacement errors on the convergence rate of an iterative approach for solving the transform domain based preconditioned system of equations is discussed by Ng and Bose [21]. They also develop a fast restoration algorithm for color images in [22]. Nguyen *et al.* have proposed circulant block preconditioners to accelerate the conjugate gradient descent method while solving the Tikhonov-regularized super-resolution problem [23].

Generally, the SR image reconstruction approach is an ill-posed problem because of an insufficient number of LR images and ill-conditioned inverse of blur operators. Procedures adopted to stabilize the ill-posed problem are called regularization. Katsaggelos *et al.* [24] proposed a multichannel regularized SR approach in which a regularization functional is used. They also calculate the regularization parameter without any prior knowledge at each iteration step. Later, Kang [25] formulated the generalized multichannel deconvolution method including the multichannel regularized SR approach. Bose *et al.* [26] pointed out the important role of the regularization parameter and proposed a constrained least square SR reconstruction method which generates the optimum value of the regularization parameter using the L-curve method [27].

The SR reconstruction problem has also been approached from a statistical estimation framework. Tom and Katsaggelos [28] proposed a maximum likelihood (ML) SR image estimation problem to estimate the subpixel shifts, the noise variances of each image, and the HR image simultaneously. The proposed ML estimation problem is solved by the expectation-maximization (EM) algorithm. Specifically, a maximum *aposteriori* (MAP) framework is developed in [29] which is an extension of the single-frame image expansion algorithm proposed in [30]. The MAP estimator in [29] uses an edge preserving Huber-Markov random field for the image prior. In [31] the authors use a 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. Cheeseman *et al.* [32] applied the Bayesian estimation with a Gaussian prior model to the problem of integrating multiple satellite images observed by the Viking orbiter. Robustness and flexibility in modeling noise characteristics and *a priori* knowledge about the solution are the major advantage of the stochastic SR approach. Assuming that the noise process is white Gaussian, a MAP estimation with convex energy functions in the priors ensures the uniqueness of the solution and hence efficient gradient descent methods can be used to estimate the super-resolved image.

An adaptive filtering approach to super-resolution restoration is described by Elad and Feuer in [33]. The adaptation enables them to incorporate linear space and time-varying blurring and arbitrary motion among the low resolution observations. They have also developed a fast super-resolution algorithm in [34] for pure translational motion and space invariant blur.

In most SR reconstruction algorithms, the blurring process is assumed to be known. In many practical situations, the blurring process of the imaging system is not known. Nguyen *et al.* [35] proposed a technique for parametric blur identification and regularization based on the generalized cross-validation (GCV) and Gauss quadrature theory. 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.

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 [36] use edge models and a local blur estimate to develop an edge-based super-resolution algorithm. In [37] 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 [38] have proposed a non-linear interpolation scheme based on a polynomial operator wherein perceptually

relevant features (say, edges) are extracted and zoomed separately. The reconstruction/restoration methods to improve the resolution of digital images while zooming have been discussed in [39]. 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 1-D structures. A robust median-based estimator is used in an iterative process to achieve super-resolution in [40]. This method discards the measurements which are inconsistent with the imaging model, thereby increasing the resolution even in regions having the outliers. Super-resolution using an edge-model based representation of Laplacian subbands is reported in [41]. In this approach, since the chosen edge primitive set is not a basis, the representation is formulated by matching the given primitives to various image regions as opposed to decomposing given regions in terms of a basis set.

The application of the SR algorithm to the compression system is also needed, since images are routinely compressed prior to transmission and storage. Reconstruction techniques for SR images from compressed video are discussed in [42, 43, 44].

There have been some learning based approaches also for multi-frame super-resolution. Authors in [45] have proposed a super-resolution technique from multiple views using learned image models . Their method uses learned image models either to directly constrain the ML estimate or as a prior for a MAP estimate. To learn the model, they use PCA applied to the image database. In [46] 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 recognition based super-resolution or hallucination. Their prior enforces the condition that the gradient at a point in the super-resolved image should be equal to the gradient in the best matching training image. Pickup *et al.* [47] combine the motion information due to sub-pixel displacements as well as motion-free information in the form of priors to propose a domain specific super-resolution using the sampled textured prior. Here they use training images to estimate a distribution rather than to learn a discrete set of low resolution to high resolution matches. A learning based method for super-resolution enhancement of a video has been proposed by Bishop *et al.* [48]. Their

approach builds on the principle of example based super-resolution for still images proposed by Freeman *et al.* [49]. This is based on a learned data set of image patches capturing the relationship between the middle and high spatial frequency bands of natural images.

The accuracy of all the methods discussed above is dependent on how precisely one can perform the registration among the sub-pixel shifted observations. In order to circumvent this difficulty there have been some efforts in the field of motion-free super-resolution where sub-pixel shifted LR observations are not required for super-resolution. In these methods the researchers use cues other than motion cue like blur, defocus, zoom, etc., thus avoiding the correspondence problem. Several new techniques for motion-free super-resolution are discussed in [50]. A MAP-MRF based SR reconstruction method has been proposed by Rajan *et al.* in [51] where the authors make use of several blurred and noisy low resolution observations. Here a known blur acts as a cue in generating the high resolution image. In [52] the authors recover both the high resolution scene intensity and the depth fields simultaneously using the defocus cue. Rajagopalan and Kiran [53] proposed a frequency domain approach for estimating the high resolution image also using the defocus cue. They derive the Cramer-Rao lower bound for the covariance of the error in the estimated super-resolved image and show that the estimate becomes better as the relative blur increases. Elad and Feuer [54] proposed a unified methodology for super-resolution restoration from several geometrically warped, blurred, noisy and down-sampled measured images by combining ML, MAP and POCS approaches. Super-resolution using zoom as a cue has been proposed by Joshi and Chaudhuri in [55]. They have also proposed a learning-based method for image super-resolution from zoomed observations [56]. They model the high resolution image as an MRF, the parameters of which are learned from the most zoomed observation. The learned parameters are then used to obtain a MAP estimate of the high resolution image.

In this section we have described some of the important SR reconstruction methods falling under the category of multi-frame super-resolution. Several other super-resolution methods can be found in the literature especially for video. For a good

review of those techniques readers are referred to [57] and [58].

2.2 Single Image Super-Resolution

In this section we review some of the works specifically on single-frame image super-resolution which is the topic of this thesis. Simple resolution enhancement methods on smoothing and interpolation techniques for noise reduction have been commonly used in image processing. Smoothing is usually achieved by applying various spatial filters such as Gaussian, Wiener and median filters. Commonly used interpolation methods such as linear and cubic spline interpolation [59], [60] smooth the image data in discontinuous regions, producing a larger image which appears rather blurred. Cubic spline interpolants also tend to overshoot sharp discontinuities, producing a ringing effect at edges. To overcome these difficulties an edge preserving nonlinear image expansion technique is proposed in [30] using Bayesian estimation technique.

There are some very early efforts on super-resolution which deal with the property of analytic continuation of a signal. These techniques infer the missing high frequency components which is actually a process of spectral extension, given a portion of the whole spectrum. The initial work in this regard is due to Harris [61] who establishes that, given a finite extent of an object and a continuous but finite portion of the spectrum of the object, the entire spectrum can be generated uniquely using the principle of analytic continuation. This leads to an exact and complete reconstruction of the object spectrum if the measurements are noise free. However, this global method becomes highly unreliable even if a small amount of noise is present in the given portion of the spectrum. A new view of the problem of continuing a given segment of the spectrum of a finite object is presented in [62]. Here the signal extrapolation is carried out by the method of alternate projections [63], iterating alternately between time and frequency domains. This method relies on the notion of reducing the 'error energy'. A dual of the same problem is solved by Papoulis [64] where the spectrum of the bandlimited object is recovered from a finite segment of the object using an iterative procedure. Here the effect of noise and the error due to aliasing are determined

and it is shown that they can be controlled by early termination of the iteration.

Recently, some learning-based methods have been proposed for single image super-resolution reconstruction. Most of these techniques make use of a set of training images to learn the best features for the given low resolution image in order to produce a high resolution image. As our work in this thesis is motivated by this new approach for image super-resolution, we now review some of the papers under this category. Freeman *et al.* have developed a single-frame image super-resolution method using a set of training images where 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 corresponding blurred images. These are then incorporated into a Markov network to learn their relationship. Bayesian belief propagation allows them to efficiently find a local maximum of the posterior probability for the high resolution patch given the low resolution one [65, 66]. But the method is somewhat dependent on the training set and hence the result is not stable and sometimes produces artifacts in real applications. Their later work [49] simplifies and and improves on this approach.

An image analogy method applied to super-resolution is is discussed in [67]. 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 original high resolution image. Here the emphasis is on learning the local statistics at finer details. This method is less effective than other super-resolution methods as no interaction between adjacent elements in the high resolution image is explicitly enforced to ensure compatibility. Candocia and Principe [68] address the ill-posedness of the super-resolution problem by assuming that the correlated neighbors remain similar across scales, and this *a priori* information is learned 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.

For face identification, especially by human, it is desirable to render a high res-

olution face image from a low resolution one which is called face hallucination, a term coined by Baker and Kanade, which implies that high frequency part of face image must be purely fabricated. There have been quite a few new approaches toward face hallucination. A learning based method for face hallucination is proposed in [69]. Here the algorithm learns a prior on the spatial distribution of the image gradient for frontal images of faces and the hallucinated image is recovered under a MAP framework. They also extend the method for the case when there are multiple low resolution images. Another technique for hallucinating faces using a two step approach is discussed in [70]. Here, first the authors derive a global linear model to learn the relationship between the high resolution face images and their smoothed and down-sampled low resolution ones. To capture the high frequency contents of faces, the residual between an original high resolution image and the reconstructed high resolution image by learned linear model is modeled by a patch-based nonparametric Markov network. By integrating both global and local models they generate photo-realistic face images. In [71] Wang and Tang develop an efficient face hallucination method using an eigen transformation algorithm. However, the method only utilizes global information without paying any attention to local details. In another approach for face hallucination presented in [72] the authors have shown that their framework integrating TensorPatch model and coupled residue compensation is capable of producing high quality super-resolution images. Compared to other approaches, this method has the advantage of preserving global structure and local features.

Recently, Wang *et al.* [73] have proposed a learning based super-resolution method in a probabilistic framework integrating the SR reconstruction constraint and the patch based co-occurrence prior. This framework is also in fact a combination of the global parametric model and the learned nonparametric model as in [70]. Here the authors perform experiments for both single image and multiple image cases. A multi-modal (such as variations in viewpoint and illumination) face image super-resolution and recognition system in tensor space is presented in [74]. Given a single modal low resolution image, they super-resolve its high resolution reconstructions across different modalities for face recognition. An image hallucination approach based on

the primal sketch priors for generic images is presented in [75]. Here primal sketch priors (edges, ridges and corners) are constructed and used to enhance the quality of the hallucinated high resolution image. A contour smoothness constraint enforces consistency of primitives in the hallucinated image by a Markov-chain based inference algorithm.

There are some other learning-based techniques which do not require a training set but require strong image priors that are hypothesized as in [30] or learned from data [76]. In [76] a generic unsupervised method for single image super-resolution is proposed. Here a dynamic tree-like architecture is used to model the high resolution data and approximate conditioning in the low resolution image is achieved through a mean field approach.

A method for single-frame image super-resolution using an unsupervised learning technique is proposed in [77]. In this approach the required prior knowledge about the high resolution images is obtained using kernel principal component analysis. The single image super-resolution method proposed in [78] is the extension of a Markov-based learning algorithm described in [66], capable of processing an LR image with unknown degradation parameters. A different method for enhancing the resolution of LR facial images using an error back projection method based on a top-down learning is proposed in [79]. Here a face is represented by a linear combination of prototypes of shape and texture. Recently, Chang *et al.* [80] have proposed a single image super-resolution method by exploring the manifold properties. Assuming LR and HR patches reside on two non-linear manifolds having similar locally linear structure, the manifold correlation is realized by a three layer Markov network and performing super-resolution with energy minimization. This method has the advantages of both recognition based and reconstruction based approaches.

It may be noted that the super-resolution problem is now getting quite matured and several journals have devoted special issues on this topic. The special issues [81] -[88] discuss some of the recent developments in the field of multi-frame as well as single-frame image super-resolution.

Chapter 3

Use of Learned Wavelet Prior

Super-resolution is the process of estimating a high resolution image from one or more low resolution observations. This process includes upsampling the image, thereby increasing the maximum spatial frequency, and removing degradations that arise during the image capture, *viz.*, aliasing and blurring. A suitable way of recovering the missing high frequency details is to learn them from an appropriate high resolution training data and use a suitable regularization to obtain an accurate solution. We show in this chapter that by using a wavelet based learning prior along with a suitable discontinuity preserving smoothness prior, an effective super-resolution can be achieved. The advantage of this method is that there is no correspondence problem. Also in many of the applications more than one low resolution observations may not be available, but may have a database of a number of similar images at a higher spatial resolution. Hence one needs to minimize the aliasing by making use of a single low resolution observation.

Here we consider having access to a set of high resolution training images to learn the 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 generation of

the super-resolved image. The learned coefficients are then used in a prior term 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 learned from the training set. In order to preserve the spatial continuity of the restored image, we use a smoothness constraint in conjunction with the learned prior to obtain the super-resolved image.

The reminder of the chapter is organized as follows. In section 3.1 we briefly review some of the recent works in super-resolution reconstruction using only wavelet based methods. We discuss the model for the formation of a low resolution image in section 3.2. Some brief background on discrete wavelet transform and the multi-resolution analysis for estimating the wavelet coefficients at the finer scales using high resolution training images are the subject matters of section 3.3. Section 3.4 discusses the regularization based approach to derive a cost function for the super-resolution estimation. We present experimental results on different types of images in section 3.5, and the chapter concludes in section 3.6.

3.1 Related Work

We have already provided a detailed literature survey on SR in the previous chapter. We now briefly review the related work on SR that uses wavelet decomposition of an image. Wavelet theory has previously been used extensively for image de-noising and de-blurring from static images, for example see [89, 90]. But there have been only a very few studies for the use of wavelets for image super-resolution. In [91], Nguyen *et al.* proposes a wavelet based method for super-resolution. But here the authors use wavelets just for the purpose of speeding up the computation. In a recent work by Bose *et al.*[92], the use of second-generation wavelets to attain mult-frame super-resolution with noise filtering is discussed. The procedure allows them to incorporate a more general projective camera motion model into the framework, instead of only displacement and rotational models. Some novel wavelet domain algorithms for high resolution image reconstruction are discussed in [93]. In this iterative approach,

the authors apply wavelet (packet) thresholding methods to de-noise the function obtained in the previous step before adding it into the new iteration resulting in better reconstructed images. But, we use wavelets in a different framework- for learning the best edges from a high resolution training data using the low resolution edge primitives and use it as a prior in the final solution. Thus the proposed method can be classified under deterministic learning based super-resolution schemes.

3.2 Low Resolution Image Formation Model

The super-resolution problem can be cast in a restoration frame-work. We use the simplified decimation model for low resolution image formation as discussed in chapter 1, assuming no blurring in the observation. For a decimation factor of q the low resolution image $f(i,j)$ can be obtained from its high resolution version $z(k,l)$ as

$$f(i,j) = \frac{1}{q^2} \sum_{k=qi}^{q(i+1)-1} \sum_{l=qj}^{q(j+1)-1} z(k,l) \tag{3.1}$$

i.e., the low resolution intensity is the average of the high resolution intensities over a neighborhood of q^2 pixels. This decimation model simulates the integration of light intensity that falls on the high resolution detector and the decimation process is represented by the matrix D which has a structure as given in equation (3.3). Quite naturally, it is assumed that the fill factor for the CCD array is unity.

Let $\mathbf{z}$ represent the lexicographically ordered high resolution image of $N^2 \times 1$ pixels. If $\mathbf{f}$ is the $M^2 \times 1$ lexicographically ordered vector containing pixels from the low resolution observation, then it can be modeled as

$$\mathbf{f} = D\mathbf{z} + \mathbf{n} \tag{3.2}$$

where D is the decimation matrix, size of which depends on the decimation factor. For a decimation factor of q, the decimation matrix D consists of q^2 non-zero elements of value $\frac{1}{q^2}$ along each row at appropriate locations and has the form [30] (using a

proper reordering of $\mathbf{z}$)

$$D = \frac{1}{q^2}\begin{bmatrix} 11\ldots1 & & \mathbf{0} \\ & 11\ldots1 & \\ & & \ddots \\ \mathbf{0} & & 11\ldots1 \end{bmatrix}. \tag{3.3}$$

As an example, for a decimation factor of $q = 2$ and with lexicographically ordered $\mathbf{z}$ of size, say 16×1, the D matrix is of size 4×16 and can be written as

$$D = \frac{1}{4}\begin{bmatrix} 1100110000000000 \\ 0011001100000000 \\ 0000000011001100 \\ 0000000000110011 \end{bmatrix}. \tag{3.4}$$

In equation (3.2) $\mathbf{n}$ is the $M^2 \times 1$ noise vector. We assume the noise to be zero mean i.i.d process. Our problem now reduces to estimating $\mathbf{z}$ given $\mathbf{f}$, which is an ill-posed, inverse problem.

3.3 Wavelet based Learning

3.3.1 Background

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 domain 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 sequence 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 by the sequence, and integrating over all times. In the case of DWT, filters of different cutoff frequencies are employed to analyze the sequence at different scales. The input sequence is

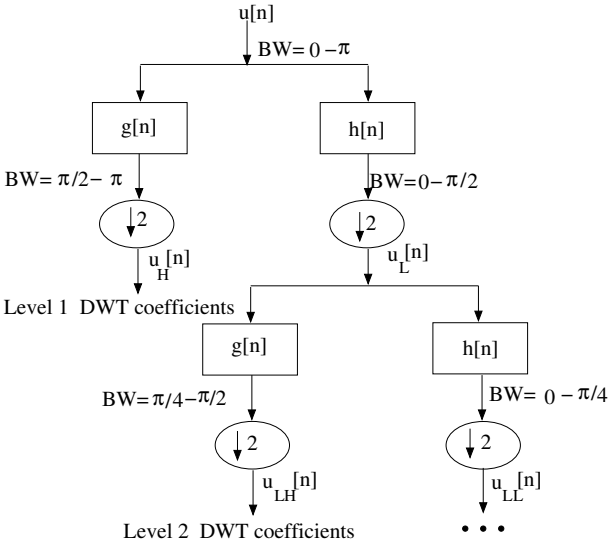


Figure 3.1: Illustration of subband wavelet decomposition. Here $u[n]$ is the original sequence to be decomposed and $h[n]$ and $g[n]$ are lowpass and highpass filters, respectively. The bandwidth of the signal is marked as "BW".

passed through a series of highpass and lowpass 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 lowpass 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 lowpass filter thus halves the resolution, but leaves the scale unchanged. The subsequent subsampling by a factor of 2, however, changes the scale. This is illustrated in Figure 3.1.

The wavelet transform for a 2-D sequence is similar to that of 1-D decomposition. A 2-D wavelet decomposition is first performed (horizontally) on the rows by applying lowpass and highpass filters. Then we perform the same operations vertically (on the columns) resulting in four subbands LL, LH, HL, HH. We repeat the operation with 'LL' as the input image for further decomposition. The readers are referred to [94, 95, 96] for further discussion on wavelet decomposition.

3.3.2 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 3.1, 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 IDWT (inverse discrete wavelet transform) 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 2-D 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 resolution, and since the wavelets are best suited for a multiresolution analysis, it motivates us to use a wavelet based approach for learning the wavelet coefficients at the finer resolution.

These wavelet coefficients at finer scales indicate the high frequency details in an image. The learning 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 3.1) are quite significant and they cannot be neglected while obtaining the high resolution image. These coefficients must be learned 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 database, 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

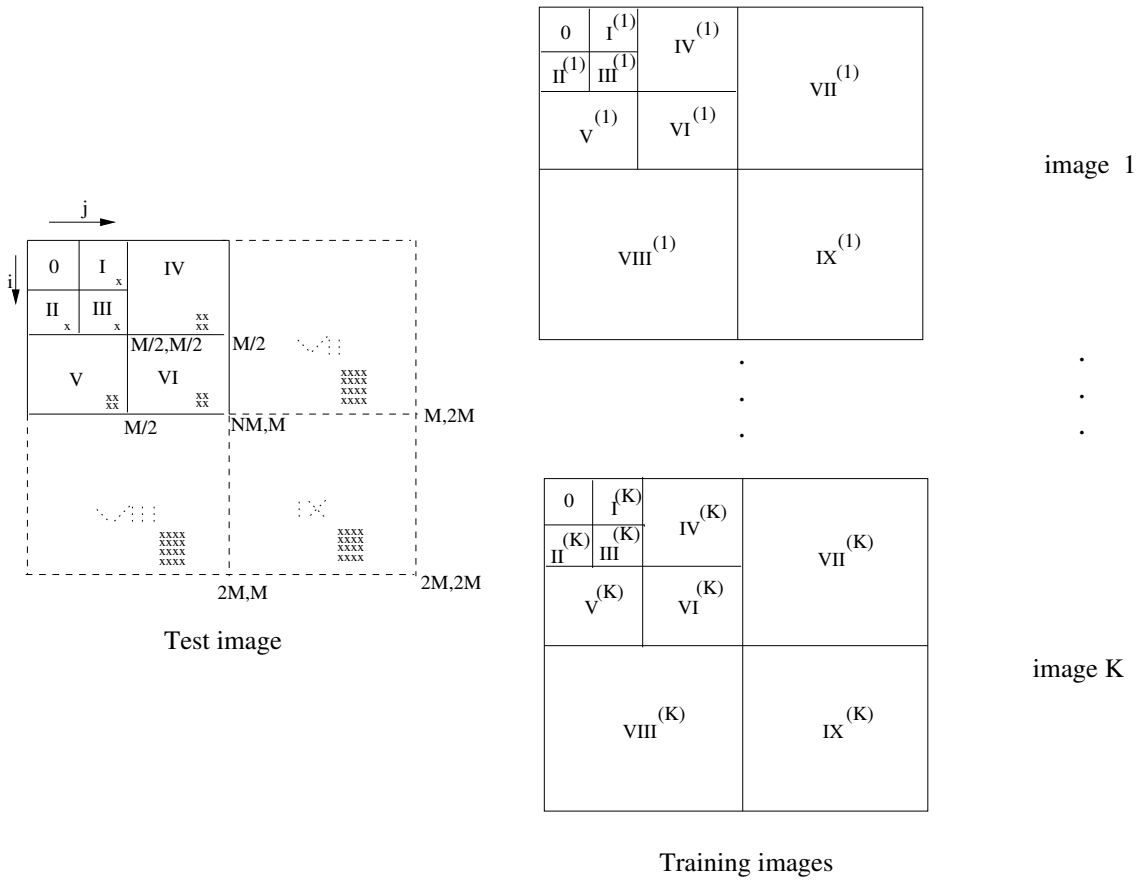


Figure 3.2: Illustration of learning of wavelet coefficients at a finer scale. (a) Low resolution image with two level wavelet decomposition. Wavelet coefficients (marked as x) in subbands shown with the dotted lines are to be estimated for bands $VII - IX$. (b) High resolution training set in wavelet domain with three level decomposition.

(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 3.2 illustrates the block schematic of how the wavelet coefficients at finer scales are learned from a set of K 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 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 3.2(a)) for the test

image. They correspond to the estimated high pass wavelet coefficients at the first 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 extrapolate a block of 4×4 wavelet coefficients in each of the subbands VII, $VIII$ and IX. In order to do this we exploit the idea from zero tree concept, i.e., in a multiresolution system, every coefficient at a given scale can be related to a set of coefficients at the next coarser scale of similar orientation [97]. 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 $d(i, j)$. Consider the range $0 \leq i, j \leq M/4$. The wavelet coefficients $d_I(i, j + M/4)$, $d_{II}(i + M/4, j)$, $d_{III}(i+M/4, j+M/4)$ corresponding to subbands $I-III$ and a 2×2 block consisting of $d_{IV}(k, l+M/2)$, $d_V(k+M/2, l)$, $d_{VI}(k+M/2, l+M/2)$, for $k = 2i : 2i+1$ and $l = 2j : 2j + 1$ 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 $d_{VII}(k, l + M)$,$d_{VIII}(k + M, l)$, and $d_{IX}(k + M, l + M)$ for $k = 4i : 4i + 3$ and $l = 4j : 4j + 3$. In order to illustrate which set of wavelet coefficients we select for learning purposes, we denote them with 'x' marks in Figure 3.2(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 3.2(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 bands $I-VI$ in MAD sense and copy the corresponding high resolution wavelet coefficients, in bands $VII - IX$ to those bands for the test image. In effect, we use the following equation

to find the minimum.

$$\begin{aligned}
\hat{p}(i,j) &= \arg\min_{p}[|d_I(i,j+M/4) - d_{I^{(p)}}(i,j+M/4)| \\
&+ |d_{II}(i+M/4,j) - d_{II^{(p)}}(i+M/4,j)| \\
&+ |d_{III}(i+M/4,j+M/4) - d_{III^{(p)}}(i+M/4,j+M/4)| \\
&+ \sum_{k=2i}^{2i+1}\sum_{l=2j}^{2j+1} |d_{IV}(k,l+M/2) - d_{IV^{(p)}}(k,l+M/2)| \\
&+ \sum_{k=2i}^{2i+1}\sum_{l=2j}^{2j+1} |d_{V}(k+M/2,l) - d_{V^{(p)}}(k+M/2,l)| \\
&+ \sum_{k=2i}^{2i+1}\sum_{l=2j}^{2j+1} |d_{VI}(k+M/2,l+M/2) - d_{VI^{(p)}}(k+M/2,l+M/2)|],
\end{aligned} \tag{3.5}$$

where $p = 1, 2, \cdots, K$. Here $d_J(p)$ denotes the wavelet coefficients for the p^{th} training image at the J^{th} band. For each (i,j) in $I - III$ of low resolution observation, a 4×4 block of wavelet coefficients in subbands $VII - IX$ from that training image given by $\hat{p}(i,j)$ which gives the minimum are then copied into subbands VII, $VIII$, IX of the observed image. In effect, equation (3.5) helps in matching edge primitives at low resolutions. Thus we have, $d_{VII}(i,j) := d^{(\hat{p})}_{VII}(i,j), d_{VIII}(i,j) := d^{(\hat{p})}_{VIII}(i,j), d_{IX}(i,j) := d^{(\hat{p})}_{IX}(i,j), \text{for}(i,j) \in (VII - IX)$ where $\hat{p}$ 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 $I - III$, we learn 16 coefficients each, for subbands $VII - IX$ from the training set.

It may be mentioned here that each 4×4 region in the low resolution image could be learned from different training images. In case the error (MAD) term in equation (3.5) 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. Now the issue of threshold selection poses a problem. If the threshold is chosen very small we may not be able to find a correct match from the database unless it consists of images very much similar to the LR input. If we do not find a match, the coresponding HR

region should be filled by zeros which indicate a loss of information. On the other hand, if the threshold is chosen very high, lot of unwanted coefficients will be learned which will give rise to spurious frequencies in the super-resolved image. Hence the selection of threshold is a tradeoff between spurious learning and information loss. In a practical senario, one can try with different thresholds and accept the one which gives the best result. We have used the same approach in selcting the threshold for different LR inputs.

The goodness of the learning depends on how extensive and useful is the training data set. In our experiments we use Daub4 wavelet for computing the discrete wavelet transform. The subband 0 corresponds to the low resolution portion 'LL' (see Figure 3.2(a)) in the wavelet decomposition and since the corresponding 'LL' portions in the training set may have different brightness averages, including 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 two level wavelet decomposition on the low resolution test image of size* $M \times M$ *and three level decomposition on 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* $IV - VI$ *of the low resolution image as well as the high resolution training set.*

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

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

STEP 5 : *Repeat steps (2 - 4) for every wavelet coefficient in bands $I - VI$ 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 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. Furthermore, there is no reinforcement of the learned coefficients through 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 super-resolution techniques.

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 better matches in the coefficient, but the localization is poor and suffers from severe blockiness. Furthermore, the requirement for the training data size goes up drastically.

3.4 Regularization And Super-Resolution Estimation

With the wavelet coefficients learned 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 learned from the training set to enforce the constraint that the wavelet coefficients of the super-resolved image should be close to the best matching wavelets learned from the training images in a least squares sense.

Let$\hat{\mathbf{Z}}_{wt}$ be the learned wavelet prior as discussed in the previous section. Then the learning prior term can be expressed as

$$C(\mathbf{z}) = \beta||\chi(\mathbf{z}) - \hat{\mathbf{Z}}_{wt}||^2. \tag{3.6}$$

where $\chi(\mathbf{z})$ is the wavelet decomposition of the unknown high resolution image and β is a weight factor for the learning term. Now, in order to enforce the smoothness constraint we make use of the fact that the image pixel intensities have a spatial correlation. This prior knowledge serves as a contextual constraint and has to be used to regularize the solution. 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 learned 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 binary variable $l_{i,j}$ connecting sites (pixel locations) (i,j) to $(i-1,j)$ aids in detecting a horizontal edge while the variable $v_{i,j}$ connecting sites (i,j) to $(i,j-1)$ helps in detecting a vertical edge. The variables $l_{i,j}$ and $v_{i,j}$ are set to 1 if $|z(i,j) - z(i-1,j)| > Threshold1$ and $|z(i,j) - z(i,j-1)| > Threshold2$, respectively. Else they are set to 0. We use the following prior for the smoothness constraint in this study.

$$U(\mathbf{z}) \;=\; \sum_{i,j}\{\mu[(z_{i,j} - z_{i,j-1})^2(1 - v_{i,j}) + (z_{i,j+1} - z_{i,j})^2(1 - v_{i,j+1})$$

$$+ \quad (z_{i,j} - z_{i-1,j})^2(1 - l_{i,j}) + (z_{i+1,j} - z_{i,j})^2(1 - l_{i+1,j})]$$
$$+ \quad \gamma(l_{i,j} + l_{i+1,j} + v_{i,j} + v_{i,j+1})\}. \tag{3.7}$$

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

$$\varepsilon = ||\mathbf{f} - D\mathbf{z}||^2 + C(\mathbf{z}) + U(\mathbf{z}). \tag{3.8}$$

The above cost function is nonconvex and also consists of terms in both spatial domain (the first and the third term) and in frequency 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 leads 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 learned wavelet coefficients is used as the initial estimate for $\mathbf{z}$.

We now explain the various terms in equation (3.8) with respect to the wavelet based learning method. The first term relates to the consistency in data fitting. If $\mathbf{z}$ is the actual HR image, we observe that $||\mathbf{f} - D\mathbf{z}||^2$ need not be zero as the chosen decimation operator D as defined in equation (3.3) need not be close to the wavelet decomposition (LL image in Figure 3.2) 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 equation (3.8) was based on the idea that each term in the equation should have comparable magnitudes when the algorithm converges to the high resolution image.

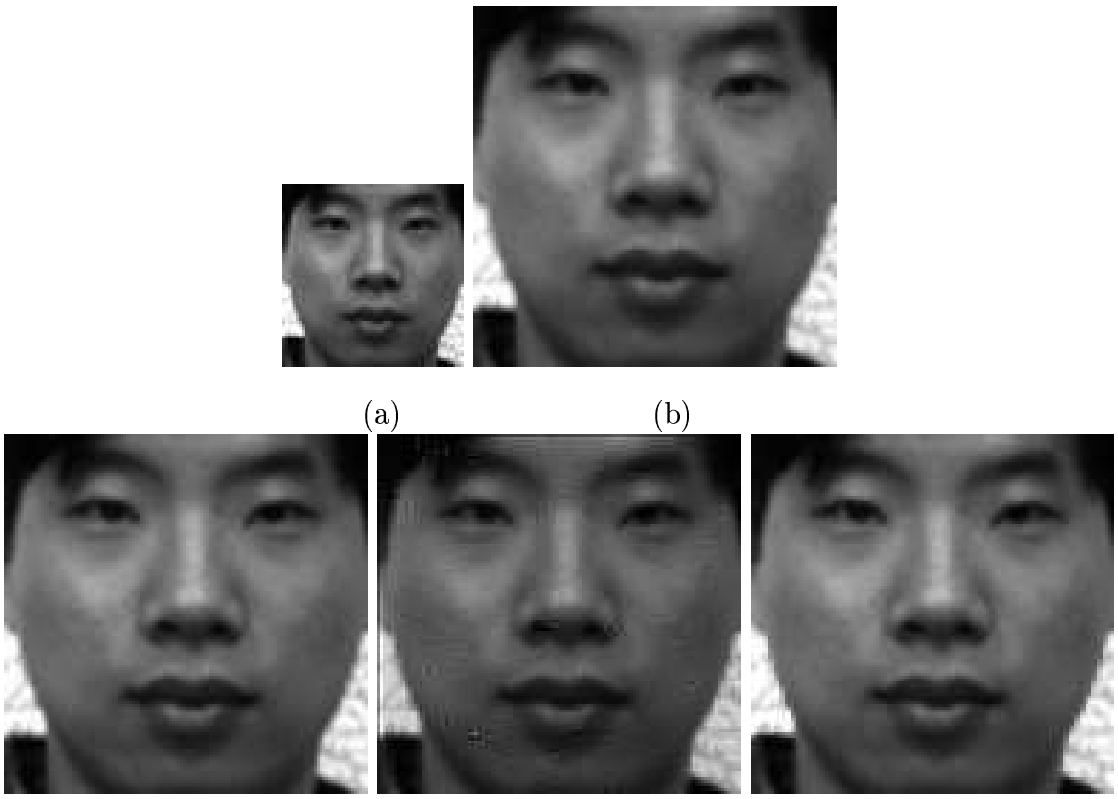
3.5 Experimental Results

In this section, we demonstrate the efficacy of the proposed technique to super-resolve a low resolution observation using the wavelet coefficients learned from a high resolution training data set. We first present the results on gray scale images and then show that it works well for color images also.

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 $K = 200$. Some representative training images are shown in Figure 3.3. The same training data set has been used in all experiments. In order to obtain a low resolution test image,

Figure 3.3: Some representative training images.

we consider a high resolution image which does not belong to the training set and downsample it by a factor of 2. Figure 3.4(a) shows one such low resolution face image



(a) (b)

(c) (d) (e)

Figure 3.4: (a) A low resolution observation (face1), (b) bicubic interpolated image, (c) interpolation using Lanczos method, (d) super-resolved image using wavelet-based learning without regularization and (e) and the super-resolved image using the proposed approach.

of size 64×64. Figure 3.4(b) shows the same image upsampled by a factor of 2 using the bicubic interpolation technique and Figure 3.4(c) shows the interpolated image using the Lanczos method. The super-resolved image using wavelet-based learning without regularization is shown in Figure 3.4(d). Here, even though the output is slightly sharper, one can observe certain artifacts which is due to spurious learning. To get rid of such artifacts we solve the problem under a regularization framework as explained in section 3.4 and the corresponding result is shown in Figure 3.4(e). A comparison of the Figures 3.4(b),(c) and (e) shows more clear details in the super-resolved image. The features such as eyes, nose and the mouth appear blurred in the interpolated images shown in Figure 3.4(b) and (c), while they are restored well in

Figure 3.4(e). Also the eye balls are sharper in the displayed super-resolved image. It has been experimentally found that the best results are obtained with the parameters $\mu = 0.01$, $\gamma = 25$, the weight for the learning term $\beta = 0.08$ and $Threshold1 = Threshold2 = 30$. These parameters were selected so that all the components in the cost function (refer to equation (3.8))have comparable contributions. We retain the same values for the parameters in all subsequent experiments.

Next we consider another experiment on face image. The low resolution observation obtained by down sampling the high resolution Lena image is shown in Figure 3.5(a). The super-resolution result obtained using the proposed approach is displayed in Figure 3.5(d), and Figure 3.5(b) shows the bicubic interpolated image and interpolation using the Lanczos method is shown in Figure 3.5(c). 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 and the lace on the hat appear more clearly. 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. Furthermore, the edge primitives are chosen over a 8×8 block in the wavelet domain. Hence the learned edges may suffer from blockiness. The smoothness constraint is supposed to take care of such jaggedness. However, the $Threshold$ value being chosen on an ad hoc basis it fails to undo the jaggedness. But this blockiness is nothing compared to the blockiness one obtains when a simple pixel replication is used. Comparing with the simple zero order hold expanded image shown in Figure 3.5(e) (in which every feature in the image appears blocky), we see that the blockiness is negligible in the proposed approach. We could have played with the parameter set in equations (3.7) and (3.8), but the various parameters for recovering the super-resolved image for this experiment were kept the same as used in the previous experiment.

In order to test our 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 3.6(b-d) with the low resolution observation depicted in Figure 3.6(a). We can clearly see that the discontinuities are better estimated in the super-resolved image shown in

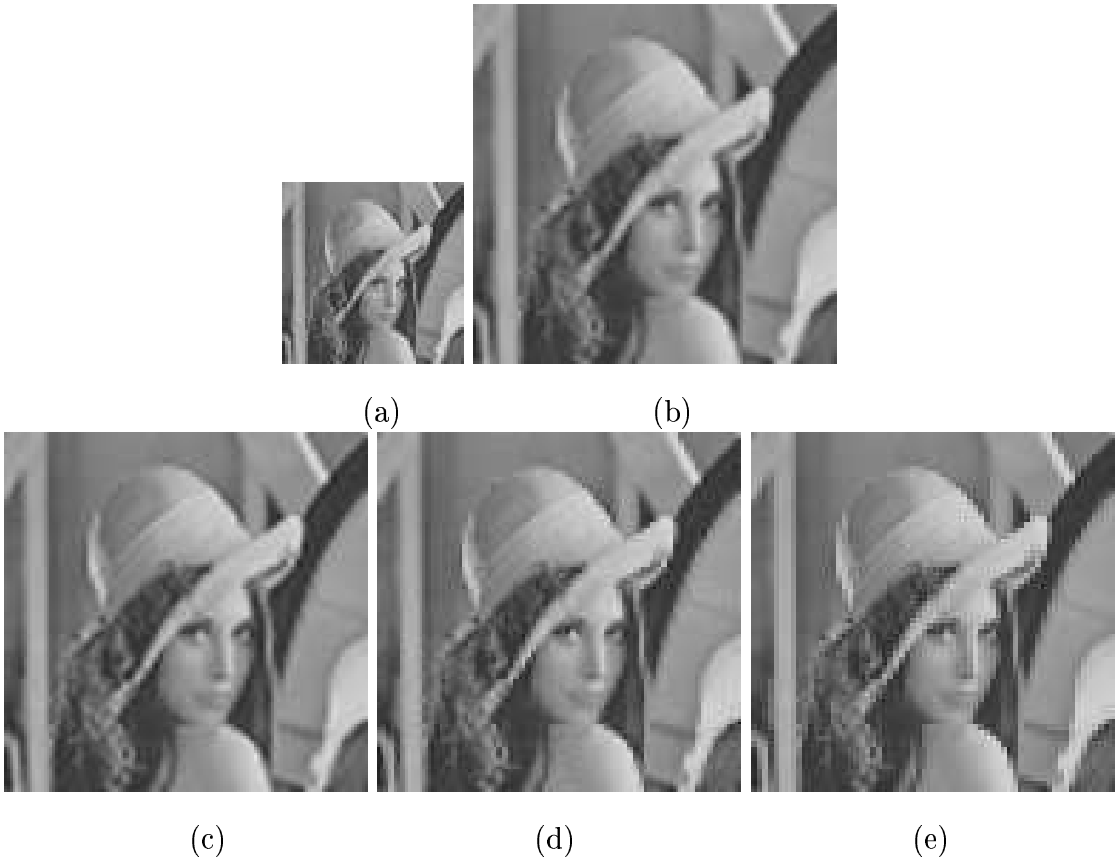


Figure 3.5: (a) Another low resolution observation (Lena), (b) bicubic interpolated image, (c) interpolated image using Lanczos method, (d) the super-resolved image, and (e) result of simple pixel replication.

Figure 3.6(d), but they appear blurred in the bicubic interpolated image (see Figure 3.6(b)). The interpolation using the Lanczos method shown Figure 3.6(c) also appears to be blurred. This substantiates our claim that the learning of wavelet coefficients does help in improving the resolutions.

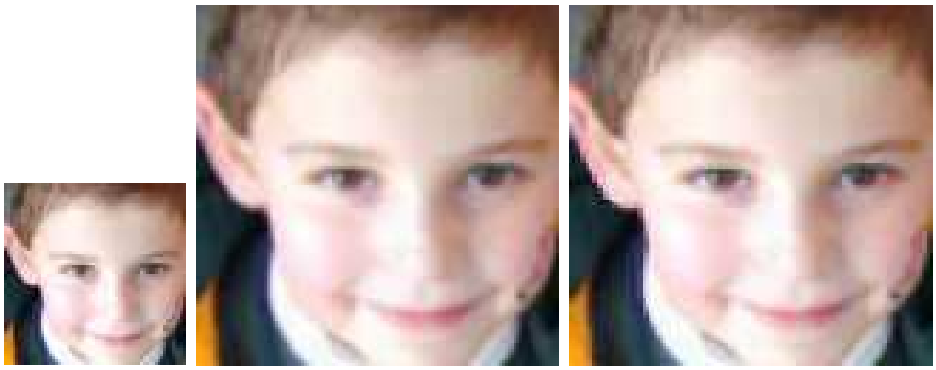
(a) (b) (c) (d)

Figure 3.6: (a) Low resolution observation of a building, (b) upsampled by bicubic interpolation, (c) interpolated image using Lanczos method, (d) and 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 resolution images. However, we refrain from doing it as any possible error in learning in any of the color planes may introduce chromatic distortions and the human vision appears to be sensitive to that.

We now show results of two experiments conducted on the color face images. Fig-

ure 3.7(c) shows the result of the proposed approach on a low resolution observation shown in Figure 3.7(a). Compare this with the bicubic interpolated image shown in Figure 3.7(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. The results for another low resolution face image are displayed in Figures 3.8(a-c). Similar conclusions can again be drawn from this experiment. Observe the eye balls, eye brows, frontal hair, and the nose shown in Figure 3.8(c) which appear sharper when compared to the bicubic interpolated image given in Figure 3.8(b). Thus we conclude that our approach works well for color images as well. As the bicubic interpolation method produces a fairly pleasing output in comparison with other interpolation methods, now onwards we restrict the comparison of our methods only with bicubic interpolation.



(a) (b) (c)

Figure 3.7: (a) A low resolution observation, (b) upsampling using the bicubic interpolation, and (c) the super-resolved image.

In order to convey the comparative edge over the conventional interpolation techniques, we show the PSNR during interpolation for the gray scale images. Table 3.1 shows the comparison of the proposed method with the standard bilinear interpolation, bicubic interpolation and the Lanczos method. 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 an increase in PSNR for

(a) (b) (c)

Figure 3.8: (a) Another low resolution face image, (b) upsampling using the bicubic interpolation, (c) the super-resolved image.

the proposed 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
Lanczos	31.70	27.72	26.75
Proposed	32.74	28.05	26.97

Table 3.1: Comparison of PSNR for different schemes.

3.6 Conclusions

We have described a method for super-resolution restoration of images using a wavelet based learning technique. The wavelet coefficients at finer scales, learned 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 proposed 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 the proposed 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 $q = 2$ 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.

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 horizontal and vertical edges properly. Hence we do not have difficulty in learning vertical and horizontal edges, but we do have some problems in learning edges oriented along arbitrary directions. This leads to blockiness in the learned edges. A better way to handle this is to use directionally selective wavelet decomposition to learn the oriented edges. We explore this issue in the next chapter.

Chapter 4

Use of Contourlets

The motivation for using wavelets for super-resolution learning as proposed in the previous chapter was its ability to capture edges across scales and the computational efficiency due to the wavelet kernels being separable. Even though wavelets are capable of capturing the vertical and horizontal edges properly, they are inferior in acquiring the geometry of edges. As a result we were forced to solve the problem under a regularization framework using the learned wavelet coefficients as a prior. This gave rise to a computationally expensive optimization procedure which made SR reconstruction problem very slow. In addition to this, the super-resolved output was slightly blocky as the learning procedure failed to acquire the directional edges as can be observed in Fig.8.5(c). In order to overcome these difficulties, in this chapter we propose a learning based super-resolution based on the contourlet transform, recently developed by Do *et al.* [6]. Here, we make use of the property of the contourlet transform to capture the geometric smoothness along contours of any two dimensional objects using directional decompositions.

The contourlet coefficients at finer scales of the unknown high resolution image are learned locally from a set of high resolution training images, the inverse contourlet transform of which recovers the super-resolved image. In effect, we learn the high resolution representation of an oriented edge primitive from the training data. Here we use the same low resolution model as in chapter 3.

4.1 Contourlet Transform

For one-dimensional piecewise smooth signals, like scan-lines of an image, wavelets have been established as the right tool, because they provide an optimal representation for these signals in a certain sense [98, 99]. In addition, the wavelet representation is amenable to efficient algorithms; in particular it leads to fast transforms and convenient tree structures. These are the key reasons for the success of wavelets in many signal processing and communication applications; for example, the wavelet transform was adopted as the transform for the new image compression standard JPEG-2000 [100]. However, natural images are not simply stacks of 1-D piece-wise smooth scan-lines; discontinuity points (i.e., edges) are typically located along smooth curves (i.e., contours) owing to smooth boundaries of physical objects. Thus, natural images contain intrinsic geometrical structures that are key features in visual information. As a result of a separable extension from 1-D bases, wavelets in 2-D are good at isolating the discontinuities at edge points, but will not see the smoothness along the contours. In addition, separable wavelets can capture only limited directional information, an important and unique feature of multidimensional signals.

In order to overcome these difficulties Do *et al.* developed a true 2-D transform called the contourlet transform [6], [101] which is an extension of the Cartesian wavelet transform in two dimensions using multi-scale and directional filter banks. The contourlet expansion of images consists of basis images oriented at various directions in multiple scales, with flexible aspect ratios. Thus the contourlet transform retains the multi-scale and time-frequency localization properties of wavelets. In addition, it also offers a high degree of directionality. Thus they are capable of capturing the geometrical smoothness of the contour along any possible direction. The contourlet transform does not use separable basis functions. The idea here is not to decompose an image into horizontal and vertical edges, but to capture the edges normal to the contour present in the image.

The contourlet transform is implemented in two stages: the subband (spectral) decomposition stage and the directional decomposition stages. For the subband de-

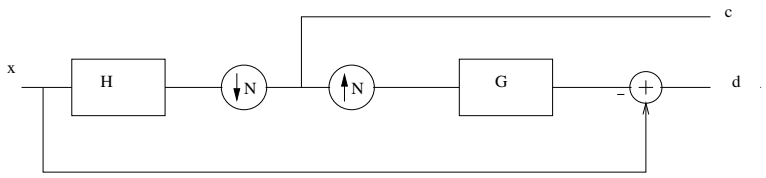


Figure 4.1: The Laplacian pyramid decomposition. The outputs are a coarse approximation c and a difference d between the original and the prediction.

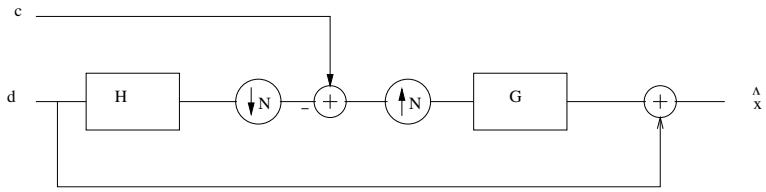


Figure 4.2: The reconstruction scheme for the Laplacian pyramid.

composition stage we use the Laplacian pyramid introduced by Burt and Adelson [102] where the decomposition at each step generates a sampled lowpass version of the original and the difference between the original image and the prediction as shown in Fig. 4.1. The input image x is first lowpass filtered using the analysis filter H and then decimated to get a coarse approximation c. This is then interpolated and passed through the synthesis filter G. The resulting image is then subtracted from the original image x to obtain the bandpass image d. The process is then iterated on the coarser version of the image c. An optimal linear reconstruction for the Laplacian pyramid is shown in Fig. 4.2. This reconstruction method differs from the usual method, where the signal is obtained by simply adding back the difference to the prediction from the coarse signal, and has significant improvement over the usual reconstruction in the presence of noise as shown in [103].

The directional filter bank (DFB) is efficiently implemented by using an t-level binary tree decomposition that leads to 2^t subbands with wedge-shaped frequency partitioning as shown in Fig. 4.3(a). The original construction of the DFB proposed in [104] involves modulating the input image and uses quincunx filter banks with diamond-shaped filters [105]. The desired frequency partitioning is obtained by following a tree expanding rule for finer directional subbands as described in [106]. In

[101], a new construction for the DFB is proposed that avoids modulating the input image and has a simpler rule for expanding the decomposition tree. A t-level tree structured DFB is equivalent to a 2^t parallel channel filter bank with equivalent filters and overall sampling matrices as shown in Fig. 4.3(b). The equivalent analysis and synthesis filters are denoted by H_k and $G_k, 0 \leq k < 2^t$ respectively corresponding to the subbands indexed as in Fig. 4.3(a). The corresponding overall sampling matrices will have the following diagonal form [101].

$$S_k = \begin{cases} diag(2^{t-1}, 2), & \text{for } 0 \leq k < 2^{t-1} \\ diag(2, 2^{t-1}), & \text{for } 2^{t-1} \leq k < 2^t. \end{cases}$$

The two sets correspond to the mostly horizontal and vertical set of directions, respectively. Thus the family $\{g_k[n - S_k l]\}_{0 \leq k < 2^t, l \in Z^2}$ obtained by translating the impulse responses of the equivalent synthesis filters G_k over the sampling lattices S_k, provide a *basis* for discrete signals in $l^2(Z^2)$. This basis exhibits both directional and localization properties. For a detailed description of contourlet decomposition, readers are referred to [101],[104] and[106].

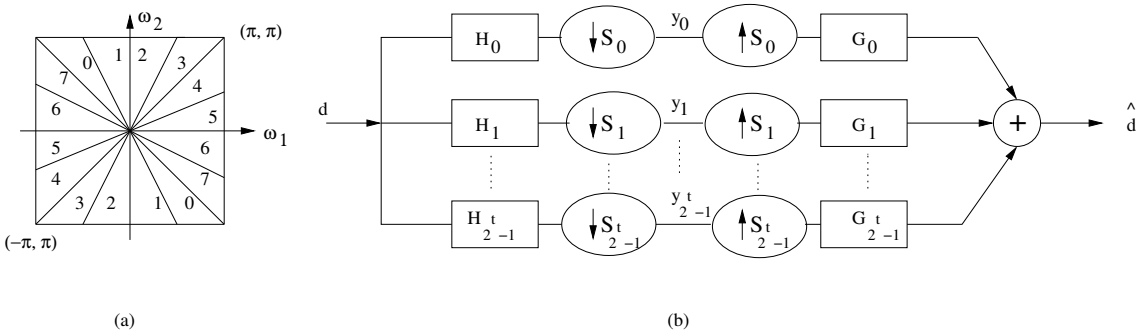


Figure 4.3: (a) An example of the directional filter bank frequency partitioning with $t = 3$. The subbands $0 - 3$ correspond to the *mostly horizontal* directions, while subbands $4 - 7$ correspond to the *mostly vertical* directions. (b) Multichannel view of an t-level tree-structured directional filter bank.

Combining the Laplacian pyramid and the directional filter bank yields the discrete contourlet transform. The multi-scale and directional decomposition stages in the contourlet transform are independent of each other and hence each scale can be

decomposed into any arbitrary power of two number of directions and different scales into different number of directions. Fig. 4.4 shows the pyramidal directional filter bank structure that implements the contourlet transform. Fig. 4.5 shows the three-scale contourlet decomposition of the "Lena" image for the purpose of illustration. It may be noted that the coefficients at the finer scale are no longer horizontally or vertically oriented as is the case in Cartesian wavelet decomposition.

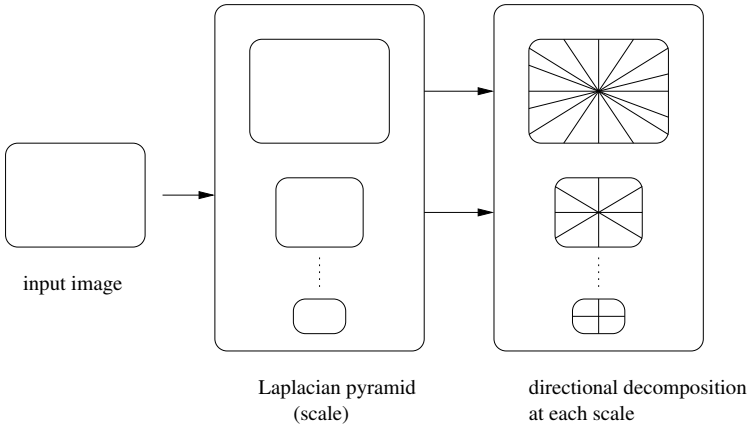


Figure 4.4: Pyramidal directional filter bank structure that implements the contourlet transform.

4.2 Learning the Contourlet Coefficients

Given a low resolution input image f, we perform a contourlet decomposition consisting of two pyramidal levels and each pyramidal level is then decomposed into four directional subbands which yield the decomposition as shown in Fig. 4.6(a). A three level decomposition is performed on all the high resolution database images and each pyramidal level is decomposed into four directional subbands resulting in the decomposition as shown in Fig. 4.6(b). Our idea is to learn the contourlet coefficients in the four directional subbands corresponding to the finest level for the given low resolution image (shown with dotted lines in Fig. 4.6(a)). After learning, effectively we have a three level decomposition for the input image, i.e, the original low level decomposition coefficients plus the learned coefficients at the finer scale. The inverse

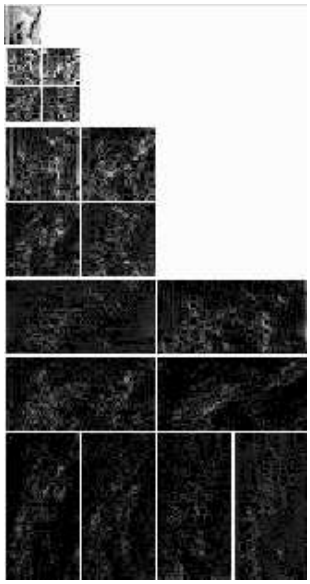


Figure 4.5: Contourlet transform of the "Lena" image using three Laplacian pyramidal levels and eight directions at the finest level and four directions at the coarser levels.

transform of this will yield the high resolution equivalent of the low resolution input.

Fig. 4.6 illustrates the block schematic of how the contourlet coefficients at finer scales are learned from a set of K training images using a two level contourlet decomposition of the low resolution test image. This is quite similar to the learning process explained in previous chapter in Fig.3.2. As explained earlier, the high resolution training images are decomposed into three pyramidal levels and the test image at each location is compared to the training images in the contourlet domain at two coarser scales to search for presence of a nearly identical edge at all possible locations. This is required for extrapolating the missing contourlet coefficients in the directional subbands $IX - XII$ for the test 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 $I - IV$ and the corresponding 2×2 blocks in the subbands $V - VIII$ we extrapolate a block of 4×4 contourlet coefficients in each of the subbands IX, X, XI, and XII. We follow the minimum absolute difference (MAD) criterion to estimate the contourlet coefficients. We take the absolute difference locally

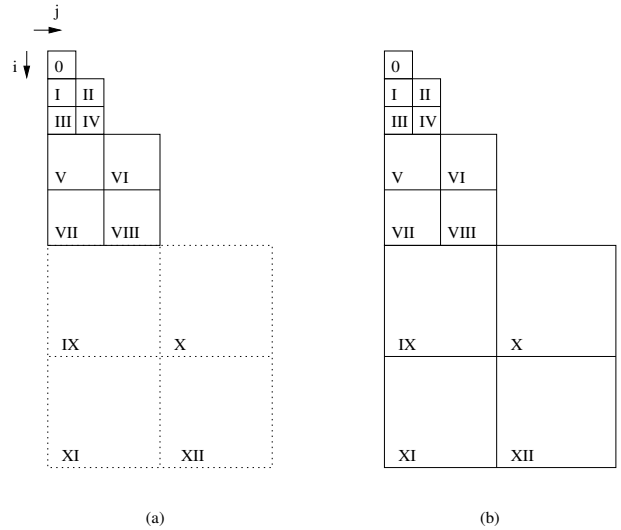


Figure 4.6: Illustration of learning the contourlet coefficients at a finer scale. (a) A low resolution image with two level decomposition. Coefficients in the dotted subbands are to be learned. (b) A representative high resolution training set in contourlet domain with three level decomposition.

between the contourlet 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 $I - VIII$ of the low resolution image. Denote the contourlet coefficient at a location (i, j) as $d(i, j)$. Consider the range $0 \leq i, j \leq M/4$. The contourlet coefficients $d_I(i, j), d_{II}(i, j + M/4),\ d_{III}(i+M/4, j),\ d_{IV}(i+M/4, j+M/4)$ corresponding to subbands $I-IV$ and 2×2 blocks consisting of $d_V(k, l), d_{VI}(k, l+M/2), d_{VII}(k+M/2, l), d_{VIII}(k+M/2, l+M/2)$ for $k = 2i : 2i + 1\ \ \&\ \ l = 2j : 2j + 1$ corresponding to subbands $V - VIII$ in the low resolution test image and all the high resolution training images are considered to learn a 4×4 contourlet block in each of the subbands $IX - XII$ consisting of unknown coefficients $d_{IX}(k, l), d_X(k, l+M), d_{XI}(k+M, l), d_{XII}(k+M, l+M)$ for $k = 4i : 4i + 3\ \ \&\ \ l = 4j : 4j + 3$. Thus, for a given set of a total of twenty contourlet coefficients in subbands $I - VIII$ in the low resolution image, we perform a search in the two coarser pyramidal levels of all the training images at all pixel locations for the best match in the MAD sense and copy the corresponding 4×4 contourlet block

in bands $IX - XII$ to those bands for the test image. In effect, we use the following equation to find the minimum.

$$\begin{aligned}\hat{p}(\hat{r},\hat{s}) = \arg\min_{p,r,s}[&|d_I(i,j) - d_{I^{(p)}}(r,s)| \\ &+|d_{II}(i,j+M_1) - d_{II^{(p)}}(r,s+M_1)| \\ &+|d_{III}(i+M_1,j) - d_{III^{(p)}}(r+M_1,s)| \\ &+|d_{IV}(i+M_1,j+M_1) - d_{IV^{(p)}}(r+M_1,s+M_1)| \\ &+S_V + S_{VI} + S_{VII} + S_{VIII}]\end{aligned} \tag{4.1}$$

where $M_1 = M/4$ and

$$\begin{aligned}S_V = &|d_V(2i,2j) - d_{V^{(p)}}(2r,2s)| \\ &+|d_V(2i,2j+1) - d_{V^{(p)}}(2r,2s+1)| \\ &+|d_V(2i+1,2j) - d_{V^{(p)}}(2r+1,2s)| \\ &+|d_V(2i+1,2j+1) - d_{V^{(p)}}(2r+1,2s+1)|\end{aligned}$$

and S_{VI}, S_{VII} and S_{VIII} are the corresponding sums for subbands VI, VII and $VIII$ respectively and $p = 1, 2, \cdots, K$ and $0 \leq r, s \leq M/4$. Here $d_J(p)$ denotes the contourlet coefficients for the p^{th} training image at the J^{th} subband. Here $\hat{p}(\hat{r},\hat{s})$ denotes the $(\hat{r},\hat{s})^{th}$ location for the $\hat{p}^{th}$ training image that best matches the test image at $(i,j)^{th}$ location in terms of contourlet coefficients.

Thus we have,

$$\begin{aligned}d_{IX}(i,j) &:= d_{IX}(\hat{p})(\hat{r},\hat{s}), \\ d_X(i,j) &:= d_X(\hat{p})(\hat{r},\hat{s}), \\ d_{XI}(i,j) &:= d_{XI}(\hat{p})(\hat{r},\hat{s}), \\ d_{XII}(i,j) &:= d_{XII}(\hat{p})(\hat{r},\hat{s}),\end{aligned}$$

for $(i,j),(\hat{r},\hat{s}) \in (IX - XII)$. This is repeated for each coefficient in subbands I, II, III and IV of the low resolution image. In effect, we find the best matching 8×8 edge primitive from the training data for a given 4×4 representation in the low resolution image through contourlet expansion.

It may be mentioned here that each 4×4 region in the low resolution image is being learned from different training images independently. In case the MAD error 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. Such spurious learning will introduce unwanted artifacts in the reconstructed image. In order to avoid such artifacts, we accept the contourlet coefficients only when the MAD is less than a chosen threshold. The threshold selection is based on the same criteria as discussed in section 3.3.2 in chapter 3. The goodness of the learning depends on how extensive and useful is the training data set. The subband 0 corresponds to the coarsest resolution (see Fig. 4.6(a)) in the contourlet decomposition and since the corresponding training set may have different average brightness, including the pixels from the 0-band 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 0^{th} band while learning.

In our experiments we used "$9-7$" biorthogonal filters [107] for the Laplacian pyramid because they are close to being orthogonal and also because of their linear phase characteristics. For the directional filter banks we used the "$23-45$" biorthogonal quincunx filters designed by Phoong *et al.*[108] and modulate them to obtain the biorthogonal fan filters. These filters are also nearly orthogonal and have linear phase response.

The complete learning-based resolution enhancement procedure is summarized below in terms of the steps involved.

STEP 1 *: Perform two level contourlet decomposition with four directional subbands on the low resolution test image of size $M \times M$ and three level decomposition on all training images each of size $2M \times 2M$.*

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

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

STEP 4 *: If MAD $<$ threshold, obtain the unknown high resolution contourlet coefficients (4×4 block) from the training image offering the best match locally in subbands $IX - XII$, else set them all zeros.*

STEP 5 *: Repeat steps (2 - 4) for every contourlet coefficient in bands $I - IV$ of the low resolution image.*

STEP 6 *: Perform inverse contourlet transform to obtain the high resolution image of the given test image.*

It may be noted that we have explained the super-resolving procedure for the special case when the image is decomposed into 4 directional components at each resolution. The same procedure remains valid if we prefer to have decomposition into 8 or 16 directional components. However, some of the notations and equations used in this section need to be properly adjusted.

4.3 Experimental Results

In this section we demonstrate the usefulness of the proposed approach to super-resolve a low resolution observation using contourlet learning. Experiments were performed for various types of face, fingerprint, natural and texture images. The training set consists of about 100 good quality images of all possible class of objects and is not specific to the class of objects to be super-resolved. Further,the training data is the same for all results displayed in this section.

First we consider experiments with a natural image. To obtain a low resolution test image and in order to be able to quantify the improvement during super-resolution, we consider a high resolution image which does not belong to the training set and downsample it by a factor 2 using the decimation matrix D in equation (3.2). Fig.

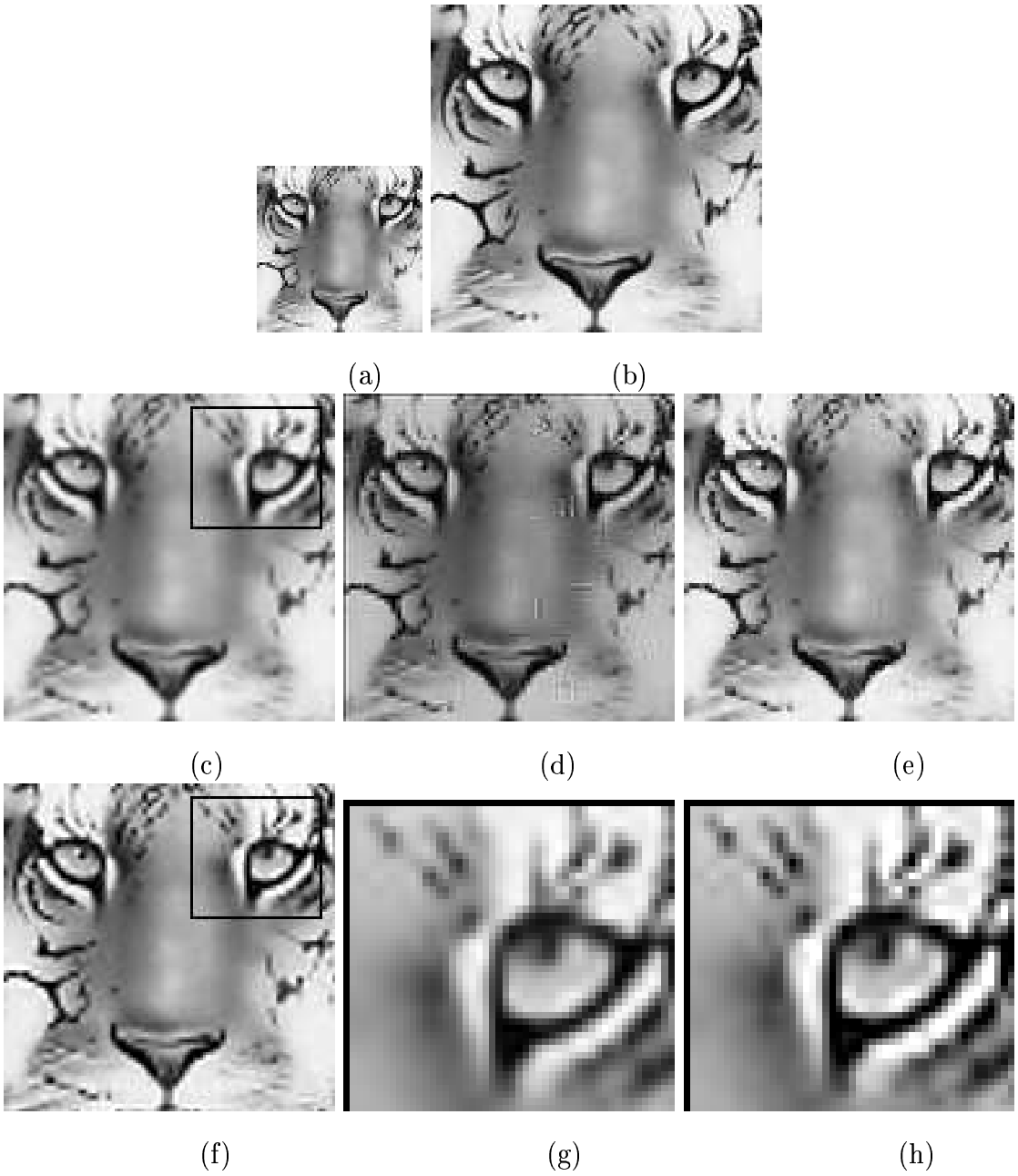


Figure 4.7: (a) A low resolution tiger image, (b) original high resolution image, (c) bicubic interpolated image, (d) super-resolution using wavelet learning without smoothing, (e) super-resolution using wavelet learning with smoothing, (f) the super-resolved image using the proposed approach, (g) zoomed up portion of the marked region from (c) and (h) zoomed up portion of the marked region from (f).

4.7(a) shows one such low resolution image of size 64×64. Fig. 4.7(b) is the original HR image. Fig. 4.7(c) shows the LR test image upsampled by a factor of 2 using the bicubic interpolation technique. The super-resolved image obtained using a wavelet-based learning described in chapter 3 is shown in Fig. 4.7(d) where we can observe lots of artifacts. If this result is further regularized using an appropriate edge preserving smoothness constraint as we did in 3, the artifacts can be minimized and the corresponding result is shown in Fig. 4.7(e). Even though, this result seems sharper than what is shown in Fig. 4.7(d), it is still slightly blocky (observe the eyes). The super-resolved image using contourlet-based learning is shown in Fig. 4.7(f). Here we can note that the artifacts as seen in Fig. 4.7(d) are almost absent. This is because of the capability of the contourlet transform to capture smoothness along contours. Note that all these results are quite sharp compared to the result of bicubic interpolation shown in Fig. 4.7(c). A comparison of the Figs. 4.7(c) and (f) shows more clear details in the super-resolved image. Figs. 4.7(g) and (h) are the zoomed up versions corresponding to the marked regions in Fig.s 4.7(c) and (f), respectively. Here one can clearly observe that the super-resolved image using contourlet learning is quite sharper than the bicubic interpolated one. The super-resolved image is quite close to the original HR image shown in Fig. 4.7(b).

Next we show the results of the experiment performed on a low resolution fingerprint image. Fig. 4.8(a) shows the LR observation of size 128×128 and Fig. 4.8(b) is the original HR image of size 256×256. The bicubic interpolated image is shown in Fig. 4.8(c). Fig 4.8(d) shows the results of wavelet-based learning with smoothing. The super-resolved image using the proposed approach is shown in Fig. 4.8(e). Figs. 4.8(f) and (g) are the zoomed up versions corresponding to the highlighted regions in Fig.s 4.8(c) and (e), respectively. It can be observed that the contours in the super-resolved image are less blurred than the bicubic interpolated image.

Fig.s 4.9(c - e) show the results of the corresponding experiments conducted on an LR textured image shown in Fig. 4.9(a). The super-resolved image using the proposed approach seems to be much sharper compared to the results of bicubic interpolation and of wavelet-based learning. In particular, the edges are better preserved

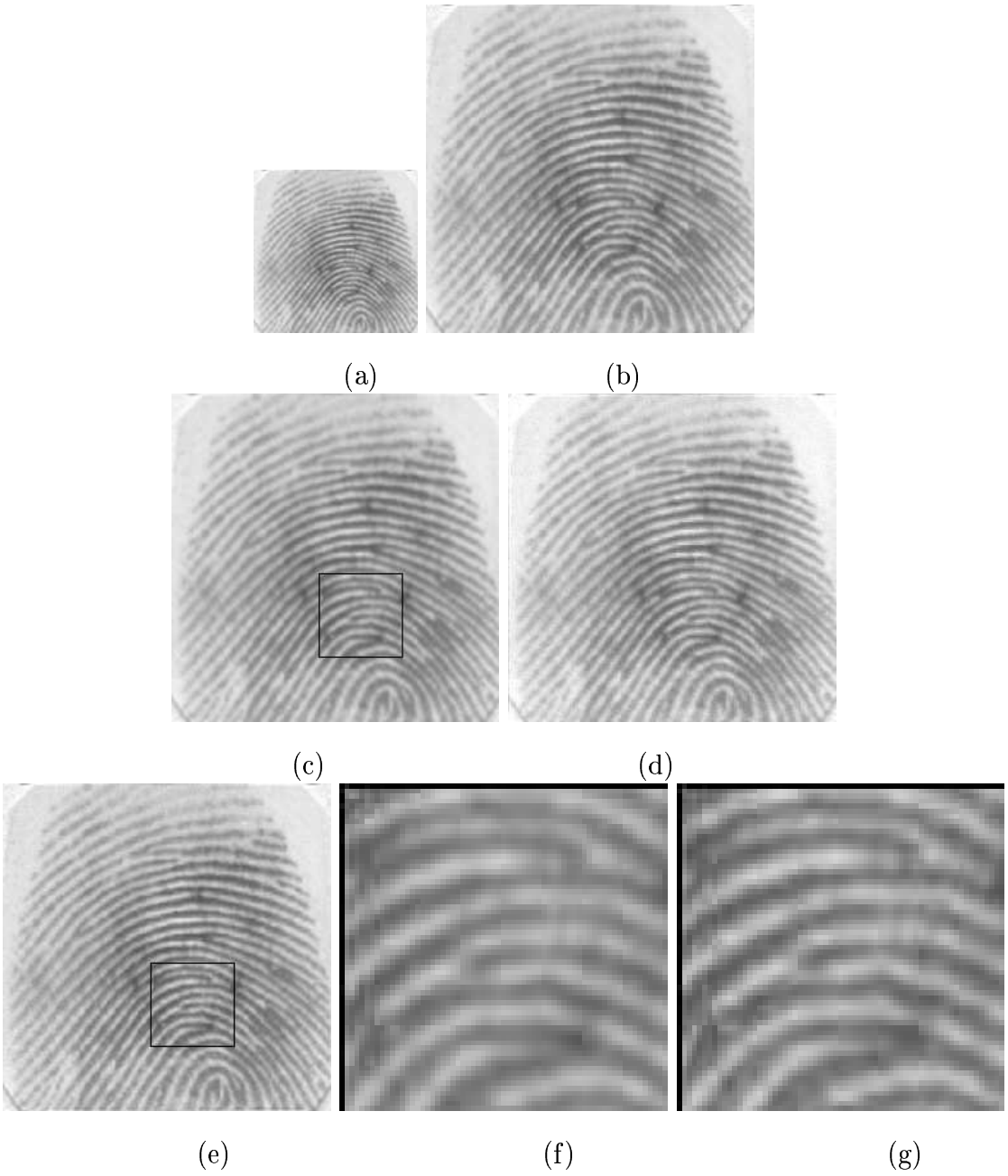


Figure 4.8: (a) A low resolution fingerprint observation, (b) original high resolution image, (c) bicubic interpolated image, (d) super-resolution using wavelet learning with smoothing, (e) the super-resolved image using the proposed approach, (f) zoomed up portion of the marked region from (c) and (g) zoomed up portion of the marked region from (f).

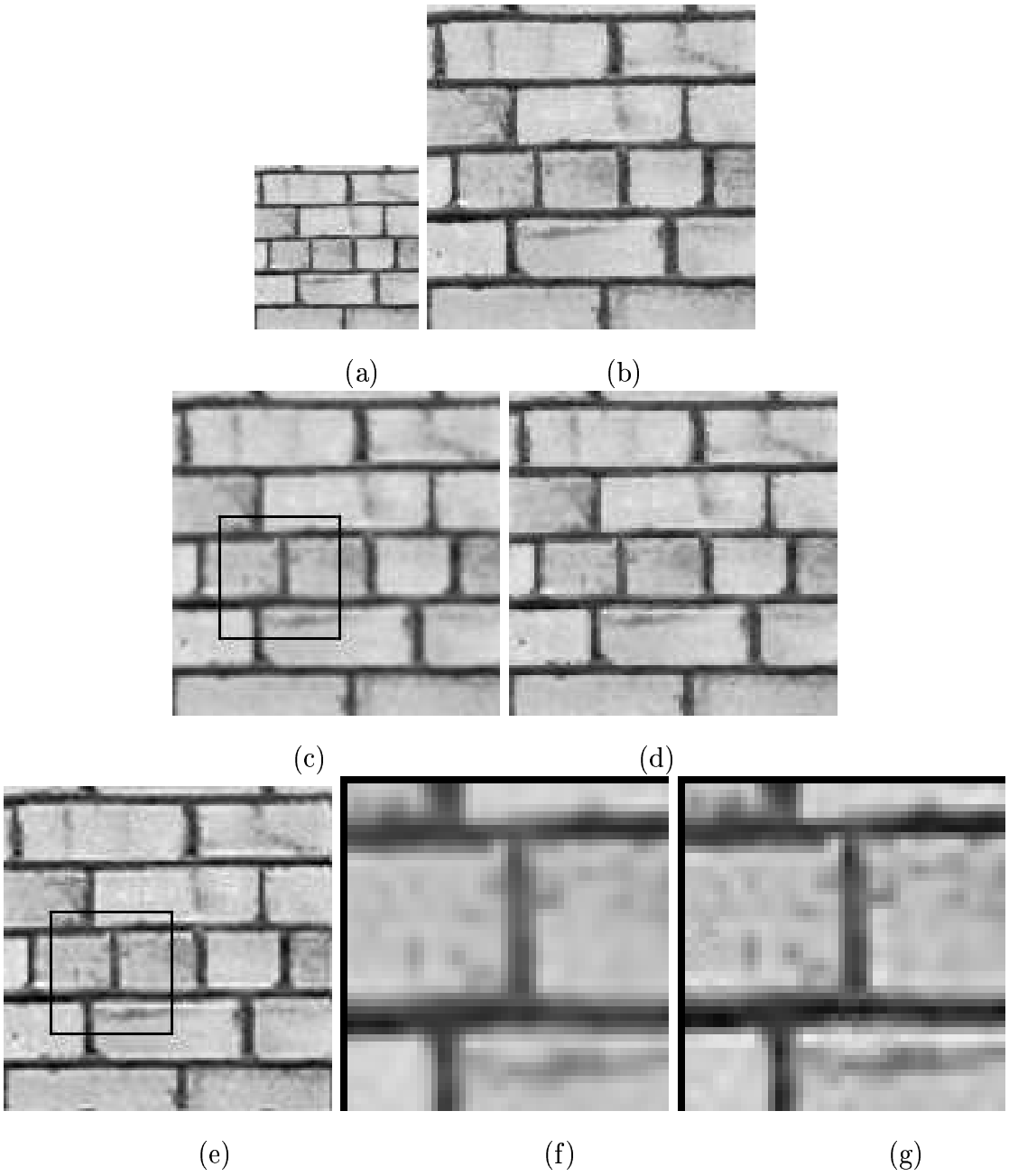


Figure 4.9: (a) A low resolution textured image, (b) original high resolution image, (c) bicubic interpolated image, (d) super-resolution using wavelet learning with smoothing, (e) the super-resolved image using the proposed approach, (f) zoomed up portion of the marked region from (c) and (g) zoomed up portion of the marked region from (f).

in the super-resolved image using contourlet learning than the bicubic interpolated image where it appears to be more blurred. The super-resolved image compared very favorably to the original high resolution image shown in Fig. 4.9(b).

Now we show the results of the experiments performed on a LR image where the aliasing is very high. The purpose of this experiment is to demonstrate the behavior of the proposed method when severe aliasing is present in the LR data. Such a low resolution image is shown in Fig. 4.10(a) and the corresponding bicubic interpolated image is shown in Fig. 4.10(c). Note that the stripes on the scarf is aliased. Fig. 4.10(d) shows the result of wavelet-based learning with smoothing. The super-resolved image using the proposed approach is shown in Fig. 4.10(e). The super-resolved image appears to be much sharper than the bicubic interpolated one. However, the proposed method was unable to remove the aliasing effect.

Finally, we obtain the PSNR of the reconstructed images in table 4.1. For the fingerprint image, we observe that it offers upto a $4dB$ gain in PSNR over the bicubic interpolation technique and over $3dB$ gain over the wavelet-based learning technique that incorporates the highly time consuming smoothness constraint. This is quite expected as the fingerprint has many curved edges which cannot be handled well with separable basis functions. The contourlet expansion provides a much better solution. The PSNR improvement is about $2dB$ when the input is quite a smooth one such as the face image, and it improves to over $3dB$ when the input image has noticeably strong edges such as in the brick texture or in the tiger image.

We end this section with a comment on how the high resolution contourlet coefficients were learned from the HR database images. For the experiment with the face image, about 24% of all the contourlet coefficients were learned from the HR database and the rest were set to zero as no matching coefficients could be found from the database for a given choice of threshold. Further, all these were picked up from various different images in the database (such as faces, textures, building images, natural textures, etc) suggesting that the learning process is not class specific.

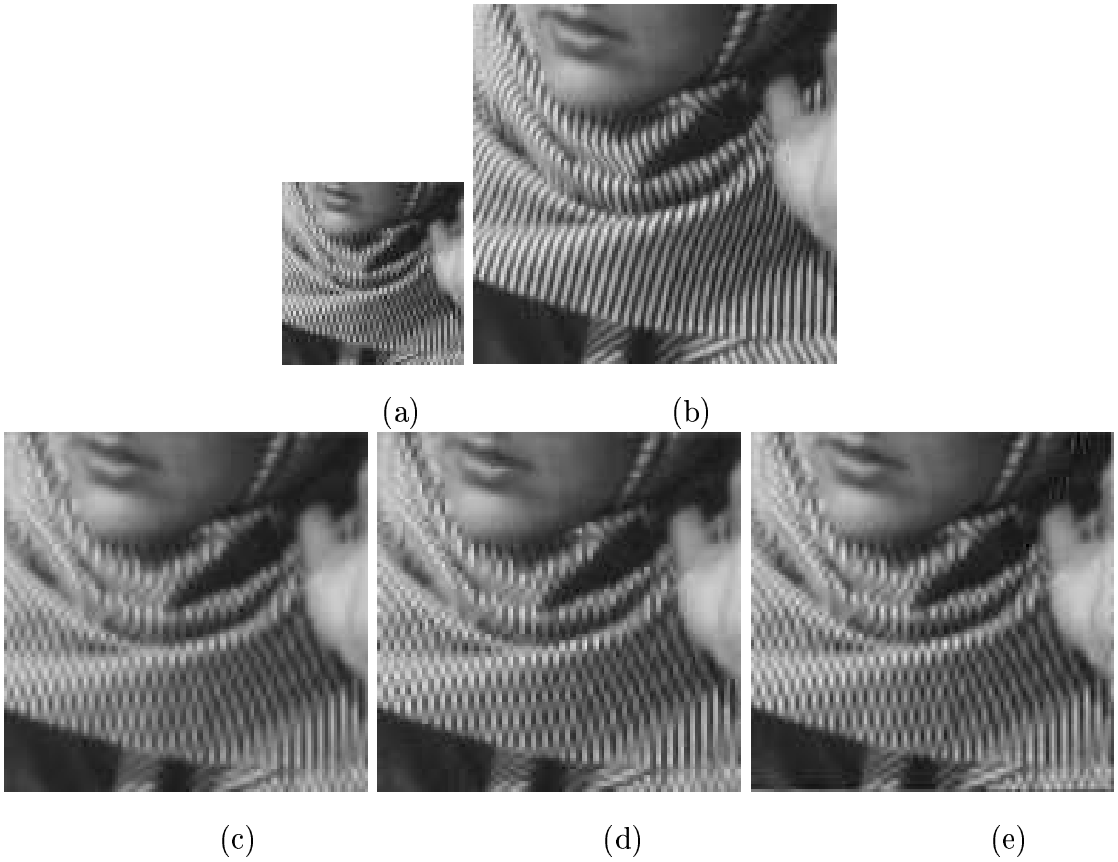


Figure 4.10: (a) A severely aliased low resolution observation, (b) original high resolution image, (c) bicubic interpolated image, (d) super-resolution using wavelet learning with smoothing and (e) the super-resolved image using the proposed approach.

Image	Bicubic inter-polation	Wavelet Learning		Contourlet learning
		without smoothing	with smoothing	
Tiger	20.88	15.41	20.18	24.18
Fingerprint	30.81	14.46	30.25	34.30
Texture	22.69	17.53	23.29	26.05
Face	19.95	17.67	19.12	21.81

Table 4.1: Comparison of PSNRs for the zoom factor $q = 2$ expressed in dB .

4.4 Conclusions

We have described a single-frame image super-resolution technique using a contourlet transform-based learning. The contourlet coefficients at finer scales, learned from a set of several high resolution training images, after proper thresholding to avoid spurious learning, are used to estimate the super-resolved image. The learning process ensures capturing the best high resolution edges from the training set given a low resolution observation, as well as captures the smoothness along contours. The results obtained for different class of images show perceptual as well as quantifiable improvements over conventional interpolation techniques. It also shows a better edge over the regularized standard wavelet-based learning described in chapter 3.

Chapter 5

Linear Combination of Edge Primitives

In this chapter we further extend the contourlet-based learning method discussed in the previous chapter, using a linear combination of edge primitives inspired by the concept of locally linear embedding (LLE) [109], a recent manifold learning method, under a supervised learning framework. The methods proposed by Freeman *et al.* [49] and Hertzmann *et al.* [67] are based on supervised learning, i.e., the training data are given as input-output pairs of low and high resolution images. In machine learning, it is generally believed that when feasible, a supervised learning approach often leads to the best results. Our work discussed in this chapter is motivated by the work of Chang *et al.* [110] where the authors propose a patch-based local learning method for super-resolution through neighborhood embedding. Here, the generation of a high resolution patch does not depend on only one of the nearest neighbors in the training set. Instead, it depends simultaneously on multiple nearest neighbors in a way similar to LLE for manifold learning. Here low resolution and the corresponding high resolution training image pairs are used to estimate the high resolution embedding.

Our method is similar to the work of Chang *et al.*, but we perform the patch-based local learning in the contourlet transform domain, having identified the capability of contourlets to capture the geometric smoothness along contours. For this supervised

learning scheme we make use of a training set consisting of pairs of LR and the corresponding HR images. The advantage of this lies in capturing the effect of the decimation process (matrix D in Eq. (3.2)) explicitly by the contourlet transform through the matched pair of LR and HR training data. In previous two chapters it has been implicitly assumed that the multiresolution decomposition using either wavelet or contourlet has the same effect as that of the decimation matrix D, which is often not true. For a given low resolution test image patch (in contourlet domain) we find k nearest neighbors from the LR training set by performing an extensive search using the Euclidean distance to define the neighborhood. Now the test patch is reconstructed from a linear combination of the k neighbors where the reconstruction weights are computed so as to minimize the reconstruction error. The linear combination of the corresponding k HR patches using the same reconstruction weights generate the required super-resolved patch in the contourlet domain.

5.1 Problem Formulation

The single image super-resolution problem that we solve in this chapter can be formulated as follows. Given a low resolution image f as input, we estimate the unknown target high resolution image z with the help of a training set consisting of low resolution images and the corresponding high resolution images. It may be noted that this method is different from the SR techniques described in the last two chapters where the training set was consisting only of high resolution images. Here also, we use the same low resolution image formation model described in chapter 3.

We assume that, among our training data, the set of HR patches and the set of corresponding LR patches respectively reside on two different nonlinear manifolds, but with similar locally linear structure. The same assumption has been made in Chang *et al.* and shown to produce stable super-resolution performance. In our case, we represent each low or high resolution image as a set of non-overlapping patches in the contourlet domain. Recently, some new manifold learning or nonlinear dimensionality reduction techniques have been proposed to automatically discover

low-dimensional nonlinear manifolds in high-dimensional data spaces and embed them onto low-dimensional embedding space, using tractable linear algebraic approaches that are not prone to local minima. These include isometric feature mapping (Isomap) [111], LLE [109, 112] and Laplacian eigenmap [113]. In our method we use an LLE based approach because of its simplicity.

5.2 LLE Method

LLE is a promising manifold learning method that has aroused a great deal of interest in machine learning. It is a way to map high dimensional data into a low dimensional space with the useful property of maintaining the neighborhood relationship. LLE assumes that each data point and it neighbors lie on a locally linear patch of the manifold. The local geometry of each patch can be characterized by the reconstruction weights with which the data point is reconstructed from its neighbors. The LLE algorithm can be summarized as follows. See [112] for details.

1. For each data point in the Ψ-dimensional data space:
   (a) Find k nearest neighbors in the same space.
   (b) Obtain the reconstruction weights of the k neighbors so as to minimize the reconstruction error.
2. Compute the low-dimensional embedding in the ψ-dimensional space (usually $\psi < \Psi$) such that the local geometry represented by the reconstruction weights is preserved.

5.3 Combination of Contourlet Primitives

We use LLE to construct the HR contourlet coefficients of the given LR edge primitive as opposed to the low dimensional embedding in standard LLE. First we perform the contourlet decomposition of the LR and HR training images and also for the given test image. For simplicity we have performed a single pyramidal level decomposition

and with four directional subbands. For a given test image patch in the contourlet domain in a specific subband, we perform an extensive search in the LR database to find k nearest neighbors. Now we express the given test patch as a linear combination of these neighbors. The reconstruction weights are computed so as to minimize the reconstruction error

$$e = ||\tilde{\mathbf{f}} - \sum_{i=1}^{k} w_i \tilde{\mathbf{f}}_\mathbf{i}||^2. \tag{5.1}$$

where $\tilde{\mathbf{f}}$ is the vector containing the contourlet coefficients of the test patch and $\tilde{\mathbf{f}}_\mathbf{i}$ represents the corresponding vector for the i^{th} nearest neighbor and w_i's are the reconstruction weights subject to the constraint that $\sum_{i=1}^{k} w_i = 1$. To solve for w_i we form the local Gram matrix $\mathcal{G}$ given by [110]

$$\mathcal{G} = (\tilde{\mathbf{f}}\,\mathbf{1}^T - \mathbf{F})^\mathbf{T}(\tilde{\mathbf{f}}\,\mathbf{1}^\mathbf{T} - \mathbf{F}). \tag{5.2}$$

where $\mathbf{1}$ is a column vector of ones and $\mathbf{F}$ is a $\Psi \times k$ matrix with its columns being the neighbors of $\tilde{\mathbf{f}}$. For a decimation factor of q, $\Psi = q^2\psi$. Now $\mathbf{w}$ can be obtained by solving the linear system of equations $\mathcal{G}\mathbf{w} = \mathbf{1}$ and finally normalizing the weights such that $\sum_{i=1}^{k} w_i = 1$. The linear combination of the corresponding k HR patches of the LR neighbors with the same reconstruction weights computed as above give the super-resolved patch for the given test patch in the contourlet domain. The procedure is repeated for all the patches and the inverse contourlet transform reconstructs the SR image. We learn the contourlet coefficients in all subbands except at the lowpass subband where we obtain the contourlet coefficients from a suitably interpolated version of the LR image. This is because the coefficients in the lowpass subband do not really contribute towards the missing high frequency details which we try to infer.

The LLE-based resolution enhancement procedure is summarized below in terms of the steps involved.

STEP 1 : Perform contourlet decomposition on the low resolution test image, and also on the LR and HR training images.

STEP 2 : For each 2×2 patch in a particular subband

1. Find k nearest neighboring patches by searching the LR database.

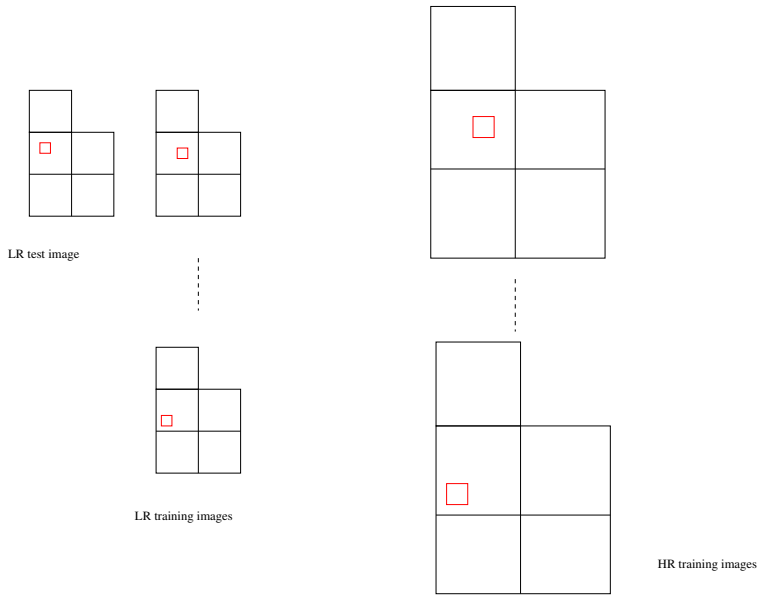


Figure 5.1: Illustration of contourlet decompositions of test, LR and HR training images for HR reconstruction. The box shows a patch in the test image, neighbors in the LR training images and the corresponding patches in the HR training images.

2. Compute the reconstruction weights of the neighbors that minimize the error of reconstructing the test patch under consideration.
3. Compute the high resolution coefficients using the linear combination HR patches corresponding to the above k LR patches with the same reconstruction weights obtained from step 2.2.

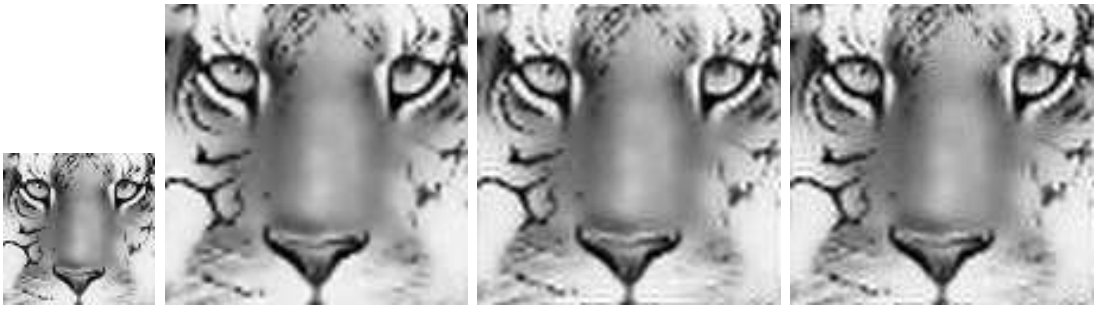
STEP 3 : Repeat the above for all the patches in all subbands.

STEP 4 : Perform inverse contourlet transform to obtain the super-resolved image.

The procedure is illustrated in Fig.5.1.

5.4 Experimental Results

In this section we present our experimental results conducted on various images. Here we use the same high resolution database as used in the previous chapter. The low



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

Figure 5.2: (a) A low resolution texture image, (b) bicubic interpolated image, (c) super-resolution using linear combination of edge primitives and (d) super-resolution using the best matching edge primitive (i.e. $k = 1$).

resolution database is constructed by downsampling the HR database. In our experiments we made use of five nearest patches ($k = 5$) corresponding to every test image patch to construct their high resolution contourlet coefficients. Figs. 5.2(a) and (b) show a low resolution texture image and its bicubic interpolated versions respectively. The super-resolved image using contourlet learning using linear combination of edge primitives is shown in Fig.5.2(c). This output looks slightly sharper compared to the bicubic interpolated one, but a little noisy at certain locations. This could be due to spurious learning at those locations and since we use multiple neighbors the effect of spurious learning can get added up making the output noisy. The super-resolved image obtained using the best matching patch ($k = 1$) for every test patch is shown in Fig.5.2(d). This output looks much blocky and noisy in comparison to that shown in Fig.5.2(c).

Next we perform the experiment on a low resolution Barbara image where the aliasing is slightly high, shown in Fig.5.4(a). The bicubic interpolated and the super-resolved image using the proposed approach are shown in Figs.5.4(b) and (c) respectively. The super-resolved image is sharper than the bicubic interpolated one. Observe the scarf. But it looks slightly noisy which is due to spurious learning. Observe the palm.

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

Figure 5.3: (a) A severely aliased low resolution observation, (b) bicubic interpolated image, (c) super-resolution using linear combination of edge primitives and (d) super-resolution using the best matching edge primitive (i.e. $k = 1$).

5.5 Conclusions

In this chapter we have proposed a local learning method for super-resolution under a supervised framework. To obtain the contourlet coefficients of the unknown high resolution image at the finer subbands, we make use of several nearest neighbors of the low resolution test patch as in locally linear embedding. In effect, we make use of a linear combination of several best matching edge primitives to obtain the edge elements of the unknown high resolution image under a given down sampling process. The resulting high resolution image is sharper than the conventional bicubic interpolated image. But the disadvantage of the method is that, if there is some spurious learning at certain locations, the resulting high resolution image at the corresponding location will be much noisy as it is contributed by several spuriously learned neighbors. To overcome this difficulty one has to carefully set the threshold in order to avoid any spurious learning.

Chapter 6

PCA-based Global Approach

In the previous three chapters we presented methods where a single observation was super-resolved by learning the high resolution wavelet or contourlet coefficients representing the edges in the image from a given set of arbitrary training images. In all these approaches we were learning the coefficients *locally* from the image database. 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). 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.

i 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 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 the previous chapters we learned 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 or contourlet based representation of the image due to their excellent 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 learned locally. Certain aspects of the input image now has to be learned 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 for an efficient utilization of the signal statistics.

Similar PCA-based super-resolution reconstruction technique for face hallucination is proposed in [71, 114]. Here the authors make use of both low and high resolution image databases to recover the high resolution image. They also add constraints to the principal components to reduce the nonface-like distortion. The method that we propose in this chapter makes use of either a low resolution or high resolution image database to obtain the super-resolved image. An eigenface-domain super-resolution reconstruction algorithm for face recognition is proposed in [115]. This is a multi-frame super-resolution method where the authors apply super-resolution to a low-dimensional face sub-space rather than to the pixel domain as in conventional SR methods, resulting in a significant decrease in the computational complexity of SR reconstruction.

The chapter is organized as follows. We discuss the concept of generalized interpolation and its application to our problem using a low resolution image database in section 6.1.The usefulness of PCA for super-resolution is discussed in section 6.2 and

the complete algorithm is presented in 6.3. In section 6.4 we discuss the interpolation using a high resolution database. We present the experimental results in section 6.5 followed by discussions in 6.6.

6.1 Generalized Interpolation

The idea of generalized interpolation was proposed by Rajan *et al.*[116] for image expansion and generation of super-resolution images. This is achieved by decomposing the image into appropriate subspaces, carrying out interpolation in individual subspaces and subsequently transforming the interpolated values back to the image domain. Consider a function $f(x, y)$ decomposed as

$$f(x, y) = g(a_1(x, y), a_2(x, y), \ldots, a_m(x, y)) \tag{6.1}$$

where $a_i(x, y),\ i = 1, 2, \ldots, m$ are different functions of the interpolating variables x and y and when they are combined by an appropriate m-variate function g, one recovers the original function f. Now the individual functions $a_i(x, y)$ are interpolated and combined using equation(6.1) to obtain the rescaled function $f(x \uparrow q, y \uparrow q)$. In [116] photometric cues were used to decompose the image assuming a Lambertian reflectance model, and the interpolants were $a_1(x, y) = p(x, y), a_2(x, y) = q(x, y)$ and $a_3(x, y) = \rho(x, y)$ where p and q are the surface normals and ρ is the albedo. It was claimed that various optical and structural properties of the image, such as 3-D shape of an object, regional homogeneity, local variations in scene reflectivity, etc., can be better preserved during the interpolation process. It was also shown that an alias free reconstruction of $f(x \uparrow q, y \uparrow q)$ is possible if the sub functions are all bandlimited. In this chapter we use the same parametric decomposition given by Eq. (6.1), however, we decompose the given function into an eigen-space containing the principal components $\bar{a}_i,\quad i = 1, \cdots, m$ and use a linear function ϕ, *i.e.*,

$$\phi(a_1(x, y), a_2(x, y), \ldots, a_m(x, y)) = \sum_{i=1}^{m} w_i \bar{a}_i. \tag{6.2}$$

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

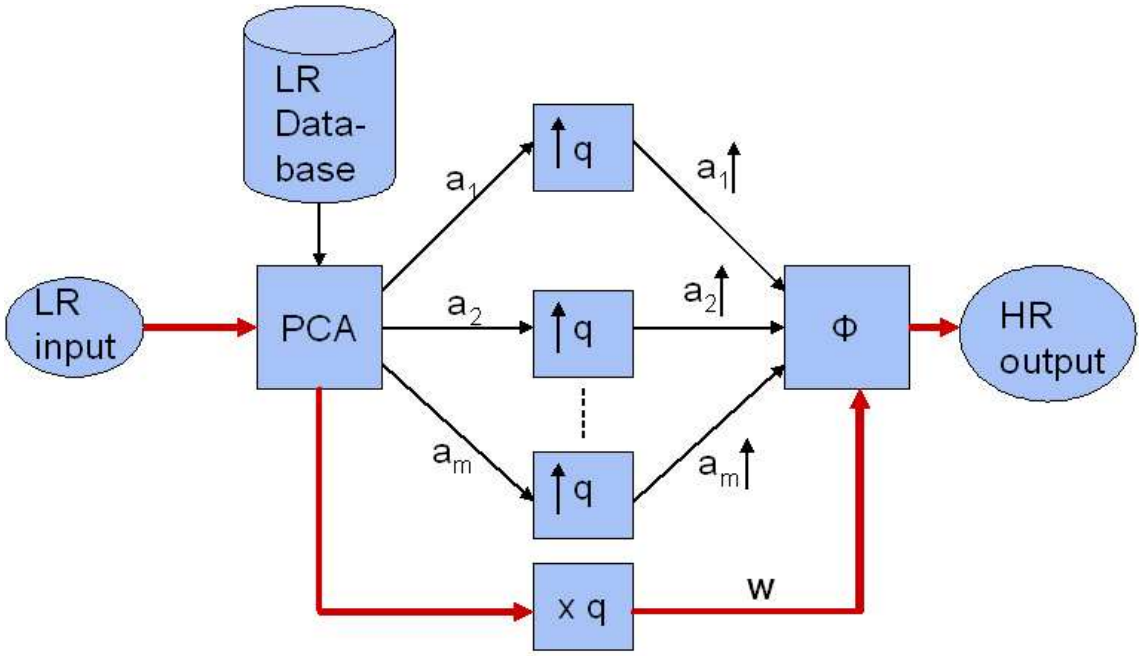


Figure 6.1: Illustration of PCA-based Generalized Interpolation.

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. The procedure is illustrated in Figure 6.1.

6.1.1 Eigen-Image Decomposition

Eigen-image decomposition for a class of similar objects is currently the most popular and actively pursued area of research in pattern recognition. An image can be reconstructed from eigen-images in the PCA representation as described in [117]. The basic procedure for computing the eigen-space is as follows: We have a dataset of K similar training images, represented by the matrix $\Lambda = [\boldsymbol{\mathcal{F}}_1, \boldsymbol{\mathcal{F}}_2, \ldots, \boldsymbol{\mathcal{F}}_K]$, 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 $M^2 \times m$. In PCA, a set of top m eigenvectors $E = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m]$, also called eigen-images, of dimension $M^2 \times m$ are computed from the covariance matrix,

$$\Sigma = \sum_{i=1}^{K} (\boldsymbol{\mathcal{F}}_i - \mathbf{m}_{\mathcal{F}})(\boldsymbol{\mathcal{F}}_i - \mathbf{m}_{\mathcal{F}})^T. \tag{6.3}$$

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

$$\mathbf{m}_{\mathcal{F}} = \frac{1}{K} \sum_{i=1}^{K} \boldsymbol{\mathcal{F}}_i. \tag{6.4}$$

Note that the size of the training database is much smaller than the dimension of the image, *i.e.*, $K << M^2$. Hence Σ in Eq. (6.3) is rank deficient. Further, one does not store all eigen-images for the non-zero eigenvalues. We retain only top m (where $m < K$) eigen-images based on the magnitude of the eigenvalues. Since $m << M^2$, 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 m suffices to reconstruct an image at good enough details.

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

$$\mathbf{w} = E^T(\mathbf{f} - \mathbf{m}_{\mathcal{F}}). \tag{6.5}$$

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

$$\hat{\mathbf{f}} = E\mathbf{w} + \mathbf{m}_{\mathcal{F}}. \tag{6.6}$$

Since m is typically much smaller than the size of the image vector, the image representation [116] through the eigen-image expansion is not complete. Hence $\hat{\mathbf{f}}$ is an approximation of $\mathbf{f}$ and the quality of approximation depends on its nearness to the class of images in the database.

6.1.2 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 m 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 $E_h = [\mathbf{e}_{1h}, \mathbf{e}_{2h}, \cdots, \mathbf{e}_{mh}]$, *i.e.*,

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

and

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

where the symbol $\uparrow q$ represents upsampling by a factor of q. One may use an appropriate upsampling factor such as $q = 2, 3, 4$, *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 q (*i.e.*, $\mathbf{w}_h = q\mathbf{w}$) to preserve the average brightness of the interpolated pictures. The high resolution image is now reconstructed using

$$\begin{aligned} \hat{\mathbf{z}} &= E_h \mathbf{w}_h + \mathbf{m}_z \\ &= q \sum_{i=1}^{m} w_i \mathbf{e}_i(\uparrow q) + \mathbf{m}_z, \end{aligned} \tag{6.7}$$

where w_i is the projection of the input image on the i^{th} eigen-image. Compare this to Eq. (6.2) and observe that Eq. (6.7) is nothing but the generalized interpolation. There are two primary differences with the way the generalized interpolation is carried out compared to the one presented in [116].

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 learned from the training images.

2. The decomposition of the original image $\mathbf{f}$ into eigen-images $\mathbf{e}_i$ is linear.

The motivation for using the generalized interpolation lies in the fact that the original function $\mathbf{f}$ 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. (6.7) are bandlimited, so is $\mathbf{f}$ due to linearity. Hence the PCA-based upsampling cannot eliminate aliasing from the given image $\mathbf{f}$. 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 [118]

$$|f - \tilde{f}| < \frac{|\delta x|^{n+1}}{(n+1)!}\max_x |(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})^{n+1} f(x)|. \tag{6.8}$$

For a thin plate fitting spline over a square grid of size h, the maximum error is bounded by [119]

$$|f - \tilde{f}| \leq ch\sqrt{|\log h|}\ \|\mathcal{L}f\|, \tag{6.9}$$

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 $\tilde{f}(x)$.

$$\tilde{f}(x) = \phi(a_1(x), a_2(x), \ldots, a_m(x)), \tag{6.10}$$

where $a_i(x),\ i = 1, 2, \ldots, m$ are different functions of the interpolating variable x and when they are combined by an appropriate m-variate function ϕ, one recovers the original function. We can now interpolate the individual functions $a_i(x)$ and combine them using Eq. (6.10) to obtain a rescaled $\tilde{f}(x)$.

The interpolation error at a point x can be written as

$$\begin{aligned} |f_g - \tilde{f}| &= |\phi(a_1(x) + \epsilon_1, \ldots, a_m(x) + \epsilon_m) \\ &\quad -\phi(a_1(x), a_2(x), \ldots, a_m(x))| \\ &\approx |\epsilon_1 \tfrac{\partial \phi}{\partial a_1} + \cdots + \epsilon_m \tfrac{\partial \phi}{\partial a_m}| \end{aligned} \tag{6.11}$$

where f_g represents the result of generalized interpolation. Here ϵ_i, $i = 1, 2, \ldots, m$ 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.*,

$$\phi(a_1(x), a_2(x), \ldots, a_m(x)) = \sum_{i=1}^{m} \alpha_i a_i(x), \;\; \alpha_i \geq 0 \; \forall i. \tag{6.12}$$

From Eq. (6.12), 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.,

$$|f_g(x) - \tilde{f}(x)| < \frac{|\delta|^{(n+1)}}{(n+1)!} \sum_{i=1}^{m} \alpha_i \max_x |(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})^{n+1} a_i(x)|. \tag{6.13}$$

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

$$|f_g(x) - \tilde{f}(x)| < \frac{|\delta|^{(n+1)}}{(n+1)!} \max_x |(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})^{n+1} f(x)|. \tag{6.14}$$

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. (6.13) and (6.14) and must prove that

$$\sum_i \alpha_i \max_x |\frac{\partial^{n+1} a_i(x)}{\partial x^{n+1}}| \leq \max_x |\frac{\partial^{n+1} f(x)}{\partial x^{n+1}}| \tag{6.15}$$

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

$$\sum_{i=1}^{K} \alpha_i \|\mathcal{L} a_i(x)\| \leq \|\mathcal{L} f(x)\|. \tag{6.16}$$

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.

6.2 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. (6.6)) 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

$$\mathbf{f}_b = \sum_i \alpha_i f(x+i) \triangleq \sum_i \alpha_i \mathbf{f}_i, \tag{6.17}$$

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. (6.5) we can compute

the projections on the eigen-images

$$\mathbf{w}_b = E^T\mathbf{f}_b = \sum_i \alpha_i (E^T\mathbf{f}_i) \triangleq \alpha_0\mathbf{w} + \sum_{i=1} \alpha_i \mathbf{w}_i, \tag{6.18}$$

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

$$\hat{\mathbf{f}}_b = E\mathbf{w}_b = \alpha_0 E\mathbf{w} + \sum_{i=1} \alpha_i E\mathbf{w}_i. \tag{6.19}$$

The error in reconstruction due to the blurred observation with respect to the previously obtained (see Eq. (6.6)) 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

$$|\hat{\mathbf{f}} - \hat{\mathbf{f}}_b| = |\sum_{i=1} \alpha_i E\mathbf{w} - \sum_{i=1} \alpha_i E\mathbf{w}_i| \quad = \quad |\sum_{i=1} \alpha_i E(\mathbf{w} - \mathbf{w}_i)|. \tag{6.20}$$

Since E form an orthonormal basis set,

$$|\hat{\mathbf{f}} - \hat{\mathbf{f}}_b| \leq \sum_{i=1} \alpha_i |(\mathbf{w} - \mathbf{w}_i)|. \tag{6.21}$$

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.

6.3 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 top m eigen-images represented by the matrix E and also obtain the mean image $m_{\mathcal{F}}$.

STEP 2: Project the given low resolution image $\mathbf{f}$ 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. (6.7).

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 de-blurring, 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.

6.4 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 6.3 to the input side before step-1 itself. The procedure is illustrated in Figure6.2.

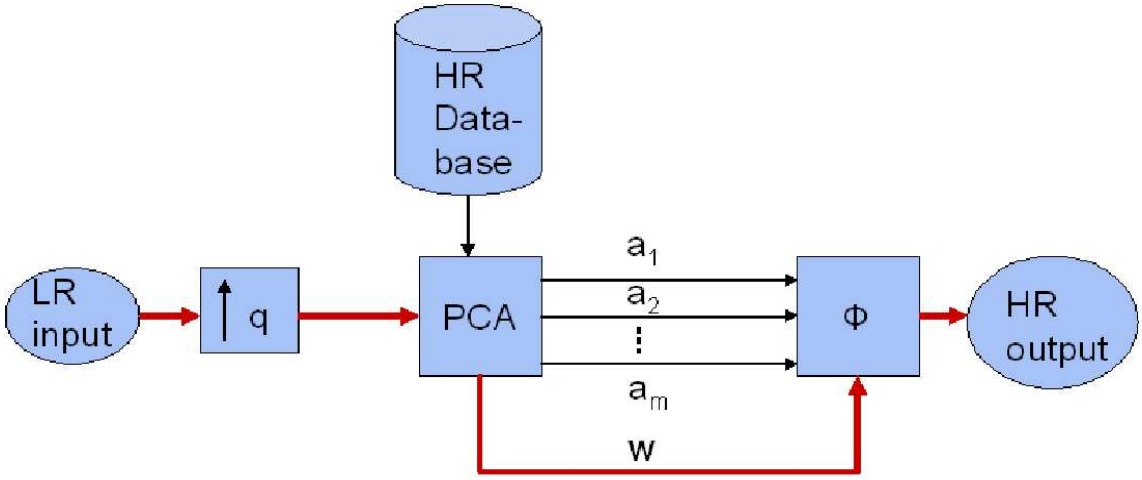


Figure 6.2: SR using high resolution database.

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 q), blurred and noisy image is first upsampled by a factor of q 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. (6.6). 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.

6.5 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 pre-processing 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 q to serve as the input image to be super-resolved. The same database is used to serve as the high resolution training data.

Figure 6.3 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 $q = 2, 3, 4,$ *etc.*) and stored for subsequent usage.

Figure 6.3: First ten eigen-images obtained from the training data set.

In Figure 6.4 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.

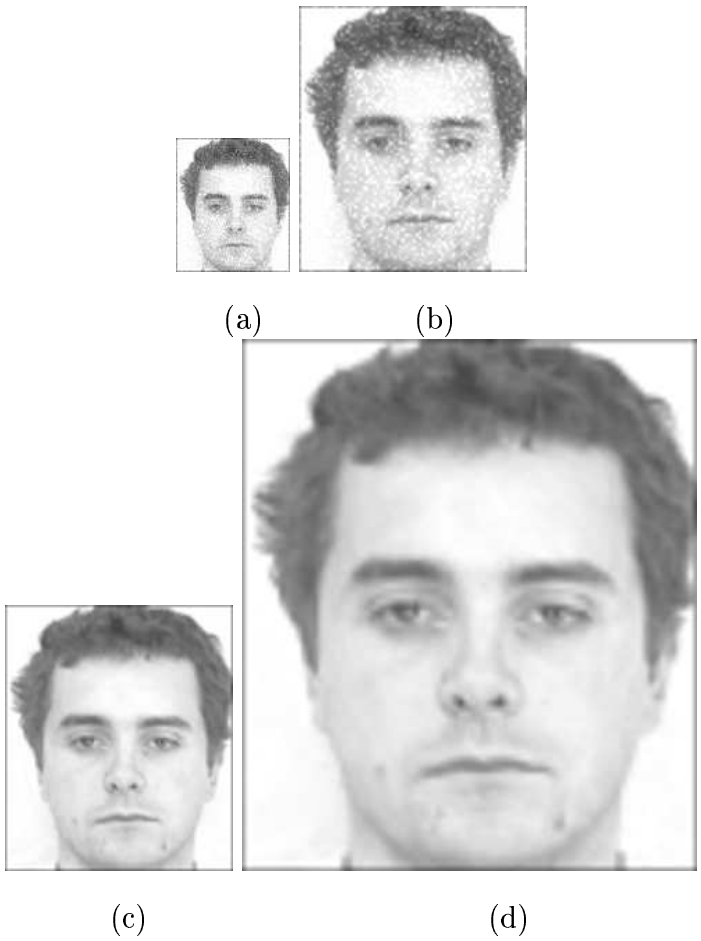


Figure 6.4: (a) A low resolution noisy observation ($\sigma = 0.1$), (b) bicubic interpolated image with $q = 2$, PCA-based restoration with (c) $q = 2$ and (d) $q = 4$.

We now experiment on what happens if the noise level is increased. In Figure 6.5, 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 Figure 6.4(c). Some artifacts are now visible on the left cheek. The performance is quantified in terms of the PSNR tabulated in table 6.1 where the PSNR for the bicubic interpolated image and super-resolved image for a zoom factor $q = 4$ and for different values of noise level are shown. As mentioned in section 6.2 it is observed that when the noise level σ is very large, the reconstructed image deviates

from the original face image.

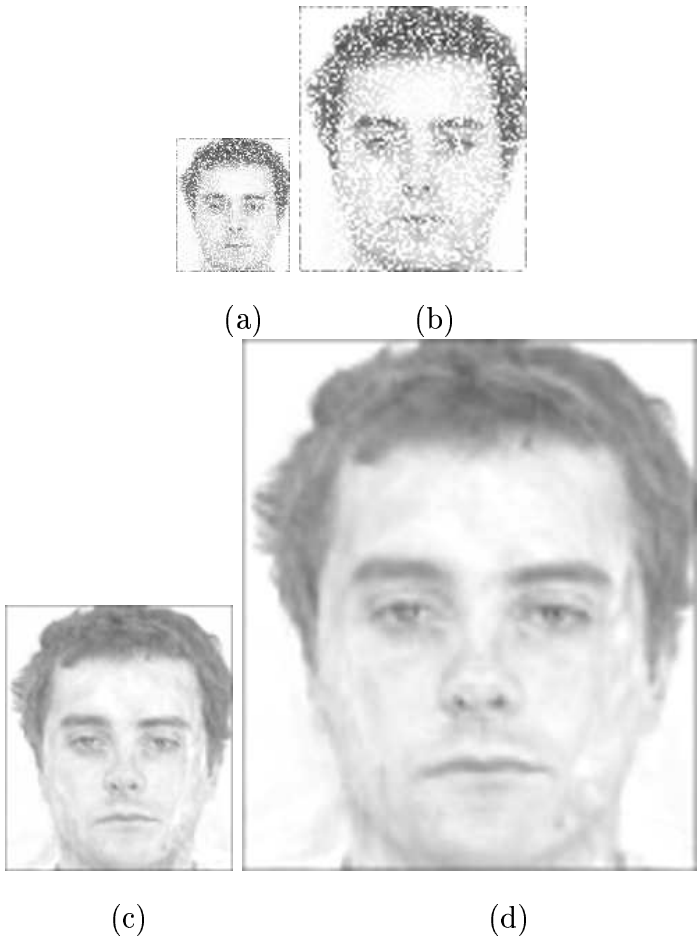


Figure 6.5: (a) A very noisy observation ($\sigma = 0.5$). (b) Result of bicubic interpolation with $q = 2$. PCA-based reconstruction for (c) $q = 2$, and (d) $q = 4$.

We now investigate the performance of the PCA-based method when the input image is severely blurred. In Figure 6.6(a) an input image which is blurred with a 7×7 Gaussian mask with a standard deviation of 2 is shown. The details on the face is almost 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 6.6(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.

(a)

(b) (c)

Figure 6.6: (a) A highly blurred low resolution observation, (b) bicubic interpolated image, with $q = 2$, and (c) the reconstructed image with $q = 2$.

Now we experiment on how many eigen-images are required for a good reconstruction. Figure 6.7 shows the reconstructed image obtained using $10, 20$ and 50 eigen-images. Here the figure given in 6.4(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 6.4(c) which was obtained using 100 eigen-images. They are nearly indistinguishable in quality.

(a) (b) (c)

Figure 6.7: PCA-based reconstructions using different numbers of eigen-images. (a) $K = 10$, (b)$K = 20$, and (c)$K = 50$.

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 6.8 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.

(a) (b) (c)

Figure 6.8: 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 $q = 2$.

(a) (b)

Figure 6.9: 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!

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 6.9 shows the reconstructed image 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.

In the next experiment we show that one can obtain a faithful reconstruction even if a portion of the LR image is missing. In Figure 6.10(a) we have a low resolution image with a portion of the face missing. The missing region is filled with the average image intensity of the LR image and the PCA-based interpolation is used to obtain the super-resolved image shown in Figure 6.10(b). As can be observed, the PCA-based algorithm almost recovers the missing region while super-resolving the LR face image. No bilateral symmetry property of the face has been used in this experiment. The recovery is possible due to the available information in the constituent eigenimages.

(a) (b)

Figure 6.10: Super-resolution reconstruction for an input face image with a portion of the image missing: (a) low resolution input and (b) super-resolved image using PCA.

We now show results of experiments on a different database. Figure 6.11 shows the poor quality input, bicubic interpolated result and the super-resolved images for zoom factors of $q = 2$ 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 the corresponding values are given in Table 6.1 for different values of noise variance. We observe a substantial improvement in PSNR for the PCA-based approach.

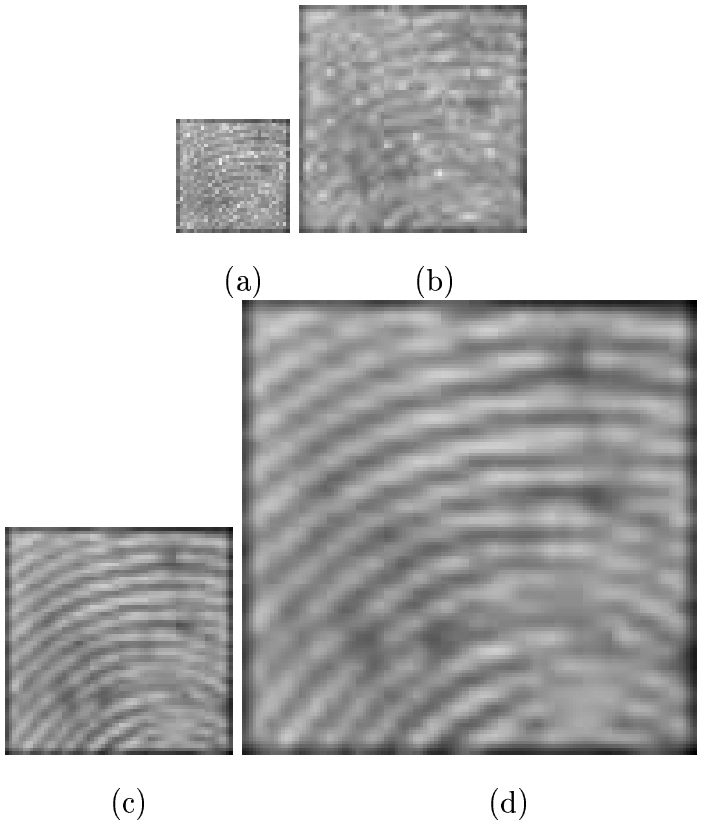


Figure 6.11: Illustration of results for a different object class. (a) A poor quality fingerprint image. Results of (b) bicubic interpolation, and PCA-based interpolation for (c) $q = 2$, and (d) $q = 4$.

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

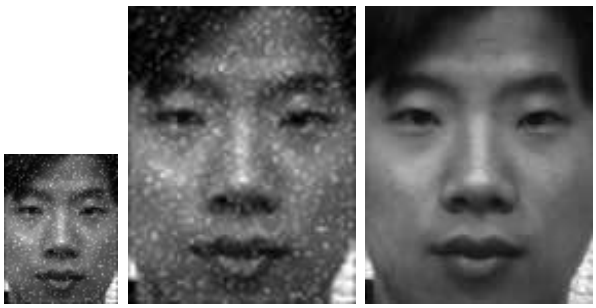
Table 6.1: Comparison of PSNRs for a zoom factor of $q = 4$ for different levels of noise.

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 6.4). 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 q to serve as the low resolution observation. This observed image is then appropriately interpolated before applying the PCA-based restoration.

In Figure 6.12(a), we show a low resolution observation. We add a white Gaussian noise with $\sigma = 0.1$ to simulate the presence of noise in the data. Figure 6.12(b) shows the result of bicubic interpolation. Quite naturally it is poor due to the presence of noise. In Figure 6.12(c) we show the result of PCA-based high resolution restoration for the upsampling factor of $q = 2$. The effect of noise is almost removed and the quality of reconstruction is very good.

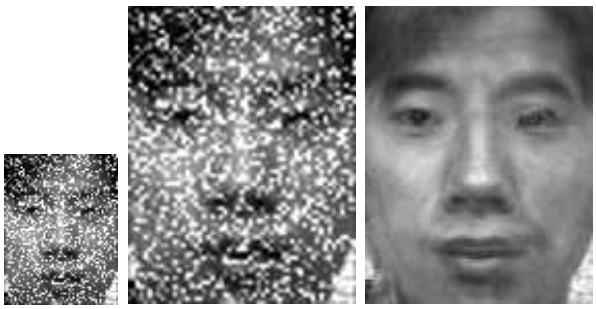
In Figure 6.13(a), the same input as shown in Figure 6.12(a) is further corrupted with additive noise. We use $\sigma = 0.5$ and the corresponding input image is of very poor quality. Hence we do not expect to view the image content in Figure 6.13(b) when the image is bicubic interpolated. The result of high resolution PCA-based restoration for $q = 2$ is shown in Figure 6.13(c). Although the corresponding reconstruction is inferior compared to what is given in Figure 6.12(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 $q = 4$. The input image having a good amount of noise corruption ($\sigma = 0.1$) is shown in Figure 6.14(a). The corresponding bicubic interpolated image with $q = 4$ is shown in Figure 6.14(b). The result of PCA-based reconstruction is shown in Figure 6.14(c). Compare this with the corresponding result for $q = 2$ given in Figure 6.12(c). We notice that reconstruction is still very good even for an upsampling factor of $q = 4$. The face is clearly recognizable and the effects of noise are no longer visible. Likewise in the case of upsampling by a factor of $q = 2$, 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 6.15(a, b). The result of PCA-based reconstruction is shown in



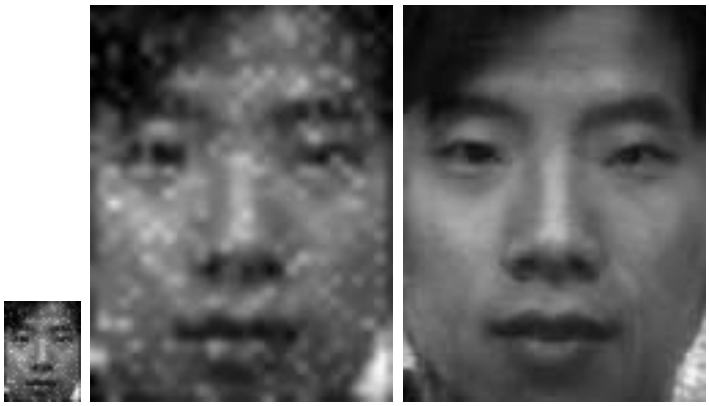
(a) (b) (c)

Figure 6.12: (a) A low resolution noisy observation ($\sigma = 0.1$), (b) bicubic interpolated image, and (c) PCA-based restoration with $q = 2$ using a high resolution training data set.



(a) (b) (c)

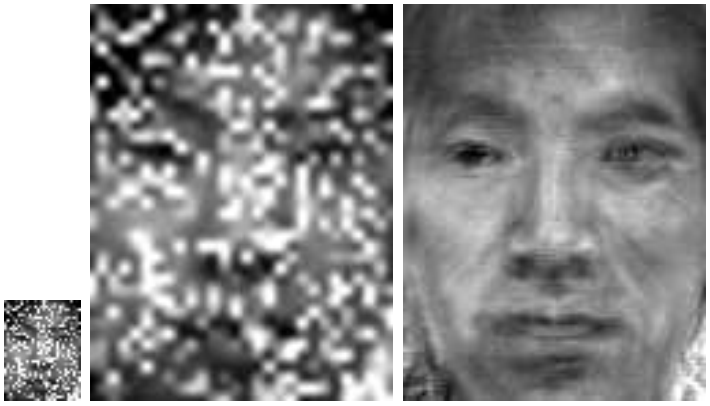
Figure 6.13: (a) A low resolution, extremely noisy observation ($\sigma = 0.5$), (b) bicubic interpolated image, and (c) the high resolution PCA-based restoration with $q = 2$.



(a) (b) (c)

Figure 6.14: (a) A low resolution noisy observation ($\sigma = 0.1$), (b) bicubic interpolated image, and (c) high resolution PCA-based restoration with $q = 4$.

Figure 6.15(c). We now notice a considerable distortion in the reconstructed image compared to what we obtained for $q = 2$ in Figure 6.13(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 q is also large, the PCA-based reconstruction method will fail.



(a) (b) (c)

Figure 6.15: (a) A low resolution very noisy observation ($\sigma = 0.5$), (b) bicubic interpolated image, and (c) high resolution PCA-based restoration with $q = 4$.

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 6.16(a)). The image, when upsampled by a factor of $q = 4$ using bicubic interpolation, shows a large amount of blur in Figure 6.16(b). In Figure 6.16(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.

In the previous set of experiments shown in Figures 6.12-6.15, a variant of the

(a) (b) (c)

Figure 6.16: (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 $q = 4$.

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 6.17(a, b) show the noise corrupted low resolution observation and its bicubic interpolation, respectively. Figure 6.17(c) shows the results of high resolution PCA-based reconstruction for the upsampling factor of $q = 4$. 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 $q = q_1$ and $q = q_2$ at both the input and the output ends, respectively.

In Figure 6.18(a) a low resolution observation of size 41×48 pixels is shown. The database training images were of dimension 82×96 pixels. Figure 6.18(b) shows the bicubic interpolated output for a zoom factor of $q_1 q_2 = 8$. 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 6.18(c).

(a) (b) (c)

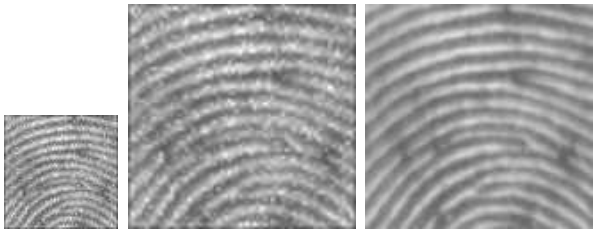
Figure 6.17: (a) A low resolution image, without any corresponding high resolution image in the database is corrupted by noise with $\sigma = 0.1$. (b) Bicubic interpolated image, and (c) high resolution PCA-based restoration with $q = 4$.

As expected, the super-resolved image is less blurred than the bicubic result.

(a) (b) (c)

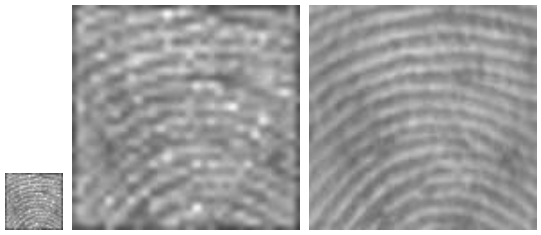
Figure 6.18: (a) A low resolution observation of different image size 41×48 pixels, (b) bicubic interpolated image with $q = 8$ and, (c) PCA-based reconstruction.

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 6.19(a) and 6.20(a) show two low resolution observations at two different levels in the resolution pyramid. The images have been corrupted with additive white Gaussian noise with $\sigma = 0.1$. In Figures 6.19(b and c), we compare the performance of the bicubic interpolation with that of the PCA-based method for an upsampling factor of $q = 2$. Figures 6.20(b and c) show the same results for $q = 4$. We can clearly



(a) (b) (c)

Figure 6.19: (a) A low resolution noisy fingerprint observation ($\sigma = 0.1$), (b) bicubic interpolated image, and (c) high resolution PCA-based restoration with $q = 2$.



(a) (b) (c)

Figure 6.20: (a) A low resolution very noisy fingerprint observation ($\sigma = 0.1$), (b) bicubic interpolated image, and (c) high resolution PCA-based restoration with $q = 4$.

observe an improvement in the picture quality when the PCA-based reconstruction is used.

6.6 Discussions

We have proposed 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 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 authentication or recognition purposes.

Chapter 7

A Hybrid Approach

The global high resolution reconstruction using a PCA based approach discussed in the previous chapter fails to do super-resolution at edges, though it removes the blur and noise present in the input low resolution image. But it very well learns the global characteristics from the image database. On the other hand, the methods discussed in chapters 3, 4 and 5 obtain the missing high frequency details by locally learning the best edges from the training images. To extract the advantages of both these methods, in this chapter we propose a hybrid approach for super-resolution by integrating the global learning using PCA and the local learning approach using wavelets or contourlets. The PCA-based global learning obtains the best statistical features from a database of similar class if images while the latter ensures the best edges from the same database. We solve the problem under a Bayesian framework incorporating the the global and local priors obtained as above, in addition to the data consistency term. We apply the method particularly for face hallucination where we synthesize the missing high frequency details of the input low resolution face image.

A similar approach for face hallucination is proposed in [70]. But here the authors separately obtain the global and local faces and add them to obtain the hallucinated face. The global high resolution image is obtained using PCA by computing the optimum eigen image coefficients which minimize the error in reconstructing the input low resolution image. The residual local portion of the super-resolved image is obtained using a patch-based learning similar to the discussions in [66, 67]. But our

method is based on the optimization of a single cost function containing separately learned global and local priors. We make use of only a high resolution database for this purpose.

7.1 Bayesian MAP Formulation

We use the same low resolution image formation model discussed in chapter 3. i.e. the low resolution image $\mathbf{f}$ is obtained by downsampling the high resolution image $\mathbf{z}$.

$$\mathbf{f} = D\mathbf{z} + \mathbf{n} \tag{7.1}$$

Based on the MAP criterion, we find the optimal solution $\hat{\mathbf{z}}$ by maximizing the posterior probability $p(\mathbf{z}|\mathbf{f})$, i.e.

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} p(\mathbf{z}|\mathbf{f}) = \arg\max_{\mathbf{z}} p(\mathbf{f}|\mathbf{z})p(\mathbf{z}). \tag{7.2}$$

Taking the logarithm of the posterior probability,

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} [\log p(\mathbf{f}|\mathbf{z}) + \log p(\mathbf{z})]. \tag{7.3}$$

Assuming the noise $\mathbf{n}$ to be i.i.d. and Gaussian, maximizing the log likelihood term $\log p(\mathbf{f}|\mathbf{z})$, will be equivalent to minimizing $||\mathbf{f} - D\mathbf{z}||^2$. For the prior term $U(\mathbf{z}) = \log p(\mathbf{z})$, we use two different priors obtained from global and local learning as explained in the following section. Thus the final solution becomes,

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} [||\mathbf{f} - D\mathbf{z}||^2 + U(\mathbf{z})]. \tag{7.4}$$

7.1.1 Global prior

Here we obtain an appropriate global prior on the desired high resolution image using the PCA-based method as discussed in section 6.4. The given low resolution face image is interpolated and projected on to the eigenimages computed from a high resolution database to obtain the eigenimage coefficients. The reconstruction is obtained using Eq. 6.6. i.e.

$$\hat{\mathbf{z}}_{\mathbf{g}} = E\mathbf{w} + \mathbf{m}_{\mathcal{F}}. \tag{7.5}$$

Here E is the matrix containing the eigenimages, $\mathbf{w}$ is the vector of eigenimage coefficients and $\mathbf{m}_{\mathcal{F}}$ is the average image. The prior term due to global learning can be expressed as

$$G(\mathbf{z}) = ||(EE^T(\mathbf{z} - \mathbf{m}_{\mathcal{F}}) + \mathbf{m}_{\mathcal{F}}) - \hat{\mathbf{z}}_{\mathbf{g}}||^2. \tag{7.6}$$

Using this prior we enforce the constraint that the PCA reconstruction of the super-resolved image should be close to the PCA reconstruction, $\hat{\mathbf{z}}_{\mathbf{g}}$ obtained as in section 6.4. This is quite reasonable as the image $\hat{\mathbf{z}}_{\mathbf{g}}$ has better learned the global characteristics from the high resolution database. Though the global features are better preserved in this learned image, this may not represent the exact face image we intend to recover as the database is exclusive of the original high resolution image. Hence the prior as shown in Eq.7.6 is not sufficient to recover the super-resolve image. So we introduce a second prior in the form of locally learned wavelet coefficients and call it a local prior.

7.1.2 Locally Learned prior

The introduction of this prior ensures the best matching edges for the given low resolution image by learning the contourlet coefficients at finer scales from a high resolution database. The learning procedure is similar to the one discussed in 4.2 in chapter 4. Let $\hat{\mathbf{Z}}_c$ be the contourlet transform of the learned image as discussed in 4.2. Now, the learning prior term can be expressed as

$$L(\mathbf{z}) = ||\Gamma(\mathbf{z}) - \hat{\mathbf{Z}}_{wt}||^2. \tag{7.7}$$

where $\Gamma(\mathbf{z})$ is the contourlet transform of the unknown high resolution image. For learning this local prior one can use wavelets or contourlets or any other representation which gives a better localization of edges. A patch-based learning as discussed in [66] will also serve as an appropriate local prior. But in this study we use a contourlet-based learning because of the advantages of the contourlet transform as discussed in section 4.1.

7.2 Super-Resolution Estimation

Incorporating the above two priors, i.e., substituting $U(\mathbf{z}) = \beta_1 G(\mathbf{z}) + \beta_2 L(\mathbf{z})$ in Eq. (7.4), the final solution becomes,

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}}[||\mathbf{f} - D\mathbf{z}||^2 + \beta_1 G(\mathbf{z}) + \beta_2 L(\mathbf{z})]. \tag{7.8}$$

The above cost function is nonconvex and also consists of terms in both spatial domain and also in the wavelet domain. 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 assumed to lead into a global minimum. We choose appropriate values for the parameters β_1 and β_2 so that all the components in the cost function in Eq. (7.8) have comparable contributions.

7.3 Experimental Results

We performed experiments on various face images using a database consisting of 150 high resolution face images of 15 people captured with different poses and different lighting conditions. An image which does not belong to the database was used as input. Figure 7.1(a) shows a low resolution face image of size 64×64. The original high resolution image from which the low resolution input was formed is shown in Figure 7.1(b). The bicubic interpolated version of Figure 7.1(a) is shown in Figure 7.1(c). Figure 7.1(d) shows the super-resolved image obtained using contourlet transform based learning as discussed in chapter 4. The best high resolution reconstruction obtained using the PCA-based global approach is shown in Figure 7.1(e). Here, though the eyes seem to be closed (as the original image does not belong to the database), all the other global features of the face are synthesized well. The image obtained using the proposed hybrid approach is shown in Figure 7.1(f). One can clearly see improvements near the eyes and lips. But, in comparison with the super-resolved image obtained using contourlet learning shown in Figure 7.1(d), we are unable to prove that the hybrid methods achieve a better edge over the local methods.

This could be due to the fact that as edges are more important in a super-resolved image, local methods always play a good role in restoring them in a better way. A proper study is required in terms of whether the global prior or the local prior is more important in a particular application, for the development of an efficient hybrid approach for super-resolution. Also, the selection of appropriate priors is equally important to reconstruct a good quality super-resolved image.

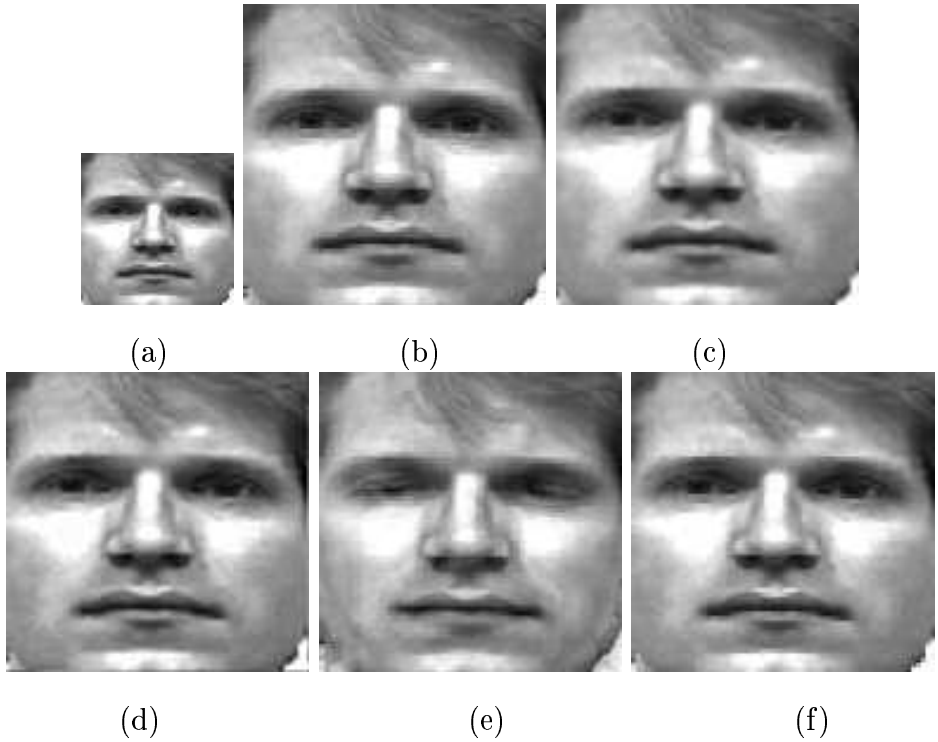


Figure 7.1: (a) A low resolution face image, (b) original high resolution image, (c) bicubic interpolated image, (d) super-resolution using contourlet learning, (e) interpolation using PCA-based learning, and (f) super-resolution using the hybrid approach.

Next we show the results of experiments performed on a female face image. The low resolution input, original high resolution image and the bicubic interpolated image are shown in Figures 7.2(a), (b) and (c) respectively. The super-resolved image obtained using contourlet learning, PCA-reconstructed output and the super-resolved image using the hybrid approach are shown Figures 7.2(d), (e) and (f) respectively. The super-resolved image looks sharper compared to the bicubic one but at the same time it is a little blocky.

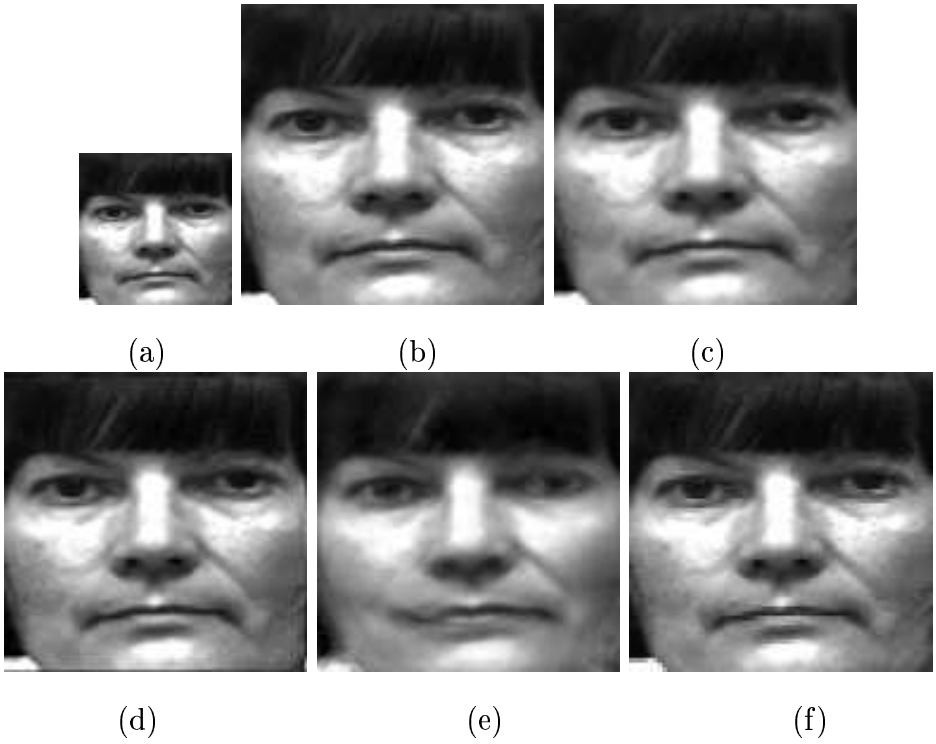


Figure 7.2: (a) A low resolution face image, (b) original high resolution image, (c) bicubic interpolated image, (d) super-resolution using contourlet learning, (e) interpolation using PCA-based learning, (f) super-resolution using the hybrid approach.

7.4 Conclusions

In order to integrate the advantages of both the local and global approaches proposed in the previous chapters, in this chapter we presented a hybrid approach for image super-resolution, particularly for face hallucination. Ideally, such a hybrid method should handle the blur and noise present in the input while performing interpolation. But our method could not properly handle blurred and noisy observations. The improvement achieved over bicubic interpolation was very substantive either. In addition, we were unable to prove that the hybrid methods achieve a better edge over the local methods. The issues of selection of appropriate global and local learning methods and whether the global prior or the local prior is more important in a particular application is crucial in developing an efficient hybrid approach for super-resolution. This requires further investigation.

Chapter 8

Alias-Free Interpolation

All the SR reconstruction methods discussed in the previous chapters are able to provide a super-resolved image, but they are unable to handle the aliasing effects. Even in the local approaches using wavelets or contourlets, though the edges are handled properly, the aliasing present in the LR image is not reduced much during the upsampling process. In this chapter we study the possibility of removing aliasing in a scene using a single observation by designing an alias-free upsampling scheme. Quite naturally this is impossible unless additional information about the HR data is available. Hence we assume that the HR data is piece-wise constant. We generate the unknown high frequency components of the given partially aliased (low resolution) image by minimizing the total variation of the interpolant subject to the constraint that part of unaliased spectral components in the low resolution observation are known precisely and under the assumption of sparsity in the data. This provides a mathematical basis for exact reproduction of high frequency components with probability approaching one, from their aliased observation.

In all the SR methods proposed in literature, the quality of the super-resolved image is measured either by means of visual inspection or using a PSNR check. It can be easily shown that the PSNR measure is heavily biased towards the lower part of the spectrum due to the fact that most of the energy is contained in this region. Hence the PSNR may not be a good measure to evaluate the performance of an SR scheme. Hence we refrain from using the PSNR measure for analysis of the quality of

results in this chapter. A different measure to evaluate the quality of visually pleasing images using singular values of the image matrix is proposed in [120]. But the issue that the reconstructed components are really the high frequency components has not really been investigated so far. Our focus in this chapter is a study in this direction. In [121], Lin and Shum determine the quantitative limits of reconstruction-based super-resolution algorithms and obtain the upsampling limits from the conditioning analysis of the coefficient matrix. But it is restricted to a perturbation analysis and not on spectral resolvability. Shahram and Milanfar in [122] study how far beyond the classical Rayleigh limit of resolution one can reach at a given signal to noise ratio using statistical analysis. Here the authors do not study the system performance in the presence of aliasing. Reconstruction of 2-D bandlimited signals from irregular samples is discussed in [123] and [124].

Rajan *et al.* have analyzed the possibility of alias-free upsampling of images in [116] through the use of a generalized interpolation. They have shown the conditions under which such an interpolation is possible. However, it requires several observations and the knowledge of a non-linear transform to achieve this. We study the issue of alias-free interpolation at a more fundamental level and restrict ourselves again to using a single observation. Our work is motivated by the work of Candes *et al.* [125] where the authors address the problem of exact signal reconstruction from incomplete frequency information. We build on the theorem developed by them to derive a method for exact removal of aliasing while interpolating an image.

The reminder of the chapter is organized as follows. We discuss in detail about the aliasing mechanism in section 8.1. A relevant theorem which we make use of in solving the problem is stated in 8.2. Some useful corollaries are also given. Section 8.2.1 explains our alias-free interpolation technique. In section 8.2.2 we discuss the computational scheme to solve the problem. We present experimental results for 1-D and 2-D cases in section 8.3, and the chapter concludes in section 8.4.

8.1 Aliasing in LR Image

It is assumed that the observed low resolution image is produced from a single high resolution image under the generative model discussed in chapter 1. Let $\mathbf{z}$ represent the lexicographically ordered high resolution image of $N^2 \times 1$ pixels. If $\mathbf{f}$ is the $M^2 \times 1$ lexicographically ordered vector containing pixels from the low resolution observation, then it can be modeled as

$$\mathbf{f} = DA\mathbf{z} \tag{8.1}$$

where D is the decimation matrix, size of which depends on the decimation factor and A is the blur matrix, assumed to be an identity matrix in this chapter for the specific task of studying the alias-removal property. It may be noted that unlike the methods proposed in previous chapters, here we do not make use of a training set to recover the upsampled image.

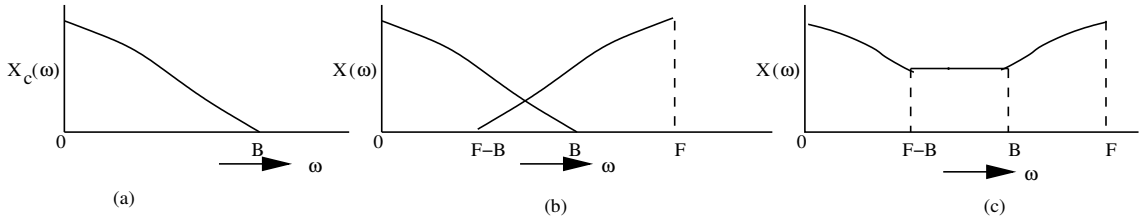


Figure 8.1: Illustration of the aliasing process: (a) Spectrum of a continuous-time signal $x_c(t)$, (b) components of the spectrum of the sampled signal $x(n)$, and (c) resultant spectrum of $x(n)$.

The LR image formed through the above process will, in general, be aliased. The aliasing mechanism is illustrated in Figure 8.1. The spectrum of a continuous-time 1-D signal $x_c(t)$ band limited to B is shown in Figure 8.1(a). The spectrum of the sampled signal $x(n)$ sampled at a rate $F < 2B$ is shown in Figure 8.1(b). Of course the spectrum will be aliased since the signal is sampled at a rate less than the Nyquist rate. The resultant aliased spectrum of the sampled signal is shown in Figure 8.1(c). As can be noted from Figure 8.1(c) the portion of the spectrum $F - B \leq \omega \leq B$ will be aliased and the rest will be alias-free. A similar form of aliasing takes place in low resolution images unless the blur matrix A in equation (8.1) is quite severe.

The knowledge about the portion of the spectrum $0 \leq \omega \leq F - B$ will be used as a constraint, as these components are free from aliasing, in the proposed method to recover the high frequency components.

Having explained the aliasing process, we now define our problem in terms of alias-free interpolation: Given an LR image $f(x, y)$ of size $M \times M$ whose spectrum is *partially* aliased, generate an interpolated image $z(x, y)$ of size $2M \times 2M$ which is *completely* alias-free under the assumption that the image consists of piece-wise constant intensity regions. The significance of the assumption will be explained in the next section.

8.2 A Relevant Theorem by Candes *et al.*

Theorem 1: Consider a discrete-time 1-D signal $f \in C^M$ and a randomly chosen set of frequencies Ω of mean size $\tau M, 0 < \tau < 1$. Then for each $\zeta > 0$, suppose that f obeys

$$\#\{n, f(n) \neq 0\} \leq \alpha(\zeta) \cdot (\log M)^{-1}.\#\Omega, \tag{8.2}$$

then with probability at least $1 - O(M^{-\zeta})$, f can be reconstructed exactly as the solution to the l_1 minimization problem

$$\min_h \sum_{n=0}^{M-1} |h(n)| \quad s.t. \;\; H(\omega) = F(\omega) \;\; \forall \omega \in \Omega \tag{8.3}$$

where $H(\omega)$ and $F(\omega)$ are the discrete Fourier transforms of $h(n)$ and $f(n)$ respectively and # refers to the count.

Here ζ is an accuracy parameter in the term $O(M^{-\zeta})$ and $\alpha(\zeta)$ has been shown to be equal to $(1 + o(1))/(29.6(\zeta + 1))$ under certain conditions in [125]. In a simple language it means that as one selects more spectral components compared to the number of non-zero elements in $f(n)$, one is likely to recover the true function $f(n)$ with a higher accuracy. Proof of the theorem can be found in [125]. This is typically known as a data sparsity problem [126], one such common example of which is inpainting [127] where one is required to reconstruct the missing data. An interesting reference to this work is by Chan *et al.* [128] where the authors investigate the reverse problem,

i.e., how much loss in data can be tolerated for a faithful reconstruction of a signal as opposed to what Candes *et al.* [125] have studied.

According to the theorem the discrete-time signal f can be reconstructed from its partial frequency samples as the solution to the constrained l_1 optimization problem as stated above. It may be noted that the reconstruction is possible if the signal consists of a limited number of spikes (Kronecker delta) only. However, most of the input signals or images do not satisfy the above condition. If we consider this function to be a derivative (forward difference in the discrete case) of the function $f(n)$, then the corresponding l_1 minimization should be performed on the derivative of $h(n)$, or in other words, one has to minimize the total variation (TV). This leads to the following corollary.

Corollary 1: A piecewise constant object can be reconstructed from incomplete frequency samples provided the number of discontinuities satisfy the above condition given in Eq. (8.2), as the solution to the minimization problem

$$\min_h \sum_{n=0}^{M-1} |h(n) - h(n-1)| \quad s.t. \;\; H(\omega) = F(\omega) \;\; \forall \omega \in \Omega \tag{8.4}$$

Corollary 2: If $f(x, y)$ is a two-dimensional object, it can be reconstructed from its incomplete frequency samples as the solution to the minimization problem

$$\min_h \sum\sum (|h_x| + |h_y|) \quad s.t. \;\; H(\omega) = F(\omega) \;\; \forall \omega \in \Omega \tag{8.5}$$

where $h_x(x, y) = h(x, y) - h(x-1, y)$ and $h_y(x, y) = h(x, y) - h(x, y-1)$. This is similar to minimizing the total variation norm of $h(x, y)$. But this is not rotationally symmetric.

It may be noted that the same solution was proposed in 1981 by Levy and Fullagar [129] in connection with the reconstruction of geophysical data. Also see the reference [130] for a similar work. Candes *et al.* have provided a theoretical footing of the existing solution. It may be noted that total variation-based image interpolation methods are also proposed in [131] and [132]. But the authors do not specifically address the issue of alias removal. For a detailed review of TV, readers are referred to [133].

8.2.1 Alias-free Interpolation

The problem addressed in [125] is a restoration problem where the discrete-time signal is reconstructed from its incomplete Fourier samples such as in computed tomography. However, the above problem does not consider the effect of aliasing on the sampled data. But our problem is a signal interpolation one, where only one LR observation f is available, which is the decimated version of the unknown HR image z as explained in section 8.1. Of course, f will be aliased. We wish to remove this aliasing completely while interpolating the image assuming the aliasing to be only partial. It may be noted that without the interpolation (use of a denser grid to represent the data), one cannot recover the aliased components. To apply the above theorem to our problem, a partial knowledge about $Z(\omega)$ should be available. We now explain how a partial knowledge of $Z(\omega)$ can be obtained from the given observation $F(\omega)$.

Our alias-free interpolation procedure is illustrated in Figure 8.2 with respect to a given 1-D LR sequence $f(n)$ of length M. Note that unlike in theorem 1, we are dealing with real valued function $f(n)$ and hence the spectrum is always conjugate symmetric and one has to consider only one half of the spectral components. Figure 8.2(a) shows the partially aliased spectrum of the LR sequence $f(n)$ of length M. We assume that $F(\omega)$ in $0 \leq \omega \leq L$ is free from aliasing and the remaining portion is aliased. This corresponds to the assumption that the continuous signal $f_c(t)$ is band limited to the normalized frequency $(1 - L/M)$, where $0 \leq L \leq M/2$. Smaller the value of L, larger is the amount of aliasing. Figure 8.2(b) shows the spectrum of the HR sequence $z(n)$ of length $2M$ to be estimated. The alias-free interpolation method should recover the frequency components in the region $L \leq \omega \leq M - L$ in $Z(\omega)$ as shown in Figure 8.2(b). From the figure, note that we have

$$Z(i) = 2F(i) \text{ for } 0 \leq i < L \tag{8.6}$$

$$Z(i) = 0 \text{ for } M - L < i \leq M \tag{8.7}$$

and using the property of aliasing (wrapping around of frequencies)

$$Z(i) + Z(M - i) = 2F(i) \text{ for } L \leq i \leq M/2 \tag{8.8}$$

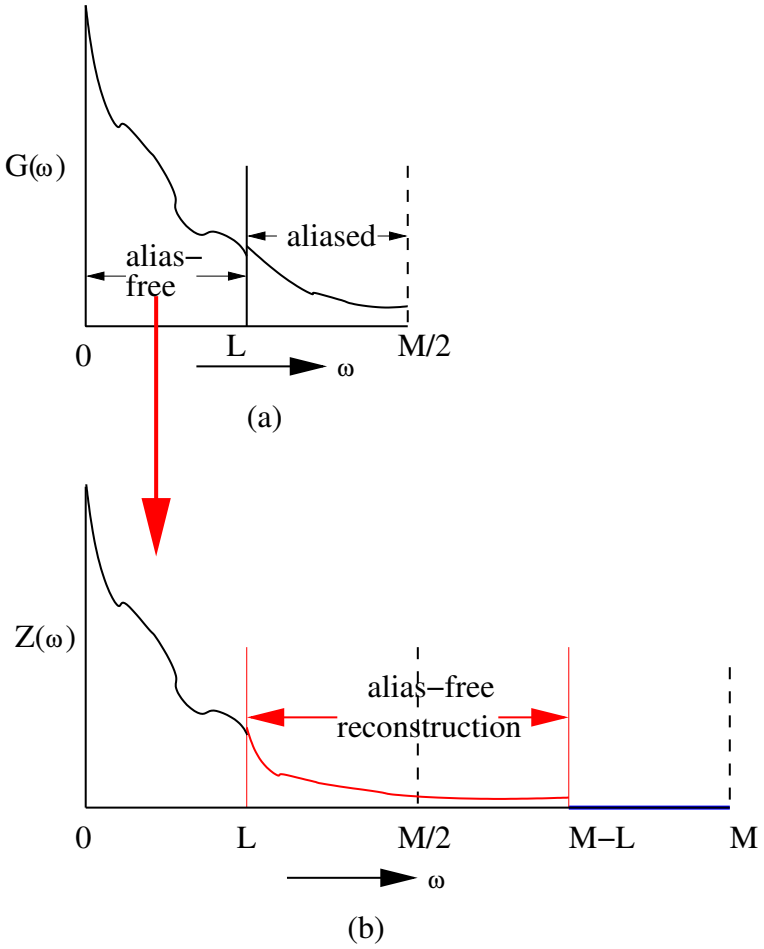


Figure 8.2: Illustration of (a) partially aliased spectrum of the LR sequence $f(n)$, and (b) spectrum of the HR sequence $z(n)$ to be estimated. Note that only half of the spectrum is shown due to conjugate symmetry.

Hence the alias-free reconstruction of the high resolution signal involves recovering the spectrum $Z(\omega)$ given Eqs. (8.6-8.8). Clearly, this cannot be done without additional constraints. Note that one needs the scale factor 2 to satisfy the energy relationship (Parseval's theorem). In order to recover $z(n)$, we need to first state the following theorem.

Theorem 2: Given a discrete-time partially aliased 1-D signal $f \in R^M$, and two distinct spectral intervals $\Omega_a = \{0 \leq \omega < L\}$ and $\Omega_b = \{L \leq \omega \leq M/2\}$ and another discrete-time signal $z \in R^{2M}$ satisfying Nyquist criterion with three distinct spectral intervals $\Omega_a{}' = \Omega_a$, $\Omega_b{}' = \{L \leq \omega \leq M - L\}$ and $\Omega_c{}' = \{M - L < \omega \leq M\}$, under conditions very similar to those defined in theorem 1, z can be recovered exactly from f as a solution to the l_1 minimization problem

$$\min_z \sum_{n=0}^{2M-1} |z(n)| \tag{8.9}$$

subject to the constraints

$$Z(\omega) = 2F(\omega) \ \ \forall \omega \in \Omega_a{}' \tag{8.10}$$

$$Z(\omega) + Z(N - \omega) = 2F(\omega) \ \ \forall \omega \in \Omega_b \tag{8.11}$$

$$Z(\omega) = 0 \ \ \forall \omega \in \Omega_c{}' \tag{8.12}$$

One can follow arguments similar to those in [125] to prove the above theorem except that the partitions are deterministic and hence it will lead to different values of the parameters α and ζ.

It may be noted that the partitions Ω_a and Ω_b correspond to the alias-free and the aliased components of the low resolution signal f, respectively. Since the partition is known, it implicitly means that we know the extent of aliasing in the observation. Furthermore, the above theorem assumes that the signal consists of a limited number of Kronecker deltas. Extending the theorem to deal with piece-wise constant signal, and also on the 2-D lattice, we realize that we should minimize the expression

$$\min_z \sum\sum(|z_x| + |z_y|) \tag{8.13}$$

instead of equation (8.9) to recover the high resolution image z. It may be noted that

if z is, indeed, piecewise constant then it cannot ideally be band limited, and hence the partition Ω_f will not be completely free from aliasing.

Now we look at the issue of the choice of the value of L for alias removal. It is assumed in theorem 2 that L is known. This is tantamount to assuming that the highest frequency component present in z is known apriori. However, one would not know L in practical super-resolution applications. We suggest that one solves the problem for different values of L and then compares the results. However, as the value of L is lowered from $M/2$ towards 0, the cardinality of the set Ω_a reduces and the reconstruction would be progressively more unreliable. It also leads to the following observation that one cannot use an interpolation factor greater than 2 as this would mean $L = 0$, implying a several fold aliasing when $\Omega_a = \{\emptyset\}$ and hence reconstruction would be very unreliable.

8.2.2 Computational Method

Theorem 2 provides a theoretical basis for obtaining the alias-free interpolated signal z. We now provide the computational tool to solve this. We obtain the solution to the above optimization problem using linear programming (LP). The objective function for the LP problem is the total variation cost as given in Eq. (8.13). The equality constraints are obtained using Eqs. (8.10), (8.11) and (8.12). The equality constraints corresponding to Eq. (8.10) can be written in the form

$$T_a z = F_a \tag{8.14}$$

where T is the $2M \times 2M$ DFT matrix with elements $T(m,n) = [e^{-j\pi/M}]^{mn}$ and T_a represents the top L rows of T. Thus T_a is an $L \times 2M$ matrix. Similarly F is the DFT of the observation $2f(n)$ and F_a corresponds to the top L elements of F. The equality constraints corresponding to Eq. (8.11) can be written as

$$T_b z = F_b \tag{8.15}$$

where T_b is an $(M/2 - L + 1) \times 2M$ matrix whose each row is obtained by summing the corresponding two rows of the DFT matrix T_a as per the indices shown in Eq. (8.8).

F_b corresponds to the spectral components $(L+1)$ to $(M/2-L)$ in F. Similarly Eq. (8.12) can be written as a linear equality

$$T_c z = \bar{0} \tag{8.16}$$

where T_0 consists of the $(M-L+1)$ to M rows of the DFT matrix T and $\bar{0}$ is a null vector. All the above three linear equations can now be compactly written as

$$\begin{bmatrix} T_a \\ T_b \\ T_c \end{bmatrix} z = \begin{bmatrix} \bar{F} \\ \bar{0} \end{bmatrix} \tag{8.17}$$

which is of the form $Cz = d$. We also know that $z(n) \geq 0 \ \ \forall n$ as $z(n)$ corresponds to an image. Also note that the above equation is meant for the first half of the spectrum. One would get an equivalent constraint for the other half based on the conjugate symmetry. Hence Eqs. (8.9) and (8.17) constitute a standard LP problem. We have explained the problem with respect a 1-D signal and it should be suitably changed to handle 2-D images.

To solve the l_1 minimization using LP equation (8.13) should be written as

$$\min_z \sum\sum((z_x^+ + z_x^-) + (z_y^+ + z_y^-)) \tag{8.18}$$

where $z_x^+ = \max(z_x, 0)$ and $z_x^- = -\min(z_x, 0)$, etc. Unfortunately, this increases the dimensionality of the unknown variables by five folds, increasing the computation significantly. Further the constraint equations involve complex numbers when the relationships have to be split into real and imaginary parts separately, increasing the computation further. We generate the appropriate constraint matrix and solve using the *linprog* routine in Matlab. But, even for a 128×128 image, the computational resources required are very large. Unfortunately, Matlab fails to allocate the necessary memory even for a small sized image. A typical option in LP is to utilize the possible sparseness properties of C matrix in Eq. (8.17). Unfortunately again, C does not have any sparseness as the DFT matrix T is not a sparse one. So we solve it as a sequential 1-D problem taking first the rows and then the columns. Hence the results obtained by this method for images in this chapter are all sub-optimal.

8.3 Experimental Results

In this section we present the results of alias-free interpolation obtained using the proposed approach. All the LR images are of size 64×64. All the results shown in this section are for interpolation factor of 2 for the reason described in section 8.2.1. Since the amount of aliasing L is not known, we show results for various choices of L.

First we show the applicability of the proposed method on a simulated 1-D data. Figure 8.3(a) shows a low resolution rectangular pulse train and the corresponding spectrum is shown in Figure 8.3(b). The signal shown in Figure 8.3(a) is superimposed with three high frequency components corresponding to the normalized frequencies $35/64, 36/64$ and $37/64$ to obtain the signal shown in Figure 8.3(c). Clearly, these three spectral components are aliased ones. Figure 8.3(d) shows the spectrum of the aliased signal. One cannot find that the signal is aliased either from Figures 8.3(c) or 8.3(d). Figure 8.3(e) shows the interpolated signal using the proposed method and its spectrum is shown in Figure 8.3(f). One can see that there are spectral components at locations beyond the normalized frequency $32/128$. These components match quite well with the introduced high frequency components. We have used $L = 26$ in this example. To further see the gain arising out of the proposed method, one can note that the spectrum of the rectangular pulse train (without the additional high frequency components) shown in Figure 8.3(b) compares very favorably with the spectrum of the interpolated signal till the normalized frequency of $32/128$. On comparing the interpolated signal in Figure 8.3(e) with the low resolution signal in Figure 8.3(c), one can clearly see that Figure 8.3(e) cannot be obtained by the linear or cubic interpolation of the original signal. (see the highlighted spectral components in Figure 8.3(f)). This confirms the utility of the proposed method. Further, to illustrate the capability of our method we compare the results of our alias free interpolation with spline interpolation in Figure 8.4. Figure 8.4(a) shows a low resolution rectangular pulse train and the corresponding spectrum is shown in Figure 8.4(b). The spline interpolated result and the corresponding spectrum are shown in Figures 8.4(c) and (d) respectively. The alias free interpolated signal and its spectrum are shown in

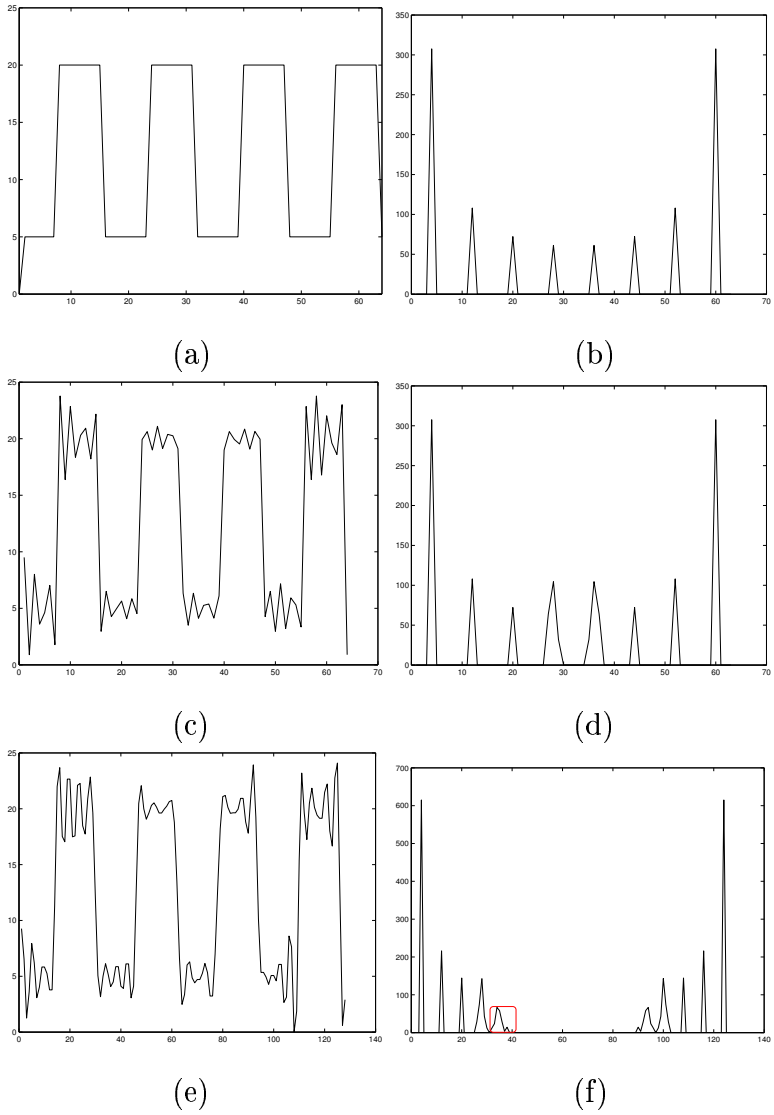


Figure 8.3: Demonstration of the proposed approach for a $1-D$ signal: (a) alias-free LR signal, (b) spectrum of (a), (c) aliased LR signal, (d) spectrum of (c), (e) interpolated signal using the proposed approach and (f) spectrum of (e).

Figures 8.4(e) and (f), respectively. As can be observed, the alias free interpolated signal is almost free from overshoot and ripples as compared to the spline interpolated one.

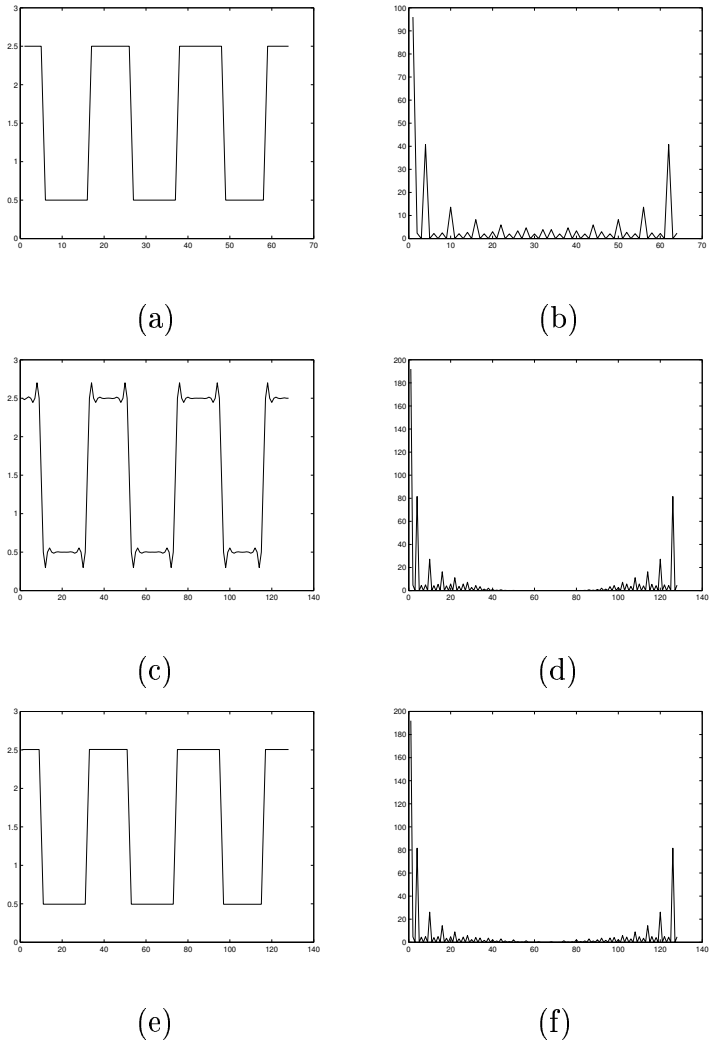


Figure 8.4: Demonstration of the proposed approach for another $1-D$ signal: (a) LR signal, (b) spectrum of (a), (c) spline interpolated signal, (d) spectrum of (c), (e) interpolated signal using the proposed approach and (f) spectrum of (e).

Figure 8.5(a) shows a partially aliased low resolution Lena image of size 64×64. Figure 8.5(b) shows the bicubic interpolated image for comparison to the proposed method. Figures 8.5(c-e) show the alias-free interpolation results obtained using the sub-optimal linear programming method. Figure 8.5(c) corresponds to the result

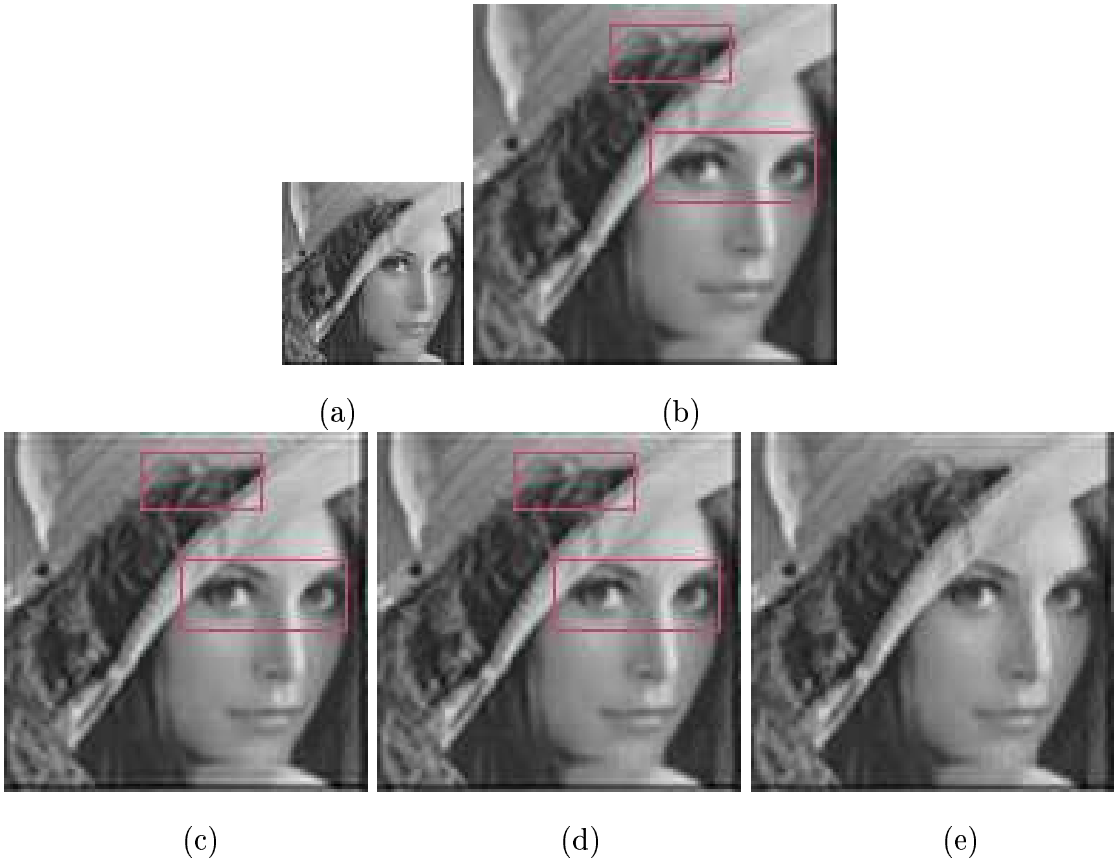


Figure 8.5: (a) A low resolution Lena image, (b) bicubic interpolated image. Interpolated images using the proposed approach generating additional (c) 10%, (d) 20% and (e) 30% high frequency components.

where 10% ($L = 29$), additional high frequency components are generated. Here we assume that the aliasing in the LR image is small, only 10% of the entire spectrum. If we assume that the aliasing in the LR observation is about 20%, the corresponding alias-free interpolated image is shown in Figure 8.5(d). This corresponds to the choice of $L = 26$. Figure 8.5(e) shows the alias-free interpolated image where we attempt to generate 30% additional high frequency components assuming that 30% of the spectrum of the LR image is aliased. As can be observed from Figures 8.5(c-e), there is a gradual reduction in the quality of the reconstructed image as the aliasing in the LR image is assumed to have increased from 10% to 30%. This is due to the fact that only a smaller subset of spectral components are known exactly. In comparison to the bicubic interpolated image, the result using the proposed approach is much

sharper. Observe the eyes, hair strands, etc. in Figure 8.5(d). Some of the regions are highlighted in the figure. We have observed that the reconstruction becomes poor when the aliasing present in the LR image is assumed to be more than 20 to 30%.

Now we perform the experiments on a severely aliased randomly textured image. The purpose of this experiment is to demonstrate that one does not get any improve-

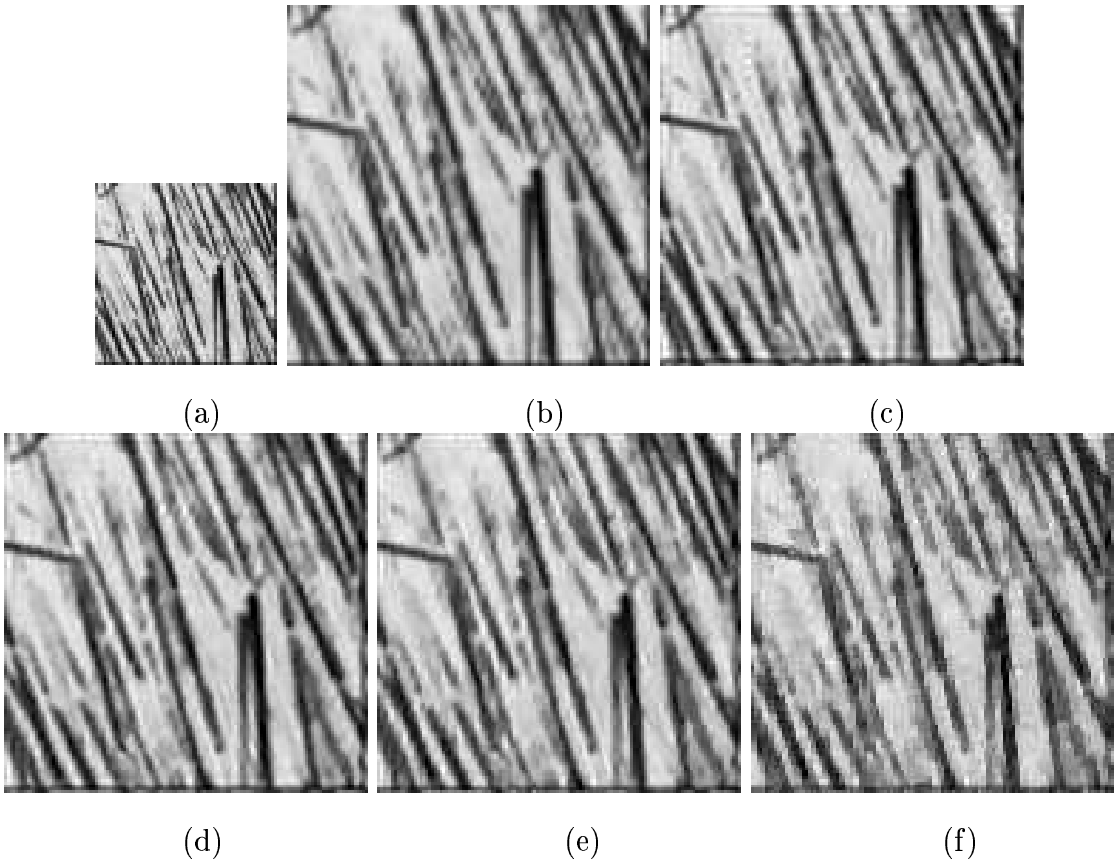


Figure 8.6: (a) A severely aliased low resolution texture image, (b) bicubic interpolated image. Interpolated images using the proposed approach generating (c) 10%, (d) 20% and (e) 30% high frequency components, (f) Interpolated image when the LR image is fully aliased.

ment during interpolation if the signal is highly aliased. The LR observation is shown Figure 8.6(a). The interpolated images using the proposed approach are shown in Figures 8.6(c-e) assuming 10%, 20% and 30% aliasing, respectively, in the given LR image. As the aliasing present in the LR image is very high, the proposed method does not give a significant edge over bicubic interpolation as can be observed from

Figures 8.6(c-e). Now we assume that the entire spectrum is aliased, ie,$\Omega_f = \{\emptyset\}$ in theorem 2($L = 1$). Figure 8.6(f) shows the corresponding interpolated result. We observe that the reconstruction is quite inferior as we do not have any of the spectral components known exactly.

8.4 Conclusion

In this chapter we have presented a method for alias-free interpolation from a partially aliased low resolution image. We have provided a theoretical basis on how an alias-free upsampling can be achieved. In order to interpolate the given LR image we generate the exact additional high frequency components assuming a knowledge of the nature of aliasing in the spectrum of the LR observation and assuming a piecewise constant intensity image. The alias-free interpolation is achieved by solving the l_1 optimization. A sub-optimal computational procedure using linear programming is also presented. Experiments performed on 1-D and 2-D signals confirm the utility of the proposed method. But a disadvantage of our method is that we put a strong restriction on the type of signals which can be faithfully interpolated using our approach. Also we specifically need the alias-free region in the low resolution signal spectrum which is practically impossible.

Chapter 9

Conclusions and Future Research Issues

9.1 Conclusions

We have addressed the problem of single image super-resolution in this thesis. Here we try to obtain a super-resolved image offering better details from a single low resolution observation. Our method is different from the conventional super-resolution techniques where one tries to recover the super-resolved image from several low resolution observations of the same scene. Most of these techniques use several sub-pixel shifted observations and extract the non redundant information among them to reconstruct the high resolution image. Also, our methods are different from standard interpolation techniques where the interpolated images usually appear blurred because of smoothing of edges.

The novelty of most of the single image super-resolution techniques compared to the conventional SR techniques as well as ordinary interpolation methods is in the usage of training images to learn the missing high frequency contents while super-resolving the low resolution input. The SR reconstruction methods proposed in this thesis also make use of training images to extract the finer details during super-resolution. The novelty of our methods lies in the way we pick up the best features from the training set compared to other learning-based methods for single image

super-resolution. In this section first we summarize the different methods proposed in this thesis.

First we proposed an SR reconstruction technique using learned wavelet coefficients using a high resolution training set. Wavelet coefficients at finer scales for the high resolution image are locally learned from the data set making use of edge primitives in the low resolution representation. The learned wavelet coefficients are used as a prior under a Bayesian formulation to solve for the unknown high resolution image. In addition to this learned prior, we also use a smoothness prior to ensure local correlation among neighboring pixels. Experiments were performed on both gray scale and color images. The reconstructed images showed improvements over conventional interpolation methods. The disadvantage of the method was that the reconstruction process was very slow as we had to use a slow stochastic optimization technique to obtain the solution, since the cost function was containing terms both in spatial domain and in the transform domain. Another disadvantage of the method is that the super-resolved image is still blocky especially when the edges lie along an arbitrary contour.

As wavelets are not capable of learning the directional edges, in the next chapter we proposed a contourlet transform-based resolution enhancement method. As contourlets by itself are capable of capturing the geometric smoothness along contours by performing directional decomposition at various scales, the direct inverse contourlet transform could give the enhanced image. Thus the problem need not be solved under a regularization frame work and as a consequence the reconstruction process is very fast. Experiments performed on various class of images showed improvements over different interpolation schemes as well as over a standard wavelet-based learning. The effectiveness of this method depends on how good is the training set and how effectively one sets the threshold in order to avoid spurious learning.

As taking the best match for a test patch from a database does not always provide a better solution, and since the process of decimation may not be captured well by the wavelet or contourlet decomposition, we further considered a locally linear embedding method under a supervised learning frame work in the contourlet domain.

Here we make use of a training set containing both low resolution images and their corresponding high resolution versions. To extrapolate the contourlet coefficients at the finer subbands, we make use of several nearest neighbors of the test patch as in locally linear embedding. The resulting high resolution image is sharper than the conventional bicubic interpolated image. But the disadvantage of the method is that, if there is some spurious learning at certain locations, the resulting high resolution image at the corresponding location will be much noisy as it is contributed by several spuriously learned neighbors. To overcome this difficulty one has to carefully set the threshold in order to avoid any spurious learning.

All the above discussed methods were based on learning the high frequency details locally from the training set. As all these methods failed to handle the de-blurring and de-noising issues in super-resolution, we proposed a global learning approach based on PCA to enhance the low resolution observation. Here we described two different methods, one based on generalized interpolation using a low resolution database and the other using a high resolution database. Both the methods could very well handle the blur and noise present in the data. An inherent disadvantage of this approach is that the method is applicable only for a specific class of images like face and fingerprint images as we learn the global features from the database. Hence method cannot be classified under a general purpose super-resolution technique.

Having identified the advantages and disadvantages of both global and local learning approaches for super-resolution, we tried to combine them into a single hybrid approach. We obtain the global prior from a PCA-based learning and the local prior is derived from a contourlet based learning. Using these two priors we solved the problem under a Bayesian framework. As we use PCA for global learning, the method again is applicable only for a specific class of images, and we apply it particularly for face hallucination. Though the hallucinated image showed better improvement over bicubic interpolation and the global learning using PCA, we couldn't establish that the hybrid approach achieves a better edge over the local approaches. This could be due to the fact that as edges are more important in a super-resolved image, local methods always play a good role in restoring them in a better way. A proper study

is required in terms of whether the global prior or the local prior is more important in a particular application, for the development of an efficient hybrid approach for super-resolution. Also, the selection of appropriate priors is equally important to reconstruct a good quality super-resolved image.

As the various local and global methods fail to handle the aliasing present in the low resolution image, in the last chapter we presented a method for alias-free interpolation from a partially aliased low resolution image. We have provided a theoretical basis to check whether the generated components are the exact high frequency components required. In order to interpolate the given LR image we generate the exact additional high frequency components assuming a knowledge of the nature of aliasing in the spectrum of the LR observation and assuming a piecewise constant intensity image. The alias-free interpolation is achieved by solving the l_1 optimization. A sub-optimal computational procedure using linear programming is also presented. Experiments performed on 1-D and 2-D signals confirm the utility of the proposed method. But a disadvantage of this method is the restriction on the type of signals which can be faithfully interpolated applying the proposed approach. Also we specifically need the alias-free region in the low resolution signal spectrum which is practically impossible. Further more, this method cannot handle a blurred or noisy observation. In spite of these disadvantages, the novelty of the method lies in providing an exact theoretical basis on how an alias-free upsampling can be achieved.

9.2 Future Research Issues

Even though, the multi-frame super-resolution problem is very well addressed in the literature, it is to be mentioned that the single image super-resolution idea is still in its infancy. The initial work on this field using training images is due to Freeman *et al.* [66] in 2000. Recently, quite a few new papers have been published in this field, particularly for face hallucination. Most of these techniques suffer from one or more disadvantages in terms of visual quality, computational complexity, need for a huge training database, etc., to name a few. Therefore, several of these issues are to be

properly tackled in the near future. In this section, we discuss the future directions in which the research could progress so as to produce good practical super-resolution.

- One of the major disadvantages with most of the currently available super-resolution algorithms is the huge computational requirements. Super-resolution in real time is, thus, the need of the hour. Even though, some efforts have been made by researchers to develop computationally efficient SR algorithms for the multi-frame case, there are several assumptions being made for the image formation process and the structure of the scene in order to reduce the computations, which actually limit the practical applications. The methods proposed in this thesis using wavelet learning and the hybrid approach are computationally too expensive for real time applications. One could think of ways to speed up these algorithms. One possibility could be by converting the cost function in to a single domain, i.e. either to the wavelet domain or to the spatial domain, and also by replacing the discontinuity preserving smoothness term by an appropriate differentiable term.

- For the wavelet and contourlet based methods, it is assumed that the low resolution image is free from blurring. This enables us to restrict the search for the best match of the edge element from the training images at the same scale. If the input image blurred, then one may have to go down the resolution pyramid of the training image to locate the best match. This appears to be a difficult problem as the scale is unknown and it may not exactly match the given levels of scales of the training image after the wavelet or contourlet decomposition. As a result, our experiments to super-resolve blurred satellite images was a failure. It will be interesting to extend the technique to deal with scale changes so as to take care of blurred observations.

- Though the hybrid method discussed in chapter 7 appears to be a more general approach, the use of PCA for global learning restricts the applicability of the method only for a particular class of images. The modification of this method using appropriate global and local learning approaches seem to be a

promising area of research. This, in effect, should take care of all the issues of super-resolution namely alias removal, de-blurring and de-noising, simultaneously giving rise to a better quality super-resolved image.

- Although the alias-free interpolation method discussed in the last chapter does address the issue of alias removal while upsampling the data, there are quite a few issues that require a thorough investigation. The appropriate choice of the extent of aliasing is one such issue. Similarly, the assumption of a piecewise constant image may not be suitable for practical applications. A more appropriate model such as a locally quadratic intensity variation would be more appropriate. This requires further investigations. Further more, this method cannot handle a blurred or noisy observation. It will be interesting to design an alias-free interpolation scheme with sufficient mathematical footing with a more general observation model.

- In all the SR methods proposed in literature, the quality of the super-resolved image is measured either by means of visual inspection or using a PSNR check. It can be easily shown that the PSNR measure is heavily biased towards the lower part of the spectrum due to the fact that most of the energy is contained in this region. Hence the PSNR may not be a good measure to evaluate the performance of an SR scheme. Also, in practical super-resolution the original high resolution image is unknown. So, it will be an interesting idea to develop an appropriate measure to check the quality of super-resolved images.

- Recently, there have been some papers where the authors mathematically determine the limits on super-resolution, i.e. by what factor one can achieve reliable super-resolution. These works are basically for reconstruction-based super-resolution particularly for the multi-frame case. It will be interesting to extend this work for learning-based super-resolution where one has to quantify how much additional information is learned from the training images and also the maximum amount of information one can learn.

Acknowledgement

It is really a great pleasure to acknowledge the help and support that has gone into the making of this dissertation.

I consider myself fortunate to have had Prof. Subhasis Chaudhuri as my thesis supervisor. Working with him was really a wonderful experience for me. His amazing attitude for research has been truly inspiring. His invaluable guidance, constant encouragement and deep insight into the subject helped me keep this research on course. I am very much thankful to you sir for all your help and support.

I am thankful to Prof. V. M. Gadre, Prof. S. N. Merchant and Prof. S. Dutta Roy for agreeing to be on my research committee. I thank them very much for their timely suggestions and advice. I thank Prof. Sharat Chandran for his critical comments and suggestions which were very useful.

I am very much thankful to my parents for their deep interest in my studies and support and prayers through out my academic career. Special thanks to my wife, our dear daughter Anjali and our new born Anjith for their care and understanding. I thank all my relatives who contributed indirectly to my research work.

I am deeply indebted to Dr. Manjunath Joshi, former student of our lab for his advice and help during the initial stages of my work. I am thankful to all my colleagues in the lab - P. Neethu, Priyam Chatterjee, Nyamati, Tanmay Pawar, Hitesh, Malay Nema and Shilpa who have been of help one way or the other. Special thanks to Vinay P. N. and Srikrishnan, my colleagues and system administrators of the lab for all their help and co-operation.

Summary

There is always a high demand for digital images with higher spatial resolution because they not only give the viewer a pleasing picture but also offer additional details that are important for further processing in many applications like medical imaging, remote sensing, robot vision, etc. The most direct solution to increase spatial resolution is to reduce the pixel size (i.e., increase the number of pixels per unit area) by sensor manufacturing techniques. As the pixel size decreases, however, the amount of light available also decreases. It generates shot noise that degrades the quality of the image captured. To reduce the pixel size without suffering the effects of shot noise, therefore, there exists the limitation of the pixel size reduction, and the optimally limited pixel size is estimated at about $40\mu m^2$ for a $0.35\mu m$ CMOS process. The current image sensor technology has almost reached this level. The high cost for high precision optics and image sensors is also an important concern in many commercial applications regarding high resolution (HR) imaging. Therefore, a new cost effective approach toward increasing spatial resolution is required to overcome these limitations of the sensors and optics manufacturing technology. A promising approach is to use post acquisition signal processing techniques to enhance the resolution. These techniques using one or more low resolution (LR) images to construct a high resolution image are called super-resolution (SR) techniques. Super-resolution techniques handle the issues of alias removal, de-blurring and de-noising while interpolating the low resolution inputs.

Simple resolution enhancement methods on smoothing and interpolation techniques for noise reduction have been commonly used in image processing. Smoothing is usually achieved by applying various spatial filters such as Gaussian, Wiener, and

median filters. Commonly used interpolation methods such as linear and cubic spline interpolation smooth the image data in discontinuous regions, producing a larger image which appears rather blurred. Cubic spline interpolants also tend to overshoot sharp discontinuities, producing a ringing effect at edges. These techniques are not considered as super-resolution techniques as they do not generate the exact high frequencies missing in the low resolution input.

A natural solution to super-resolution reconstruction is to make use of multiple sub-pixel shifted low resolution observations of the same scene and extract the non redundant information among them. This is termed as multi-frame super-resolution in the literature. This technique involves exact registration between the different frames followed by interpolation and finally restoration to remove blur and noise. Most of the literature available on SR is based on multi-frame super-resolution using the motion cue. There have also been some work on motion free super-resolution where the authors use cues other than motion - like blur, defocus, zoom, etc. for super-resolution reconstruction. These techniques also use several low resolution observations to produce the super-resolved image.

Recently, there have been some works on single frame image super-resolution where only one low resolution observation is used to construct the super-resolved image. These techniques are based on statistical and machine learning approaches where the missing high frequency details in the low resolution image is learned from an appropriate image database. Super-resolution methods under this category are also known as learning based super-resolution techniques. In this thesis we propose some new approaches for recovering a high resolution image from a single low resolution observation making use of a training set containing several high resolution images. In super-resolution one always aims at generating the missing high frequency components. To this end, we also propose a method to quantify that the generated details are really the high frequency details missing in the given low resolution observation.

As the effectiveness of the use of wavelets for image super-resolution has already been established in the literature, first we propose a single frame image super resolution method using a learned wavelet prior on the high resolution image. Wavelet

coefficients at locations having edges in an image have significant value at finer resolutions. On the other hand, at homogeneous regions they are negligible at finer resolutions. Hence we learn the wavelet coefficients at finer scales of the unknown high resolution image locally from a set of high resolution training images using the low resolution edge primitives in the input image. As the direct inverse wavelet transform of the learned coefficients cannot produce a good quality super-resolved image, the learned image in the wavelet domain is used a prior and we solve the super-resolution problem under a regularization framework. We use an appropriate smoothness prior with discontinuity preservation in addition to the wavelet based constraint to estimate the super-resolved image. Thus we minimize a cost function containing the data consistency term, the smoothness constraint and the learned prior in the wavelet domain. The smoothness term ensures the spatial correlation among the pixels whereas the learning term chooses the best edges from the training set. Because this amounts to extrapolating the high frequency components, the proposed method does not suffer from over smoothing effects. Experiments performed on both gray scale and color images showed perceptual as well as quantifiable improvements over conventional interpolation techniques even though the method is computationally expensive.

One of the major difficulties 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 gave rise to certain artifacts in the reconstructed image and in order to get a good quality super-resolved image we were forced to solve the problem under a Bayesian framework using an appropriate discontinuity preserving smoothness constraint in addition to the learned wavelet prior. Thus we ensure spatial correlation among pixels using the smoothness constraint, as well as obtain the best matching edges from the training set using wavelet learning. As the two priors are in two different domains, it required a stochastic optimization technique to obtain the solution, which made the

reconstruction process very slow.

A better way to handle the above situation is to use directionally selective wavelet decomposition to learn the oriented edges where the reconstruction problem need not be solved under a regularization framework, resulting in a much faster solution. This motivated us to use the contourlet transform which is capable of catching the smoothness along contour naturally. The contourlet transform is an extension of the Cartesian wavelet transform in two dimensions using multi-scale and directional filter banks. They offer a much richer set of directions and shapes, and thus, they are more effective in capturing smooth contours and geometric structure in the image. The contourlet coefficients at finer scales of the unknown high resolution image are learned locally from a set of high resolution training images, the inverse contourlet transform of which recovers the super-resolved image. In effect, we learn the high resolution representation of an oriented edge primitive from the training data. Our experiments show that the proposed approach outperforms standard interpolation techniques as well as a standard wavelet-based learning both visually and in terms of the PSNR values, especially for images with arbitrarily oriented edges. Also the method is very fast compared to the wavelet based regularized approach.

We further extend the contourlet-based learning method using a linear combination of edge primitives inspired by the concept of locally linear embedding (LLE), a recent manifold learning method, under a supervised learning framework. Here the generation of a high resolution patch does not depend on only one of the nearest neighbors in the training set. Instead, it depends simultaneously on multiple nearest neighbors in a way similar to LLE for manifold learning. Here we make use of the low resolution and the corresponding high resolution training image pairs to learn the oriented edges to estimate the high resolution embedding. The advantage of this lies in capturing the effect of the decimation process explicitly by the contourlet transform through the matched pair of LR and HR training data. In the previous methods it has been implicitly assumed that the multiresolution decomposition using either wavelet or contourlet has the same effect as that of the decimation process, which is often not true. For a given LR test image patch (in contourlet domain) we find k

nearest neighbors from the LR training set by performing an extensive search. Now the test patch is reconstructed from a linear combination of the k neighbors where the reconstruction weights are computed so as to minimize the reconstruction error. The linear combination of the corresponding k HR patches using the same reconstruction weights generate the required super-resolved patch in the contourlet domain. The process is repeated for all patches in the test image. A major disadvantage of the method is that, if there is some spurious learning at certain locations, the resulting high resolution image at the corresponding location will be much noisy as it is contributed by several spuriously learned neighbors. To overcome this difficulty one has to carefully set the threshold thereby avoiding any spurious learning.

All the above methods are based on learning the features locally and are good at enhancing the edges during super-resolution, but are poor in handling noisy data. Further, they are unable to undo the effect of aliasing. But in certain applications one need to super-resolve an image which is severely blurred and noisy. For example, in investigative criminology one has available face and fingerprint databases which are often taken at controlled environment. 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 ? To solve this problem we propose a global learning approach using principal component analysis (PCA) unlike the above methods where the learning was local. In this global approach, we learn the statistics of a good quality image database and use them to enhance the given LR input. As the learning is global it imposes the constraint that all the training images should be globally similar, *i.e.* they should represent a similar class of objects.

Here we propose two approaches to enhance the low resolution input. In the first method we use the idea of generalized interpolation. A few significant eigenimages of a database of several similar low resolution images are computed and the given low resolution image is projected onto the eigenimages to compute the projection coefficients. The high resolution image is reconstructed using these coefficients and the HR eigenimages obtained by performing generalized interpolation on the LR eigenimages. In the second approach, we use a high resolution database and compute the principal

eigenimages. The only difference here is that the upsampling process is shifted to the input side rather than performing generalized interpolation of the eigenimages. These methods are applicable to images of a particular class, for example, face and fingerprint images. The algorithm offers a significant advantage when the input image is blurred and noisy. 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. However, this global upsampling method does not provide an alias-free reconstruction. Further, the method has limited applicability-mostly in enhancing face or fingerprint images.

We realize that the global high resolution reconstruction using generalized interpolation fails to do exact super-resolution, though it removes the blur and noise present in the input low resolution image. But it very well learns the global characteristics from the image database. On the other hand, the local learning methods using wavelets or contourlets obtain the missing high frequency details by locally learning the best edges from the image database. To extract the advantages of both these methods, we now propose a hybrid approach for super-resolution by integrating the global learning using PCA and the local learning approach using wavelets or contourlets. The former extracts the best statistical features from a database of similar class of images while the latter ensures the best edges from the database. We solve the problem under a Bayesian framework incorporating the global and local priors, in addition to the data consistency term. We apply the method particularly for face hallucination where we synthesize the missing high frequency details of the input low resolution face image. Ideally, a good hybrid approach should handle the blur and noise present in the input while performing interpolation. But our method could not properly handle blurred and noisy observations. The improvement achieved over bicubic interpolation was very substantive either. In addition, we were unable to prove that the hybrid methods achieve a better edge over the local methods. This could be due to the fact that as edges are more important in a super-resolved image,

local methods always play a good role in restoring them in a better way. A proper study is required in terms of whether the global prior or the local prior is more important in a particular application, for the development of an efficient hybrid approach for super-resolution. Also, the selection of appropriate priors is equally important to reconstruct a good quality super-resolved image.

The final contribution of the thesis is in developing an alias-free interpolation scheme. All the local and global methods discussed so far are able to provide a super-resolved image, but they are unable to handle the aliasing effects. Even in the local approaches, though the edges are handled properly, the aliasing present in the LR image is not reduced much during the upsampling process. Here we study the possibility of removing aliasing in a scene by designing an alias-free upsampling scheme. Our work is motivated by the work of Candes *et al.* where the authors address the problem of exact signal reconstruction from incomplete frequency information. We build on the theorem developed by them to derive a method for exact removal of aliasing while interpolating an image. Quite naturally this is impossible unless additional information about the HR data is available. Hence we assume that the HR data is piece-wise constant. We generate the unknown high frequency components of the given partially aliased (low resolution) image by minimizing the total variation of the interpolant subject to the constraint that part of unaliased spectral components in the low resolution observation are known precisely and under the assumption of sparsity in the data. This provides a mathematical basis for exact reproduction of high frequency components with probability approaching one, from their aliased observation.

In all the SR methods proposed in literature, the quality of the super-resolved image is measured either by means of visual inspection or using a PSNR check. It can be easily shown that the PSNR measure is heavily biased towards the lower part of the spectrum due to the fact that most of the energy is contained in this region. Hence the PSNR may not be a good measure to evaluate the performance of an SR scheme. Even though, several different measures to evaluate the quality of visually pleasing images are proposed in the literature, the issue that the reconstructed components

are really the missing high frequency components has not really been investigated so far. Our focus in this work is a study in this direction. We study the issue of alias-free interpolation at a more fundamental level and restrict ourselves again to using a single observation without using any training database unlike the other methods. A sub-optimal computational procedure using linear programming is also presented. Experiments performed on 1-D and 2-D signals confirm the utility of the proposed method. But a disadvantage of our method is that it is best suited only for piecewise constant objects. Also we specifically need the alias-free region in the low resolution signal spectrum which is impossible in practice.

We provide a detailed description of all proposed techniques in the dissertation. We have also experimented on a number of images and shown the effectiveness of the proposed techniques.

Bibliography

[1] 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.

[2] T. Komatsu, K. Aizawa, T. Igarashi, and T. Saito, "Signal Processing based Method for Acquiring Very High Resolution Image with Multiple Cameras and its Theoretical Analysis," in *Proc. IEE-I*, 1993, pp. 19–25.

[3] Zeev Zalevsky and David Mendlovic, *Optical Superresolution*, Springer, 2004.

[4] Nicholas Fang, Hyesog Lee, Cheng Sun, and Xiang Zhang, "Sub-Diffraction-Limited Optical Imaging with a Silver Superlens," *Science*, vol. 308, pp. 534–537, 2005.

[5] 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," *Intl. J. Computer Vision*, vol. 30, no. 3, pp. 175–190, December 1998.

[6] M. N. Do and M. Vetterli, "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation," *IEEE Trans. on Image Processing*, vol. 14, no. 12, pp. 2091–2106, December 2005.

[7] 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.

[8] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive Reconstruction of High Resolution Image from Noisy Undersampled Multiframes," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, June 1990.

[9] S. P. Kim and W. Y. Su, "Recursive High Resolution Reconstruction of Blurred Multiframe Images," *IEEE Trans. Image Processing*, vol. 2, no. 4, pp. 534–539, October 1993.

[10] S.H.Rhee and M.G.Kang, "Discrete cosine transform based regularized high-resolution image reconstruction algorithm," *Opt.Eng.*, vol. 38, no. 8, pp. 1348–1356, August 1999.

[11] 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.

[12] A. Papoulis, "Generalized Sampling Theorem," *IEEE Trans. on Circuits and Systems*, vol. 24, pp. 652–654, November 1977.

[13] 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.

[14] 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 *Proc. IEEE Conf. Acoustics, Speech and Signal Processing*, Minneapolis, MN, 1993, pp. 269–272.

[15] M.S. Alam, J.G. Bognar, R.C. Hardie, and B.J. Yasuda, "Infrared Image Registration and High-Resolution Reconstruction using Multiple Translationally Shifted Aliased Video Frames," *IEEE Trans. on Instrum. Meas.*, vol. 49, pp. 915–923, October 2000.

[16] N. Nguyen and P. Milanfar, "An Efficient Wavelet-based Algorithm for Image Superresolution," in *Proc. Int. Conf. Image Processing*, 2000, pp. 351–354.

[17] M. Irani and S. Peleg, "Improving Resolution by Image Registration," *CVGIP:Graphical Models and Image Processing*, vol. 53, pp. 231–239, March 1991.

[18] M. Irani and S. Peleg, "Motion Analysis for Image Enhancement : Resolution, Occlusion, and Transparency," *VCIR*, vol. 4, pp. 324–335, December 1993.

[19] 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.

[20] 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, 2000.

[21] M. K. Ng and N. K. Bose, "Analysis of Displacement Errors in High Resolution Image Reconstruction with Multisensors," *IEEE Trans. Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 6, pp. 806–813, June 2002.

[22] M. K. Ng and N. K. Bose, "Fast Color Image Restoration with Multisensors," *International Journal of Imaging Systems and Technoloy*, vol. 12, no. 5, pp. 189–197, 2002.

[23] N. Nguyen, P. Milanfar, and G. Golub, "A Computationally Efficient Superresolution Reconstruction Algorithm," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 573–583, April 2001.

[24] M.C. Hong, M.G. Kang, and A.K. Katsaggelos, "An Iterative Weighted Regularized Algorithm for Improving the Resolution of Video Sequences," in *Proc. Int. Conf. Image Processing*, 1997, pp. 474–477.

[25] M. G. Kang, "Generalized Multichannel Image Deconvolution Approach and its Applications," *Opt. Eng.*, vol. 37, no. 11, pp. 2953–2964, November 1998.

[26] N.K. Bose, S. Lertrattanapanich, and J. Koo, "Advances in Superresolution using l-curve," in *Proc. Int. Symp. Circuits and Systems*, Sydney, 2001, pp. 433–436.

[27] P.C. Hansen and D. Prost O'Leary, "The Use of the L-curve in the Regularization of Discrete Ill-posed Problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, November 1993.

[28] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a High-Resolution Image by Simultaneous Registration, Restoration, and Interpolation of Low-Resolution Images," in *Proc. IEEE Int. Conf. on Image Processing*, Washington,DC, 1995, pp. 539–542.

[29] 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.

[30] 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.

[31] 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.

[32] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson, *Super-resolved Surface Reconstruction from Multiple Images*, NASA Ames Research Center, Moffett Field, CA, Tech. Rep. FIA-94-12, 1994.

[33] 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.

[34] M. Elad and Y. Hel-Or, "A Fast Super-resolution Reconstruction Algorithm for Pure Translation Motion and Common Space-Invariant Blur," *IEEE Trans. on Image Processing*, vol. 10, no. 8, pp. 1187–1193, August 2001.

[35] N. Nguyen, P. Milanfar, and G. Golub, "Efficient Generalized Cross-Validation with Applications to Parametric Image Restoration and Resolution Enhancement," *IEEE Trans. Image Processing*, vol. 10, no. 9, pp. 1299–1308, September 2001.

[36] M. C. Chiang and T. E. Boult, "Local Blur Estimation and Super-Resolution," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Puerto Rico, USA, 1997, pp. 821–826.

[37] N. Kaulgud, J. Karlekar, and U. B. Desai, "Compressed Domain Video Zooming : Use of Motion Vectors," in *Proc. of Eighth National Conf. on Communications*, Mumbai, India, Jan. 2002, pp. 120–124.

[38] S. Thurnhofer and S. K. Mitra, "Edge-preserving Image Zooming," *Optical Engineering*, vol. 35, no. 7, pp. 1862–1870, 1996.

[39] 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.

[40] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust Super-resolution," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Kauai, HI, USA, 2001, pp. 645–650.

[41] M. K. Nema and S. Rakshit, "Edge-Model Based Representation of Laplacian Subbands," in *Proc. ACCV (LNCS vol.3851)*, Hyderabad, India, 2006, pp. 80–89.

[42] C. A. Segall, R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian Resolution Enhancement of Compressed Video," *IEEE Trans. Image Proc.*, vol. 13, no. 7, pp. 898–911, July 2004.

[43] L. D. Alvarez, J. Mateos, R. Molina, and A. K. Katsaggelos, "High Resolution Images from Compressed Low Resolution Video: Motion Estimation and Observable Pixels," *Int. Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 58–66, August 2004.

[44] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High Resolution Images from Low Resolution Compressed Video," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 37–48, May 2003.

[45] 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*, 2001, pp. II:627–634.

[46] 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.

[47] L. C. Pickup, S. J. Roberts, and A. Zisserman, "A Sampled Texture Prior for Image Super-Resolution," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds., pp. 1587–1594. MIT Press, Cambridge, MA, 2004.

[48] 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.

[49] 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.

[50] S. Chaudhuri and M. V. Joshi, *Motion-Free Super-Resolution*, Springer, 2005.

[51] D.Rajan and S.Chaudhuri, "Generation of super-resolution images from blurred observations using an MRF model," *J.Math.Imaging Vision*, vol. 16, pp. 5–15, 2002.

[52] D. Rajan and S. Chaudhuri, "Simultaneous Estimation of Super-Resolved Intensity and Depth Maps from Low Resolution Defocussed Observations of a Scene," in *Proc. IEEE Int. Conf. on Computer Vision*, Vancouver Canada, 2001, pp. 113–118.

[53] 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.

[54] 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.

[55] 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.

[56] 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*, Indian Statistical Institute, Kolkata, India, 2003, pp. 179–182.

[57] S. Chaudhuri, *(Ed.), Super-Resolution Imaging*, Kluwer Academic Publisher, Boston, 2001.

[58] S. C. Park and M. K. Park ang M. G. Kang, "Super-Resolution Image Reconstruction: A Technical Overview," *IEEE Signal Processing Magazine, Special issue of Super-Resolution Image Reconstruction*, vol. 20, no. 3, pp. 21–36, May 2003.

[59] H. H. Hou and H. C. Andrews, "Cubic Splines for Image Interpolation and Digital Filtering," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-26, no. 6, pp. 508–517, 1978.

[60] R. G. Keys, "Cubic Convolution Interpolation for Digital Image Processing," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-29, no. 6, pp. 1153–1160, 1981.

[61] J. L. Harris, "Diffraction and resolving power," *J. Opt. Soc. Am.*, , no. 54, pp. 931–936, 1964.

[62] R. W. Gerchberg, "Super-resolution through error energy reduction," *Opt. Acta*, , no. 21, pp. 709–720, 1974.

[63] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall of India Private Limited, 2001.

[64] A. Papoulis, "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," *IEEE Trans. on Circuits and Systems*, vol. CAS-22, no. 9, pp. 735–742, September 1975.

[65] W. T. Freeman and E. C. Pasztor, "Learning Low-level Vision," in *Proc. Seventh IEEE Int. Conf. on Computer Vision*, Kerkyra, Greece, 1999, pp. 1182–1189.

[66] W. T. Freeman, E. C.Pasztor, and O. T. Carmichael, "Learning Low -level Vision," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.

[67] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image Analogies," in *Proc. SIGGRAPH*, 2001, pp. 327–340.

[68] 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.

[69] S. Baker and T. Kanade, "Hallucinating Faces," in *Proc. Fourth Int. Conf. on Automatic Face and Gesture Recognition*, 2000.

[70] C. Liu, H. Y. Shum, and C. S. Zhang, "A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Kauai, HI, USA, 2001, pp. 192–198.

[71] X. Wang and X. Tang, "Hallucinating Face By Eigentransformation with Distortion Reduction," in *Proc. Int. Conf. on Biometrics Authentication*, Hong Kong, 2004.

[72] W. Liu, D. Lin, and X. Tang, "Hallucinating Faces: TensorPatch Super-Resolution and Coupled Residue Compensation," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, San Diego,CA, USA, 2005, pp. 478–484.

[73] Q. Wang, X. Tang, and H. Shum, "Patch Based Blind Image Super Resolution," in *Proc. Tenth Int. Conf. on Computer Vision*, Beijing, China, 2005, pp. 709–716.

[74] K. Jia and S. Gong, "Multi-Modal Tensor Face for Simultaneous Super-Resolution and Recognition," in *Proc. Tenth Int. Conf. on Computer Vision*, Beijing, China, 2005, pp. 1683–1690.

[75] J. Sun, N. N. Zheng, H. Tao, and H. Y. Shum, "Image Hallucination with Primal Sketch Priors," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Wisconsin, 2003, pp. 729–736.

[76] A. J. Storkey, "Dynamic Structure Super-Resolution," in *Advances in Neural Information Processing Systems 15*, S.Becker, S. Thrun, and K. Obermayer, Eds., pp. 1295–1302. MIT Press, Cambridge, MA, 2003.

[77] K. I. Kim, M. O. Franz, and B. Schölkopf, "Kernel Hebbian Algorithm for Single-Frame Super-Resolution," in *Statistical Learning in Computer Vision , ECCV 2004 Workshop*, A. Leonardis and H. Bischof, Eds., pp. 135–149. Prague, Czech Republic.

[78] I. Begin and F. P. Ferrie, "Blind Super-Resolution Using a Learning-Based Approach," in *Proc. Int. Conf. on Pattern Recognition*, Cambridge, U. K., 2004, pp. 85–89.

[79] J. S. Park and S. W. Lee, "Enhancing Low-Resolution Facial Images Using Error Back-Projection for Human Identification at a Distance," in *Proc. Int. Conf. on Pattern Recognition*, Cambridge, U. K., 2004, pp. 346–349.

[80] T. L. Chang, T. L. Liu, and J. H. Chuang, "Direct Energy Minimization for Super-Resolution on Nonlinear Manifolds," in *Proc. ECCV (LNCS vol.3954)*, Graz, Austria, 2006, pp. 281–294.

[81] S. vanHuffel, P. C. Hansen, and P. Maas, *(Eds.), Linear Algebra and its Applications : Special Issue on Linear Algebra in Signal and Image Processing*, vol. 391, 2004.

[82] Michael Ng, Edmund Lam, and Chong-Sze Tong, *(Eds.), Multidimensional Systems and Signal Processing : Special Issue on Superresolution imaging: Theory, Algorithms and Applications*, No 2-3, 2007.

[83] Michael Ng, Tony Chan, Moon Gi Kang, and Peyman Milanfar, *(Eds.), EURASIP Journal on Advances in Signal Processing: Special Issue on Super-Resolution Imaging: Analysis, Algorithms, and Applications*, vol. 2006, 2006.

[84] Russell C. Hardie, Richard R. Schultz, and Kenneth E. Barner, *(Eds.),EURASIP Journal on Advances in Signal Processing: Special Issue on Super-Resolution Enhancement of Digital Video*, vol. 2007, 2006.

[85] Nirmal K. Bose, Raymond H. Chan, and Michael K. Ng., *(Eds.),International Journal of Imaging Systems and Technology : Special Issue on High-Resolution Image Reconstruction. I*, vol. 14, Issue 2, 2004.

[86] Nirmal K. Bose, Raymond H. Chan, and Michael K. Ng., *(Eds.),International Journal of Imaging Systems and Technology : Special Issue on High-Resolution Image Reconstruction.II*, vol. 14, Issue 3, 2004.

[87] M. G. Kang and S. Chaudhauri, *(Eds.), IEEE Signal Processing Magazine : Special Issue on Super-Resolution Image Reconstruction*, Vol. 20, no:3, 2003.

[88] Rafael Molina and Aggelos Katsaggelos, *(Eds.), The Computer Journal : Special Issue on Super-Resolution in Imaging and Video*, Oxford University Press, 2007.

[89] A. Chambolle, R. A. DeVore, N. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Proc.*, vol. 7, no. 3, pp. 319–335, March 1998.

[90] M. K. Ng, C. K. Sze, and S. P. Yung, "Wavelet Algorithms for Debluring," *Int. J. Imaging Syst. Tech.*, vol. 14, no. 3, pp. 113–121, 2004.

[91] N. Nguyen and P. Milanfar, "A wavelet-based interpolation-restoration method for superresolution," *Circuits, Systems, and Signal Processing*, vol. 19, no. 4, pp. 321–338, August 2000.

[92] N. K. Bose and M. B. Chappalli, "A Second Generation Wavelet Frame-work for Superresolution with Noise Filtering," *Int. J. Imaging Syst. Tech.*, vol. 14, no. 2, pp. 84–89, 2004.

[93] R. Chan, T. Chan, L. Shen, and Z. Shen, "Wavelet algorithms for high-resolution image reconstruction," *SIAM J. Sc. Computing.*, vol. 24, no. 4, pp. 1408–1432, 2003.

[94] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, New Jersy, 1998.

[95] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pennsylvania, 1992.

[96] A. Jensen and A. Cour-Harbo, *Ripples in Mathematics: The Discrete Wavelet Transform*, Springer-Verlag, 2001.

[97] H. M. Shapiro, "Embedded Image Coding," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, December 1993.

[98] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data Compression and Harmonic Analysis," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2435–2476, October 1998.

[99] A. Mallat, *A Wavelet Tour of Signal Processing*, Academic: New York, 1999.

[100] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 Still Image Compression Standard," *IEEE Signal Process. Mag.*, vol. 18, no. 6, pp. 36–58, September 2001.

[101] M. N. Do, "Directional Multiresolution Image Representations," Ph.D. thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland, 2001.

[102] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. on Communication*, vol. 31, no. 4, pp. 532–540, April 1983.

[103] M. N. Do and M. Vetterli, "Framing Pyramids," *IEEE Trans. Signal Process.*, , no. 9, pp. 2329–2342, September 2003.

[104] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images," *IEEE Trans. on Signal Processing*, vol. 40, no. 4, pp. 882–893, April 1992.

[105] M. Vetterli, "Multidimensional subband coding: Some theory and algorithms," *Signal Proc.*, vol. 6, no. 2, pp. 97–112, February 1984.

[106] S. Park, M. J. T. Smith, and R. M. Mersereau, "A new directional filterbank for image analysis and classification," in *Proc.IEEE Int.Conf. on Acoustics, Speech, and Signal Processing*, 1999, pp. 1417–1420.

[107] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. on Signal Processing*, vol. 40, no. 9, pp. 2207–2232, September 1992.

[108] S.M.Phoong, C.W.Kim, P.P.Vaidyanathan, and R.Ansari, "A new class of Two-channel Biorthogonal Filter banks and Wavelet bases," *IEEE Trans. on Signal Processing*, vol. 43, no. 3, pp. 649–665, March 1995.

[109] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, , no. 290, pp. 2323–2326, 2000.

[110] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-Resolution Through Neighbor Embedding," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Washington, 2004, pp. 275–282.

[111] J. B. Tenenbaum, V. de silva, and J. C. Langford, "A Global Geometric Framework for nonlinear Dimensionality Reduction," *Science*, , no. 290, pp. 2319–2323, 2000.

[112] L. K. Saul and S. T. Roweis, "Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds," *Journal of Machine Learning Research*, vol. 4, pp. 119–155, 2003.

[113] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[114] X. Wang and X. Tang, "Hallucinating Face by Eigentransformation," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 35, no. 3, pp. 425–434, August 2005.

[115] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, 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.

[116] 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.

[117] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[118] G. F. Carey, *Computational Grids: Generation Adaptation and Solution Strategies*, Taylor and Francis, 1997.

[119] M. J. D. Powell, "The Uniform Convergence of Thin Plate Spline Interpolation in Two Dimensions," *Numerishce Mathematik*, pp. 107–128, August 1994.

[120] M. B. Chappalli and N. K.Bose, "Simultaneous noise filtering and super-resolution with second-generation wavelets," *Signal Proc. Letters*, vol. 12, no. 11, pp. 772–775, November 2005.

[121] 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.

[122] M. Shahram and P. Milanfar, "Imaging Below the Diffraction Limit: A Statistical Analysis," *IEEE Trans. on Image Processing*, vol. 13, no. 5, pp. 677–689, May 2004.

[123] S. P. Kim and N. K.Bose, "Reconstruction of 2-D bandlimited discrete signals from nonuniform samples," *IEE Proceedings*, vol. 137,pt. F, no. 3, pp. 197–204, June 1990.

[124] T. Strohmer, "Computationally Attractive Reconstruction of Bandlimited Images from Irregular Samples," *IEEE Trans. on Image Processing*, vol. 6, no. 4, pp. 540–548, April 1997.

[125] E. Candes, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, February 2006.

[126] D. L. Donoho and M. Elad, "Maximal sparsity Representation via l_1 Minimization," *the Proc. Nat. Aca. Sci.*, vol. 100, pp. 2197–2202, March 2003.

[127] M. Bertalmio, V. Caselles G. Sapiro, and C. Ballester, "Image Inpainting," in *Proc. SIGGRAPH*, New Orleans, USA, 2000.

[128] T. Chan and S. H. Kang, "Error Analysis for Image Inpainting," *UCLA CAM report 04-72*, December 2004.

[129] S. Levy and P. K. Fullagar, "Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution," *GEOPHYSICS*, vol. 46, no. 9, pp. 1235–1243, September 1981.

[130] F. Santosa and W. W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM J. Sci. Statist. Comput.*, vol. 7, pp. 1307–1330, 1986.

[131] F. Guichard and F. Malgouyres, "Total Variation Based Interpolation," in *Proc. European Signal Processing Conference*, 1998, pp. 1741–1744.

[132] F. Malgouyres and F. Guichard, "Edge Direction Preserving Image Zooming: A Mathematical and Numerical Analysis," *SIAM J. Numerical Analysis*, vol. 39, no. 1, pp. 1–37, 2001.

[133] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *PHYSICA D*, vol. 60, pp. 259–268, 1992.

Publications

1. C.V. Jiji, M.V. Joshi and S. Chaudhuri, "Single Frame Image Super-Resolution Using Learned Wavelet Coefficients", Intl. J. Imaging Systems and Tech., (special issue on high resolution image reconstruction), vol-14, no-3, pp 105-112, Sep 2004.

2. C.V. Jiji and S. Chaudhuri, "Single Frame Image Super-Resolution Through Contourlet Learning", EURASIP J. Applied Signal Processing, vol-2006, pp 1-11, 2006.

3. C.V. Jiji and S. Chaudhuri, "PCA-based Generalized Interpolation for Image Super-Resolution", Proc. Indian Conf. on Vision, Graphics and Image Processing (ICVGIP), Kolkata, Dec. 2004.

4. C.V. Jiji, P. Neethu and S. Chaudhuri, "Alias-Free Interpolation", Proc. ECCV 2006 (LNCS vol-3953) Graz, Austria, May 2006.

5. C. V. Jiji, S. Chaudhuri and P. Chatterjee, "Super-Resolution: Should we process Locally or Globally ? " Multidimensional Systems and Signal Processing vol-18 (2 & 3) Sept. 2007, pp 123-152.