Some New Techniques for Motion-free Super-resolution

by

Prakashchandra Purushottamdas Gajjar

(200521004)

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree

of

Doctor of Philosophy

in

Information and Communication Technology

to

Dhirubhai Ambani Institute of Information and Communication Technology

JULY 2010

DECLARATION

This is to certify that

1. the thesis comprises my original work towards the degree of Doctor of Philosophy in Information and Communication Technology at DA-IICT and has not been submitted elsewhere for a degree,

2. due acknowledgment has been made in the text to all other material used.

Signature of Student

CERTIFICATE

This is to certify that the thesis work entitled "Some New Techniques for Motion-free Super-resolution" has been carried out by Prakashchandra Purushottamdas Gajjar (200521004) for the degree of Doctor of Philosophy in Information and Communication Technology at this Institute under my supervision.

Thesis Supervisor Prof. Manjunath Joshi

Acknowledgments

First and above all, I praise God, the almighty for providing me this opportunity and granting me the capability to proceed successfully. Although the work described here was performed independently, I never would have been able to complete it if not for the support of many wonderful people. I would therefore like to offer my sincere thanks to them.

I would like to express my sincere gratitude to my supervisor Prof. Manjunath V. Joshi with whom I have learned immensely and has had a strong influence in my development as a researcher. Thank you sir for the patience you had to supervise and direct my work, and for all the hard job of guidance, fruitful discussions, and helping me throughout all the different steps of my doctoral research endeavor for the past few years. I greatly appreciate the freedom and collegial respect you have given me. As my supervisor, Prof. Joshi has constantly forced me to remain focused on achieving my goal. His observations, guidance, and comments helped me to establish the overall direction of the research and to move forward with investigation in depth. Working with Prof. Joshi has enriched my life not only in the academic sense. His achievements, his work ethics, and his keen eye for every important detail have been an inspiration throughout all the years I have worked with him.

On a broader note, I wish to acknowledge all the professors of DA-IICT who have inspired me directly or indirectly. I would like to thank faculty members V. P. Sinha, Suman Mitra, and Asim Banerjee who have shaped my thinking about this research direction during the initial phases of thesis work. It is always nice remembering the excellent times of the lessons by Prof. V. P. Sinha and how this sweet experience moved me toward the initiating research work. Perhaps most importantly, he has taught us how to think critically and independently for carrying out research.

I express my sincere gratitude to Director, Prof. S.C. Sahasrabudhe, Registrar, Dean-AP, Convener-PGC and all other faculties who directly or indirectly helped me throughout my time at DA-IICT. I would like to thank administrative and technical staff members of DA-IICT who have been kind enough to advise and help me in their respective roles. My gratitude also extend to staff of DA-IICT resource-center.

I would like to express my gratitude to Prof. Subhasis Chaudhuri (Indian Institute

of Technology, Bombay) for all his technical support, valuable input and remarks at several occasions. I also thank Prof. S. Ravishanker (Amrita School of Engineering, Banglore) for generously sharing his time and knowledge that influenced the work and resulted in cooperative joint publication. I wish to acknowledge Prof. Pradipta Kumar Nanda (Institute of Technical Education and Research, Bhubaneshwar) for his help and discussions on parameter estimation.

I am extremely grateful to Prof. N. M. Santoki, Head of Electronics and Communication Department at Government Polytechnic, Palanpur for being a constant source of inspiration to me while I was serving at Government Polytechnic, Palanpur. Without his warm encouragement and support, I would not have gone back to school to pursue my PhD. A special word of thanks to Kishor Upla, PhD Scholar, DA-IICT and Ex. Lecturer (GP, Palanpur) for academic help while I was at GP, Palanpur and technical help while at DA-IICT. He has been always ready to help me since the first day I came to DA-IICT. I also express my thanks to faculties of EC department, Government Polytechnic, Ahmedabad for their support and patience when I needed them.

I made a lot of new friends at DA-IICT, who helped me in many steps of my study. I believe that I would not be able to name every colleague and friend separately but I thank all of them for everything that they did for me.

I thank Junping Zhang (Fudan University, Shanghai, China), Kim (Max-Planck-Institute for biologische Kybernetik,Germany) and Kwon (Korea Advanced Institute of Science and Technology, Taejon, Korea) for availing their code for comparing our results on super-resolution. Moreover, I want to express my gratitude to the anonymous reviewers and the editors for their constructive suggestions that have greatly improved the published articles.

I would love to cordially express my gratitude to my father whose understanding and help led me to where I am, my mother whose compassion and determination helped me to become who I am. I warmly thank my in-laws, sisters and their families for their understanding, encouragement, and support in many aspects of my life.

And finally, I know that you did not want to be named, my wife, Bhavna, without your supports and encouragements, I could not have finished this work, it was you who kept the fundamental of our family, and I understand it was difficult for you. I see myself unable to even express my feelings about the love and patience that I observed from you.

I can just say thanks for everything and may God give you all the best in return. Special thanks to my wonderful children Riddhi and Vatsal, the true meaning of family. For the past six months, they gave up their favorite thing, the laptop, in favor of completing the thesis and kept on reminding me: "Pappa, take this laptop and write your thesis"!

DA-IICT, Gandhinagar
July, 2010

Prakash P. Gajjar
200521004

Abstract

Digital image processing has exhibited a tremendous growth during past decades in terms of both the theoretical developments and applications. At present, image processing and computer vision are the leading technologies in a number of areas that includes digital communication, medical imaging, the Internet, multimedia, manufacturing, remote sensing, biometrics and robotics. The recent increase in the widespread use of cheaper digital imaging terminals such as personal digital assistants, cellular phones, digital camera, high definition TV and computers in consumer market has brought with it a simultaneous demand for higher-resolution (HR) images and video. Since the high resolution images and video carry more details and subtle gray level transitions, they offer pleasant views of the pictures and videos on these devices. In commercial and industrial applications the high resolution images are desired as they lead to better analysis, interpretation and classification of the information in the images. High resolution images provide better details that are critical in many imaging applications such as medical imaging, remote sensing, surveillance.

The resolution of the image captured using a digital camera depends on the number of the photo detectors in the optical sensors. Increasing the density of the photo detectors leads to high resolution images. The current hardware approach to capture images with high resolution relies on sensor manufacturing technology that attempts to increase the number of pixels per unit area by reducing the pixel size. The cost for such sensors and related high-precision optics may be prohibitively high for consumer and commercial applications. Further, there is a limitation to pixel size reduction due to shot noise encountered in the sensor itself. Since the current sensor manufacturing technology has almost reached this limit, the hardware approach is no more helpful beyond this limit. One promising solution is to use signal processing approaches based on computational, mathematical, and statistical techniques. Because of the recent emergence of these key-relevant techniques, resolution enhancement algorithms have received a great deal of attention. Super-resolution is an algorithmic approach to reconstruct high resolution image using one or more low resolution images. The main advantages of the approach are that it costs less, it is easy to implement and the existing low resolution imaging systems can be used without any additional expense. The application of such algorithms will

certainly grow in situations where high quality optical imaging systems are too expensive to utilize.

The motion based super-resolution approaches produce a high resolution image using non-redundant information from the multiple sub-pixel shifted low resolution observations. The difficulty in these approaches is the estimation of motion between the low resolution frames at a sub-pixel accuracy. Motion-free super-resolution techniques alleviate this problem by using cues other than motion cue. The additional observations are generated without introducing relative motion among them.

In this thesis, we present learning based approaches for motion-free super-resolution. First we solve the super-resolution problem using zoom cue. The observations of a static scene are captured by varying the zoom setting of a camera. The least zoomed image containing the entire scene is super-resolved at the resolution of the most zoomed image which contains a small area of the entire scene. Generally, the decimation process is modeled as the averaging process and the aliased pixel in the low resolution image is obtained by averaging the corresponding pixels in the high resolution image. However, aliasing depends on several factors such as zooming and camera hardware. This motivates us to estimate the aliasing. Since a part of the scene is available at high resolution in the most zoomed images, we make use of the same to estimate the aliasing on the lesser zoomed observations. The aliasing is estimated using the most zoomed image and the lesser zoomed images. We represent the super-resolved image using Markov random fields (MRF) and obtain super-resolution using maximum *a posteriori* technique. We demonstrate the application of proposed aliasing learning technique to the fusion of remotely sensed image. While experimenting, the MRF prior model parameters were adjusted on trial and error basis. A better solution can be obtained using the parameters estimated from the observations themselves. The estimation of the parameters requires the computation of the partition function. Since it is a computationally intensive technique, we use autoregressive (AR) model to represent the super-resolved image. The AR prior model parameters are obtained from the most zoomed observation. We apply this technique to the fusion in remotely sensed images.

The spatial features of a low resolution image are related to its high resolution version. The analytical representation of the relationship of the spatial features across the scales is difficult. This motivates us to prepare a database of low resolution images and its high

resolution versions all captured using same real camera and use this database to obtain high frequency details of the super-resolved image. We propose wavelet based new learning approach using this database and obtain close approximation to the super-resolved image. The close approximation is used as an initial estimate while minimizing the cost function. We employ a prior model that can adapt to the local structure of the image and estimate the model parameters as well as the aliasing from the close approximation. The proposed approach is extended to super-resolve color images. We learn the details of the chrominance components using wavelet based interpolation technique and super-resolve the luminance component using the proposed approach. We show the results for gray level images and for color images and compare the them with existing techniques.

In most current image acquisition systems in handheld devices, images are compressed prior to digital storage and transmission. Since, the discrete cosine transform (DCT) is the basis of many popular codecs such as JPEG, MPEG and H.26X, we consider the DCT for learning. The use the DCT for learning alleviates the limitations of wavelet based learning approach that it can not recover the edges oriented along arbitrary directions. We propose a learning based approach in the discrete cosine transform domain to learn the finer details of the super-resolved image from the database of low resolution (LR) images and their high resolution (HR) versions. Regularization using the homogeneous prior model imposes the smoothness constraint everywhere in the image and leads to smooth solution. To preserve edges and finer details, we represent the super-resolved image using nonhomogeneous AR prior model and solve the single frame super-resolution problem in regularization framework.

Finally, we readdress the zoom based super-resolution problem using discontinuity preserving MRF prior in order to prevent the distortions across the edges while optimization. We obtain the close approximation for the super-resolved image using the learning based approach and use it to estimate the model parameters and the aliasing. Since the cost function consists of a linear term and a non-linear term, it cannot be optimized using simple gradient descent optimization technique. The global optimization technique such as simulated annealing can be employed. Since it is computationally taxing, we propose the use of particle swarm optimization technique. We show the computational advantage of the proposed approach.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"One picture is worth a thousand words." This proverb correctly expresses the amount of information contained in a picture. Pictures are the most effective form of information representation for mass communication. The tremendous volume of optical information and the need for its processing, storage and transmission paved the way to image processing by digital computers. In past few decades, the image processing has exhibited a remarkable growth and created important technological impact in several areas such as medicine, forensic science, telecommunications, printing technology, TV broadcasting, remote sensing, film industry and many more.

Digital image processing concerns the transformation of an image to a digital format and its processing by digital computers. Image processing covers a broad scope of techniques that manipulates images to meet the needs of numerous applications. These techniques can enhance an image, highlight certain features of an image, improve the resolution of an image, create a new image from portions of other images, embed a secret code in an image for security, restore an image that has been degraded during acquisition or transmission and so on. Since both the input and output of a digital image processing system are digital images the images processing algorithms are classified as *low-level vision* algorithms.

Digital image formation is the first step in any digital image processing application. Generally, an image formation system consists of three subsystems; an optical system, a sensor and a digitizer. The optical system collects the light signal reflected from the real world objects under observation and concentrate the signal to fall on the sensor. The sensor receives this optical signal and converts it to an electrical signal. This elec-

trical signal is fed to the digitizer which finally transforms it to a digital image. During the entire process of capturing a digital image of a scene, each subsystem introduces a degradation or deformation to the digital image. These degradations may include geometrical distortions, nonlinear transformation, decimation, blur, and noise. The image processing techniques attempt to recover the digital image by removing or reducing the deformations and degradations introduced by the subsystems while capturing the image. The mathematical modeling of the image formation system is very important in order to have precise knowledge of the degradation introduced. The accurate modeling of the degradation helps to improve the performance of the image processing system and finally leads to better solution.

In almost all imaging applications, high quality images are demanded. The high quality images that provide more detailed information are crucial in applications such as clinical diagnosis, biometrics, industrial inspection, surveillance, remote sensing and machine vision. The images with more details help accurate localization of the diseased cells in medical images, faster and correct identification of criminals, more pleasing view in an high definition television and better classification and interpretation of the contents of the image. The amount of the details available in an image captured using an image acquisition system depends on the sensor's ability to detect smallest optical signal. The acquisition of images with more information require complex optical system with high quality sensors. Specifically, the imaging systems and optical components necessary to capture high quality images become prohibitively expensive for scientific as well as commercial applications. Further there exist a technical limit on manufacturing the sensors that can give the best quality images rich in the finer details. The current image sensor manufacturing technology has almost reached this limit. Thus there is a clear need for finding a relatively inexpensive and an easy to implement solution to this problem using signal processing tools. The applications of such algorithms will certainly be useful in situations where high quality optical imaging systems cannot be incorporated or are too expensive to utilize.

1.1 What is Image Resolution?

The term '*resolution*' refers to the smallest measurable physical quantity. It is used to quantify the quality of various physical instruments. The high resolution of an instrument enables one to measure the quantity with more precision. The resolution of an imaging system is defined as the ability of the system to record fine detail in a distinguishable manner [6]. The term '*image resolution*' can be defined as the smallest measurable detail in a visual presentation. In image processing, it is measure of the amount of detail provided by an image or a video signal. The term image resolution is classified into different types.

- *Spatial resolution:* A digital image is represented using a set of picture elements. These picture elements are called 'pixels' or 'pels'. A pixel at any location in an image carry the information regarding the image intensity at that location in the image. An image represented using a large number of pixels conveys more information as compared to the same image when represented using less number of pixels. The spatial resolution refers to the spacing of the pixels in an image and is measured in pixels/inch (ppi). High spatial resolution allows for sharp details and fine intensity transitions across all directions. The representation of an image with sharp edges and subtle intensity transition by a spatially less dense set of pixels, gives rise to blocky effects. On other hand, the images with spatially dense set of pixels enable the viewer the perception of finer details and offer pleasing view.

- *Brightness resolution:* Pixels carry information of the image intensity in form of binary digits called 'bits'. The intensity at any location in a real world scene may vary from zero to infinity. However in digital image it is not possible to represent this entire range. In practice this range is divided into a finite levels and the real world intensity is quantized and assigned the nearest finite level. The brightness resolution refers to the smallest change in brightness that can be represented in an image. Each brightness level is assigned a binary code. The increase in the brightness resolution requires more number brightness levels and hence more number of bits for each level. A binary image has two levels; black and white, hence requires only one bit for each level. A gray scale image is usually quantized using 256 grey levels with each level represented using 8 bits. Similarly, if each color plane of an RGB image

requires 8 bits then at least 24 bits are needed for representing each pixel.

- *Spectral resolution:* Spectral resolution refers to the frequency or spectral resolving power of a sensor and is defined as the smallest resolvable wavelength difference by the sensor. The spectral resolution plays important role in satellite imaging. High spectral resolution images captured by remote sensing camera provide more detailed information about mineral resources and geographical structures of the earth or any other planet under observation.

- *Temporal resolution:* The term temporal resolution is related to video signals. A video of an event is a sequence of images (frames) captured at regular and short time interval between them. Temporal resolution, also known as frame rate, is the measure of the capability of displaying smallest movement/ motion of the moving objects in the video. Thus it refers to the number of frames captured per second. A video captured with low temporal resolution exhibits abrupt and discontinuous transitions of the moving objects in the scene/event. With high temporal resolution, the movement of the moving objects appears smooth and continuous. For a given duration of time, a high temporal resolution video requires more memory for storage and large bandwidth for transmission. The standard television signals use a typical frame rate of 25 frames per second.

In this work we address the problem of increasing the spatial resolution of given low spatial resolution images. In the rest of the thesis the term resolution is explicitly used to mean spatial resolution unless specified otherwise.

1.2 Limitations of Optical Imaging Systems

The optical sensor is the most important component of any digital imaging system. This sensor convert optical energy into an electrical signal. In modern digital cameras, charge coupled devices (CCD) and CMOS sensors are widely used to capture digital images. These sensors consist of an array of photo-detectors. Each of these detectors generates an output voltage signal proportional to light falling on it. The spatial resolution of an image captured using a camera is determined by the number of photo-detector elements in the sensor. A sensor with less number of photo-detectors samples the scene with a low

sampling frequency and causes aliasing effect. Such sensors produce low resolution image with blocky effect. The direct solution for enhancing spatial resolution of the image is to increase the number of photo-detectors in the sensor. As the number of photo-detectors increases in the sensor chip, the size of the chip increases. This leads to an increase in capacitance [7]. Since, the increase in the capacitance causes limitation on charge transfer rate, this approach is not considered effective. Alternate to this solution is to increase the photo-detector density by reducing the size of photo-detectors. As the photo-detector size decreases, the amount of light falling on each photo-detector also decreases. At certain size of the detector, this amount of light reaches to such a low level that light signal is no more prominent as compared to the noise. This generates shot noise that degrades the image quality severely [8, 9]. Thus there exists a lower limit on reducing the size of photo-detector element. The optimal size photo-detector to generate the light signal without suffering from the effects of shot noise is estimated at about 40 μm^2 for a 35 μm CMOS process. The current image sensor technology has almost reached this level. Therefore, new approaches towards increasing spatial resolution are required to overcome the inherent limitations of the sensors and optical imaging systems. The high cost for high precision optics and image sensors is also an important factor in many commercial applications. Hence, a promising approach is to use algorithmic approaches based on digital signal processing techniques to construct a high resolution image from one or more available low resolution observations.

1.3 Super-resolution Reconstruction

In the process of capturing an image using a digital image acquisition system, there is a natural loss of spatial resolution. There are many factors that contribute to this effect. These include optical distortions such as out of focus, diffraction limits; motion blur due to limited shutter speed or relative movement between camera and the object; noise occurring within the sensor or during transmission and insufficient sensor density. The resulting image usually degraded due to blur, noise and aliasing effects. Figure 1.1 illustrates the degradations introduced in the image at various stages of the image capturing process. Aliasing occurs when the image is sampled at low spatial sampling rate and it causes the distortions in the high frequency contents of the image. Resolution

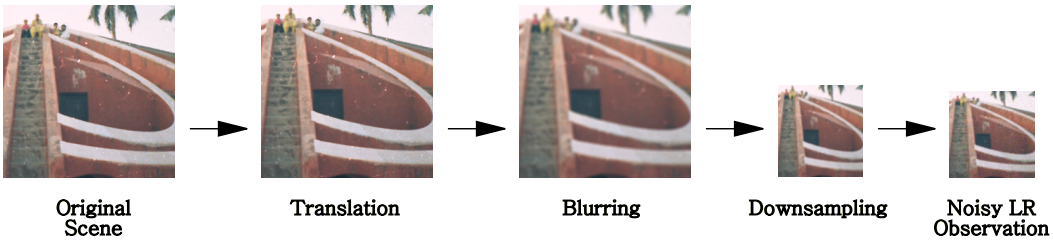


Figure 1.1: Loss of information during the process of recording an image using a digital camera. Here, LR stands for low resolution.

improvement by applying tools from digital signal processing technique has been a topic of great interest. The term '*super-resolution*' refers to such signal processing techniques that reconstruct high spatial resolution image from one or more low resolution images. The goal of super-resolution techniques is to recover the high frequency content that lost during image acquisition process. In effect, although the main concern of the super-resolution algorithms is to reconstruct high resolution images from undersampled low resolution observations, it produce high quality images from noisy, blurred and degraded images. The word 'super' in super-resolution represents very well the characteristics of the technique overcoming the inherent resolution limitation of low resolution imaging systems. The advantages of super-resolution approach are that the existing low resolution imaging systems can be still utilized without any additional hardware and it costs less and offers flexibility.

The Super-resolution reconstruction problem is closely related to image restoration problem. The goal of image restoration is to recover an image from degradations such as blur and noise, and is does not increase the size of the image. Thus, for image restoration problem, the size of the restored image is the same as that of the observed image while it is different in image super-resolution depending on the decimation factor of the super-resolved image. Image interpolation is an another problem related to super-resolution. It increases the size of image using a single and aliased low resolution observation. Since the single image can provide no more non-redundant information, the quality of the interpolated images is very much limited. Some of the interpolation methods convolves image with a filter designed to boost high frequency contents. The drawback of such methods is that it also amplifies any noise in the image and degrades the quality. The quality of an image interpolated from an aliased low resolution image is inherently limited even though the ideal sinc function is employed. The single image interpolation techniques

cannot recover the high frequency components lost or degraded during low resolution sampling process. For this reason, image interpolation methods are not considered as super-resolution techniques.

1.3.1 Observation Model

In order to apply a super-resolution algorithm, a detailed understanding of how images are captured and of the transformations they undergo is necessary. The common digital image acquisition systems (digital camera, video camera/camcorders) consist of focusing lens, optical sensors, processor chip, electronic circuits and other mechanical subsystems. In the process of capturing an image of a scene using such camera, the high resolution image goes through a sequence of degradations, including a blur, down-sampling, and additive noise. There is a natural loss of spatial resolution caused by optical distortions such as insufficient sensor density, out-of-focus and diffraction limit. The observations may be blurred due to causes like optical aberration, relative motion between camera and object, limited shutter speed and atmospheric turbulence. Furthermore, the images could be degraded by various types of noise occurring within the sensor or during transmission. The frames captured using video camera may be rotated and scaled due to camera motion like zooming, panning, tilting. In this case, blur may also be introduced due to relative motion between the observations. Thus the observed images are degraded versions of the high resolution images. In order to analyze the super-resolution reconstruction problem, it is required to formulate a mathematical model that represents the image acquisition process. This model, known as observation or forward model, relates the original high resolution image to the observed low resolution image(s). The correct formulation of the observation model plays an important role in the success of any super-resolution approach. The most commonly used forward models for super-resolution reconstruction incorporate translation, blur, aliasing and noise in the formulation. A typical forward model is shown in Figure 1.2.

1.3.2 Super-resolution: An Ill-posed Inverse Problem

The Super-resolution algorithms attempt to reconstruct the high resolution image corrupted by the limitations of the optical imaging systems. The algorithms estimate the

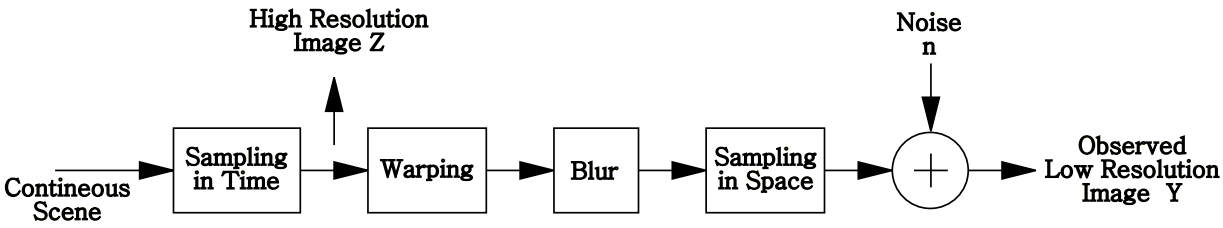


Figure 1.2: A typical image formation model showing relationship between low reolution image and high resolution image.

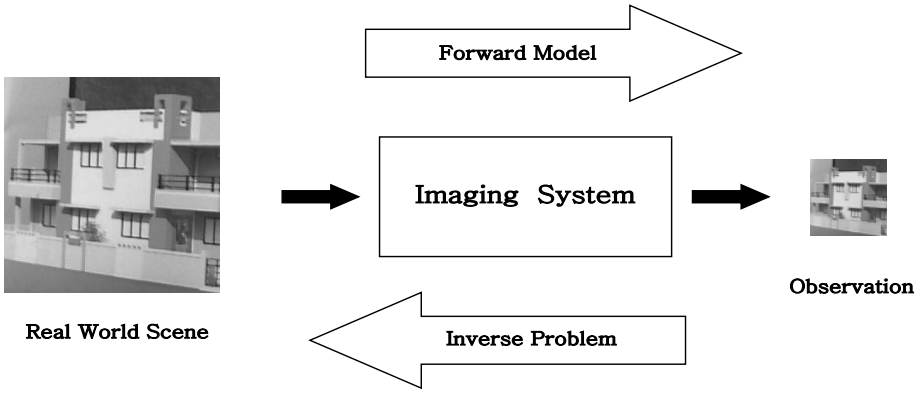


Figure 1.3: Schematic representation of inverse problem. The forward model is a mathematical description of the image degradation process. The inverse problem addresses the issue of reconstructing the original scene from one or more observations.

high resolution image from the one or more degraded low resolution images. This is an inverse problem wherein the original information is retrieved from the observed data. A schematic representation of the inverse problem is shown in Figure 1.3. Solving the inverse problem requires the inverting the effects of the forward model. It is difficult to invert the forward model without amplifying the noise in the observed data. Since, in super-resolution, the forward model of high resolution (HR) to low resolution (LR) transformation is reduced to matrix manipulations, it is logical to formulate the restoration problem as matrix inversion. The problem is worsened by the fact that the model matrix is singular. For a singular matrix, there is an infinite space of solutions. In other words, there exist infinite high resolution images which are consistent with the original data. Thus the super-resolution problem in an ill-posed inverse problem. While solving the ill-posed inverse problems, knowing the forward model alone is not sufficient to obtain satisfactory results. Some form of constraints on the space of solutions must be included. Procedures adopted to stabilize the inversion of ill-posed problems is called regularization. The regularization based approach solves the ill-posed inverse problems by making them

better-posed using the prior information about the solution. It is a systematic method for adding more information to the reconstruction system. Bayesian super-resolution reconstruction approach is commonly employed for solving ill-posed inverse problems. This method is used when a posterior probability density function of the original image can be established. Bayesian estimation distinguish between possible solutions by using a priori image model. The major advantages of the Bayesian super-resolution approach are its robustness and flexibility in modeling noise characteristics and a priori knowledge about the solution. Since, the convex cost function ensures the uniqueness of the solution, an efficient gradient descent method can be used to estimate the high resolution image.

1.4 Applications of Super-resolution

Images with high spatial resolution are desired in many imaging applications. High resolution images leads to better analysis, classification and interpretation. Synthetic zooming of region of interest [10] is an important application in medical, forensic, surveillance, and satellite imaging. The medical industry has long been a user of image processing technology. The images captured using computer aided tomography, magnetic resonance imaging, positron emission tomography etc are stored in memory. These can be enhanced using super-resolution so that doctors can focus on the area of interest and make better diagnosis of diseases. In forensic investigations it is often required to magnify objects in the scene such as the fingerprints or face of a criminal. Super-resolving the areas of interest in such images facilitates the investigation process and finally leads to quick identification of criminal. In video surveillance, when any unfair incident happens, the analysis of the stored video is performed to locate anything suspicious. It is possible that resolution of the region of interest in the scene of the incident may not be sufficient for proper interpretation. In this case the frames can be super-resolved to obtain a good description of the incident [11]. In satellite imaging applications such as remote sensing, the super-resolution technique can be considered to improve the resolution of the land area. The satellite images are widely used to estimate the production of the crops in regional land area. The super-resolve images of land fields can lead to better classification of crops and hence more accurate estimation of each crop. The super-resolve images of geographical land area help in better segmentation of regions containing forests, rivers,

roads and other geographical structures. In military applications, images taken by aircraft, over geographical areas of interest, provides the warfighter with the most up-to-date information. The data can be used to determine potential targets, to locate and determine the number of vehicles and/or buildings in an area, as well as a variety of other information. The accuracy of this information depends upon the resolution of the images. Super-resolution approaches can be helpful to obtain higher resolution images that lead to more accurate information. Holography is a widely used technique for recording and reconstructing three-dimensional information of an object. Digital holography technique uses a CCD camera to record holographic patterns. The low resolution of CCD sensors significantly limits the size of objects that can be recorded. To overcome this problem, super-resolution image reconstruction can be used to increase the resolution of holograms and eliminate the aliasing effects caused by undersampling [12]. By doing so, larger objects can be recorded. Recently, high definition television (HDTV) are rapidly getting popular due to superior perceptual quality of the displayed video. The HDTV signals are high resolution video signals generated by compliant video acquisition systems. A non-HDTV signal fed to HDTV results in degraded picture with visual artifacts. At present, there is a clear need to display a non-HDTV signals such as PAL, NTSC signals on the HDTV without visual artifacts. Scaling lower-quality signals to fit a TV's higher-resolution screen is often called upconversion. The super-resolution technique can be applied for upconversion from a non-HDTV (low resolution) video to an HDTV video [13, 14]. Similarly, multimedia terminals such as digital cameras, cellular phones, PDAs (personal digital assistants), and computers are prevalent these days. People would like to capture, manipulate, and display multimedia content more flexibly with heterogeneous terminals so that this content can be shared with friends and peers effortlessly. For this purpose, format conversion and content adaptation techniques are needed to compensate for the differences in these devices [13]. The super-resolution techniques can be employed in such conversion so that multimedia contents captured by portable devices can be viewed on higher resolution terminals such as PC or TV screens at home. Some of the other applications include ultrasound image enhancement [15, 16], text enhancement for optical character recognition [17, 18, 19], digital video enhancement [20, 21], salient stills [22] and remote sensing image enhancement [23, 24].

In many applications it may not be feasible to capture the high resolution images even

if the camera is capable of it. In low cost applications, the super-resolution technique can be used for saving resources such as memory, bandwidth, power and cost. Consider a wild life sensor network. The high resolution imagery in this application requires installation of large number of high end cameras at different locations in forest. The transmission of the large volume of data generated by these cameras need large transmission bandwidth as well as large amount of power. Super-resolution technique can help in such situations. Low cost cameras can be employed to capture the video/images and the data can be transmitted over the channel having limited bandwidth. At base station super-resolution techniques can be applied to the received images to reconstruct high resolution images.

1.5 Contributions of the Thesis

Super-resolution algorithms attempt to extract non-redundant information from different sources and fuse the same in the observed data to obtain super-resolved image. Motion cue based approaches extract non-redundant information from multiple observations that have sub-pixel shift between them. These approaches require the task of registration of the observations. The quality of the super-resolved image is highly dependent on the accurate registration. This task is very complex and is computationally intensive. Researchers have proposed approaches that use cues other than motion cue and avoid the problem of accurate registration. The zoom cue based approaches use observations captured at different zoom setting of the camera and extract non-redundant information from the zoomed observations. We in this thesis propose new learning based techniques for motion-free super-resolution. The contributions of this thesis are summarized as under.

- The image formation process used in super-resolution problem is generally modeled using a linear model. The decimation process in this model is represented by a matrix having a few entries that have constant and equal values. This decimation matrix simulates a low resolution pixel intensity as an average of the intensities of corresponding pixels in the high resolution image. However, in practice, intensity at a pixel in the low resolution image depends on various factors such as camera gain, illumination condition, zoom factor, noise etc. Hence the aliased low resolution pixel intensity of an image point is not always equally weighted sum of the high resolution

intensities. We propose to estimate the aliasing (decimation) from the observations captured using different zoom settings. We then address the zoom based approach for super-resolution using decimation estimation. The multiple observations of a scene are captured by varying the camera zoom setting. The image captured at the least zoom setting covers the entire scene area. It is sampled with low sampling rate and is represented by limited number of pixels. The most zoomed image, captured at highest zoom setting, covers a small portion of the scene. This portion is represented by relatively large number of pixels and sampled with high sampling rate. Since we capture the images at different resolutions using zoom camera and the most zoomed observation is assumed to be alias free, we learn the aliasing from the most zoomed observation. We then solve this ill-posed inverse problem using regularization technique. We model represent the super-resolved image using Markov random fields and arrive at convex cost function. We optimize it using simple gradient descent technique.

The MRF model prior involves very high computational complexity as one requires to compute the partition function in order to estimate the true parameters. This motivates us to employ a computationally less taxing model for representing the spatial dependencies. We consider the linear dependency of a pixel in a super-resolved image to its neighbors and represent the same using autoregressive process model. We estimate the AR parameters from the most zoomed observation and use them while regularization.

We then show the application of this approach to multiresolution fusion in remotely sensed images. Since, the panchromatic image is captured at high spatial resolution and the multispectral images are captured at low spatial resolution, we use panchromatic image to learn aliasing on the multispectral images.

- In many applications, the multiple observations of the image to be super-resolved are not available. We propose a fast approach to the learning based single frame super-resolution where only one observation is used to construction the super-resolution. The learning based approaches for single frame super-resolution proposed in literature use a database consisting of high resolution images only. Low resolution and high resolution image pairs are generated by down sampling the high resolution

images. The correspondence between the spatial features in the observation and its super-resolved version cannot be effectively learned from such database. We construct a database consisting of the low resolution images and their high resolution versions all captured using a real camera. We capture the images of a scene at different resolution by varying the zoom setting of the camera. These pairs of images truly represents the spatial features correspondence across the varying resolutions. We learn the spatial features of the super-resolved image from this database and obtain the close approximate to the super-resolved image. As real images contain smooth areas, textures and edges, they may not be efficiently represented by homogeneous models. This motivates us to consider an inhomogeneous prior which can adapt to the local structure of the image and provide a better reconstruction. We use an inhomogeneous Gaussian MRF to represent the image field. It is adaptive to the local structures in an image and eliminates the need for two separate priors i.e. edge preserving prior and smoothness prior.

- We propose a new learning approach based on the discrete cosine transform (DCT) and apply it to single frame super-resolution. The motivation behind this is that the DCT is the basis of many popular video codec such as JPEG, H.263 and MPEG. Consequently, the low resolution frame does not need to be decoded prior to super-resolving a video. The advantage of the DCT over wavelet transform is that it helps to preserve the edges in arbitrary directions unlike the wavelet transform that captures discontinuities in limited directions only. We learn the high frequency DCT coefficients of the super-resolve image from the database and obtain the close approximation to the super-resolution. In this experiment we employ nonhomogeneous AR model to represent the image field and alleviate the drawbacks of the homogeneous models for the prior.

- Finally, we readdress the problem of zoom based super-resolution using the new learning based approaches and particle swarm optimization (PSO) technique. We obtain the close approximation of the super-resolved image by learning the details from the database consisting of low resolution images and their high resolution versions. In order to preserve the edges, we incorporate discontinuity preserving prior and arrive at non-convex cost function. The optimization of such cost function

requires the computationally taxing tools such as simulated annealing. We employ PSO technique for optimizing the non-convex cost function consisting of a data fitting term and a discontinuity preserving prior term. We obtain the particles using existing interpolation techniques as well as using the proposed learning based approaches. The effectiveness of the proposed approach is shown using mean squared error as well as using another recently proposed measure called structural similarity (SSIM) which is derived based on the human visual perceptual capability.

1.6 Organization of the Thesis

In this thesis, we address the problem of motion-free super-resolution using new learning based approaches. We develop learning based techniques for super-resolution using zoom cue as well as single frame super-resolution. We propose a technique to estimate the aliasing on the low resolution observation and use the learnt aliasing in the forward model. We develop new learning based techniques using discrete wavelet transform and discrete cosine transform and obtain close approximation to the super-resolved image. We solve the super-resolution problems using regularization based approaches and suggest use of different homogeneous and nonhomogeneous models for prior. We propose a novel application of particle swarm optimization tool in the super-resolution problem. The thesis is organized as follows.

Many researchers have attempted super-resolution problem using different methodologies such as super-resolution using motion cue, motion-free super-resolution and example based techniques. chapter 2 provides a review of the existing super-resolution approaches proposed in literature.

In chapter 3, we address the problem of super-resolution using zoom cue. We suggest a novel technique for estimating aliasing from the observations and obtain super-resolution using MAP-MRF framework. In order to avoid the computational complexity of the MRF model for prior, we use autoregressive model. Along with the aliasing we learn the AR model parameters from the most zoomed observation and regularize the solution. We further show the application of estimating aliasing on multispectral (MS) images in the remotely sensed images.

We introduce a new learning technique based on the discrete wavelet transform (DWT)

and propose a single frame super-resolution approach in chapter 4. We construct a new database consisting of low resolution images and their high resolution versions and use it to capture the relationship of spatial features among the images at different resolutions. We reconstruct the super-resolved image using an inhomogeneous Gaussian Markov random field. We derive prior model parameters from the close approximation obtained using the DWT based learning technique.

We describe the super-resolution technique for a single frame super-resolution using non-homogeneous autoregressive prior in chapter 5. We develop a new learning technique based on the DCT to obtain a close approximation to the super-resolution image and obtain super-resolution using regularization framework.

In chapter 6, we readdress the zoom based super-resolution with discontinuity preserving MRF model prior and propose a novel and fast approach for super-resolution using particle swarm optimization as a tool for minimizing non-convex cost function. In PSO, the optimization process begins with many initial solutions known as particles. We obtain these particles using different techniques including the learning techniques based on the DWT and the DCT. In this chapter, the quality of a super-resolved image is defined as the similarity of the super-resolved image with the original high-resolution image. We use structural similarity index and mean squared error to measure the quality of results.

Finally, we summarize our work and conclude in chapter 7. We discuss the further challenges and directions for future research.

Chapter 2

Literature Review

Images captured using a camera having CCD or CMOS image sensors, provide good quality images. Although, these sensors are suitable for the most imaging applications, the current resolution level will not satisfy future demand. The most direct solution to obtain high resolution images is to increase the density of the photodetectors in the optical sensor by improving the sensor manufacturing technique. However, due to hardware cost and fabrication technological limitations, it is expensive and difficult to acquire the image with higher resolution than that with the current resolution level of the sensors. Thus not possible to increase the current resolution level by improving hardware performance. The magnification of an image captured with the current resolution level of the sensors introduces visible artifacts. Resolution improvement by applying tools from digital signal processing technique has, therefore, been a topic of very great interest [25, 26, 27, 28, 29]. Super-resolution (SR) is an algorithmic technique that produces an image or video with a resolution higher than those of any of the input images or frames. Super-resolution has attracted a growing interest as a purely computational means to increase imaging sensors performance. The pioneer work of super-resolution reconstruction may go back to 1984 by Tsai and Huang [30]. Since then a variety of approaches for solving the super-resolution problem have been proposed [31, 32, 33]. The approaches include deterministic regularization approaches, stochastic methods, nonuniform interpolation, projection onto convex sets, iterative backprojection, adaptive filtering and learning based approaches. The difference among these approaches is subject to what type of the reconstruction method is used, which observation method is used, in which particular domain is used, how the observations are captured and so on. We categorize SR reconstruction methods into three

main divisions: motion based SR reconstruction, motion free super-resolution and single frame super-resolution. Motion based technique use the relative motion between different low resolution observations as a cue in estimating the high resolution image, while motion free super-resolution techniques may use cues such as blur, zoom, and defocus. Single frame super-resolution approaches attempt to reconstruct the super-resolution using a single low resolution observation. Since, only a single undersampled and degraded input image is available, the task of obtaining a super-resolved image comprises of recovering the additional spatial data from available database of high resolution images.

2.1 Motion Based SR Reconstruction

Motion based SR approaches obtain super-resolution from several subsampled and mis-registerd low resolution images of the desired scene. The low resolution images can be either obtained as a sequence taken over a time, or taken at the same time with different sensors. In particular, camera and scene motion lead to multiple observations containing similar, but not identical information. Since, the observations have subpixel shifts and are aliased, new information in each of the observations can be exploited to construct a high resolution image. The most of the SR approaches proposed consist of three stages: registration, nonuniform interpolation and restoration [34, 35, 36]. In the registration, motions between observations are estimated with sub-pixel accuracy [37, 38, 39]. Accurate estimation of motion is an important factor in the success of these approach. The interpolation based approach is the most intuitive method for super-resolution. This approach takes relatively low computational load and makes real time applications possible. However, since the errors at the interpolation process is not accounted for during the de-convolution, it does not guarantee an optimal solution. Furthermore, this approach applies only to the case when the blur and the noise effects are constant over the lower resolution images. Hence, the use of degradation models is limited in this approach.

The frequency based approach reconstruct HR images using the aliasing in each of the LR images. The relationship between LR images and the HR image is demonstrated in the frequency domain. Frequency domain methods are based on three fundamental principles: i) the shifting property of the Fourier transform (FT), ii) the aliasing relationship between the continuous Fourier transform (CFT) and the discrete Fourier

transform (DFT), iii) the original scene is band-limited. These properties allow the formulation of a system of equations relating the aliased DFT coefficients of the observed images to samples of the CFT of the unknown scene. These equations are solved yielding the frequency domain coefficients of the original scene, which may then be recovered by inverse DFT. Formulation of the system of equations requires knowledge of the translational motion between frames to sub-pixel accuracy. Each observation image must contribute independent equations, which places restrictions on the inter-frame motion that contributes useful data. Tsai and Huang [30] assume that the desire HR image is bandlimited and derive analytical relationship between observations and the desired HR image. They obtain super-resolution using shifting property of the Fourier transform and the aliasing relationship between continuous Fourier transform of HR image and the discrete Fourier transform of the observed LR images. In [40], authors extend this approach using weighted least squares formulation for blurred and noisy images having same blur and noise characteristics. Other approaches using frequency domain techniques include [41, 42, 43]. Approaching the super-resolution problem in the frequency domain makes a lot of sense because it is relatively simple and computationally efficient. The capability of parallel implementation of these techniques makes the hardware less complex. However, there are some problems with a frequency domain formulation. It restricts the inter-frame motion to be translational because the DFT assumes uniformly spaced samples. The observation model is restricted only to global translational motion. Another disadvantage is that prior knowledge that might be used to constrain or regularize the super-resolution problem is often difficult to express in the frequency domain. In other words, the frequency domain approach make it difficult to apply spatial domain prior for regularization. Since the super-resolution problem is fundamentally ill-posed, incorporation of prior knowledge is essential to achieve good results.

A variety of techniques exist for the super-resolution problem in the spatial domain. These solutions include interpolation, deterministic regularized techniques, stochastic methods, iterative back projection, and projection onto convex sets among others. The primary advantages to working in the spatial domain are support for unconstrained motion between frames and ease of incorporating prior knowledge into the solution. Non-uniform interpolation approach maps pixels from the low-resolution images onto a common plane and then interpolate over a more finely sampled grid to obtain a high resolution

image. Unfortunately, this technique generally works very poorly because of the inherent assumption that camera sensors do not act as impulse functions, but instead spatially average the incident light across each pixel. Since the solution obtained using non-linear interpolation lacks high frequency recovery, it looks only marginally better than bilinear interpolation. Since there are limited number of observed images and ill-conditioned blur operators, the super-resolution problem is an ill-posed inverse problem. When presented with an ill-posed problem it becomes necessary to impose prior knowledge on the solution space in order to obtain a unique solution. Regularization procedures are adopted to stabilize the inversion of ill-posed problems.

Reconstruction-based algorithms can be classified in two categories: deterministic approach and stochastic approach. Deterministic approach encodes knowledge of what the highresolution image should look like as priors and regularize the solution using constrained least squares method [44, 45, 46, 47]. The common approach is to impose a smoothness prior via regularization on top of a least-squares optimization. The presence of the regularization term guarantees a convex and differentiable optimization function. Thus, a unique optimal solution can be computed using a number of standard methods like gradient descent. The results of this approach are certainly an improvement over the low-resolution image, but enforcing smoothness is not always the best option, especially if other priors can be formulated that preserve high-frequency details better.

Stochastic methods which treat SR reconstruction as a statistical estimation problem have rapidly gained prominence since they provide a powerful theoretical framework for the inclusion of a-priori constraints necessary for satisfactory solution of the ill-posed SR inverse problem. The statistical techniques explicitly handle prior information and noise. Inclusion of prior knowledge is usually more natural using a stochastic approach. The stochastic SR reconstruction using Bayesian approach provides a flexible and convenient way to model a priori knowledge about the final solution. This method can be applied when a posteriori probability density function of the original image can be estimated. The Maximum *a posteriori* (MAP) approach to estimating the super-resolution seeks the estimate for which the *a posteriori* probability is a maximum. It is common to utilize Markov random field (MRF) image models as the prior term. Under typical assumptions of Gaussian noise the prior may be chosen to ensure a convex optimization enabling the use of descent optimization procedures.

Schultz and Stevenson propose SR reconstruction from LR video frames using MAP technique [48]. They employ discontinuity preserving Huber-Markov Gibbs prior model and use constrained optimization. The authors in [49] use the MAP framework for jointly estimating the registration parameters and the high-resolution image from severely aliased observations. They try to dealias the low resolution images by utilizing the phase difference among the low resolution images. Elad and Feuer in [50] propose a unified methodology to super-resolve an image from several observations which are geometrically warped, blurred, noisy and downsampled. They combine maximum likelihood, MAP and projection onto convex sets approaches. Capel and Zisserman [51] have employed fusion of information from several planer views for mosaicing and super-resolution. Recently, the authors in [52] propose a joint MAP formulation combining motion estimation, segmentation, and super-resolution together. They solve the super-resolution problem by a cyclic coordinate decent process that treats the motion and the segmentation fields as well as the HR image as unknowns and estimates them jointly using the available data. In the letter [53], the authors present an approach to reconstruct high spatial-resolution and high-dynamic-range images from multiple and differently exposed images simultaneously. They propose a stochastic super-resolution reconstruction algorithm that models nonlinear camera response function, exposure time, sensor noise, and quantization error in addition to spatial blurring and sampling. The authors in [54] propose a technique to enhance license plate numbers of moving vehicles in real traffic videos. They obtain a high-resolution image of the number plate by fusing the information derived from multiple, sub-pixel shifted, and noisy low-resolution observations. They model the super-resolved image as a Markov random field and estimate it using a graduated non-convexity optimization procedure. The article [55] reviews a variety of super-resolution methods. Farsiu *et al.* [56] propose a unified approach of demosaicing and super-resolution of set of low resolution color images. They employ bilateral regularization of the luminance term for reconstruction of sharp edges, and that of the chrominance term and intercolor dependencies term to remove the color artifacts from the HR estimate. They use L_1 norm for the data error term to make the method robust to errors in data and modeling. The authors in [57] propose a 3D structure preserving super-resolution technique using disparity map. Most common methods for color image super-resolution involve application of super-resolution algorithm to each of the color components independently or

transform the problem to a different color space where chrominance layers are separated from luminance and super-resolution is applied to luminance component only. The major advantages of the stochastic approaches are robustness and flexibility in modeling noise characteristics, incorporation of priori knowledge of the solution in formulation, applicability of the simple optimization technique with the assumption of white Gaussian noise process and simultaneous estimation of motion and HR image field.

Set theoretic methods, especially the method of projection onto convex sets (POCS), are popular as they are simple, utilize the powerful spatial domain observation model, and allow convenient inclusion of *a priori* information. In set theoretic methods, the space of SR solution images is intersected with a set of (typically convex) constraint sets representing desirable SR image characteristics such as positivity, bounded energy, fidelity to data, smoothness etc., to yield a reduced solution space. POCS refers to an iterative procedure which, given any point in the space SR images, locates a point which satisfies all the convex constraint sets. Main problems with the POCS approach are non-uniqueness of solution, difficulty in defining the projections onto convex sets, dependence of the solution on the initial guess, slow convergence and high computational cost. POCS reconstruction methods have been successfully applied to sophisticated observation and degradation models [58, 59].

Irani and Peleg [60] proposed a super-resolution algorithm based on iterative back projection (IBP). The key idea is that the error between the observed low-resolution images and the corresponding low-resolution images formed using an estimate of the SR image can be used to iteratively refine the estimated SR image. This approach begins by guessing an initial HR image. This initial HR image can be generated from one of the LR images by decimating the pixels. This initial HR image is then downsampled to simulate the observed LR images. The simulated LR images is subtracted from the observe LR images. If the initial HR image was the real observed HR image, then the simulated LR images and the observed LR images would be identical and their differences zero. Hence, the computed differences can be "back-projected"' to improve the initial guess. The back-projecting process is repeated iteratively to minimize the difference between the simulated and the observed LR images, and subsequently produce a better HR image. While, iterative back projection is relatively easy to understand, the method does not directly address the ill-conditioning of the problem and incorporation of a priori

constraints is difficult.

The primary factor that controls the quality of the super-resolved image is the extremely precise alignment of the low-resolution frames. Park et al.[61] has shown by example that small error in registration can considerably effect the super-resolution results. Many researchers have proposed algorithms for registration of the images [62, 63, 64, 65]. A survey of different registration methods is provided in [66]. Most of the registration algorithms tend to be sensitive to illumination, blur variations and noise. The authors in [67], propose registration algorithm that uses the local phase information, which is robust to the above degradations. They derive the theoretical error rate of the estimates in presence of non-ideal band-pass behavior of the filter and show that the error converges to zero over iterations. Approaches that use frequency domain processing to compute the registration parameters are relatively stable under various image artifacts. However, they are limited in the class of transformations that can be estimated between two images [64]. The registration algorithm in [65] deal with the illumination variation at the SR phase and assumes accurate registration. Solutions that deal with registration error by treating it as noise [68] during the SR phase, and a combined optimization of SR and registration [49] have been tried. In [69], Suresh and Rajgopalan present a computationally efficient method for super-resolution using a discontinuity adaptive MRF model to provide robustness to errors in motion and blur estimates. A discontinuity adaptive regularizer is proposed in which the degree of interaction between pixels across edges is adjusted adaptively to preserve discontinuities.

2.2 Motion-free Super-resolution

All the above approaches use motion as a cue for solving the super-resolution problem. These approaches require accurate registration between the low resolution observations. Establishing a dense point correspondence among the observations is an important step and a prerequisite for successful SR implementations. Any error in establishing the correspondence among the observations affects the quality of super-resolution. It is a difficult and computationally taxing procedure. It is shown that the super-resolution is also possible from the observations captured without relative motion between them. There has been work on the spatial resolution enhancement by using cues which do not

involve a motion among low resolution observations. The new approaches based on the cues other than motion such as blur, defocus[70] and zoom[1, 71] are known as motion-free super-resolution approaches. The authors in [72] describe an MAP-MRF based super-resolution technique for SR reconstruction from several blurred and noisy low resolution observations. In [73], the authors recover both the high-resolution scene intensity and the depth fields simultaneously using defocus cue. Rajagopalan and Kiran in [70] propose a frequency domain approach for SR reconstruction using the defocus cue. They also show that the estimation of the HR image improves as the relative blur increases.

In [74, 75] the authors show the estimation of super-resolved image and depth map using photometric cue. They model the surface gradients and albedo as the Markov random fields and use line fields for discontinuity preservation. Since they use simulated annealing for minimization, the approach is computationally very taxing. Joshi and Chaudhuri in [1] demonstrate the use of zoom cue for super-resolution. The obtain the low resolution observation by varying the zoom setting of a camera and obtain super-resolution using MAP-MRF framework. They also propose a learning based approach for SR reconstruction from zoomed observations [76]. They represent the HR image using MRF prior model and learn the model parameters from the observations. The authors in [77] propose shape from focus method to super-resolve the focused image of 3D objects. Using the observations in the shape from focus stack and the depth map of the object, they reconstruct super-resolution by magnification factors of 2 or greater using a MAP-MRF technique.

2.3 Single Frame Super-resolution

Multi-frame image super-resolution discussed in previous sections refers to the case where multiple images of the scene are available. In general, changes in these low-resolution images caused by camera or scene motion, camera zoom, focus and blur allow one to recover extra data for reconstructing an output image at a resolution above the limits of the original camera or other imaging device. The super-resolved output image captures more of the original scenes details than any one of the input images was able to record.

Single frame super-resolution approaches attempt to obtain super-resolution using a single observation. In comparison to the multiple image case, this problem is more

severely underconstrained as less information about the scene is provided. Furthermore, single frame super-resolution can be more general as it might include magnifying images which do not have underlying ground truth. In this case, estimating sensor characteristics might be less meaningful and the objective becomes the generation of visually plausible images rather then reconstructing the underlying scene. Accordingly, for single frame super-resolution, one has to inevitably rely on very strong prior information. This prior information is available either in the explicit form of a distribution or energy functional defined on the image class [78, 79, 80, 81, 82, 83], and/or in the implicit form of example images which leads to example/learning-based super-resolution [84, 85, 86, 87, 88, 89].

A variety of linear and non-linear tools are available which try to address the problem of single frame super-resolution. A detailed mathematical analysis of regularization based schemes has been provided by Malgouyres and Guichard in [90]. The authors in [91] provide an interpolation method under the total variation regularization scheme. In their paper they start off with a higher resolution image formed by zero-padded interpolation of the LR image. A constrained gradient descent algorithm is presented where the authors minimize the gradient energy of the image which conforms to a linear smoothing and sampling process. The use of TV for super-resolution has also been demonstrated by Aly and Dubois in [92]. In their method they modify the data fidelity term to closely model the assumed image acquisition model. Their iterative algorithm then makes use of the back projection technique introduced by Irani and Peleg [60] for data fidelity in a regularization framework. The authors then present an algorithm that converges to a unique solution irrespective of the starting interpolated image. However, the resultant image depends upon the choice of the image formation model. The dependence of the result on the selection of the proper mathematical model that captures the downsampling process for such regularization based methods has been discussed in [93]. Other approaches using TV for super-resolution are presented in [94]. Jiji et al. in [95] propose an interpolation technique where the aliasing present in the LR image is used. They assume knowledge of the bandwidth and the amount of aliasing in a given observation and use a signal processing approach to perform super-resolution. Vandewalle *et al.* [96, 97, 98, 99] have proposed a fast super-resolution reconstruction based on a nonuniform interpolation using a frequency domain registration. This method has low computation and can be used in the real-time system but the degradation models are limited therefore this algorithm can

apply on few applications.

Because the richness of real-world images is difficult to capture analytically, researchers are exploring a learning-based approach for super-resolving images. Given a training set, the learning based algorithm learns the fine details that correspond to different image regions seen at a low-resolution and then uses those learned relationships to predict fine details in other images. A number of learning-based SR algorithms have been studied in [85, 89, 100, 101, 102, 103, 84, 104, 105, 3]. These algorithms use a learning scheme to capture the high-frequency details by determining the correspondence between LR and HR training images. Compared to traditional methods, which basically process images at the signal level, learning-based SR algorithms incorporate application dependent priors to infer the unknown high resolution image. The input LR image is split into either overlapping or non-overlapping patches. Then, for each LR patch from the input image, either one best-matched patch or a set of the best-matched LR patches is selected from the training set. The corresponding HR patches are used to reconstruct the output HR image.

Freeman *et al.* [85] propose an example based super-resolution technique. They estimate missing high-frequency details by interpolating the input low-resolution image into the desired scale and then search the high spatial frequency patches from the database. They embed two matching conditions into a Markov network. One is that the LR patch from the training set should be similar to the input observed patch, while the other condition is that the contents of the corresponding HR patch should be consistent with its neighbors. The super-resolution is performed by the nearest neighbor based estimation of high-frequency patches based on the corresponding patches of input low-frequency image. The authors in [106] present a learning based approach to recognize digits of the vehicle registration plates. They super-resolve and restore the image patches using undirected graphical model. The learning-based image hallucination technique is proposed in [107]. Here the authors use primal sketch priors for primitive layers (edges and junctions) and employ patch-based learning using large image database. In [105], Baker and Kanade proposed a hallucination technique based on the recognition of generic local features. These local features are then used to predict a recognition-based prior rather than a smoothness prior as is the case with most of the super-resolution techniques. The authors in [108] present a learning-based method to super-resolve face images using

a kernel principal component analysis (PCA) based prior model. They regularize the solution using prior probability based on the energy lying outside the span of principal components identified in a higher-dimensional feature space. Ni and Nguyen utilize support vector regression (SVR) in the frequency domain and pose the super-resolution problem as a kernel learning problem [109]. The drawback of SVR is that it increases the computational complexity. Based on the framework of Freeman *et al.* [84], Kim and Kwon investigate a regression-based approach for single-image super-resolution [4]. Here the authors generate a set of candidates for each pixel using patch-wise regression and combine them based on the estimated confidence for each pixel. In the post processing step, they employ regularization technique using discontinuity preserving prior. Brandi *et al.* [110] propose an example-based approach for video super-resolution. They restore the high-frequency information of an interpolated block by searching in a database for a similar block, and by adding the high frequency of the chosen block to the interpolated one. A novel method with manifold learning is proposed in [111]. In this paper, neighbor embedding with training images is adopted to recover the super-resolution image. One disadvantage of the approach is that the recovery of super-resolution image is easily affected by the training image which needs to be selected within related contents manually. Based on Super-Resolution through Neighbor Embedding algorithm Chang and Zhang [112] propose an improved super-resolution approach to choose more reasonable training images using histogram matching. Jia et al. combined the robust global models for a face space with the local ones for recovering image details to get a generalized face super-resolution [113]. Karl et al. used support vector regression to learn the relationship of DCT coefficients between the low- and high-resolution images [114].

It is reasonable to assume an edge smoothness prior without any other prior knowledge on the image. This prior is also consistent with human perception, which seems to also favor smooth curves in natural images. Based on this assumption, a lot of algorithms have been proposed to obtain smooth edges. Various techniques are investigated in the literature to obtain smooth image boundaries, such as level-set [115], multiscale tensor voting [116], and snake-based vectorization [117] techniques. An edge smoothness prior is favored since it is able to suppress the jagged edge artifact effectively. However, it is difficult to obtain analytical forms for evaluating the smoothness of the soft edges with gradual intensity transitions. The authors in [118] propose soft edge smoothness

measure that can approximate the average length of all level lines in an intensity image and use it as a prior. They obtain super-resolution using a novel combination of this soft edge smoothness prior and the alpha matting technique for color image SR by adaptively normalizing image edges according to their alpha-channel description. By introducing the soft edge smoothness prior, Dai et al. combined it with alpha matting technique for color image super-resolution [80]. Wei and Yeung [119] proposed an image hallucination using neighbor embedding over visual primitive manifolds. In their method, the maximum filter responses are extracted as features. In [120] authors propose a Neighbor embedding based super-resolution through edge detection and Feature Selection (NeedFS). They propose a combination of appropriate features for preserving edges as well as smoothing the color regions. The training patches are learned with different neighborhood sizes depending on edge detection. In [121], the authors address the problem of super-resolution from a single image using multi-scale tensor voting framework. They consider simultaneously all the three color channels to produce a multi-scale edge representation to guide the process of high-resolution color image reconstruction, which is subjected to the back projection constraint. Liu and Shum [122] present a two-step hybrid approach for super-resolving face image by combining Freeman's image primitive technique [85] and PCA model-based approach. They propose a global parametric model called "global face image" carrying the common facial properties and a local nonparametric model called "local feature image" that records the local individualities. The high resolution face image is obtained by composition of the global face image and the local feature image. The disadvantage of all the above approaches is that they either obtain the LR images in the database by downsampling the high resolution images i.e., simulate the LR images or use an interpolated version of the LR image while searching. Such a database do not represent the true spatial features relationship between LR-HR pairs as they do not correspond to the images captured by a real camera. Edge-preserving regularization terms [123, 78, 124] are designed to address the over-smoothness problem at image boundaries. Jiji *et al.* demonstrate super-resolution of a single frame gray scale image using a training database consisting of high resolution images downloaded from the internet [3]. They learn the high frequency details of the SR image from the database and obtain regularized solution by employing Markov Random Field prior model and a wavelet prior in order to preserve edges. PCA based approaches for super-resolving face images are presented in [125, 126].

The super-resolution reconstruction is one of the most spot lighted research area and is now quite matured problem. Several journals have published special issues on various approaches for multi-frame resolution and single frame super-resolution [127, 128, 129, 130, 131, 132, 133]. Similarly many books are published on the super-resolution techniques [134, 135, 136].

Chapter 3

Decimation Estimation and Super-resolution Using Zoom Based Approach

A high-resolution image is indispensable in many applications including health diagnosis and monitoring, military surveillance, and terrain mapping by remote sensing. Due to hardware cost, size, and fabrication complexity limitations, imaging systems like charge-coupled device (CCD) detector arrays often provide only multiple low resolution degraded images. The distortions are introduces in an image due to undersampling and loss of high frequency details due to various reasons such as outof focus optical blurring, sensor blur. The resolution enhancement from a single observation using image interpolation is of limited application because of the aliasing present in the low resolution image. Super-resolution refers to the process of producing a high spatial resolution image from several low-resolution observations. Super-resolution algorithms attempt to reconstruct high resolution image by predicting the missing high frequency details. In order to predict the missing information of the high resolution image, one needs non-redundant information from the observations. In zoom based super-resolution approach, the multiple observations captured by varying the zoom setting of a digital camera are used to extract the missing information of the super-resolved image.

In this chapter, we propose a method for estimating the aliasing from the zoomed observations and use the same for super-resolving the least zoomed observation consisting

of the entire scene at the resolution of the most zoomed observation. We then show the application of the proposed technique of estimating decimation to the multiresolution fusion of the multi-spectral images and panchromatic images in remote sensing. We estimate the decimation from the panchromatic image and use it to obtain high resolution for multispectral images.

3.1 Decimation and its Estimation

In the process of capturing an image of a scene using a real camera, the high resolution image goes through a sequence of degradations, including a blur, down-sampling, and additive noise. Hence the observed images are degraded versions of the high resolution images. In order to analyze the super-resolution reconstruction problem, it is required to formulate a mathematical model that represents the image acquisition process. This model, known as observation or forward model, relates the original high resolution image to the observed low resolution images.

The most commonly used image formation models for super-resolution reconstruction incorporate warping, blur, aliasing and noise in the formulation. Let the observed image Y be of size $M \times M$ pixels and $\mathbf{y}$ be the lexicographically ordered vector of size $M^2 \times 1$, which contains the pixels from image Y. Similarly, let $\mathbf{z}$ be an HR image. The general forward model can be given as [61],

$$\mathbf{y} = DHW\mathbf{z} + \mathbf{n}, \tag{3.1}$$

where W is a warping matrix, H is a blur matrix, D is a decimation matrix. The decimation matrix D takes care of aliasing. For an integer decimation factor of q, the decimation matrix D consists of q^2 non-zero elements along each row at appropriate locations. Here $\mathbf{n}$ is the independent and identically distributed (i.i.d.) noise vector with zero mean and variance σ_n^2. It has same size as $\mathbf{y}$. The LR image formed through above process will, in general, be aliased.

The decimation process relates to the fact that the low resolution data is due to the integration of light falling on the photosensor array of suitable area compared to the desired high resolution images. In literature, a simple model for decimation (aliasing)

is used. Generally the decimation model to obtain the aliased pixel intensities from the high resolution pixels has the form [48],

$$D = \frac{1}{q^2}\begin{pmatrix} 1\ 1\ldots1 & & & \mathbf{0} \\ & 1\ 1\ldots1 & & \\ & & & \\ \mathbf{0} & & & 1\ 1\ldots1 \end{pmatrix}. \tag{3.2}$$

As an example, consider an observation of size 2×2. For the decimation factor of $q = 2$, the size of $\mathbf{z}$ becomes 4×4. $\mathbf{z}$ can be represented as lexicographically ordered vector having 16 elements. The Decimation matrix D is of size 4×16 and it can be expressed with reordering of $\mathbf{z}$ as,

$$D = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{3.3}$$

In other words the aliased pixel intensity at a location (i, j) of a low resolution image for a zoom factor of $q = 2$ is given by,

$$y(i,j) = \frac{1}{4}z(2i,2j) + \frac{1}{4}z(2i,2j+1) + \frac{1}{4}z(2i+1,2j) + \frac{1}{4}z(2i+1,2j+1) + n(i,j). \tag{3.4}$$

Here $(2i, 2j), (2i, 2j+1), (2i+1, 2j)$ and $(2i+1, 2j+1)$ are corresponding 4 pixel locations in the higher resolution image and $n(i, j)$ is the noise at the pixel (i, j). The decimation matrix in equation (3.2) indicates that a low resolution pixel intensity $y(i, j)$ is obtained by averaging the intensities of q^2 pixels corresponding to the same scene in the high resolution image and adding noise intensity $n(i, j)$ (refer to equation (3.1)). In other words, all q^2 high resolution intensities are weighted equally by $\frac{1}{q^2}$ ($\frac{1}{4}$ for $q = 2$) to obtain the distorted or aliased pixel.

3.1.1 Proposed Decimation Model

The decimation model in equation (3.2) simulates the integration of light intensity that falls on the high resolution detector. This assumes that the entire area of a pixel acts

as the light sensing area and there is no space in the pixel area for wiring or insulation. In other words, fill factor for the CCD array is unity. However, in practice, the observed intensity at a pixel captured due to low resolution sampling depends on various factors such as camera gain, illumination condition, zoom factor, noise etc. The aliased low resolution pixel intensity of an image point cannot always be an equally weighted sum of the high resolution intensities and hence it has to be estimated.

The decimation matrix of the form shown in equation (3.2), can now be modified as [137],

$$D = \begin{pmatrix} a_1 \;\; a_2 \;\; \ldots \;\; a_{q^2} & & \mathbf{0} \\ & a_1 \;\; a_2 \;\; \ldots \;\; a_{q^2} & \\ \mathbf{0} & & a_1 \;\; a_2 \;\; \ldots \;\; a_{q^2} \end{pmatrix}, \tag{3.5}$$

where $|a_i| \leq 1, i = 1, 2, \ldots q^2$. The decimation matrix D for the considered example of 2×2 observation can be expressed with reordering of $\mathbf{z}$ as,

$$D = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_1 & a_2 & a_3 & a_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_1 & a_2 & a_3 & a_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_1 & a_2 & a_3 & a_4 \end{pmatrix}. \tag{3.6}$$

The aliased pixel intensity at a location (i, j) for a zoom factor of $q = 2$ is now given by,

$$y(i,j) = a_1 z(2i, 2j) + a_2 z(2i, 2j+1) + a_3 z(2i+1, 2j) + a_4 z(2i+1, 2j+1) + n(i,j). \tag{3.7}$$

When we use equation (3.2) each of the a_i has a value of $\frac{1}{q^2}$. However, in the above equation the estimates of a_i are obtained using the most zoomed observation and corresponding portions in the other LR observations. Thus the estimated a_i are more accurate as compared to using $a_i = \frac{1}{q^2}$ in equation (3.2) and are closer to the true values for the chosen model.

Y_1 Y_2

Figure 3.1: Two images of a scene, captured with different integer zoom settings. The zoom factor between Y_1 and Y_2 is 2.

3.1.2 Least Squares Approach for Decimation Estimation

In [138], authors discuss the spatial interaction model and choice of neighbors and use the same for texture synthesis. They model every pixel in an image as a linear combination of neighboring pixels considering neighborhood system. They obtain the initial estimate the model parameters using the Least Squares (LS) estimation approach. In this work, we use their approach for estimating the weights in decimation matrix.

Let Y_1 and Y_2 be two images of a scene captured using different zoom settings. Assume that Y_1 is a low resolution image and Y_2 is the high resolution image and zoom factor between these two images is 2 (see Figure 3.1). A pixel at site $s = (i, j)$ in the low resolution image Y_1 corresponds to a 2×2 block of 4 pixels at sites $\{s_{hr} = (2i + k, 2j + m)_{k=0,m=0}^{k=1,m=1}\}$ in the high resolution image Y_2. The vector of the decimation weights $\mathbf{a}$ can be estimated in the Least squares sense using,

$$\mathbf{a} = \Big[\sum_{s \in Y_1} \mathbf{y}_2(s_{hr})\mathbf{y}_2(s_{hr})^T\Big]^{-1}\Big(\sum_{s \in Y_1} \mathbf{y}_2(s_{hr})y_1(s)\Big), \tag{3.8}$$

where $\mathbf{y}_2(s_{hr})$ is a vector representation of the high resolution pixels at sites s_{hr} that correspond to a low resolution pixel $y_1(s)$ at site s. The pseudocode of the proposed technique of decimation estimation for $q = 2$ is given in Algorithm 3.1.

The estimates of a_i are obtained using the HR observation and LR observation. Thus the estimated a_i are more accurate as compared to using $a_i = \frac{1}{q^2}$ in equation (3.2) and are closer to the true values for the chosen model. It may be noted that this form of decimation matrix D implicitly contains moving average (space invariant) blur in the downsampling process.

We now present the application of the decimation estimation to the zoom based super-

Algorithm 3.1: Decimation Estimation for $q = 2$

Data: a high resolution image Y_2 and its low resolution version Y_1
Result: decimation weight vector $\mathbf{a} = [a_1\ a_2\ a_3\ a_4]$
initialize $S = [0\ 0\ 0\ 0]$;
initialize $U = [0\ 0\ 0\ 0]^T$;
foreach *pixel* (i, j) *in the LR image* Y_1 **do**
  let $Z_s = [y_2(2i, 2j);\ y_2(2i, 2j+1);\ y_2(2i+1, 2j);\ y_2(2i+1, 2j+1)]$;
  calculate $Z = Z_s \times {Z_s}^T$;
  calculate $Y = Z_s \times y_1(i, j)$;
  $S = S + Z$;
  $U = U + Y$;
end
$\mathbf{a} = S^{-1}U$;

resolution where the multiple observations are obtained by varying the zoom setting of a digital camera. Decimation is estimated using LR observation and corresponding part in the most zoomed observation. We obtain super-resolution using regularization framework. A prior is introduced to avoid solutions which are subjectively very implausible to the human viewer. We employ the Maximum *a Posteriori* approach, considering general image priors commonly selected for image super-resolution. We then show the application to the fusion in the remotely sensed images. In this case the decimation is estimated using available Panchromatic image and low resolution MS image.

3.2 Zoom Based Super-resolution Using Decimation Estimation

In [1], the authors propose a technique for super-resolution using zoom cue. The authors capture the observations of a static scene at different zoom factors by varying the zoom setting of a digital camera. When one captures the images with different zoom settings, the amount of aliasing is different in differently zoomed observations. This is because the least zoomed entire area of the scene is represented by a very limited number of pixels, i.e., it is sampled with a very low sampling rate and the most zoomed image with a higher sampling frequency. Therefore, larger scene coverage will have lower resolution with more aliasing effect. By varying the zoom level, one observes the scene at different levels of aliasing and blurring. Thus, one can use zoom as a cue for generating high-resolution images at the lesser zoomed area of a scene. They model the super-resolution image as

a Markov random field and use a MAP estimation method to derive cost function. They optimize the cost function using a gradient descent technique and obtain super-resolution. In this method they use a simplified decimation model of the form in equation (3.2) for low resolution observations formulation. The intensity of a low resolution pixel intensity is represented as an average of intensities of corresponding pixels in the high resolution image. Since, the observed intensity at a pixel captured due to low resolution sampling depends on various factors such as camera gain, illumination condition, zoom factor, noise etc., the aliasing has to be estimated. We propose the use of decimation estimation and obtain super-resolution. The super-resolution image is modeled as an MRF. It is assumed that the high resolution image at the most zoom setting is super-resolved. We propose to estimate the decimation (aliasing) matrices from the most zoomed observation and lesser zoomed observation. We then use MAP-MRF formulation to obtain super-resolved image for the entire scene.

3.3 Problem Formulation

The zoom based super-resolution problem can be cast in a restoration framework. There are p observed images Y_i, $i = 1, \ldots, p$, each captured with different zoom setting and are of size $M_1 \times M_2$ pixels each. Figure 3.2 illustrates the block schematic of how the low-resolution observations of a scene at different zoom settings are related to the high-resolution image. Here we consider that the most zoomed observed image of the scene Y_p

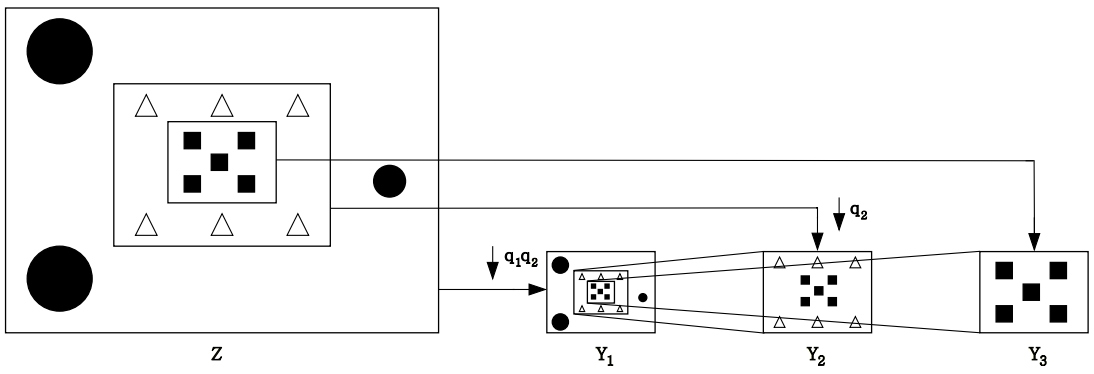


Figure 3.2: Illustration of observations at different zoom levels, Y_1 corresponds to the least zoomed and Y_3 to the most zoomed images. Here Z is the high resolution image of the scene.

($p = 3$) has the highest spatial resolution. We are assuming that there is no rotation about the optical axis between the observed images taken at different zooms. Since different zoom settings give rise to different resolutions, the least zoomed scene corresponding to entire scene needs to be upsampled to the size of $(q_1 q_2 \dots q_{p-1}) \times (M_1 \times M_2)$ pixels ($= N_1 \times N_2$ pixels), where $q_1, q_2, \dots, q_{p-1}$ are the corresponding zoom factors between two successively observed images of the scene Y_1Y_2, Y_2Y_3, ..., $Y_{p-1}Y_p$ respectively. Given Y_p, the remaining $(p-1)$ observed images are then modeled as decimated and noisy versions of this single high-resolution image of the appropriate region in the scene. The most zoomed observed image will have no decimation. The low resolution image observation model is shown in Figure 3.3.

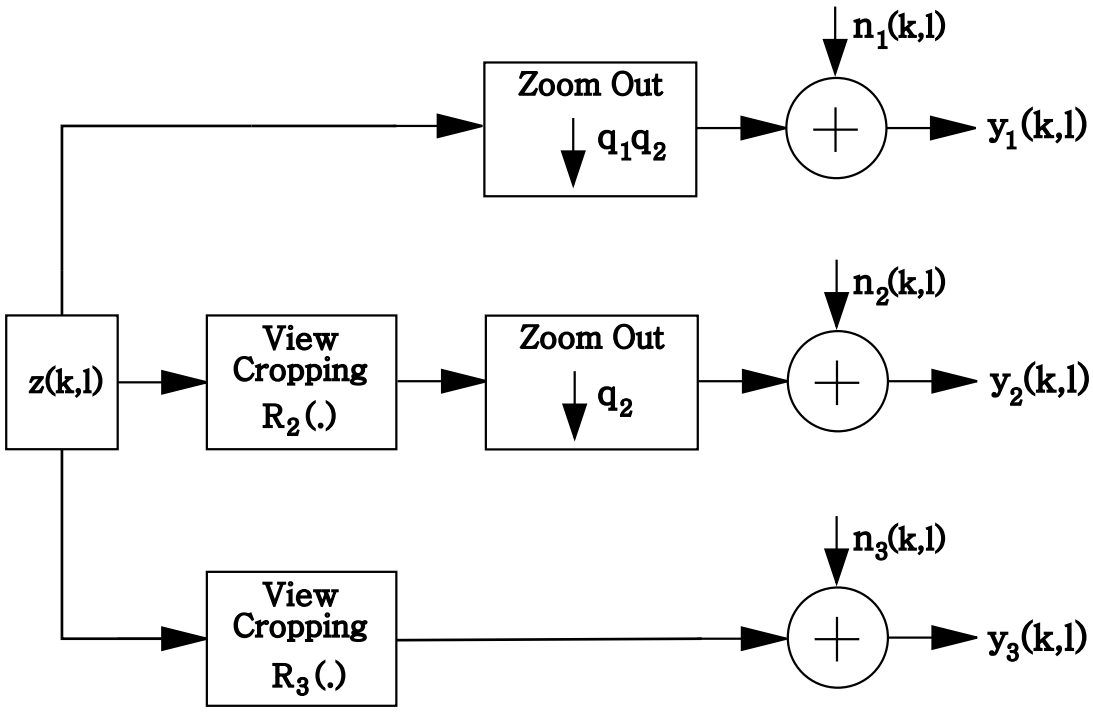


Figure 3.3: Low-resolution image formation model for three different zoom levels [1]. View cropping block just crops the relevant part of the high resolution image Z as the field of view shrinks with zooming.

Let $\mathbf{y}_m$ represent the lexicographically ordered vector of size $M_1 M_2 \times 1$, which contains the pixels from differently zoomed images Y_m and $\mathbf{z}$ be the super-resolved image. The observed images can be modeled as,

$$\mathbf{y}_m = D_m \mathbf{C}_m (\mathbf{z} - z_{\alpha_m}) + \mathbf{n}_m, \qquad m = 1, \cdots, p, \tag{3.9}$$

where D is the decimation matrix which takes care of aliasing present while zooming.

The subscript m in D denotes that the amount of decimation depends on the amount of zoom for m^{th} observation, size of which depends on the zoom factor. $\mathbf{C}_m$ is a cropping operator with $z_{\alpha_m} = z(x - \alpha_{m_x}, y - \alpha_{m_y})$ and $\alpha_m = (\alpha_{m_x}, \alpha_{m_y})$ representing the lateral shift of the optical shift during zooming process for the m^{th} observation. The cropping operation is analogous to a characteristic function which crop outs the $\lfloor q_1 q_2 \dots q_{m-1} N_1 \rfloor \times \lfloor q_1 q_2 \dots q_{m-1} N_2 \rfloor$ pixel area from the high resolution image $\mathbf{z}$ at an appropriate position. $\mathbf{n}_m$ is the i.i.d noise vector with zero mean and variance σ_n^2. It is of the size, $M_1 M_2 \times 1$. For an integer zoom factor of q, the decimation matrix D consists of q^2 non-zero elements along each row at appropriate locations.

The multivariate noise probability density is given by,

$$P(\mathbf{n}_m) = \frac{1}{(2\pi\sigma_n^2)^{\frac{M_1 M_2}{2}}} e^{-\frac{1}{2\sigma_n^2}\mathbf{n}_m^T \mathbf{n}_m}. \tag{3.10}$$

Our problem is to estimate $\mathbf{z}$ given $\mathbf{y}_m$s, which is an ill-posed inverse problem. It may be mentioned here that the observations captured are not blurred. In other words, we assume identity matrix for blur. The inherent difficulty with inverse problems is the challenge of inverting the forward model. There exist many expanded images which satisfy the equation (3.10). Regularization plays a vital role in stabilizing such problems by introducing constraints on the solution that restricts the solution space and makes the problem better-posed. Stochastic approach, typically Bayesian approach provides a flexible and convenient way to model a priori knowledge about the super-resolved image. We model the scene in such a way that the model parameters can be used as priors for obtaining regularized solution. First we estimate the decimation matrices using the procedure described in section 3.1. Since we capture the images at different resolutions using zoom camera and the most zoomed image is assumed to be alias free, we estimate the decimation matrix entries from the most zoomed region. These entries are obtained by considering the most zoomed image and corresponding portion in the lesser zoomed images. We estimate 4 weights for a zoom factor of 2 and 16 for a zoom factor of 4. The estimated weight vectors are then used in equation (3.9) for forming D matrix to get the observation model. It may be noted that for a given zoom factor, we are not estimating different weights for each location. Since the average brightness of each observation varies due to AGC of camera, we used mean correction to maintain average brightness

of the captured images approximately the same and use these observations for the D matrix estimation as well as for experimentation. Mean correction for Y_2 is obtained by subtracting it's mean from each of its pixel and adding the mean of corresponding portion in Y_1. Similarly, for Y_3, it is obtained by subtracting from each pixel, its mean and adding the mean of corresponding portion in Y_1. (Refer to Figure 3.2.) We represent the high resolution image field using the linear models and optimize the cost function to obtain the solution.

3.4 Image Field Modeling

In order to obtain a robust solution of the problem using regularization, a good assumption about the nature of true image is required. Images contain random changes and noise. Hence they are statistical in nature. It is convenient to regard an image as a sample of a stochastic process. Desired information about an image like image content, redundancy can be extracted using probability distributions and correlation functions. The obvious fact about the natural images is that the neighboring pixels in an image are highly dependent of each other. This dependence can be described either by means of a correlation function or a power spectral density or by means of Markov model. In the Markov model approach the relation between different image pixels is described by Markov random fields.

3.4.1 MRF Prior Model for the Super-resolved Image

The MRF provides a convenient and consistent way of modeling context dependent entities. This is achieved through characterizing mutual influence among such entities using conditional probabilities for a given neighborhood [139]. The practical use of MRF models is largely ascribed to the equivalence between the MRF and the Gibbs Random Fields (GRF). We assume that the high-resolution image can be represented by an MRF. This is justified because the changes in intensities in a scene is gradual and hence there is a local dependency.

Let Z be a random field over an regular $N \times N$ lattice of sites $L = \{(i, j)|1 < i, j < N\}$.

From the Hammersley-Clifford theorem for MRF-GRF equivalence, we have,

$$P(Z = z) = \frac{1}{Z_p} e^{-U(z)}, \tag{3.11}$$

where $\mathbf{z}$ is a realization of Z, Z_p is a partition function given by $Z_p = \sum_z e^{-U(z)}$ and $U(z)$ is energy function given by $U(z) = \sum_{c \in C} V_c(z)$. $V_c(z)$ denotes the potential function of clique c and $\mathbf{C}$ is the set of all cliques. The lexicographically ordered high resolution image $\mathbf{z}$ satisfying Gibbs density function is now written as,

$$P(\mathbf{z}) = \frac{1}{Z_p} e^{-\sum_{c \in C} V_c(\mathbf{z})}. \tag{3.12}$$

We consider pair wise cliques on a first-order neighborhoods consisting of the four nearest neighbors for each pixel and impose a quadratic cost which is a function of finite difference approximations of the first order derivative at each pixel location. i.e.,

$$\sum_{c \in C} V_c(\mathbf{z}) = \lambda \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} \left[(z_{k,l} - z_{k,l-1})^2 + (z_{k,l} - z_{k-1,l})^2 \right], \tag{3.13}$$

where λ represents the penalty for departure from the smoothness in $\mathbf{z}$.

3.4.2 The Autoregressive (AR) Model

In previous section, we formulated regularization framework using Markov random field model prior. Although MRF model for prior captures the spatial dependencies very well, the computational complexities with these models is very high. The estimation of true parameters needs the computation of partition function. Further, it requires the use of computationally taxing stochastic relaxation technique for obtaining global minima. This motivates use to consider a different model for prior that is computationally less intensive.

Recently, Auto-regressive (AR) models have attracted a lot attention of image/video applications. The AR model is an efficient and compact description of random process. It is able to have desirable performance for the linear prediction [140]. Authors in [141, 138] propose a multiresolution autoregressive model for the image and use the extracted parameters for texture classification and image segmentation. In [142], the authors propose a model-based approach to multiresolution fusion in remotely sensed images. They utilize

the spatial correlation of each of the high resolution multispectral channels by using an AR model.

An autoregressive (AR) model is a way of characterizing a single, stationary dynamical regime, for which much of the methodology was set out by Box and Jenkins [143]. It is based on the idea that the value y_t of a series can be explained in terms of the preceding values $y_{t-p}, \ldots, y_{t-1}$. It is therefore a regression based on the variables own history rather than on any external variables. The current value of a series is dependent on the previous p values. The resulting series is referred as an $AR(p)$ process. The scalar $AR(p)$ model is described by the equation,

$$y_t - \mu = \sum_{k=1}^{p} \theta_k (y_{t-k} - \mu) + n_t, \tag{3.14}$$

in which n_t is white noise with variance σ_n^2, and μ is the mean of the process. The values of the θ_k's and the variance σ_n^2 characterize the behavior of the process. A useful property of autoregressive processes is that they can approximate any power spectrum to within an arbitrary accuracy, given a high enough autoregression order p.

In this section, we propose the use of autoregressive model for the problem of zoom based super-resolution. We assume a linear dependency of a pixel to its neighbors in a high spatial resolution image and represent it with an AR model. We exploits the available most zoomed observation to learn the spatial relationships for the unknown high resolution images using an AR structure. The parameters of the prior model are unknown as the true high resolution image is unavailable and have to be estimated. To solve this ambiguity, we propose to use the available most zoomed observation for the learning of the parameter values, so that they can be used to improve the solution. The AR parameters learnt from the most zoomed observation are used as the AR parameters for the super-resolved image in order to enhance its spatial resolution. Thus we use a homogeneous model and derive a set of parameters for the entire image. Once we estimate the values of the AR model parameters, we use them for the super-resolving the least zoomed image.

Let $z(s)$ be the gray level value of the image pixel at site $s = (i, j)$ in an $M_1 \times M_2$ lattice, where $i = 1, 2, \cdots M_1$ and $j = 1, 2, \cdots M_2$. The AR model for $z(s)$ can be expressed

as [138],

$$z(s) = \sum_{r \in \mathcal{N}_s} \theta(r) z(s+r) + \sqrt{\rho} n(s), \tag{3.15}$$

where $\mathcal{N}_s$ is the neighborhood of pixel at s. The $\theta(r)$, r being a neighborhood index with $r \in \mathcal{N}_s$, and ρ are unknown parameters, and $n(.)$ is an i.i.d noise sequence with zero mean and unit variance. Here ρ is the variance of the white noise that generates the specified data for the given AR parameters. It is worth noting that an MRF model can represent a large variety of spatial inter-relationship among pixels locally, out of which the AR model is only a subset representing linear dependencies. We use a third order neighborhood that requires a total of 8 parameters θ (Refer Figure 3.4). We estimate the

X(0,0)	X(0,1)	X(0,2)	X(0,3)	X(0,4)
X(1,0)	X(1,1)	X(1,2)	X(1,3)	X(1,4)
X(2,0)	X(2,1)	X(2,2)	X(2,3)	X(2,4)
X(3,0)	X(3,1)	X(3,2)	X(3,3)	X(3,4)
X(4,0)	X(4,1)	X(4,2)	X(4,3)	X(4,4)

Figure 3.4: Third order neighborhood for the pixel located at $(2, 2)$. Shaded pixels are the neighboring pixels of the pixel at location $(2, 2)$.

AR model parameters by considering the image as a finite lattice model using the scheme described in section 3.4.3. We use,

$$\sum_{s=i,j} \left(z_m(s) - \sum_{r \in \mathcal{N}_s} \theta(r) z_m(s+r) \right)^2, \tag{3.16}$$

as a prior term in cost function.

3.4.3 Estimation of AR Model Parameters

Let $y(s)$ be the gray level value of the image pixel at site $s = (i, j)$ in an image Y and $\mathcal{N}_s$ be the neighborhood of pixel at s. The vector of the neighboring pixels at site s can

be represented using $\mathbf{y}_r$, r being a neighborhood index with $r \in \mathcal{N}_s$. The AR model parameter vector $\boldsymbol{\theta}$ for the image Y can be computed using,

$$\boldsymbol{\theta} = \Big[\sum_{s\in Y} \mathbf{y}_r \mathbf{y}_r^T\Big]^{-1} \Big(\sum_{s\in Y} \mathbf{y}_r y(s)\Big), \tag{3.17}$$

The pseudocode of the method of AR parameter estimation for third order neighborhood is given in Algorithm 3.2.

Algorithm 3.2: AR model parameter estimation for third order neighborhood.

Data: an image Y
Result: AR model parameter vector $\boldsymbol{\theta}$=$[\theta_1\ \theta_2\ \ldots\ \theta_8]$
initialize $S = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$;
initialize $U = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^T$;
foreach *pixel* (i, j) *in the image* Y **do**
  let $\mathbf{y}_r = [y(i-1, j-1);\ y(i-1, j);\ y(i-1, j+1);\ y(i, j-1);\ y(i, j+1);\ y(i+1, j-1);\ y(i+1, j);\ y(i+1, j+1)]$;
  calculate $Z = \mathbf{y}_r \times \mathbf{y}_r{}^T$;
  calculate $X = \mathbf{y}_r \times y(i, j)$;
  $S = S + Z$;
  $U = U + X$;
end
$\boldsymbol{\theta}$=$S^{-1}U$;

3.5 Super-resolving the Scene

The problem of reconstructing the high resolution image is an ill-posed inverse problem and some form of regularization is necessary. We derive an optimal estimate of the HR image as the MAP estimate. The MAP framework allows us to impose a priori constraints on the HR image. Since statistical models can encode contextual constraints in images in a natural way, we consider Markov random field model and AR model for the original HR image.

3.5.1 Super-resolution Using MAP-MRF Estimation

The MAP estimate of the super-resolution image comes about by an application of Bayes theorem,

$$P(z|y) = \frac{P(y|z)P(z)}{P(y)}. \tag{3.18}$$

The left hand side is known as the posterior distribution over z and y represents observed data. If is held constant, then $P(y)$ may be considered as a normalization constant. We apply this to our problem. Given the ensemble of images $\mathbf{y}_i$, $i = 1$ to p, at different resolutions, the MAP estimate $\hat{\mathbf{z}}$, using Bayesian rule, is given by,

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\text{argmax}} \quad P(\mathbf{z}|\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p) = \underset{\mathbf{z}}{\text{argmax}} \quad P(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p|\mathbf{z})P(\mathbf{z}). \tag{3.19}$$

Taking the log of the posterior probability we can write,

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\text{argmax}} \quad \Big[\sum_{m=1}^{p} \log P(\mathbf{y}_m|\mathbf{z}) + \log P(\mathbf{z})\Big], \tag{3.20}$$

since n_m are independent. The above MAP formulation allows us to incorporate prior knowledge about $\mathbf{z}$ for improving robustness during reconstruction.

Now using equation (3.9) and (3.10), we get,

$$P(\mathbf{y}_m|\mathbf{z}) = \frac{1}{(2\pi\sigma_n^2)^{\frac{M_1 M_2}{2}}} e^{-\frac{\|\mathbf{y}_m - D_m C_m(\mathbf{z} - z_{\alpha_m})\|^2}{2\sigma_n^2}}. \tag{3.21}$$

The final cost function is obtained as,

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\text{argmin}} \quad \Big[\sum_{m=1}^{p} \frac{\|\mathbf{y}_m - D_m C_m(\mathbf{z} - z_{\alpha_m})\|^2}{2\sigma_n^2} + \sum_{c \in C} V_c(\mathbf{z})\Big]. \tag{3.22}$$

The above cost function is convex and is minimized using the gradient descent technique. The initial estimate $\mathbf{z}^{(0)}$ is obtained as follows. Pixels in the zero order hold of the least zoomed observation corresponding to the entire scene is replaced successively at appropriate places with zero order hold of the other observed images with increasing zoom factors. Finally, the most zoomed observed image with the highest resolution is copied at the appropriate location (see Figure 3.2.) with no interpolation.

3.5.2 Super-resolution Using AR Model

Having defined the AR prior, we use the MAP estimator to restore the high-resolution field $\mathbf{z}$. Given the ensemble of images $\mathbf{y}_i$, $i = 1$ to p, at different resolutions, the MAP

estimate $\hat{\mathbf{z}}$, using Bayesian rule, is given by,

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\text{argmax}} \quad P(\mathbf{z}|\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p) = \underset{\mathbf{z}}{\text{argmax}} \quad P(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p|\mathbf{z})P(\mathbf{z}). \tag{3.23}$$

From this, the cost function is derived as,

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\text{argmin}} \quad \Big[\sum_{m=1}^{p} \frac{\|\mathbf{y}_m - D_m C_m(\mathbf{z} - z_{\alpha_m})\|^2}{2\sigma_n^2} + \sum_{i,j} \left(z_m(s) - \sum_{r \in \mathcal{N}_s} \theta(r) z_m(s+r) \right)^2 \Big]. \tag{3.24}$$

The above cost function is convex and is minimized using the gradient descent technique.

3.6 Experimental Results

In this section, we present the results of the proposed method of obtaining super-resolution using decimation estimation. All the experiments were conducted on real images taken by a zoom camera and known integer zoom factors. It assumed that the lateral shift during zooming is known. In each experiment, we consider three low resolution observations Y_1, Y_2, Y_3 of an image. Each observed image is of size 72×96. Zoom factor q between Y_1 and Y_2 is 2 and that between Y_1 and Y_3 is 4. The super-resolved images for the entire scene are of size 288×384.

First, we estimate decimation matrices D_1 and D_2 using observations. We use observations Y_1 and Y_3 to estimate D_1 and use Y_2 and Y_3 for estimating D_2. The observation Y_3 is captured at the highest resolution and our aim is to super-resolve the observation Y_1 at the resolution of Y_3. The centermost 18×24 region in the observation Y_1 represents the observation Y_3. Figure 3.6(a) shows the highlighted region of Y_1, which is used for estimation of D_1. Each pixel in this region corresponds to a 4×4 block in Y_3. Since Y_3 is available at the highest resolution, the decimation matrix D_1 is estimated using Y_3 and the centermost 18×24 region in Y_1. It may be noted that D_1 contains 16 non-zero entries. Similarly, the centermost 36×48 region in the observation Y_2 represents the observation Y_3 (see Figure 3.6(b)). Each pixel in this region corresponds to a 2×2 block in Y_3. The decimation matrix D_2 is estimated using Y_3 and the centermost 36×48 region in Y_2. D_2 contains 4 non-zero entries.

We obtain super-resolution and compare the results obtained using decimation matrix

(a) (b) (c)

Figure 3.5: Observed images of 'Nidhi' captured with three different integer zoom settings. The zoom factor between (a) and (b) is 2 and between (b) and (c) is also 2.

Y_1 Y_2 Y_3
(a) (b) (c)

Figure 3.6: Regions of zoomed observations used for decimation estimation. D_1 is estimated using Y_3 and highlighted rectangular area in Y_1. Similarly D_2 is estimated using Y_3 and highlighted rectangular area in Y_2.

of the form in equation (3.2) consisting of equal weights. We used the quantitative measures Mean Squared Error (MSE) and Mean Absolute Error (MAE) for comparison of the results. The MSE used here is,

$$MSE = \frac{\sum_{i,j}[f(i,j) - \hat{f}(i,j)]^2}{\sum_{i,j}[f(i,j)]^2} \tag{3.25}$$

and MAE is,

$$MAE = \frac{\sum_{i,j}|f(i,j) - \hat{f}(i,j)|}{\sum_{i,j}|f(i,j)|}, \tag{3.26}$$

where $f(i,j)$ is the original high resolution image and $\hat{f}(i,j)$ is estimated super-resolution image. In order to use high resolution image for the entire scene the most zoomed image was captured with entire scene content. However, while experimenting only a portion of it was used. The estimated D matrices are used in the cost function.

3.6.1 Results for MRF Model

In the first experiment, we considered three low resolution observations of a girl image 'Nidhi' shown in Figure 3.5, where the observed images have less intensity variations. Figure 3.7 shows zoomed 'Nidhi' image obtained using bicubic interpolation. Figure 3.8(a) shows super-resolved 'Nidhi' image obtained by using the decimation matrix of

Figure 3.7: 'Nidhi' image zoomed by bicubic interpolation.

(a) (b)

Figure 3.8: Super-resolved 'Nidhi' image. (a) using equal weights decimation matrix and (b) using estimated weights for decimation matrix (proposed approach).

(a) (b) (c)

Figure 3.9: Observed images of a house captured with three different integer zoom settings.

Figure 3.10: House image zoomed by successive pixel replication.

the form in equation (3.2) and Figure 3.8(b) shows super-resolved 'Nidhi' image obtained by proposed method by using the estimated decimation matrix of the form in equation (3.5). The comparison of the images show more clear details in the regions like cheeks and forehead in the image obtained by the proposed method. The eyes look sharper and the lines on the shirt appear clearly. The seam is clearly visible in Figure 3.8(a).

In the second experiment, we considered low resolution observations having arbitrary texture with moderate amount of high frequency content. We consider a scene of a house as shown in Figure 3.9. Zoomed house image obtained by successive pixel replication is shown in Figure 3.10 and Figure 3.11 shows super-resolved house images obtained using the two different methods. The comparison of the figures show that there is less blockiness in the super-resolved image obtained by the proposed method. The textures of grills in the window pane are more clearly visible in the image super-resolved using proposed approach. We can also see the improvement in the branches of trees opposite to windows.

In order to consider images with significant texture, we experimented by capturing zoomed images of a house with natural surroundings. The observed images are displayed in Figure 3.12. Figure 3.13 shows super-resolved scene images. The small house near the

(a) (b)

Figure 3.11: Super-resolved house image. (a) using equal weights decimation matrix and (b) using estimated weights for decimation matrix (proposed approach).

(a) (b) (c)

Figure 3.12: Observed images of a scene captured with three different integer zoom settings.

center of image appear sharper in the image super-resolved using the proposed approach. Blockiness is less in the image obtained using the proposed approach as compared that obtained using the other approach. The image super-resolved using the proposed approach appear to be visually pleasant. These results testify the efficiency of our approach for a wide range of data sets.

Table 3.1. shows the quantitative comparison of the our results with the one obtained using equal weights for decimation matrix. It can be seen that for all the three experiments, MSE and MAE of the super-resolved images obtained by using estimated

Image	MSE		MAE	
	Fixed decimation (Equal weights)	Estimated decimation (proposed)	Fixed decimation (Equal weights)	Estimated decimation (proposed)
Nidhi	0.0514	0.0484	0.0525	0.0489
House	0.6733	0.6671	0.6751	0.6678
Scene	0.3056	0.2732	0.3082	0.2741

Table 3.1: Comparison of performance of the two methods of super-resolution.

(a) (b)

Figure 3.13: Super-resolved scene image. (a) using equal weights decimation matrix and (b) using estimated weights for decimation matrix (proposed approach).

decimation matrices is lower than those obtained by fixed decimation matrix entries showing improvement in the quantitative measures.

3.6.2 Results for AR Model

In the first experiment, we considered three low resolution observations of an image 'Divya' shown in Figure 3.14(a)-(c). The images were scanned from a film negative using a scanner with different resolution settings. Figure 3.14(a) is an image of size 72×96 pixels scanned with resolution of 300 dots/inch (dpi). To obtain second observation shown in Figure 3.14(b), the same region was scanned at 600 dpi resolution and cropped the centermost region of the size 72×96. Similarly, the third observation Figure 3.14(c) was obtained by scanning the region at 1200 dpi resolution and cropping the centermost region of the size 72×96. Figure 3.14(d) shows 'Divya' image of size 288×384 expanded by successive bicubic interpolation and Figure 3.14(e) shows super-resolved 'Divya' image of same size obtained using MRF model and Figure 3.14(f) shows super-resolved 'Divya' image obtained by proposed method. The comparison of the images show more clear details in the regions containing edges in the image obtained by our method. The seam is clearly visible in Figure 3.14(d). Boundary of the head and eyes in the image super-resolved using proposed approach look sharper. The proposed method has less smoothing effect. In the second experiment, we considered low resolution observations of 'Pool' image shown in Figure 3.15(a)-(c). The observation images were obtained in the manner used in the first experiment. Zoomed 'Pool' images obtained by successive bicubic interpolation

Figure 3.14: (a)-(c) Observed 'Divya' images, (d) 'Divya' image expanded using bicubic interpolation, (e) 'Divya' image super-resolved using MRF model, and (f) 'Divya' image super-resolved using the proposed approach.

is shown in Figure 3.15 (d). Figure 3.15(e) shows the super-resolved image obtained using MRF model and, (f) shows the super-resolved image obtained using AR model and estimated decimation matrix. Many details previously aliased in the LR image are now visible in the super-resolved image using the proposed approach. In Figure 3.15(d), seam is visible in the region of head and face of the girl on right, where as in Figure 3.15(f) the seam is not visible. Improvement is clearly seen in textured regions such as eyes and discontinuities in the image super-resolved using proposed approach. However, we observe some blockiness along the boundary of head. The most zoomed image has not enough texture for learning decimation and AR parameters. Hence, the super-resolved image may suffer from blockiness. We see that the high frequency details are better preserved in the other images. The reason for the better solution using the AR model as compared to MRF model is that we are using a larger neighborhood with more number of parameters for the model representation. This is enables the algorithm to capture the prior better then the MRF model as we are constrained to use a very few cliques during the MRF modeling for reasons of computational difficulties in learning these model parameters.

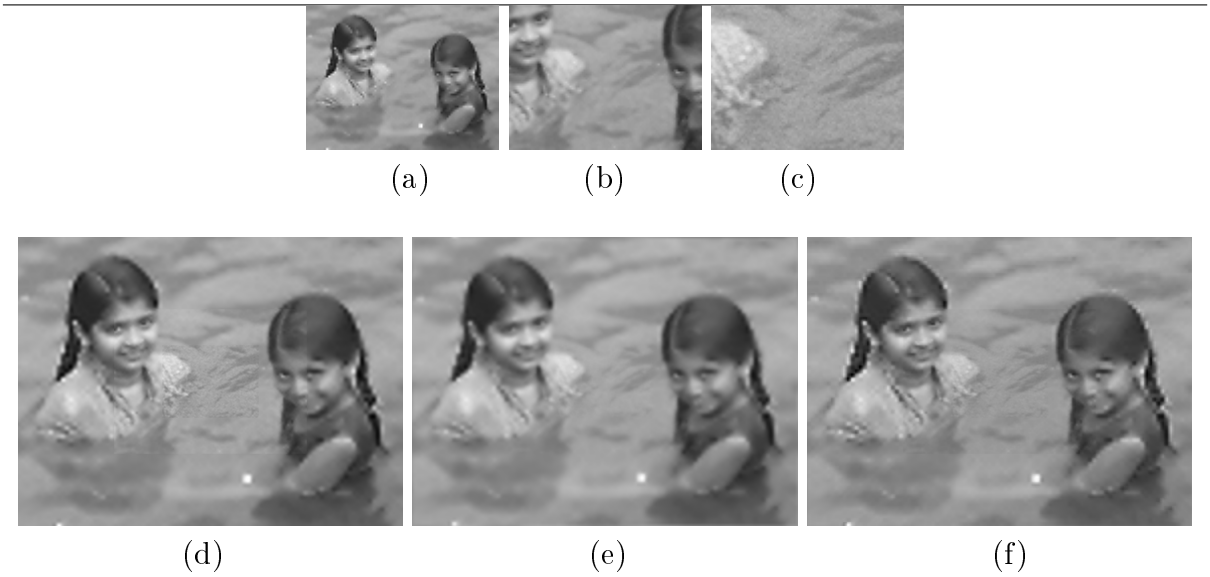


Figure 3.15: (a)-(c) Observed 'Pool' images, (d) 'Pool' image expanded using bicubic interpolation, (e) 'Pool' images super-resolved using MRF model and, (f) super-resolved using the proposed approach.

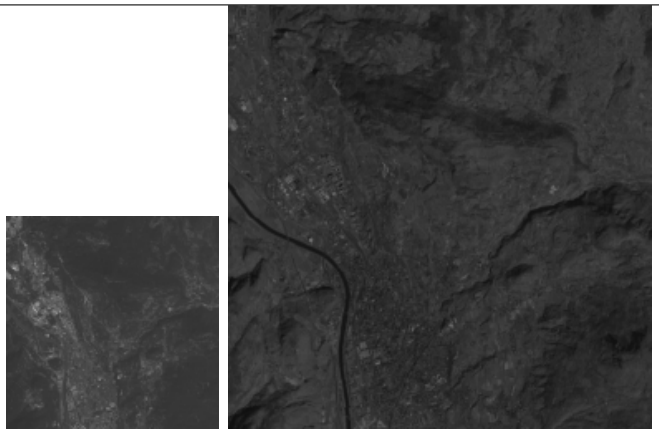
3.7 Application of Decimation Estimation to Multiresolution Fusion in Remote Sensing

The growth in agriculture, increased urbanization and natural processes all contribute to the changing nature of land use and land cover around the globe. Remote sensing has been identified as a critical tool in understanding changes on a large and small scale. Many corporations and national governments operate satellite remote sensing systems specifically designed for observation of the earth surface to collect information concerning topics such as crops, forests, water bodies, land use, cities, and minerals. Satellite sensors provide systematic observation of large areas with fine details. Image sensors have a fixed signal-to-noise ratio that is a function of the hardware design. The energy reflected by the target must have a signal level large enough for the target to be detected by the sensor. The signal level of the reflected energy increases if the signal is collected over a larger instantaneous field of view (IFOV) or if it is collected over a broader spectral bandwidth. Collecting energy over a larger IFOV reduces the spatial resolution while collecting it over a larger bandwidth reduces its spectral resolution. Thus, there is a tradeoff between the spatial and spectral resolutions of the sensor. A high spatial resolution can accurately discern small or narrow features like roads, automobiles, shallow reefs, individual trees in an orchard, etc. A high spectral resolution provides detailed information on such diverse

areas as the quality of the road surfaces, the depth of the ocean, and the health of plants. A high spectral resolution allows the detection of minor spectral changes, like those due to vegetation stress or molecular absorption.

The remote sensing systems are equipped with two kinds of sensors: Panchromatic sensors that capture the images of the earth with high spatial resolution but with lower spectral and radiometric details; and multispectral sensors that capture images that provide finer details concerning spectral characteristics of the earth but less spatial resolution. The panchromatic (Pan) images are characterized by high spatial resolution whereas multispectral (MS) images are characterized by high spectral resolution. Examples of an MS image and a Pan image are shown in Figure 3.16. In the multi-spectral images, most urban objects can be visually recognized according to the color difference, but they cannot be clearly delineated due to the lack of spatial resolution. In the panchromatic images, however, the shape of most individual objects can be clearly identified, but many of them cannot be classified due to the lack of spectral information. The process of combining Pan and MS data to produce MS images characterized by both high spatial and spectral resolutions is known as multiresolution fusion. Since, the fusion process utilizes both spectral information from multi-spectral images and spatial information from panchromatic images, one can overcome the limitations of information obtained from individual sources and obtain a better understanding of the observed scene. Many researchers have addressed the problem of multiresolution image fusion for remote sensing applications [144, 145, 146, 147, 148, 149, 150, 142, 151, 152].

In this section we show the application of the proposed technique to multispectral fusion in remotely sensed images. Since the Pan image has high spatial resolution and MS images have lower spatial resolution, we estimate the aliasing on MS images by using the Pan image. The same Pan image is used to estimate the aliasing on each of MS images. The authors in [142] obtain multiresolution fusion using AR parameters estimated from the entire PAN image. However, they do not employ decimation estimation. In addition, better results can be obtained if PAN and MS images are divided into blocks and fusion of each block is achieved using AR parameters estimated from each of the block separately. We address the problem using the proposed technique of decimation estimation. Further we divide the MS image into several blocks and learn AR model parameters for each block separately.



(a) (b)

Figure 3.16: (a) An MS image and (b) a Pan image.

Image	Spectral bandwidth (μm)	Spatial resolution (m)	Color
MS Band 1	0.45-0.52	2.44	Blue
MS Band 2	0.52-0.60	2.44	Green
MS Band 3	0.63-0.69	2.44	Red
MS Band 4	0.76-0.89	2.44	Near infrared
Panchromatic	0.45-0.89	0.61	Natural

Table 3.2: Spectral and spatial resolutions of Quickbird images.

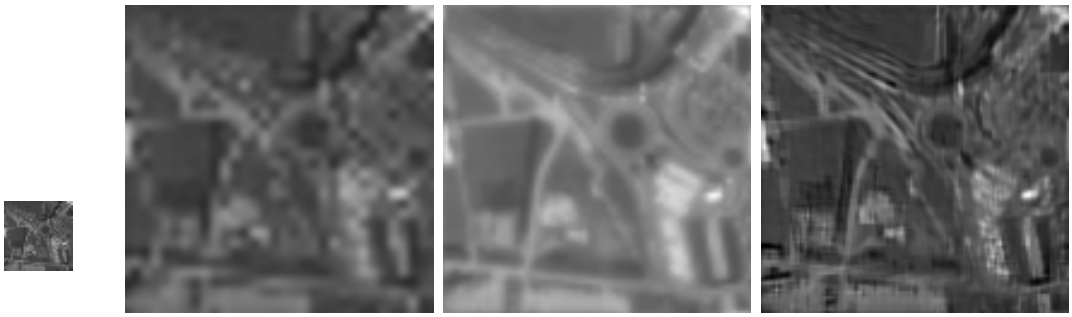
For the experiment, we consider LANDSAT-7 Enhanced Thematic Mapper Plus (ETM+) images acquired by Quickbird satellite. The Quickbird satellite is tailored to acquire fine detail imagery using a Panchromatic band and four multispectral bands. Table 3.3 shows the spectral and spatial resolutions of the Landsat 7 ETM+ sensor. We conducted experiments on Quickbird data set consists of four MS images of size 128×128 pixels at a spatial resolution of $2.4m \times 2.4m$ and a coregistered PAN image of size 512×512 pixels. The images were captured over Malpensa city in Italy. The PAN image decimated by a factor of 4 was used to learn the AR parameters, and the original MS images were used as reference (true) data in order to make a quantitative comparison. We model the MS image formation process using simple decimation process described by,

$$\mathbf{y}_m = D_m \mathbf{z}_m + \mathbf{n}_m, \qquad m = 1, \cdots, p. \tag{3.27}$$

It may be noted that the subscript m in the equation represents the MS Band number. Each of the observed low resolution MS images is modeled as decimated and noisy version

of its high resolution version. Our problem is to estimate high resolution MS images $\mathbf{z}_m$ given observed MS images $\mathbf{y}_m$ and a Pan image. Since the D_m matrix is not invertible, our problem is ill-posed and there are infinite solution to it. We solve the problem using regularization framework. Obtaining a solution using this framework requires a reasonable assumption about the nature of the true image. MAP estimation with convex priors implies a globally convex optimization, ensuring solution existence and uniqueness allowing the application of efficient descent optimization methods. Once a-priori constraint on the solution is included, the obtained good solution depends on the model parameters. We represent the contextual dependencies in each unknown high resolution fused MS image using an AR model.

In our experiment, first we estimate the decimation. The estimation of the decimation requires the true high resolution MS image. Since the true high resolution MS image is unavailable, we use Pan image and observed MS image to estimate the same. The PAN image is used for estimating aliasing matrices for all the MS images as the aliasing depends on difference in spatial resolution between high resolution and low resolution images. We divide the an MS image into 16 blocks and model the spatial dependencies in each unknown high resolution fused block using an AR model. However, the parameters of the AR prior model are unknown as the true high resolution MS images are unavailable and have to be estimated. We propose to use the Pan image for the learning of the parameter values, so that they can be used to improve the solution. The required AR parameters are estimated from corresponding block in the Pan image, and are used as coefficients for linear dependencies in the AR model of the fused MS block. This corresponds to injecting in the MS image the geometrical properties learnt from the high resolution Pan observations. The model parameters learnt from the Pan image are used in a suitable regularization framework to obtain high spatial and spectral resolution MS images. We arrive at the cost function given by equation (3.24) for each of MS images separately and minimize using simple gradient descent technique. It may be noted that the $\mathbf{z}$ in the equation has to be replaced by $\mathbf{z}_m$, where $m = 1, 2, \ldots, p$. Figure 3.17(a) shows observed MS image (Band 2). The zoomed images obtained by bicubic expansion and super-resolved images obtained using MRF model and proposed method are shown in Figure 3.17(b), (c) and (d) respectively. The roads and the traffic island are clearly visible in the Figure 3.17(d). From the figure it is clear that the fused image obtained



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

Figure 3.17: (a) Observed MS image (Band 2), (b) Zoomed MS (Band 2) image obtained using bicubic expansion, (c) Fused MS (Band 2) image obtained using MRF prior model, and (d) Fused MS (Band 2) image obtained using AR based prior nodel (proposed approach).

Image	MSE		
	Bicubic expansion	MRF model	AR model
Divya	0.003526	0.005277	0.002438
Pool	0.006508	0.005280	0.005827
MS (Band 2)	0.023230	0.192880	0.022506

Table 3.3: Performance comparison of different methods.

using the proposed method has high spatial resolution with negligible spectral distortion as compared to images obtained using the other two methods. Table 3.3. shows the quantitative comparison of performance of our method with that of the other two methods.

3.8 Conclusion

We have presented a technique to estimate decimation from the observations and recover the super-resolution intensity field from a sequence of zoomed observations by using estimated decimation matrices. The estimation of decimation requires true high resolution image and low resolution image. Since, we have multiple observations captured using different zoom setting of a camera, we use the most zoomed observation and corresponding small regions in the lesser zoomed observations to estimate decimation. The resolution of the entire scene (least zoomed observation) is obtained at the resolution of the most zoomed observation. We considered MRF model to enforce smoothness constraint while regularizing. The experimental results shows that super-resolution with decimation esti-

mation yields better results as compared to those obtained without decimation estimation. We selected the MRF model parameters using trial and error basis. Better results can be obtained using the MRF parameters estimated from the most zoomed observation. Since estimating the MRF model parameters itself is a computationally taxing task, we considered AR model to represent the spatial dependency in the super-resolved image. We have a part of the super-resolved image already available in form of the most zoomed observation, it is possible to obtain true AR parameters. We estimated AR parameters from the most zoomed observation and used it to obtain regularized solution.

We have applied the proposed technique for decimation estimation to multiresolution fusion in remotely sensed images. High spatial resolution Pan image and low resolution MS image were used to estimate the decimation. We obtained high spatial resolution MS images using regularization framework. The high resolution MS image is modeled as an AR model. Subsequently, a suitable regularization scheme is employed for the AR model. The AR prior parameters were obtained from the Pan image. Fusion results obtained using the AR prior model show better edges and textures over those obtained using MRF prior model. The qualitative illustrations of the experimental results demonstrate the effectiveness of the proposed approach.

The proposed super-resolution approach uses multiple low-quality images to produce the super-resolved image. The higher frequencies in the resulting image, which represent the newly introduced details, are in fact available in the observations in an aliased form. The proposed approach recovers these high frequencies by exploiting the various given images, each exhibiting a different aliasing effect. However, for this technique to succeed, sufficient number of low-resolution images are needed, so as to enable the recovery of the aliased frequencies. Based on the this reasoning, one might be led to the conclusion that SR based on a single measured image is impossible. Is it indeed so? The answer depends on the availability of the information that the reconstruction process needs to obtain super-resolution. One fascinating and promising alternative is to use a database of high resolution training images to recover the finer detail of the super-resolved image from it. In next chapter, we explore the learning based super-resolution technique for single image super-resolution.

Chapter 4

Learning Using LR-HR Pairs of Images

Images captured using a high spatial resolution camera provide better details that are critical in imaging applications such as medical imaging, remote sensing, surveillance. The resolution of an image, captured by a camera is determined by the spatial density of photo sensitive detectors in the camera. High resolution (HR) images can be captured using a camera fitted with high precision optics and image sensor comprising high density detectors. The cost of such a camera is prohibitively high and hence is an important concern in many commercial applications requiring HR imaging. In many imaging applications like remote surveillance, wildlife sensor network and remote sensing it is not feasible to capture the high resolution images even if the camera is capable of. This is mainly due to application specific limitations such as memory, transmission bandwidth, power and camera cost. The low cost cameras have limited optical zoom and are fitted with low memory. With such a camera one is forced to capture the images and video at limited resolution. The transmission and processing of high resolution images may still cost enough.

Consider an example of video surveillance. Surveillance cameras are used to monitor shoppers at stores, ATM users, potential interlopers, vehicular traffic and pedestrians etc. When a crime occurs, the recorded footage is often used to identify the perpetrator, the location of the incident and the sequence of events that took place. Meanwhile, video cameras are commonly embedded in mobile devices such as cell phones and PDAs, enabling users to opportunistically produce recordings of incidents that may subsequently

serve as evidence in legal proceedings. Obviously, the better the quality of a video recording, the greater its value to an investigation and subsequent legal proceedings. However, for a variety of reasons, including camera features, distance and recording speed, storage and transmission, a video recording may be of lesser quality, requiring it to be enhanced to provide adequate detail. Algorithmic approaches to obtain the high resolution image using the given low resolution (LR) observation can be useful in such applications. The low spatial resolution video captured from different cameras can be super-resolved by the use of already available LR-HR database for analysis at a later time.

In this work, we demonstrate that it is possible to super-resolve a low resolution image captured using a low cost camera. This is definitely advantageous as one can obtain high resolution images/video by using a low resolution camera fitted with a limited memory. Present day high end cameras have options to capture images and video at different spatial resolution. Hence it is possible to click a large number of LR-HR images offline and store them on a computer. This is a one time operation. This database can be used to learn the true relationship of the spatial features of an image across the scales. We propose new learning technique and solve the single frame super-resolution problem using stochastic SR reconstruction method. The proposed technique can also be used in image/video compression for transmission over a channel with limited bandwidth. One can transmit the compressed LR images and obtain high resolution at the receiver end by using a set of training pairs.

The natural scene may contain regions with a wide variety of textures and may have continuously changing textures. Since, the homogeneous MRF model prior tends to over-smooth the super-resolution reconstructions, it has to be replaced by a different model in favor of preserving edges and textures. Inhomogeneous Gaussian random fields have been investigated by Aykroyd [153] using simple simulated examples. In this chapter we propose an inhomogeneous Gaussian Markov random field as a prior model for super-resolution reconstruction. The simplicity of the Gaussian model allows rapid calculation, and the flexibility of the spatially varying prior parameter allows varying degrees of spatial smoothing. We demonstrate that the proposed procedures lead to more accurate reconstruction than edge-preserving homogeneous alternatives discussed in previous chapter. The inhomogeneous model allows greater flexibility; small features are not masked by

the smoothing, and constant regions obtain sufficient smoothing to remove the effects of noise.

4.1 Related Work

In recent times, learning-based super-resolution algorithms have attracted much attention. In these algorithms the priori information is derived from the training database [85], [106], [107], [105], [108], [4], [121], [120], [110]. Based on the framework of Freeman *et al.* [84], Kim and Kwon investigate a regression-based approach for single-image super-resolution [4]. Here the authors generate a set of candidates for each pixel using patch-wise regression and combine them based on the estimated confidence for each pixel. In the post processing step, they employ regularization technique using discontinuity preserving prior. The disadvantage of all the above approaches is that they either obtain the LR images in the database by downsampling the high resolution images i.e., simulate the LR images or use an interpolated version of the LR image while searching. Such a database do not represent the true spatial features relationship between LR-HR pairs as they do not correspond to the images captured by a real camera. Jiji *et al.* demonstrate super-resolution of a single frame gray scale image using a training database consisting of high resolution images downloaded from the internet [3]. They learn the high frequency details of the SR image from the database and obtain regularized solution by employing Markov Random Field (MRF) prior model and a wavelet prior. Our work in this paper is based on their work. However, we use a different approach for learning as well as for regularization.

In this work, we present a learning based approach for super-resolving an image using single observation. First, we learn fine details of the super-resolved image from a database and obtain initial estimate of super-resolved image. We construct the database by capturing both the low resolution as well as their high resolution versions using a real camera. Thus we make use of the true transformation that exists between the LR and HR images while learning. We use discrete wavelet transform (DWT) based method to learn the high frequency contents. We then model the unknown high resolution image as an inhomogeneous Gaussian MRF (IGMRF) and estimate the model parameters using the initial HR estimate. The aliasing (decimation) matrix entries used in the im-

age formation model are also estimated using the same initial estimate. The final HR estimate is obtained by using the MAP formulation. The method is extended to color image super-resolution where we super-resolve the luminance component using proposed learning based approach and then interpolate the chrominance components in the wavelet domain in order to obtain super-resolved color image.

4.2 Block Diagram Description of the Approach

The proposed technique of learning based super-resolution is illustrated by the block diagram shown in Figure 4.1. Given a low resolution observation (test image), we learn its high frequency contents from a database consisting of a set of low resolution images and their high resolution versions. It may be noted that the LR images are not constructed by downsampling the HR images as is done by most of the learning based approaches. Instead they are captured by using a real camera comprised of various resolution settings and hence represent the true LR-HR versions of the scenes. In order to learn the high frequency components, we consider discrete wavelet transform based method. The transform coefficients corresponding to the high frequency contents are learned from the database and an initial estimate for the high resolution version of the observation is obtained by taking the inverse DWT. This initial HR estimate is used for decimation estimation as well as for estimating the IGMRF parameters. The estimated decimation models the aliasing due to undersampling and the IGMRF parameters inject the geometrical properties in test image corresponding to the high resolution image. We then use an MAP estimation to arrive at a cost function consisting of data fitting term and the prior term. A suitable optimization is exploited to minimize the cost function. The minimization leads final super-resolved image. We extend this method to color image super-resolution where we super-resolve the luminance component using the proposed method and use the interpolation in wavelet domain for chrominance components. The luminance and chrominance components are then combined to obtain the super-resolution. We mention here that although we use a large number of LR-HR images in the database, it is not possible to capture the true spatial features for the high resolution image using the learning. Also the use of wavelet transform limits the learning of edges in the horizontal, vertical and diagonal directions only. Hence we need regularization in order to obtain a better

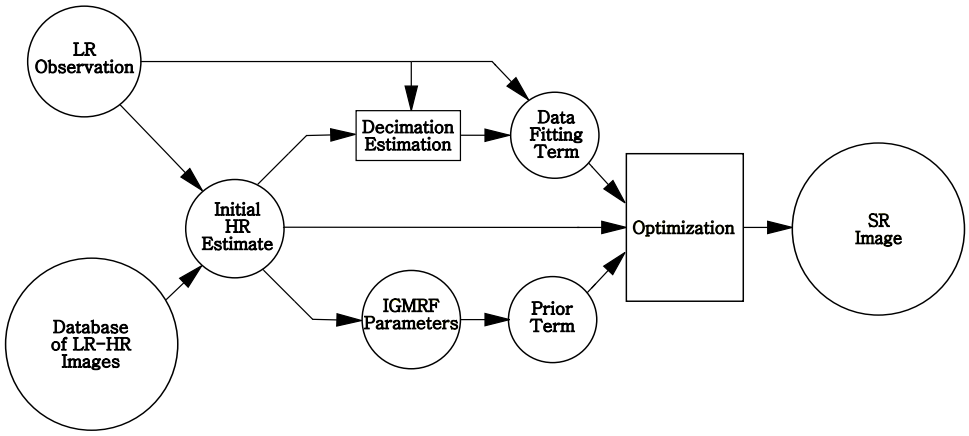


Figure 4.1: Schematic representation of proposed approach for image super-resolution [2]. Here LR, HR and SR stand for Low Resolution, High Resolution and Super-Resolution, respectively. Here, IGMRF represents Inhomogeneous Gaussian Markov Random Field.

solution.

The main highlights of the approach are as under:

- We learn the wavelet coefficients that correspond to the high frequency contents of the super-resolved image from a database consisting of a set of low resolution images and their high resolution versions. Since the construction of the database is a one time and offline operation, we can use the computer memory for the storage of database images. This allows us to capture a large number of images, even when the memory of the camera is limited.

- The database consists of LR and HR images both captured by varying the resolution setting of a camera. Such pairs truly represent the spatial features relationship between low resolution image and its high resolution versions.

- An inhomogeneous Gaussian MRF is used to model the super-resolved image field. The advantage in using this model is that it is adaptive to the local structures in an image and hence eliminates the need for a separate edge preserving prior.

- We estimate the aliasing matrix entries and the IGMRF parameters using the initial HR estimate. Since we use true LR-HR pairs for learning the initial HR estimate we expect that the estimated parameters are close to their true values.

- For color image super-resolution, we apply the proposed method for super-resolving luminance component and suggest a wavelet domain interpolation approach for the

chrominance components.

- While using the edge preserving IGMRF prior we employ a simple gradient descent approach and thus avoid the use of computationally taxing optimization techniques such as simulated annealing.

4.3 Learning the Initial HR Estimate

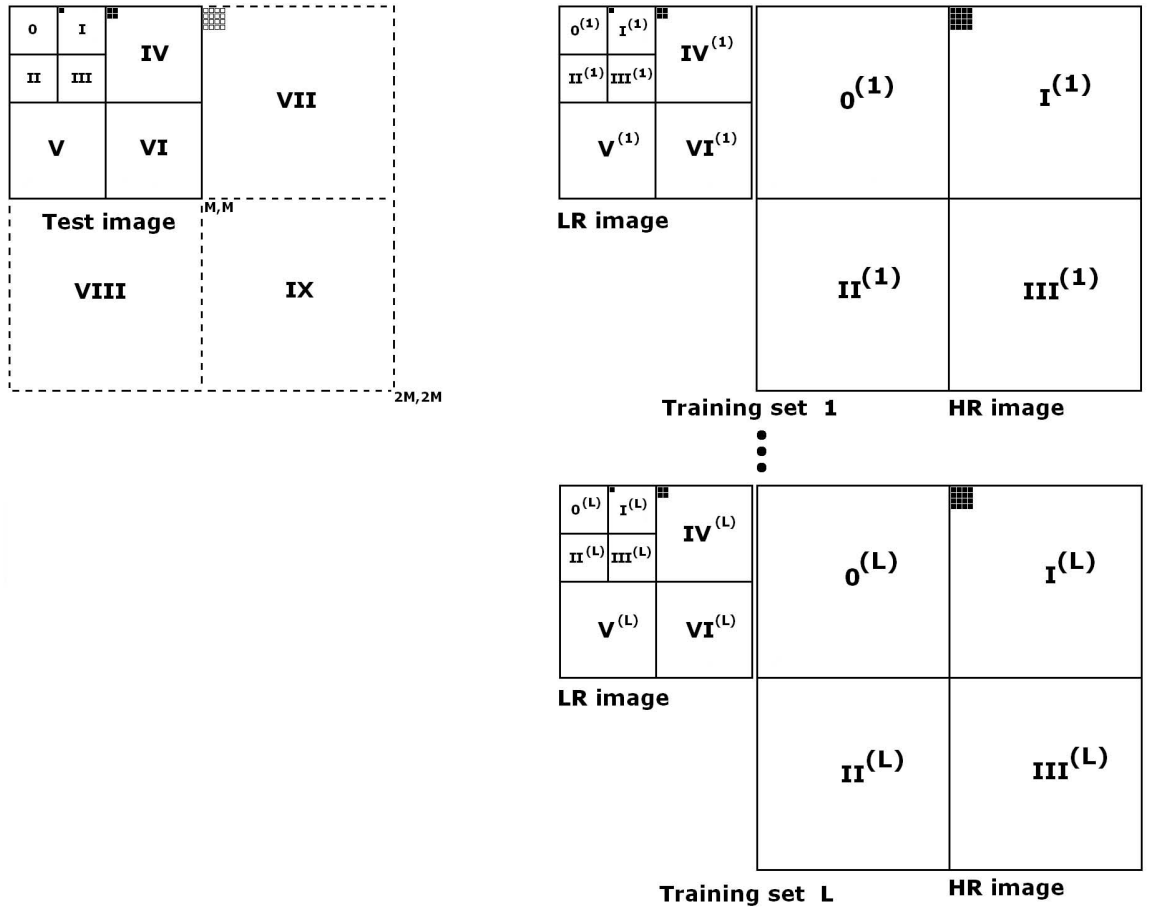
In this section we discuss the new learning technique in order to obtain the initial HR estimate. It may be noted that the learning based approach proposed in [3] uses a database consisting of high resolution images only and these images are downloaded from the internet. The main drawback of this approach is that the use of the images downloaded from the internet does not guarantee that these images indeed represent high resolution database. They may represent the upsampled versions of the low resolution images obtained using standard interpolation techniques. Also, these may represent the collection of images captured using different cameras with different hardware configurations. All this contributes for errors in the SR estimation. Our approach differs from their approach. We use a training set of LR-HR images covering a wide range of scenes. These images are captured by adjusting the resolution setting of a real camera. It is of interest to note that our database do not contain the LR images synthesized using the downsampling operation as used by the other learning based approaches. Truly, this makes the algorithm capable of super-resolution for the cases when the HR ground truth is not available.

In this work, we learn the transform coefficients for the initial estimate of the super-resolved image for a decimation (upsampling) factor of $q = 2$ and $q = 4$. The database for leaning consists of a large number sets of LR and HR images covering indoor scenes as well as outdoor scenes taken at different times and with different lighting conditions. For a decimation factor $q = 2$, a set consists of two images (LR and HR) for each of the captured scenes. Similarly, for a decimation factor of $q = 4$, there are three images in a set for each scene having different resolutions. Before we describe the method for learning it is important to mention the following points. For a decimation factor of $q = 4$, we first learn the initial estimate for $q = 2$ using the database consisting of LR-HR pairs with a resolution factor of 2. We then use this estimate as the test image for $q = 4$. We thus apply the single octave learning algorithm in two steps in order to

obtain image super-resolution for $q = 4$. The reason for the two step operation is as follows. In a multi-resolution system, every coefficient at a given scale can be related to a set of coefficients at the next coarser scale of similar orientation [154]. The Lipschitz property states that near sharp edges, the wavelet coefficients changes exponentially over the scales [155][156]. Hence the error between the coefficient in test image and its best matching coefficient from the LR training image increases exponentially when one learns the wavelet coefficients using a database of LR-HR pairs with a resolution difference of four. In the proposed single octave approach this error adds up linearly over the scales. Thus the error propagation in the proposed approach is linear and the learning of wavelet coefficients is more accurate in comparison with the learning using one step operation. It may be noted here that the two step operation of the single octave algorithm restricts the super-resolution to powers of 2 only.

4.3.1 Learning Discrete Wavelet Transform (DWT) Coefficients

Now we describe the approach for DWT based learning for a decimation factor of 2 ($q = 2$). We use two level wavelet decomposition of the test image for learning the wavelet coefficients at the finer scale. The LR training images in the database are also decomposed into two levels, while their HR versions are decomposed into one level. The reason for taking one level decomposition for HR training images is as follows. With one level decomposition the subband LL represents the scaled version of the HR image and the subbands LH, HL, and the HH represent the detailed coefficients (vertical, horizontal and diagonal edges) at the high resolution. This means that for $q = 2$, both the LR image (subband LL) and the edge details at finer scales (subbands LH, HL, and HH) are available in HR transformed image. This motivates us to compare the edges in the test image with those present in the LR training set and choose the best matching wavelet coefficients from the HR images. Thus given an LR test image, we learn its edges (high frequency content) at finer scale using these LR-HR training images. Figure 4.2 illustrates the block schematic for learning the wavelet coefficients of the test image at finer scales using a set of L training image pairs for a decimation factor of 2. Figure 4.2(a) shows the subbands $0 - VI$ of the low resolution test image and Figure 4.2(b) displays the subbands $0^{(m)} - VI^{(m)}$, $m = 1, \ldots, L$ of the LR training images and subbands $0^{(m)} - III^{(m)}$, of



(a) (b)

Figure 4.2: Illustration of learning of wavelet coefficients at a finer scale using a database of LR-HR image pairs. (a) Test image (LR observation) with a two level wavelet decomposition. Wavelet coefficients (marked as hollow squares) are to be estimated for the subbands shown with the dotted lines. (b) A training set of LR and HR images in the wavelet domain (LR training images are decomposed into two level and the HR training images into one level).

the HR training images. We compare the coefficients in subbands I to VI of the test image with those in subbands $I^{(m)}$ to $VI^{(m)}$ of the LR training images and obtain the best matching coefficients for subbands VII, $VIII$ and IX of the test image. Here the test image and the LR training images are of size $M \times M$ pixels. The HR training images have a size of $2M \times 2M$ pixels. L is the number of sets of LR-HR training images in the database.

The learning procedure is as follows. Let $\psi(i, j)$ be the wavelet coefficient at a location (i, j) in subband 0, where $0 \leq i, j < M/4$. The wavelet coefficients $\psi(i, j + M/4)$, $\psi(i + M/4, j)$ and $\psi(i + M/4, j + M/4)$ are the wavelet coefficients corresponding to subbands I, II and III, respectively. These wavelet coefficients and a 2×2 block

consisting of $\{\psi(k,l+M/2)_{k=2i,l=2j}^{k=2i+1,l=2j+1}\}$, $\{\psi(k+M/2,l)_{k=2i,l=2j}^{k=2i+1,l=2j+1}\}$ and $\{\psi(k+M/2,l+M/2)_{k=2i,l=2j}^{k=2i+1,l=2j+1}\}$ in each of the subbands IV-VI are then considered to learn a 4×4 wavelet block in each of the subbands VII-IX. For each of the subbands IV-VI in the test image, we need to learn $4\times 4 = 16$ coefficients for corresponding location in subbands $I-III$. Thus for every location (i,j) in subband 0, we learn a total $16\times 3 = 48$ coefficients for the subbands VII-IX. We search for the LR training image that has a bast match with the test image by comparing the wavelet coefficients in the subbands I-VI in the minimum absolute difference (MAD) sense. The corresponding wavelet coefficients from the subbands I-III of the HR training image are then copied into the subbands VII-IX of the test image. For a given location (i,j) the best matching LR training image in the subbands $I-III$ is found by using the following equation for MAD.

$$\begin{aligned}
\hat{c}(i,j) &= \underset{m}{argmin}\Bigg[\Big|\psi(i,j+\frac{M}{4})-\psi^{(m)}(i,j+\frac{M}{4})\Big| \\
&+ \Big|\psi(i+\frac{M}{4},j)-\psi^{(m)}(i+\frac{M}{4},j)\Big| \\
&+ \Big|\psi(i+\frac{M}{4},j+\frac{M}{4})-\psi^{(m)}(i+\frac{M}{4},j+\frac{M}{4})\Big| \\
&+ \sum_{k=2i}^{k=2i+1}\sum_{l=2j}^{l=2j+1}\Big|\psi(k,l_1)-\psi^{(m)}(k,l_1)\Big| \\
&+ \sum_{k=2i}^{k=2i+1}\sum_{l=2j}^{l=2j+1}\Big|\psi(k_1,l)-\psi^{(m)}(k_1,l)\Big| \\
&+ \sum_{k=2i}^{k=2i+1}\sum_{l=2j}^{l=2j+1}\Big|\psi(k_1,l_1)-\psi^{(m)}(k_1,l_1)\Big|\Bigg],
\end{aligned} \tag{4.1}$$

where $k_1 = k+\frac{M}{2}$, $l_1 = l+\frac{M}{2}$, and $\hat{c}(i,j)$ is an index to the best matching LR image in the database for the location (i,j) and $1\le\hat{c}(i,j)\le L$. Here $\psi^{(m)}(i,j)$, $m=1,2,\ldots,L$, denotes the wavelet coefficient for m^{th} training image at location (i,j). For each location in subband $I-III$ of the low resolution observation, a best fit 4×4 block of wavelet coefficients in subbands $I-III$ from the HR image of the training pairs given by $\hat{c}(i,j)$

are copied into subbands $VII, VIII$ and IX of the test image. Thus we have,

$$\begin{aligned}
\{\psi(s,t_1)_{s=4i,t=4j}^{s=i_1,t=j_1}\} &:= \{\psi^{(\hat{c}(i,j))}(s,t_1)_{s=4i,t=4j}^{s=i_1,t=j_1}\} \\
\{\psi(s_1,t)_{s=4i,t=4j}^{s=i_1,t=j_1}\} &:= \{\psi^{(\hat{c}(i,j))}(s_1,t)_{s=4i,t=4j}^{s=i_1,t=j_1}\} \\
\{\psi(s_1,t_1)_{s=4i,t=4j}^{s=i_1,t=j_1}\} &:= \{\psi^{(\hat{c}(i,j))}(s_1,t_1)_{s=4i,t=4j}^{s=i_1,t=j_1}\}.
\end{aligned} \tag{4.2}$$

Here $s_1 = s + M$, $t_1 = t + M$, $i_1 = 4i + 3$, and $j_1 = 4j + 3$. This completes the learning process. The inverse wavelet transform of the learned HR image then gives initial HR estimate. This HR estimate provides a sufficiently good approximation of the ground truth and its properties enable robust estimation of decimation and adaptive parameters needed for regularization.

4.3.2 Analysis of Computational Complexity

We now analyze the computational complexity of the proposed approach for single octave operation. We learn the wavelet coefficients corresponding to high frequency using a database consists of L pairs of LR-HR images. Let the test image of size $M \times M$ be decomposed into W levels of wavelet transform. As a result of this decomposition, the transformed image has $3W$ subbands of details and one subband of the approximation. There are W subbands for each of vertical, horizontal and diagonal details. Consider the subbands corresponding to vertical details. We learn the vertical details at finer scale using the vertical details across all coarser scales. In other words, we use wavelet coefficients in all the vertical details subbands to find best matching coefficients from the database. The size of a subband at W^{th} level is $\frac{M}{2^W} \times \frac{M}{2^W}$ and it consists of $(\frac{M}{2^W})^2$ coefficients. We need to learn finer coefficients for each of the coefficients for the coarse subband. For each of the coefficients in this subband, the best matching coefficients at finer level can be searched by comparing corresponding $\sum_{p=0}^{p=W-1} 2^{2p}$ coefficients in the vertical detail subbands. These coefficients are compared with coefficients at corresponding locations in the each of the LR images in the database. Thus number of comparison required for learning vertical details is,

$$L(\frac{M}{2^W})^2 \sum_{p=0}^{p=W-1} 2^{2p}. \tag{4.3}$$

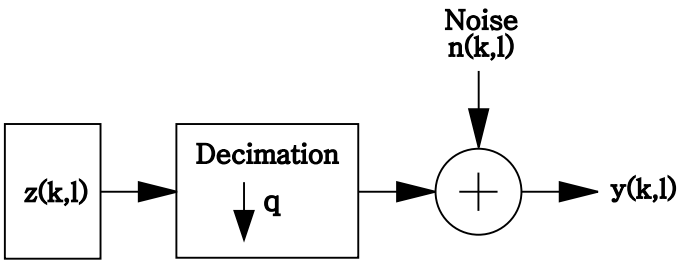


Figure 4.3: Image formation model. Here $\downarrow$ symbol represents decimation and q represents the decimation factor.

Similar comparisons are used to find best matching coefficients at finer level of horizontal and diagonal subbands. Hence total number of comparison operations for learning all the details amounts to,

$$3L(\frac{M}{2^W})^2 \sum_{p=0}^{p=W-1} 2^{2p}. \tag{4.4}$$

In our experiment, we use two level decomposition ($W = 2$) of the test image of size 64×64 and learn the coefficients using the database of 750 pairs ($L = 750$) of LR-HR images. In this case, the number of comparisons required are $3 \times 750 \times (\frac{64}{2^2})^2 \times (2^0 + 2^2)$ $= 2\,880\,000$. Although this involves number of computations, it is not computationally taxing to present day high performance computers as the process is not iterative.

4.4 Forward Model and Decimation Estimation

We propose a super-resolution algorithm that attempts to estimate high resolution image from single low resolution observation. This is an inverse problem. Solving such a problem needs a forward model that represents the image formation process. We consider a linear forward model for the image formation which can be written as,

$$\mathbf{y} = D\mathbf{z} + \mathbf{n}. \tag{4.5}$$

Here, the observed image Y is of size $M \times M$ pixels and $\mathbf{y}$ represents the lexicographically ordered vector of size $M^2 \times 1$, which contains the pixels from image Y. Similarly, $\mathbf{z}$ is actual HR image. The decimation matrix D takes care of aliasing. For an integer decimation factor of q, the decimation matrix D consists of q^2 non-zero elements along each row at appropriate locations. Here $\mathbf{n}$ is the independent and identically distributed

(i.i.d.) noise vector with zero mean and variance σ_n^2. It has same size as $\mathbf{y}$. The pictorial representation of this observation model is shown in Figure 4.3. The multivariate noise probability density is given by,

$$P(\mathbf{n}) = \frac{1}{(2\pi\sigma_n^2)^{\frac{M^2}{2}}} e^{-\frac{1}{2\sigma_n^2}\mathbf{n}^T\mathbf{n}}. \tag{4.6}$$

Our problem is to estimate $\mathbf{z}$ given $\mathbf{y}$, which is an ill-posed inverse problem.

Generally, the decimation model to obtain the aliased pixel intensities from the high resolution pixels, for a decimation factor of q, has the form [48],

$$D = \frac{1}{q^2}\begin{pmatrix} 1 \;\; 1 \ldots 1 & & & \mathbf{0} \\ & 1 \;\; 1 \ldots 1 & & \\ & & & \\ \mathbf{0} & & & 1 \;\; 1 \ldots 1 \end{pmatrix}. \tag{4.7}$$

The decimation matrix in equation (4.7) indicates that a low resolution pixel intensity $y(i,j)$ is obtained by averaging the intensities of q^2 pixels corresponding to the same scene in the high resolution image and adding noise intensity $n(i,j)$. Due to different lighting conditions and different scene contents the aliasing differs from image to image. We cannot consider the aliasing effect as averaging effect. Hence we estimate the same. Since the initial high resolution estimate is available we make use of the same to obtain the decimation matrix entries and thus learn the aliasing better. The decimation matrix of the form shown in equation (4.7), can be modified as,

$$D = \begin{pmatrix} a_1 \;\; a_2 \;\; \ldots \;\; a_{q^2} & & & \mathbf{0} \\ & a_1 \;\; a_2 \;\; \ldots \;\; a_{q^2} & & \\ & & & \\ \mathbf{0} & & & a_1 \;\; a_2 \;\; \ldots \;\; a_{q^2} \end{pmatrix}, \tag{4.8}$$

where $|a_i| \leq 1, i = 1, 2, \ldots q^2$. When we use equation (4.7) each of the a_i has a value of $\frac{1}{q^2}$. However, in the above equation the estimates of a_i are based the LR observation and the initial HR estimate. Thus the estimated a_i are more accurate as compared to using $a_i = \frac{1}{q^2}$ in equation (4.7) and are closer to the true values for the chosen model. It

may be noted that this form of decimation matrix D implicitly contains moving average (space invariant) blur in the downsampling process. The estimation of aliasing requires low resolution image and its high resolution version. The main difficulty of alias estimation comes from the fact that we cannot access the ground truth, i.e. the original high resolution image. The initial HR estimate obtained using DWT based learning approach proposed in section 4.3 provides a sufficiently good approximation of the ground truth, we use it to estimate decimation.

4.5 Image Field Model

As discussed in section 1.3.2 super-resolution is an ill-posed inverse problem. There are infinite solutions to equation (4.5). A reasonable assumption about the nature of the true image makes the ill-posed problem into better posed and this leads to a better solution. Selection of the appropriate model as the prior helps us to obtain a better solution.

Over recent years, the use of homogeneous prior models in image processing has become widely accepted. The homogeneous MRF model prior tends to oversmooth the super-resolution reconstructions. In the MRF model prior based HR reconstruction schemes, there is a fundamental trade-off between smoothness of the super-resolved image and the amount of noise or visually unappealing artifacts. This occurs because the solution penalizes discontinuities in the image. Discontinuities can be outlier pixels, restoration artifacts, or noise, but they can also be real edges that are present in the image. Since the simplest Gaussian model tends to oversmooth reconstructions, it has been rejected in favor of various edge-preserving alternatives. The problem is not with the Gaussian family, but rather with the assumption of homogeneity. A more efficient model is one which considers that only homogeneous regions are smooth and that edges must remain sharp. This motivates us to consider an inhomogeneous prior which can adapt to the local structure of the image in order to provide a better reconstruction.

4.5.1 Inhomogeneous Gaussian Markov Random Field Prior Model

In this chapter, we propose an inhomogeneous Gaussian Markov random field as a prior model for super-resolution. Inhomogeneous Gaussian random fields have been investigated by Aykroyd [153]. The simplicity of the Gaussian model allows rapid calculation,

and the flexibility of the spatially varying prior parameter allows varying degrees of spatial smoothing. The inhomogeneous model allows greater flexibility; small features are not masked by the smoothing, and constant regions obtain sufficient smoothing to remove the effects of noise. This also helps to eliminate the need for separate priors to preserve edges as well as smoother regions in an image. Here, we propose an inhomogeneous Gaussian Markov random field as a prior model for super-resolution reconstruction. We demonstrate that the proposed procedures lead to more accurate reconstruction than edge-preserving homogeneous alternatives discussed in previous chapter.

The authors in [157] model the super-resolved image by an inhomogeneous Gaussian MRF with an energy function that allows us to adjust amount of regularization locally. They define corresponding energy function as,

$$U(\mathbf{z}) = \sum_{i,j} \left[b^x_{i,j} (\mathcal{D}_x \mathbf{Z})^2_{i,j} + b^y_{i,j} (\mathcal{D}_y \mathbf{Z})^2_{i,j} \right], \tag{4.9}$$

where $\mathcal{D}_x$ and $\mathcal{D}_y$ are first order derivative operators with respect to rows and columns and $\mathbf{Z}$ is the super-resolved image. Here, $b^x_{i,j}$ and $b^y_{i,j}$ are the IGMRF parameters at location (i, j) for vertical and horizontal directions, respectively. In the above energy function, the authors model the spatial dependency at a pixel location by considering a first order neighborhood thus considering edges occurring in the horizontal and vertical directions only. However, in practice there may be diagonal edges in the reconstructed image. In order to take care of these edges we consider a second order neighborhood and modify the energy function as follows.

$$U(\mathbf{z}) = \sum_{i,j} \left[b^x_{i,j} (\mathcal{D}_x \mathbf{Z})^2_{i,j} + b^y_{i,j} (\mathcal{D}_y \mathbf{Z})^2_{i,j} + b^g_{i,j} (\mathcal{D}_g \mathbf{Z})^2_{i,j} + b^h_{i,j} (\mathcal{D}_h \mathbf{Z})^2_{i,j} \right]. \tag{4.10}$$

Here $b^g_{i,j}$ and $b^h_{i,j}$ are the IGMRF parameters at location (i, j) for diagonal directions. A low value of b indicates the presence of edge between two pixels. These parameters help to obtain a solution which is less noisy in smooth areas and preserve sharp details in other areas. Now, in order to estimate the IGMRF parameters we need the true super-resolved image which is not available and has to be estimated. Therefore an approximation of $\mathbf{Z}$ has to be accurately determined if we want the parameters obtained from it to be significant for regularization. This is why we choose to use the learning based approach

to compute the close approximation of $\mathbf{Z}$. Since the results of this approach exhibit sharp textures and are sufficiently close to the original image it enables us to estimate the adaptive parameters from it.

4.5.2 Estimation of IGMRF Parameters

The maximum likelihood estimate on complete data with respect to the original image $\mathbf{Z}$ is,

$$\hat{b}_{i,j}^{x,y,g,h} = \underset{b_{i,j}^{x,y,g,h}}{argmax} \left[logP(\mathbf{Z}|b^x, b^y, b^g, b^h) \right], \tag{4.11}$$

and the log-likelihood derivatives are,

$$\frac{\partial logP(\mathbf{Z}|b^x, b^y, b^g, b^h)}{\partial b_{i,j}^{x,y,g,h}} = E_{\mathcal{Z}}\left[(\mathcal{D}_{x,y,g,h}\mathcal{Z})_{i,j}^2 \right] - (\mathcal{D}_{x,y,g,h}\mathbf{Z})_{i,j}^2, \tag{4.12}$$

where $\mathcal{Z}$ corresponds to the maximum *a posteriori* estimate of the high resolution image. Therefore the estimation problem consists of solving system,

$$\{E_{\mathcal{Z}}\left[(\mathcal{D}_{x,y,g,h}\mathcal{Z})_{i,j}^2 \right] = (\mathcal{D}_{x,y,g,h}\mathbf{Z})_{i,j}^2)\}. \tag{4.13}$$

It can be seen that the expectation term only depends on the parameters and the other term only depends on $\mathbf{Z}$. This simplifies the estimation problem. It is sufficient to compute the variance of each pixel difference with respect to the prior law $E_{\mathcal{Z}}\left[(\mathcal{D}_{x,y,g,h}\mathbf{Z})_{i,j}^2 \right]$. The authors in [157] propose the simplest approximation of the local variance. The variance of the gradient $(\mathcal{D}_{x,y,g,h}\mathbf{Z})_{i,j}$ is equal to the variance of the same gradient in the homogeneous i.e. when all the parameters are equal to the corresponding $b_{i,j}^{x,y,g,h}$. Since the covariance matrix of the homogeneous prior distribution is diagonalized by a Fourier transform, this variance can be calculated and is equal to $\frac{1}{4b}$ [158]. This gives,

$$\hat{b}_{i,j}^{x,y,g,h} = \frac{1}{4(\mathcal{D}_{x,y,g,h}\mathbf{Z})_{i,j}^2}. \tag{4.14}$$

Since the true high resolution image $\mathbf{Z}$ is not available we use the close approximation $\mathbf{Z}_0$ obtained using the DWT based learning approach and obtain the parameters using,

$$\hat{b}_{i,j}^{x,y,g,h} = \frac{1}{4(\mathcal{D}_{x,y,g,h}\mathbf{Z}_0)_{i,j}^2}. \tag{4.15}$$

The refined estimates of the IGMRF prior parameters are obtained using the following equations.

$$\begin{aligned}
\hat{b}^{x}_{i,j} &\simeq \frac{1}{8[(z_0(i,j) - z_0(i-1,j))^2]}, \\
\hat{b}^{y}_{i,j} &\simeq \frac{1}{8[(z_0(i,j) - z_0(i,j-1))^2]}, \\
\hat{b}^{g}_{i,j} &\simeq \frac{1}{8[(z_0(i,j) - z_0(i-1,j+1))^2]}, \\
\hat{b}^{h}_{i,j} &\simeq \frac{1}{8[(z_0(i,j) - z_0(i-1,j-1))^2]},
\end{aligned} \tag{4.16}$$

where $z_0(i,j)$ is the pixel intensity of the initial estimate at location (i,j). Thus we estimate four parameters at each pixel location. These parameters cannot be approximated from degraded versions of original image. The parameters estimated from the blurred image have high values which leads to oversmooth solution and the parameters estimated from the noisy image are of very low values that leads to noisy solutions. Hence we use the already learned high resolution estimation in order to obtain a better estimate of these parameters. In order to avoid computational difficulties, we set an upper bound $\hat{b} = \frac{1}{8}$ whenever the gradient becomes zero, i.e. whenever the neighboring pixel intensities are the same. Thus we set a minimum spatial difference of 1 for practical reasons. This avoids obtaining high regularization parameter that would slow down the optimization. It ensures that the pixels with zero intensity difference are weighted almost same as those with small intensity difference (in this case with a pixel intensity difference of one).

4.6 Super-resolution Estimation

We now explain how an MAP estimation of the dense intensity field (super-resolved image) can be obtained. The IGMRF model on the super-resolved image serves as the prior for the MAP estimation in which the prior parameters are already known. The data fitting term is derived from the forward model which describes the image formation process. The data fitting term contains the decimation matrix estimated using the initial HR image and the test image. In order to use maximum *a posteriori* estimation to

super-resolve the test image, we need to obtain the estimate as,

$$\hat{\mathbf{z}} = \underset{z}{^{argmax}} P(\mathbf{z}|\mathbf{y}). \tag{4.17}$$

Using the Bayes' rule we can write,

$$\hat{\mathbf{z}} = \underset{z}{^{argmax}} \frac{P(\mathbf{y}|\mathbf{z})P(\mathbf{z})}{P(\mathbf{y})}. \tag{4.18}$$

Since the denominator is not a function of $\mathbf{z}$, equation (4.18) can be written as,

$$\hat{\mathbf{z}} = \underset{z}{^{argmax}} P(\mathbf{y}|\mathbf{z})P(\mathbf{z}). \tag{4.19}$$

Now taking the *log* we can write,

$$\hat{\mathbf{z}} = \underset{z}{^{argmax}} [logP(\mathbf{y}|\mathbf{z}) + logP(\mathbf{z})]. \tag{4.20}$$

Finally, using (4.5) and (4.10), the final cost function to be minimized can be expressed as,

$$\hat{\mathbf{z}} = \underset{z}{^{argmin}} \left[\frac{\| \mathbf{y} - D\mathbf{z} \|^2}{2\sigma_n^2} + U(\mathbf{z}) \right]. \tag{4.21}$$

In (4.21), the first term ensures the fidelity of the final solution to the observed data through the image formation model. The second term is inhomogeneous smoothness prior. Since this cost function is convex, it can be easily minimized using a simple gradient descent optimization technique, which quickly leads to the minima. This optimization process is an iterative method and the choice of initial solution fed to the optimization process determines the speed of convergence. Use of a close approximate to the solution as an initial estimate speed-up the optimization process. In order to provide good initial guess, we use the already learned HR estimate.

4.7 Applying the Algorithm to Color Images

Different image processing systems use different color models for different reasons. The RGB color space consists of three additive primary colors; red, green and blue. The RGB model is widely used in color monitors, computer graphic systems, electronic displays and

digital storage. Although this model simplifies the design of computer graphic systems, it is not suitable for super-resolution algorithms. The reason behind this is that the red, green and blue color components are highly correlated. This makes it difficult to difficult to apply the monochrome super-resolution technique to each of the R, G and B color components with maintaining the natural correspondences between the color components in the solution. The imbalanced correspondences between the color components produces certain artifacts in super-resolved color image. In addition, applying super-resolution techniques to each of these components separately increases the computational burden.

4.7.1 YC_bC_r Color Space

In order to avoid the drawbacks of RGB color space, we separate the luminance channel and chrominance components by applying color space transformation and represent the color image in YC_bC_r color space. The YC_bC_r color model represents the color image using separate luminance component and chrominance (color) components. The luminance is encoded in the Y and the blueness and redness are encoded in C_b and C_r, respectively. Following equations show the conversion from an RGB image to a YC_bC_r image.

$$\begin{aligned} Y &= 0.29900R + 0.58700G + 0.11400B, \\ C_b &= -0.16874R - 0.33126G + 0.50000B, \\ C_r &= 0.50000R - 0.41869G - 0.08131B. \end{aligned} \tag{4.22}$$

Similarly a YC_bC_r image can be converted to an RGB image using,

$$\begin{aligned} R &= 1.00000R + 0.40200C_r, \\ G &= 1.00000R - 0.34414C_b - 0.71414C_r, \\ B &= 1.00000R + 1.77200C_b. \end{aligned} \tag{4.23}$$

Since the human eye is more sensitive to the details in the luminance component of an image than the details in the chrominance component, we super-resolve the luminance

component Y using the proposed approach and expand the chrominance components using a simple interpolation technique. The authors in [159] propose an interpolation technique for gray scale images by interpolating the wavelet coefficients at finer scale. We apply their approach for expanding the chrominance components C_b and C_r. The frequency domain interpolation of these components leads to enhanced edges as compared to spatial interpolation methods like bilinear interpolation or bicubic interpolation. We use the super-resolved luminance component and the interpolated chrominance components to obtain super-resolved color image by converting YC_bC_r to RGB color space. In the following section, we describe the frequency domain interpolation of the chrominance components of a color image.

4.7.2 Interpolation of the Chrominance Components

Now we describe frequency domain interpolation of the chrominance components. We take two level wavelet decomposition of the chrominance components as shown in Figure4.4 and interpolate wavelet coefficients for finer scale i.e. subbands $VII - IX$. We exploit the idea from zero tree concept, i.e., in a multi-resolution system, every coefficient at a given scale can be related to a set of coefficients at the next coarser scale of similar orientation [154]. This can be used to find the wavelet coefficients at finer scale. Thus to interpolate the wavelet coefficients in subband VII, we relate the coefficients in subbands I and IV and calculate the ratios $d_k(i,j)$, $k = 1, \ldots, 4$ as described below.

Let $\psi(i,j)$ be the wavelet coefficient at a location (i,j), where $0 \leq i,j < M/4$. The wavelet coefficients $\psi(i, j+M/4)$, $\psi(i+M/4, j)$ and $\psi(i+M/4, j+M/4)$ are the wavelet coefficients corresponding to subbands I, II and III respectively. Consider subbands I and IV for interpolating coefficients in subband VII. The four wavelet coefficients $\{\psi(k, l+M/2)\}_{k=2i,l=2j}^{k=2i+1,l=2j+1}$ in subband IV are related to a single coefficient $\psi(i, j+M/4)$ in subband I. We define the four ratios d_k, $k = 1, \ldots, 4$, of the four coefficients in subband

IV and related one wavelet coefficient in subband I as,

$$\begin{aligned} d_1(i,j) &= \frac{\psi(2i, 2j + \frac{M}{2})}{\psi(i, j + \frac{M}{4})}, \\ d_2(i,j) &= \frac{\psi(2i, 2j + 1 + \frac{M}{2})}{\psi(i, j + \frac{M}{4})}, \\ d_3(i,j) &= \frac{\psi(2i + 1, 2j + \frac{M}{2})}{\psi(i, j + \frac{M}{4})}, \\ d_4(i,j) &= \frac{\psi(2i + 1, 2j + 1 + \frac{M}{2})}{\psi(i, j + \frac{M}{4})}. \end{aligned}$$

A single coefficient $\psi(2i, 2j + \frac{M}{2})$ in subband IV corresponds to four coefficients $\{\psi(k, l + M)\}_{k=4i,l=4j}^{k=4i+1,l=4j+1}$ in subband VII. We calculate these four coefficients by multiplying $\psi(2i, 2j + \frac{M}{2})$ with four ratios $d_k(i,j)$, $k = 1, \ldots, 4$ as,

$$\begin{aligned} \psi(4i, 4j + M) &= d_1(i,j) \times \psi(2i, 2j + \frac{M}{2}), \\ \psi(4i, 4j + 1 + M) &= d_2(i,j) \times \psi(2i, 2j + \frac{M}{2}), \\ \psi(4i + 1, 4j + M) &= d_3(i,j) \times \psi(2i, 2j + \frac{M}{2}), \\ \psi(4i + 1, 4j + 1 + M) &= d_4(i,j) \times \psi(2i, 2j + \frac{M}{2}). \end{aligned} \tag{4.24}$$

Similarly, the wavelet coefficients in subbands $VIII$ and IX are interpolated by calculating $d_k(i,j)$ from the respective subbands and multiplying them with corresponding coefficients in subbands V and VI, respectively. Once the coefficients in all three subbands $VII - IX$ are interpolated, we take inverse wavelet transform to get the expanded image. We apply this interpolation technique to both the components, i.e., C_b and C_r.

4.8 Experimental Results

In this section, we demonstrate the efficacy of the proposed method to super-resolve a low resolution observation. We first show results of our learning based techniques to obtain the initial estimate of the super-resolved image. We then illustrate the results of the optimization using the initial HR estimate for super-resolving the gray scale images

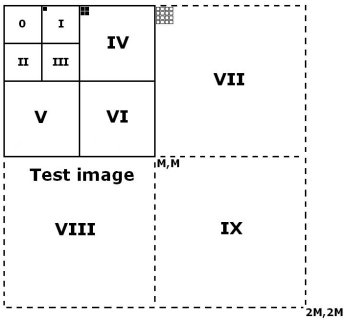


Figure 4.4: Interpolation of the wavelet coefficients at finer scales. The wavelet coefficients in subbands $VII - IX$ are interpolated using the zero tree concept.

as well as the color images. All the experiments are conducted on real world images. The test images are of size 64×64 and the super-resolution is shown for upsampling (decimation) factors of $q = 2$ and $q = 4$, respectively. Thus the size of the super-resolved images are 128×128 and 256×256, respectively. We use '$DB4$' wavelets in all our experiments while estimating the initial HR image and also while interpolating the chrominance components. In order to compare the results using quantitative measure, we use the mean squared error (MSE) as the criteria. In order to compare the results based on MSE, we choose an LR image from the database as a test image and hence the true high resolution image is available for comparison. It has to be mentioned here that the HR images of the test images are removed from the database during learning process. We also conduct an experiment when the observed image is taken from a low resolution camera. For this experiment we show the qualitative assessment only. All the experiments were conducted on a computer with Pentium M, 1.70 GHz processor.

4.8.1 Experimental Results on Initial Estimates

Since we use the initial HR image to estimate the aliasing as well as the IGMRF parameters it is important to discuss on the quality of the learned HR image. In this section we show the results on the initial estimates obtained using our new learning strategy. We compare these results with the learning based approach presented in [3] as they use a learning strategy based on the DWT. In this paper the authors learn the high frequency details from the database consisting of HR images and obtain the super-resolution in a regularization frame work. They use an MRF prior and a wavelet prior while using regularization. The training images used for learning are downloaded from the internet.

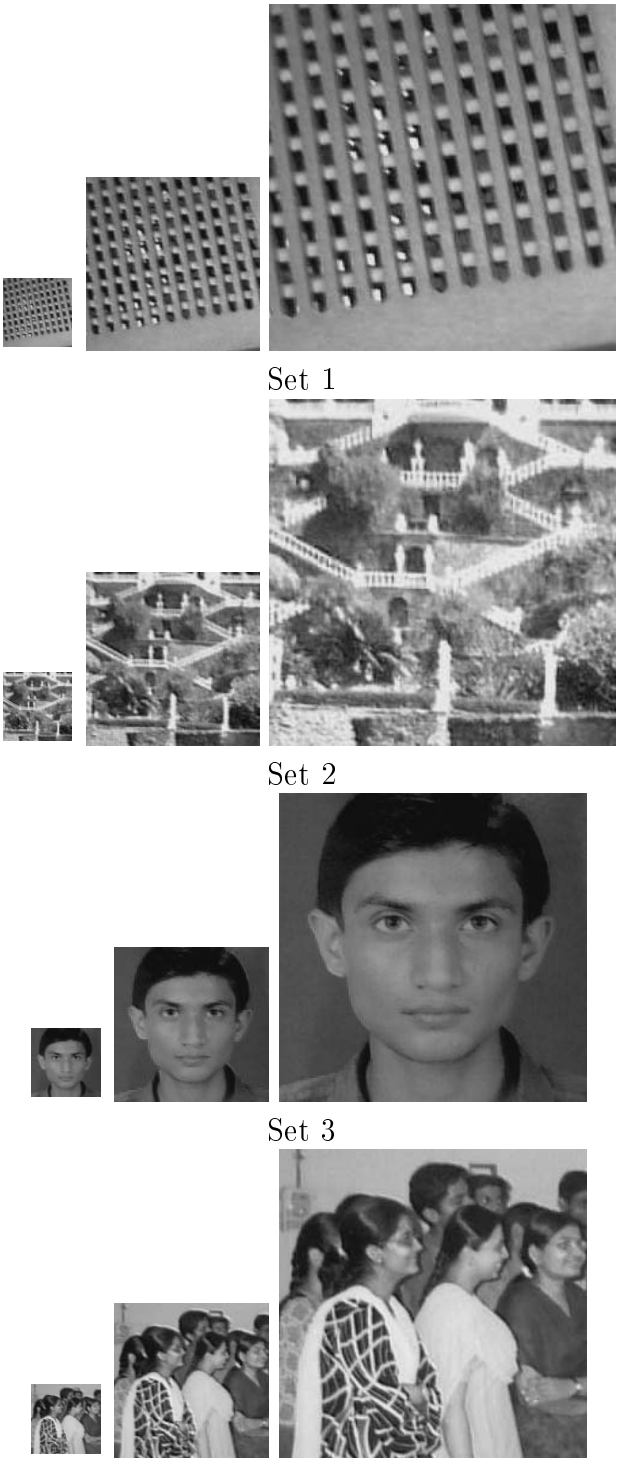


Figure 4.5: Randomly selected sets of training images in the database. (a) Low resolution images. (b) High resolution images with upsampling factor of 2 ($q = 2$), and (c) high resolution images with upsampling factor of 4 ($q = 4$).

In our approach, we construct a database of LR images and their HR versions. It is worth to mention again that our database do not contain simulated images generated using downsampling or upsampling. The construction of this database is one time and offline operation. For each scene there are 2 or 3 images depending on upsampling factor. For example, for an upsampling factor of 2 ($q = 2$), each scene has two images, an LR image and its HR version. If $q = 4$, then there are 3 images, an LR image and two HR versions of the same. All the images in the database are of real world scenes. A computer controlled camera was used to capture the images for the database. In order to avoid motion of the camera while capturing images of a scene at different resolutions, a stable and isolated physical setup was used. The camera was triggered by a MATLAB program at successive but three different time instances for capturing images at three different resolutions. The resolution setting of the camera was changed by the program before each trigger. The time duration between two successive triggers was less than a millisecond. The images of live subjects were captured under controlled environment. We assume that the motion of subjects like human and other moving objects during one thousandths of a second is negligible. We applied mean correction for compensating the intensity variations among the images of each scene. Once the database is ready it can be used for super-resolving images captured by the same camera or by a different camera having low resolution. Our database consists of images of 750 scenes. LR-HR images in the database include indoor as well as outdoor scenes captured at different times. The LR and HR images are of sizes 64×64, 128×128, and 256×256, respectively and the test image (to be super-resolved) is of size 64×64. It may be noted here that size of the test image need to be an integer power of 2 and should not exceed that of the LR training images. We have a total of $3 \times 750 = 2250$ images used for learning the initial HR estimate. Since we show the results for $q = 2$ and $q = 4$, we make use of $2 \times 750 = 1500$ and $3 \times 750 = 2250$ for $q = 2$ and $q = 4$, respectively. Here the multiplication factors 2 and 3 correspond to number of images of a scene. Figure 4.5 shows randomly selected training images in the database. It may be mentioned here that we make use of images in columns (a) and (b) while learning for $q = 2$ and use images in column (b) and (c) for $q = 4$. The gray scale test images used in the experiments are displayed in Figure 4.6. These test images are made up of different textures and contain sharp edges as well as smooth areas. The size of test images Image 1, Image 2 and Image 3 is 64×64 and that

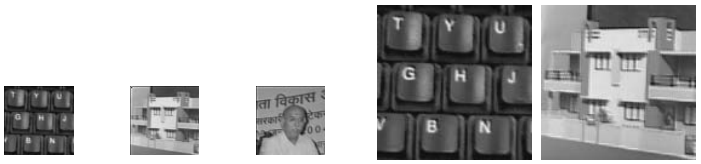


Image 1 Image 2 Image 3 Image 4 Image 5

Figure 4.6: Low resolution observed images (test images). The size of Image 1, Image 2 and Image 3 is 64×64 and that of Image 4 and Image 5 is 128×128.

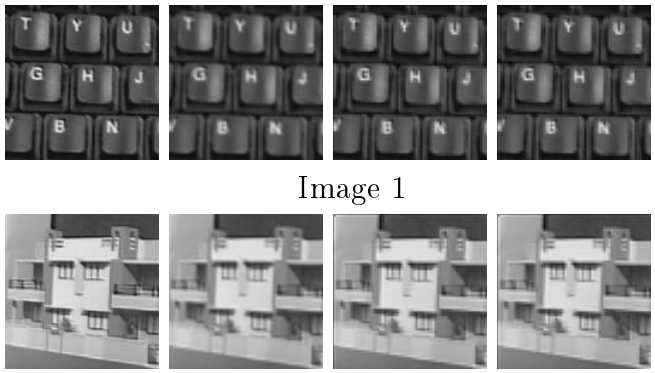


Image 1

Image 2

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

Figure 4.7: Results of learning the initial high resolution estimates for the test images of size 64×64 shown in Figure 4.6 ($q = 2$). (a) Ground truth image, (b) Images expanded using the bicubic interpolation, (c) initial HR estimates obtained using the approach proposed in [3], and (d) initial HR estimates obtained using the proposed learning technique.

of Image 4 and Image 5 is 128×128.

Figure 4.7 and Figure 4.8 show the results of the proposed learning technique for the upsampling factor $q = 2$. Figure 4.7 displays results for the test images of size 64×64. Figure 4.7(a) shows ground truth images and Figure 4.7(b) shows images expanded using bicubic interpolation. Blurry edges can be observed with the result of bicubic interpolation. Figure 4.7(c) and (d) show the images upsampled using the Jiji *et al.* [3] approach and the proposed learning technique, respectively. Similar results for the test images of size 128×128 are shown in Figure 4.8. It can be seen that characters on the keys in image 1 and image 4 upsampled using proposed approach are clearly visible. Similarly window panes in image 2 and image 5 in Figure 4.7(d) and 4.8(d) appear sharp. Compared with Jiji *et al.* approach, sharp boundaries can be achieved by our method, thus making the result look natural. Now, we show the results of the proposed learning technique for the upsampling factor $q = 4$. We use Image 1, Image 2 and Image 3 displayed in Figure 4.6 as the test images and obtain the initial HR estimates of size 256×256. The results

Image 4

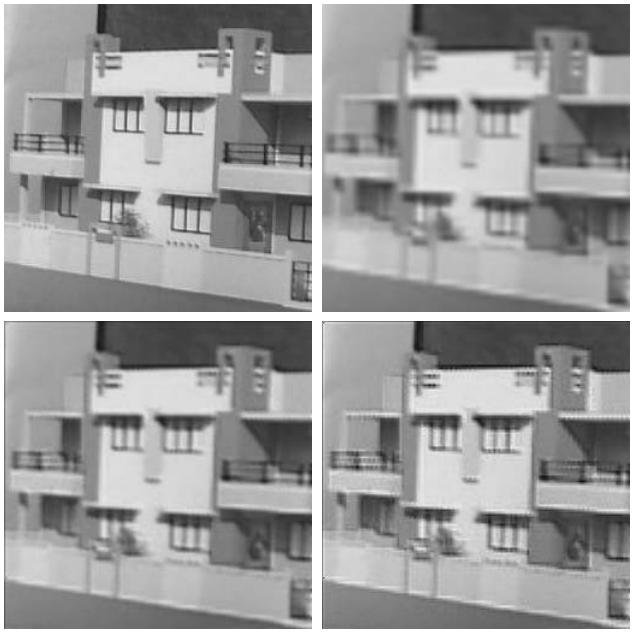


Image 5

(a) (b)

(c) (d)

Figure 4.8: Results of learning the initial high resolution estimates for the test images of size 128×128 shown in Figure 4.6 ($q = 2$). (a) Ground truth image, (b) Images expanded using the bicubic interpolation, (c) initial HR estimates obtained using the approach proposed in [3], and (d) initial HR estimates obtained using the proposed learning technique.

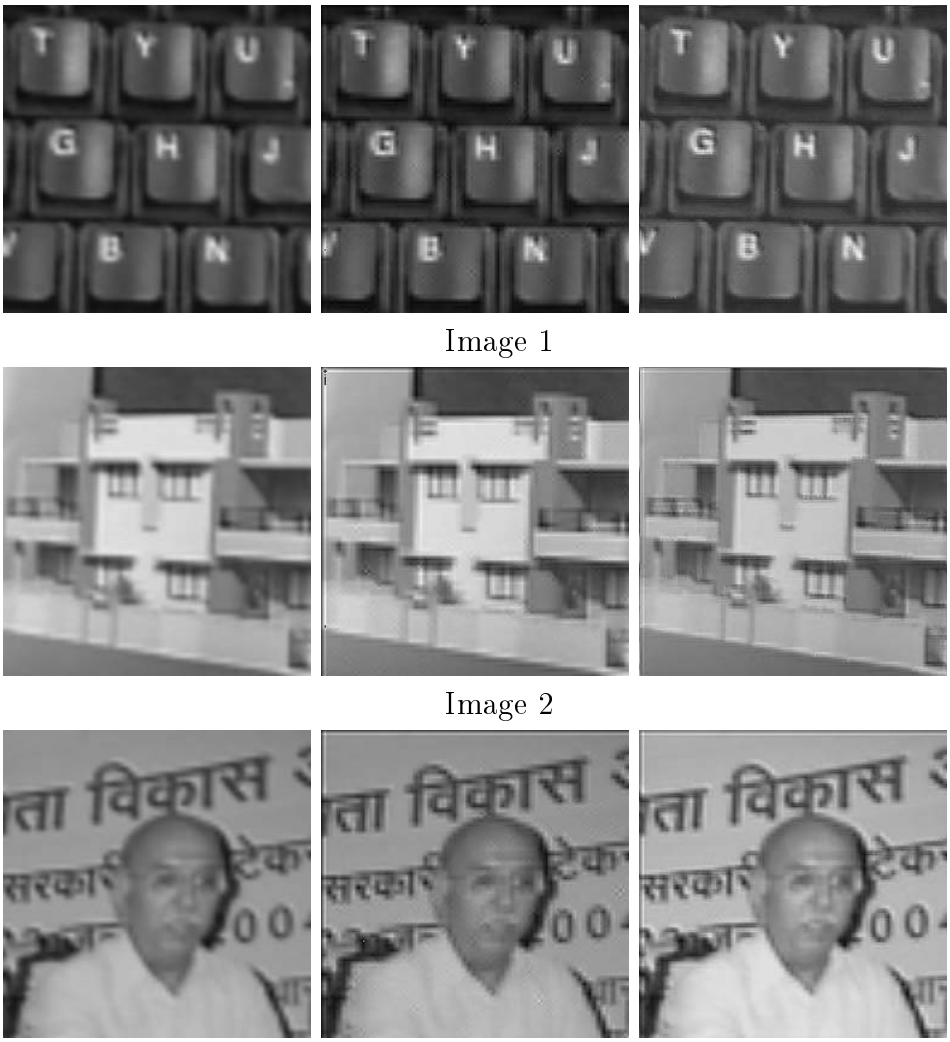


Image 1

Image 2

Image 3

(a) (b) (c)

Figure 4.9: Results of learning the initial high resolution estimates for the test images of size 64×64 shown in Figure 4.6 ($q = 4$). (a) Image expanded using bicubic interpolation, (b) initial HR estimates obtained using the approach proposed in [3], and (c) initial HR estimates obtained using the proposed wavelet based learning technique.

Image	MSE for $q = 2$			
	Size of Test image	Bicubic interpolation	Jiji *et al.* approach [3]	Proposed approach
1	64×64	0.023297	0.023067	0.021457
2	64×64	0.007912	0.011166	0.007977
4	128×128	0.044262	0.043209	0.043193
5	128×128	0.110873	0.110061	0.107358
	MSE for $q = 4$			
1	64×64	0.152321	0.157115	0.150864
2	64×64	0.020358	0.024838	0.013204
3	64×64	0.021795	0.024328	0.017989

Table 4.1: Mean squared error comparison for the initial HR estimate obtained using different techniques.

Image	Learning Time (in seconds)	
	$q = 2$	$q = 4$
1	91.350	688.508
2	91.332	867.630
3	-	691.854

Table 4.2: The computational complexity of the proposed algorithm in terms of time required for learning the initial HR estimate for $q = 2$ and $q = 4$.

are shown in Figure 4.9. In this Figure, the first column (a) shows the images expanded using bicubic interpolation. The columns (b) and (c) show the images learned using Jiji *et al.* approach and the proposed approaches, respectively. Comparison of these images shows that the learned images using proposed technique exhibit sharp edges and better texture. This is expected as we learn the edges using a data set that has true LR-HR images. The boundary regions in the images are not reconstructed well as we assume zero values for the pixels outside the image. This can be reduced by performing mirroring on the boundaries prior to processing. The quantitative comparison of these images is shown in Table 4.1. It can be clearly observed that the MSE for the proposed learning approach is less than that for the approach of Jiji *et al.*[3]. The results obtained for both $q = 2$ and $q = 4$ shows perceptual as well as quantifiable improvements over bicubic interpolation and Jiji *et al.* approach.

Table 4.2 shows the computational complexity of the proposed learning algorithm in terms of the time required for learning the initial HR estimates for $q = 2$ and $q = 4$.

From the values given in table we observe that for $q = 4$ the time for learning is higher as compared to that for $q = 2$. This is because we use two step operation for $q = 4$ while learning. The use of one step operation reduces the computation time. However, as already pointed out it suffers from the exponential error propagation while learning. Thus we compromise speed for better accuracy.

4.8.2 Experimental Results on Super-resolution

Let us now see how well we can super-resolve a given low resolution observation. First we present the results on super-resolution for gray scale images. The observed images are super-resolved using the proposed regularization framework formulated as an MAP estimate. The regularization makes use of the decimation matrix entries and the IGMRF parameters, both of which are estimated using the learned initial HR estimate. We use the same low resolution observations as used in the previous experiment. Once again the results obtained are compared with the approach proposed in [3]. The comparison is also shown with the use of bicubically interpolated image as the initial HR estimate. While using the bicubic interpolation as the initial estimate for regularization we use the decimation and the IGMRF parameters estimated by the same. Figure 4.10 shows the results for the upsampling factors of 2 ($q = 2$) and 4 ($q = 4$). In Figure 4.10(a)-(c) we show the results for $q = 2$, and in Figure 4.10(d)-(f) we show the results for $q = 4$. Figure 4.10(a) and Figure 4.10(d) show the super-resolved images using the MRF and wavelet priors [3]. In Figure 4.10(b) and Figure 4.10(e) we display the upsampled images by using the bicubic interpolation as the initial HR estimate while in Figure 4.10(c) and Figure 4.10(f) correspond to the results obtained using the learned image as the initial HR estimate. We can observe that in image 1, the characters G and H on the keyboard are clearly visible in Figure 4.10(c) and (f). The window grills are well preserved in the image 2 for the proposed method. Similarly we can see improvements in image 3 when we use the proposed method. We can clearly observe that the eyes and the mouth regions have better details. Also the text behind the person is clearly readable. It can be seen from the comparison of images in Figure 4.10(b) and (c) as well as images in Figure 4.10(e) and (f) that the learned initial HR estimate produces better final solution when compared to using the bicubically interpolated image used as the initial estimate.

Image 1

Image 2

Image 3

(a) (b) (c)
(d) (e) (f)

Figure 4.10: Results for gray scale images, (a)-(c) for $q = 2$ and (d)-(e) for $q = 4$. Images in (a) and (d) correspond to super-resolved (SR) images using MRF prior and wavelet prior as in [3]. Images in (b) and (e) correspond to the upsampled images using the bicubic interpolated as initial HR estimate obtained while regularization. Images in (c) and (f) correspond to super-resolved (SR) images using the proposed approach.

Image	MSE for $q = 2$		
	MRF and Wavelet priors	IGMRF prior with initial estimate obtained using	
		bicubic interpolation	proposed learning based approach
1	0.020006	0.022300	0.018699
2	0.007295	0.007524	0.007222
3	0.008429	0.008722	0.008185
	MSE for $q = 4$		
1	0.143968	0.151608	0.143925
2	0.018288	0.019993	0.013204
3	0.018367	0.021021	0.017989

Table 4.3: Performance comparison in terms of mean squared error for the gray scale image super-resolution.

This is because the decimation entries and the IGMRF parameters estimated from the learned initial HR estimate are close to the true values when compared to those estimated from the bicubically interpolated initial estimate. The comparison of the performance in terms of the mean squared error is shown in Table 4.3. We see that the MSE using the proposed learning based approach is less when compared to that obtained using the bicubic interpolation as the initial estimate as well as with that proposed in [3].

We now show the experiments for color image super-resolution. As we have discussed earlier the luminance component of the color image is super-resolved using the proposed learning based method and the chrominance components are interpolated in the wavelet domain. The super-resolved color image is then obtained by applying color space transformation on the super-resolved luminance component and interpolated chrominance components. The results are compared with the bicubic interpolation, the approach proposed in [3], and the single frame super-resolution approach proposed by Kim and Kwon [4]. In the later method for super-resolution, the database consists of HR images and their downsampled versions. For this experiment we show the results for an upsampling factor of 4 only. Figure 4.11 shows the observed color images. Image 1, Image 2 and Image 3 are chosen as test images from the database hence the true high resolution image is available for comparison. Image 4 is captured using a low resolution camera other than the one used to capture the images for the database. Figure 4.12(a) shows the images obtained using bicubic expansion and Figure 4.12(b) shows the super-resolved images obtained using the MRF and wavelet priors proposed in [3]. The results of super-resolution

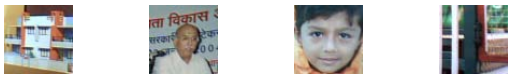


Image 1 Image 2 Image 3 Image 4

Figure 4.11: Low resolution observed color images. Image 1, Image 2 and Image 3 correspond to LR database images while Image 4 is captured using a different low resolution camera.

using the method proposed in [4] is shown in Figure 4.12(c). Finally, in Figure 4.12(d) we display the results obtained using our approach. Comparison of the figures show more clear details in the super-resolved images using the proposed approach. Figure 4.12(b) shows that the algorithm in [3] generates smoothed edges and perceptually distracting color artifacts. In the images shown in Figure 4.12(c), the color artifacts are disappeared, however edges are smoothed. It can be seen from Figure 4.12(d) that the edges in window panes in the image 1 are well preserved while using the proposed approach. Once again as in the gray scale image super-resolution the text behind the person in image 2 is clearly readable as compared to that in Figure 4.12(a)-(c). Also in the same image the spectacle glasses and the lips look clearer. In image 3, the eyes and nostrils look sharper in Figure 4.12(d). The smooth region like cheeks look better. As mentioned earlier, Image 4 is captured using a low resolution camera other than that used to construct the database. In this image the horizontal and vertical bars of the gate appear clear as compared to that in Figure 4.12(a)-(c). From this result it can be seen that the algorithm works for images captured using different cameras as well. In each of the images super-resolved using the proposed method the natural correspondences between the R, G and B components are in balance and there are no color artifacts. Among the images super-resolved using the proposed approach, Image 1, Image 2 and Image 4 are better reconstructed as compared to Image 3 is. It appears that on these three images, the high frequencies are well learned by the algorithm. The reason behind it is that these images contain more edges in vertical, horizontal, and diagonal directions. Since the wavelets capture the edges in these directions well, better super-resolution is achieved on these images. Poor reconstruction of the Image 3 is due to the inherent limitation of the wavelets that they are inferior in acquiring the geometry of edges. The comparison using the mean squared error for color super-resolution is shown in Table 4.4. We see from the table that the super-resolved images using the proposed technique have comparable MSE with the other methods. The

Figure 4.12: Results on the color images for $q = 4$. Images in each column correspond to (a) image expanded using bicubic interpolation, (b) super-resolved (SR) image using MRF prior and wavelet prior as in [3], (c) SR image using example based single image super-resolution proposed by Kim and Kwon in [4], and (d) SR image using the proposed method.

Image	MSE			
	bicubic interpolation	Jiji *et al.* approach in [3]	Kim and Kwon approach in [4]	Proposed
1	0.312001	0.610422	0.318010	0.307506
2	0.089783	0.064304	0.125707	0.075376
3	0.186941	0.198895	0.180958	0.169560

Table 4.4: Performance comparison in terms of mean squared error for the color super-resolution ($q = 4$). We have not shown MSE comparison for image 4 because the actual HR image is not available.

proposed algorithm performs better in terms of quality with lower quantitative errors, less blocky regions and better high spatial-frequency information preservation such as sharp edges.

4.9 Conclusion

We have presented a new approach for super-resolution restoration of a single image using wavelet based learning technique. The missing high frequency details are learned from a database consisting of low resolution images and their high resolution versions all captured by varying resolution settings of a real camera. Since the database does not contain synthetic training images, the learning algorithm effectively captures the true relationship of the spatial features between LR images and HR images. The super-resolved image is obtained using an MAP estimate using an inhomogeneous Gaussian MRF model as a prior. Both the model parameters as well as the decimation are estimated using the learned high resolution estimate. The use DWT based learning technique increases the sharpness of the images by selecting the best high resolution edges from the database, while the use of the locally adaptive IGMRF prior ensures a proper spatial correlation among pixel intensities along with discontinuity preservation. The another advantage of using IGMRF model for super-resolved image lies in having a differentiable cost function where a gradient-based optimization method can be used. The quality of the super-resolved gray scale images is quite good. We have extended the algorithm to super-resolve the color images, where the luminance component is super-resolved using proposed technique and the chrominance components are interpolated using the wavelet transform. The super-resolved gray image as well as the color images using the proposed

method are less noisy in constant areas and preserve the textures and sharp details in other regions. Results on real world images show this approach to be superior to existing SR in terms of fidelity to the ground truth high-resolution images. Through the proposed super-resolution approach, we are able to obtain a high resolution of an image captured using commercially available low resolution camera fitted with limited memory.

It is necessary to point out that the proposed approach have some drawbacks. An inherent limitation of the learning algorithm is that it has a strict separating line for upsampling factors. The learning process is very much resolution dependent, restricting the super-resolution to powers of 2 only. Since the wavelet transform decompose an image into the number of subbands that is a power of 2, one can not perform learning for super-resolution for upsampling factors that are not powers of 2, such as 3, 5 etc. Another drawback of the learning approach arises due to the use of the multi-step operation for learning. Learning for upsampling factor 2, requires 2 training images in each set in the database, while that for upsampling factor 4, requires 3 training images in each set. Thus each octave increase in the upsampling factor requires an additional high resolution image in each set of the training images.

Failure of wavelets in efficiently handling the intrinsic geometrical structures in natural images is an another major difficulty with the proposed learning algorithm. Although the wavelet transform has been proven to be powerful in learning edges in horizontal, vertical and diagonal directions, wavelets are not optimal in capturing the edges in arbitrary directions. A better way is to use a different image representation that can efficiently handle geometric structures in real world images. The work presented in the next chapter involve modifications to the learning approach through the use discrete cosine transform for the SR image to improve results.

Chapter 5

DCT Based Learning Technique

A variety of handheld devices such as digital cameras, cellular phones, personal digital assistants are prevalent these days. These personal and/or home electronic devices provide accessibility and convenience for users in many applications, such as communicating with other people, learning the latest information, acquiring multimedia content. The demand for audiovisual capabilities in these devices has arisen from consumers. However, while the race for extra functionalities in a single device is on, the performance of these extensions are often not very satisfactory. Due to limited computational power, and storage, the quality of multimedia contents in these integrated products is not comparable to that of dedicated ones. The image/video size is small and the compression is often severe. The discrete cosine transform (DCT) is widely used for image and video compression in these devices. Since the DCT is the basis of many popular codecs such as JPEG, MPEG and H.26X, we investigate the task of improving the resolution of an undersampled image using the DCT for learning.

The motivation behind using the discrete wavelet transform in the learning techniques in the previous chapter was its ability to capture the edges across scales and computational advantage offered by the property of the separability of the wavelet kernels. The drawback for wavelets in two dimensions is their limited ability in capturing directional information. Since the DCT based learning algorithm can capture the intrinsic geometrical structures in natural images, it helps to overcome this deficiency. The proposed learning algorithm searches the fine details of the super-resolved image from the database of LR-HR images and recover the geometrical structures in natural image. We first obtain the close approximation to the high resolution image using a new learning tech-

nique based on DCT and then use regularization framework to obtain the solution. The true scene may contain regions with a wide variety of textures and may have continuously changing textures. The homogeneous models will not be appropriate in such cases. This motivates us to consider an inhomogeneous prior which can adapt to the local structure of the image. In this chapter, we explore the use of non-homogeneous autoregressive model to represent the super-resolved image field and reconstruct the super-resolution. We represent the super-resolved image using and derive the model parameters from the approximation and use the same while minimizing the cost function.

5.1 Previous Work

In [76], the authors propose a super-resolution technique using zoom cue. Here the SR image is modeled as a homogeneous AR model and the model parameters are obtained using the most zoomed image. The drawback of the proposed method is that they assume that the entire SR region is homogeneous which is not true for real images. In [160] Krishna and Joshi propose a model based multiresolution fusion in which the regularized solution is obtained using the nonhomogeneous AR parameters estimated from the Panchromatic image. In chapter 4, we first learn an initial HR estimate and use a inhomogeneous Markov random field (IGMRF) as the prior for regularization. A wavelet based learning was used to obtain the initial estimate and the IGMRF model parameters were estimated using the local gradient as the standard deviation. Although the method works well and has advantages when compared to using a homogeneous AR prior, it has the following drawbacks:

1. The initial high resolution estimate obtained using wavelet based approach assumes that a primitive edge element is confined to a local region which is not true in practice. Also the learned edges are limited to horizontal, vertical and diagonal directions. This leads to an initial estimate that may not be a close approximation to the super-resolved image.

2. The IGMRF parameters estimated at every location are based on the approach proposed in [157]. Here the authors use a Maximum Likelihood (ML) estimate and use a simple approximation of the local variance for parameter estimation in

order to reduce the computational complexity. This leads to estimated IGMRF parameters much different from their true values. Also since these parameters are estimated using the initial HR derived from the wavelet based learning it causes larger errors in estimated parameters.

5.2 Learning Using Discrete Cosine Transform

In previous chapter we used the discrete wavelet transform for learning high frequency information. Although the wavelet transform is proved to be powerful in many signal and image processing applications, the drawback for wavelets in two dimensions is their limited ability in capturing directional information. Since the wavelets are good at capturing edges in vertical, horizontal and diagonal directions only, the wavelet based learning algorithms fail to search the best matching high frequency details corresponding to the directional edges and geometrical structures in other directions.

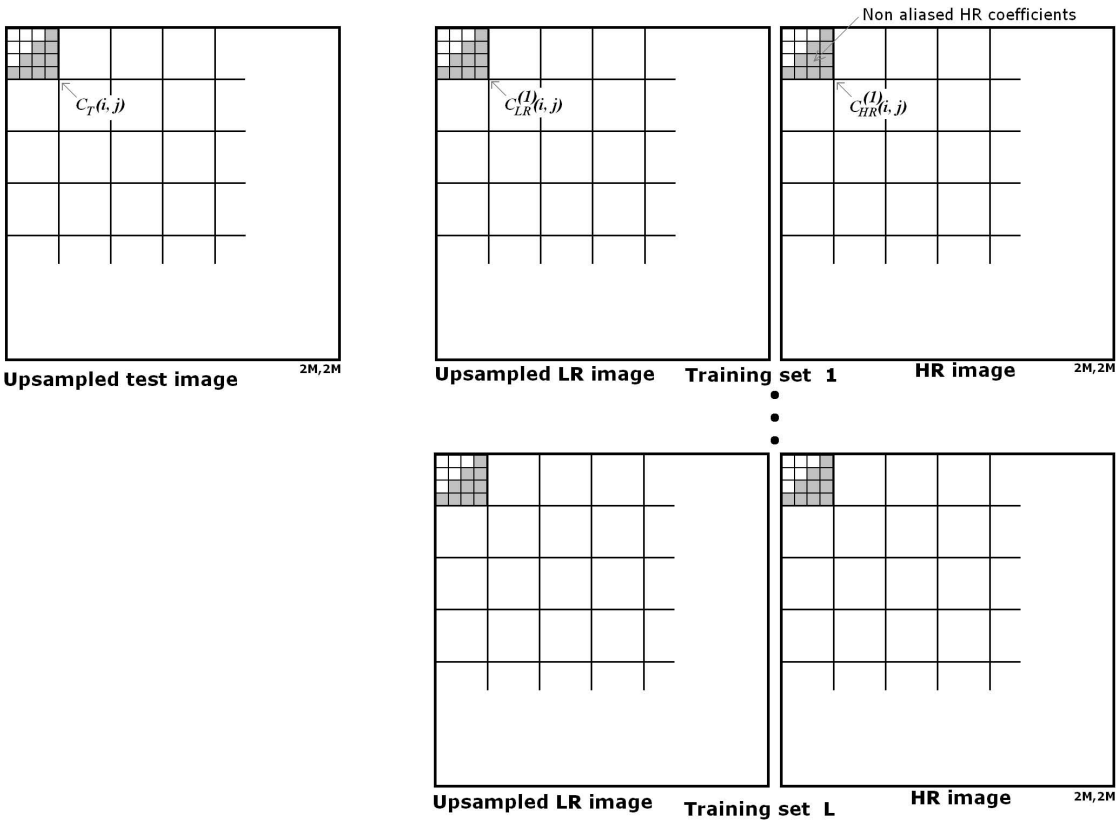
In this chapter, we explore the use of the DCT for predicting the missing high-frequency information. We use a database of images with high-quality texture to predict the missing high-frequency information in the LR inputs. We obtain the close approximation to the high resolution image using a new learning technique based on DCT. We derive the nonhomogeneous AR model parameters from the this approximation and use the same while minimizing the cost function.

5.2.1 Discrete Cosine Transform

The discrete cosine transform was introduced by Ahmed and his colleagues in 1974 [161, 162, 163]. It is extensively used orthogonal transform used in various schemes of digital image coding. The DCT outperforms the other orthogonal transforms due to its good properties such as separability, invertibility and energy compaction. The DCT is used in JPEG and MPEG standards for still image compression and video compression, respectively. Considering this fact, we exploit the use of DCT in the learning approach. The proposed learning approach can readily be extended for super-resolving compressed images/video so that it does not require decoding of low-resolution images/video prior to learning the close approximation of the high resolution video.

5.2.2 The Approach

In this section, DCT based approach to learn high frequency details for the super-resolved for a decimation factor of 2 ($q = 2$) is described. Each set in the database consists of a pair of LR image and its HR version. The test image and LR training images are of size $M \times M$ pixels. Corresponding HR training images have size of $2M \times 2M$ pixels. We first upsample the test image and all LR training images by factor of 2 and create images of size $2M \times 2M$ pixels each. A standard interpolation technique can be used for the same. We divide each of the images, i.e. the upsampled test image, upsampled LR images and their HR versions, in blocks of size 4×4. The motivation for dividing into 4×4 block is due to the theory of JPEG compression where an image is divided into 8×8 blocks in order to extract the redundancy in each block. However, in this case we are interested in learning the non aliased frequency components from the HR training images using the aliased test image and the aliased LR training images. This is done by taking the DCT on each of the block for all the images in the database as well as the test image. Figure 5.1(a) shows the DCT blocks of the upsampled test image whereas Figure 5.1(b) shows the DCT blocks of upsampled LR training images and HR training images. For most images, much of the signal energy lies at low frequencies (corresponding to large DCT coefficient magnitudes); these are relocated to the upper-left corner of the DCT array. Conversely, the lower-right values of the DCT array represent higher frequencies, and turn out to be small. We compare the DCT coefficients in each block of the upsampled test image with that of the blocks of LR training images and find best matching block. We then copy the DCT coefficients from the corresponding block in the HR version of the LR image to the block under consideration. Thus for a particular block in the test image, we make use of all the blocks in every upsampled LR training image for learning HR DCT coefficients. It is reasonable to assume that when we interpolate the test image and the low resolution training images to obtain $2M \times 2M$ pixels, the distortion is minimum in the lower frequencies. Hence we can learn those DCT coefficients that correspond to high frequencies (already aliased) and now distorted due to interpolation. Let $C_T(i, j)$, $1 \leq i, j \leq 4$, be the DCT coefficient at location (i, j) in a 4×4 block of the test image. Similarly, let $C_{LR}^{(m)}(i, j)$ and $C_{HR}^{(m)}(i, j)$, $m = 1, 2, \ldots, L$, be the DCT coefficients at location (i, j) in a block in the m^{th} upsampled LR image and m^{th} HR image, respectively. Here



(a) (b)

Figure 5.1: Learning DCT coefficients from a database of sets of LR-HR images for $q = 2$. (a) Upsampled test image and (b) sets of upsampled LR images and HR images for different scenes. DCT coefficients for the shaded locations in the upsampled test image are copied from corresponding locations of the best matching HR image.

L is the number of the training sets in the database. Now the best matching HR image block for the considered test image block (upsampled) is obtained as,

$$l = \underset{m}{^{argmin}} \sum_{i+j>Threshold} \|C_T(i,j) - C_{LR}^{(m)}(i,j)\|^2. \tag{5.1}$$

Here, l is the index for the training image which gives the minimum for the block. Those non aliased DCT coefficients from the corresponding block in best matching HR image are now copied in the block of the upsampled test image. In effect, we learn non aliased DCT coefficients for the test image blocks from the set of LR-HR images. The coefficients that corresponds to low frequencies are not altered. Thus at location (i,j) in the block, we have,

$$C_T(i,j) = \begin{cases} C_{HR}^{(l)}(i,j) & \text{if } (i+j) > Threshold, \\ C_T(i,j) & \text{else.} \end{cases}$$

This is repeated for every block in the test image. After learning the DCT coefficients for entire test image, we take inverse DCT transform to get high spatial resolution image and use it as the close approximation to the HR image. The pseudocode of the proposed technique is provided in Algorithm 5.1.

In order to fix the threshold, we conducted experiment using different $Threshold$ values. We begin with $Threshold = 0$ where all the 16 coefficients in each 4×4 block are learned. The first image of the first row in Figure 5.2 show the learnt image. The image is highly degraded. We then set $Threshold$ to 2, where all the coefficients except the DC coefficient $(i = j = 1)$ in each block are learned, and obtain the image shown in the second image of the first row in the same figure. There is a little improvement but still it is hard to interpret the content if the image. The subsequent increase in the $Threshold$ value (up to $Threshold = 4$) reduces the degradation. The fourth image in first row, obtained by with $Threshold = 4$ is the most clear. As can be seen from the images in the second row, further increase in the $Threshold$ value introduces blockiness in the learned image. Reason for such behavior is that for the values of $Threshold$ beyond 5, the number of DCT coefficients learned in each block of 4×4 decreases from 5 for $Threshold = 6$ to 0 for $Threshold = 8$. Thus, for the higher values of $Threshold$ the contribution of learned DCT coefficients reduces and hence there is little or no improvement in the interpolated image. The mean squared errors of the images learned for different $Threshold$ values are

Algorithm 5.1: Learning the HR approximation using DCT

Data: low resolution test image and a database of LR-HR training images
Result: close approximation to the HR image
upsample the test image;
divide the test image into blocks of size 4×4 pixels;
perform DCT transform on each block in the test image;
initialize an array *cost* for each block in the test image;
initialize a matrix *hrdct* for reconstructed DCT image;
foreach *LR training image in the database* **do**
    upsample the LR training image;
    divide the upsampled LR training image and its HR version into blocks of size 4×4 pixels;
    perform DCT transform on each block in both the training images;
end
foreach *block in the test image* **do**
    foreach *LR training image in the database* **do**
        calculate *sum* of the squared differences between high frequency DCT coefficients in the block of the test image and those in the corresponding block in LR training image;
        if $sum < cost$ **then**
            $cost = sum$;
            copy the high frequency DCT coefficients from the corresponding block in the HR training image to *hrdct*;
        end
    end
    copy the low frequency DCT coefficients from the block in the test image to *hrdct*;
end
obtain the close approximation to the HR image by taking inverse DCT of *hrdct* ;

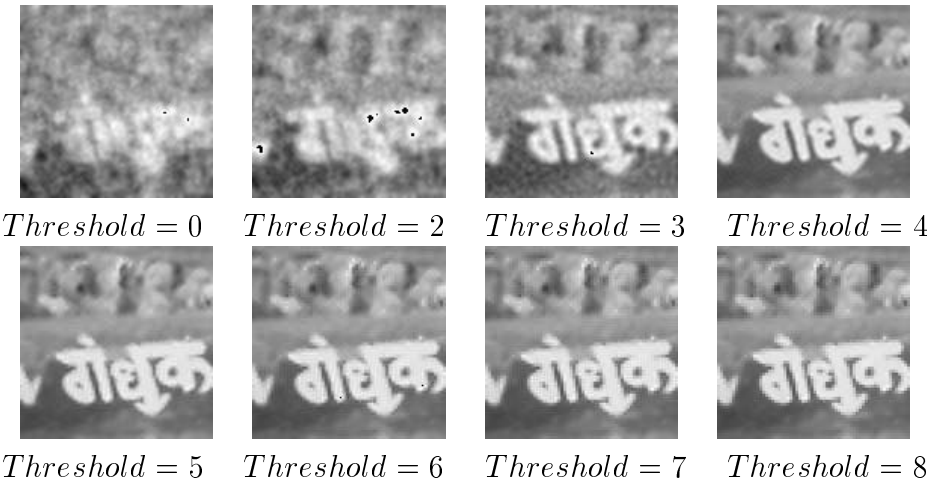


Figure 5.2: Learnt images with different *Threshold* values.

Threshold	MSE
0	0.036874
2	0.028089
3	0.017775
4	0.017483
5	0.017900
6	0.018458
7	0.019079
8	0.019866

Table 5.1: MSE Comparison of images learned with different $Threshold$ values.

given in Table 5.1. It can be seen that the best results are obtained when the $Threshold$ is set to 4 that corresponds to learning a total of 10 coefficients in a block.

5.2.3 Results on DCT Based Learning Technique

In order to evaluate the performance of the proposed learning technique, we conducted the experiment on a couple of real world images. The test images were of size 64×64 and the learning technique was applied for upsampling (decimation) factors of $q = 2$. Thus the size of the learned high resolution images was 128×128. The database includes LR-HR images of 750 indoor as well as outdoor scenes captured at different times under varying conditions. Thus the proposed learning technique uses a total of $750 \times 2 = 1500$ images, while learning. LR-HR pairs were captured by varying the zoom setting of a camera. The results of the proposed technique are compared with that of the DWT based learning technique as well as that of bicubic interpolation.

We display the results of the proposed learning approach in Figure 5.3. Figure 5.3(a) and (b) show the LR observations and corresponding ground-truth images, respectively. The standard bicubic interpolation (Figure 5.3(c)) gives fairly smooth results. High resolution images learned using DWT and DCT are shown in Figure 5.3(d) and (e), respectively. Visually, the differences between the images are subtle. The texture and curvatures in the images displayed in the rightmost column are better defined. Images learned using DWT exhibit sharp edges along horizontal, vertical and diagonal directions only. It can be seen from these images that the edges appear slightly artificially crisp. True enhancement is achieved in images obtained using DCT based learning technique. As is expected the proposed technique manages to reconstruct most of the primary edges.

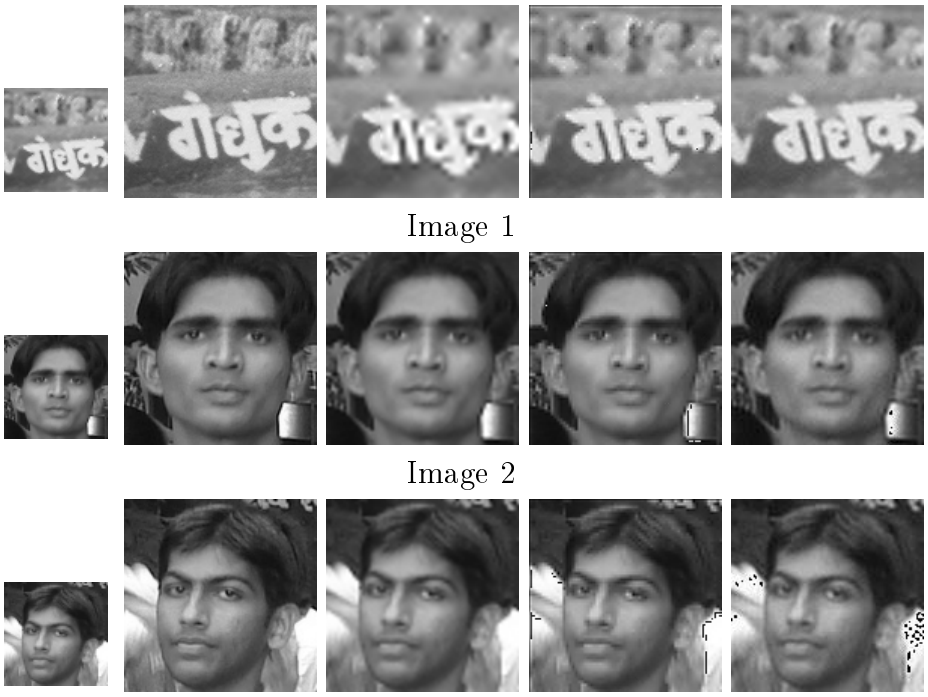


Figure 5.3: Qualitative comparison of different techniques for image expansion. (a) observed images, (b) original high resolution images, (c) expanded images using bicubic interpolation (d) learnt images using a wavelet based technique and (e) learnt images using proposed DCT based approach.

The curves along the Devnagari text in image 1 and curves of eyes and noses in Image 2 and Image 3 look sharper. From these figures, it can be seen that the textures better retrieved by wavelets are all characterized by dominant vertical, horizontal, or diagonal directions, which are the only directions captured well by wavelets. In contrast, the textures better retrieved by the DCT exhibit more diverse directional components. This shows the superiority of the DCT in capturing directional information. The quality of the reconstruction is assessed by means of the mean squared error (MSE) between reconstructed and original. The MSE values for the learned HR images are summarized in Table.5.2. The numerical results obtained using DCT for learning are superior than those obtained using DWT for learning.

Image	MSE		
	Bicubic Expansion	Wavelet Based Learning	DCT Based Learning
Image 1	0.020649	0.019470	0.017483
Image 2	0.005415	0.004573	0.004398
Image 3	0.003354	0.003133	0.002845

Table 5.2: Performance comparison of the DCT based learning technique with the other methods.

5.3 Proposed Approach for Super-resolution

The super-resolution problem is an ill-posed problem. Given an LR image, the estimation of the HR image maximizing any of the conditional distributions that describe the image formation model is a typical example of an ill-posed problem. Therefore, we have to regularize the solution or, using statistical language, introduce a priori models on the HR image. We continue with the image formation model described in section 4.4. We first apply DCT based learning technique to learn the high frequency content of the super-resolved image using a database of training images consisting of LR-HR pairs. While learning the DCT coefficients we look for the non aliased frequency components using HR training images. The learned HR image is used as a close approximation to the final solution. We use a nonhomogeneous AR model as prior and estimate the prior parameters using the learned HR image. The final cost function, consisting of data fitting term and a prior term is minimized using gradient descent optimization.

5.3.1 Segmentation of the Learned Image

Accurate estimates of the prior model parameters are critical for the super-resolution problem. The parameters estimated by considering the entire image as a homogeneous region is less accurate as compared to true parameters. Better estimation of the parameters can be obtained by considering the image as a combination of several homogeneous regions and estimating the parameters for each of the regions separately. Further, regularizer using the homogeneous prior model imposes the smoothness constraint everywhere in the image. It determines the constant interaction between neighboring points and leads to overall smoothing. The homogeneous or isotropic application of the smoothness constraint inevitably leads to oversmoothing at discontinuities at which the derivative is

Image 1 Image 2 Image 3

Figure 5.4: Segmented images. (Each image is segmented into 7 homogeneous regions.)

infinite. Since the natural image consists of smooth regions, texture regions and edge it has to be represented as nonhomogeneous region. We consider the nonhomogeneous region as a combination of a number of homogeneous regions and expect the images to be smooth within homogeneous regions. We divide the learned HR image into a number of homogeneous regions and use the segmented image to learn nonhomogeneous AR prior model parameters.

In order to segment the learned image, we employ Gaussian mixture model (GMM) and estimate the model parameters for each region in terms of the mean and variance. We use Expectation Maximization algorithm to assign label to each pixel in the image. The resulting segmented images are shown in Figure 5.4. We then estimate AR parameters for each of the segmented regions.

5.3.2 Texture Modeling

As described in section 5.3.1, the HR image is segmented into a number of homogeneous regions. In a homogeneous region, the linear dependency of a pixel to its neighbors can be represented using an AR model. Let the image $Z = \{Z_k | k = 1, \ldots, K\}$ be an image consisting of K homogeneous regions. The intensity of the pixel at location $s = (i, j)$ in k^{th} homogeneous region can expressed as the linear combination of the intensities of the pixels in the neighborhood r of that pixel. Mathematically, the AR model to the pixel at location s can applied as,

$$z_k(s) = \sum_{r \in \mathcal{N}_s} \theta_k(r) z_k(s + r) + \sqrt{\rho} n_k(s), \tag{5.2}$$

where $\mathcal{N}_s$ is the neighborhood of pixel at s, r is a neighborhood index with $r \in \mathcal{N}_s$, θ_k are the AR parameters for k^{th} region and $n_k(.)$ is an i.i.d. noise with zero mean and unit variance. We use fifth order neighborhood for every region. This requires a total of 24 parameters to be estimated for every region. We estimate these parameters using simple least squares method.

5.4 Super-resolving the Image

Since we have introduced both prior and conditional distributions, we can apply the Bayesian paradigm in order to find the maximum of the posterior distribution of HR image given the observation. Our goal is to find $\hat{\mathbf{z}}$ that satisfies,

$$\hat{\mathbf{z}} = \underset{z}{argmin} \left[\frac{\| \mathbf{y} - D\mathbf{z} \|^2}{2\sigma_n^2} + \lambda C_{AR} \right], \tag{5.3}$$

where C_{AR} is the prior term which can be represented as,

$$C_{AR} = \sum_k \sum_{i,j \in k} \left(z_k(s) - \sum_{r \in \mathcal{N}_s} \theta_k(r) z_k(s+r) \right)^2, \tag{5.4}$$

and λ is a suitable weight for the same. Since both the data fitting term and the nonhomogeneous AR prior term in above cost function are linear terms, it can be minimized using simple gradient descent optimization technique. processors

5.5 Experimental Results

In this section, we present the results of the proposed method for the super-resolution. We compare the performance of the proposed method with other methods on the basis of perceptual quality as well as using a quantitative measure. All the experiments were conducted on real images. All the test images are of size 128×128 pixels and the super-resolved images are of size 256×256 pixels. We show the quantitative comparison using mean squared error (MSE).

Figure 5.5 and Table 5.3 show the perceptual and quantitative comparison of performance of the proposed approach with that of bicubic interpolation method and super-

Image	MSE		
	bicubic interpolation	homogeneous AR model	Proposed
Image 1	0.069449	0.068504	0.064889
Image 2	0.062587	0.060049	0.059856
Image 3	0.079001	0.074469	0.070992

Table 5.3: Mean squared error comparison.

resolution approach using homogeneous AR prior. The LR observed images are displayed in Figure 5.5(a) and images expanded using bicubic interpolation are shown in Figure 5.5(b). The images super-resolved using homogeneous AR technique and using proposed technique are shown in Figure 5.5(c) and Figure 5.5(d), respectively. As to be expected, bicubic interpolation has the highest MSE, because it doesn't add any resolution, but rather averages out the observed information. The visual comparison shows that bicubic interpolation produces smooth images and the super-resolution techniques preserve textures and edges in the images. In Image 1, the edges of tiles and the stone texture appear sharp as compared to that obtained using homogeneous AR technique. From the Table, it can be seen that the MSE of the Image 1 super-resolved using proposed approach is significantly reduced. In Image 2 super-resolved using proposed approach, the improvement is clearly visible in the text written on the black background. Edges of the characters in this text are better preserved. The MSE is also less compared to that of the images obtained using other methods. Although, Image 3 super-resolved using our approach look perceptually similar to that obtained using homogeneous AR technique, the improvement in MSE can be seen from Table 5.3. Qualitatively, the results for proposed super-resolution approach are significantly better than the other approaches.

5.6 Conclusion

We have presented a technique to super-resolve a single image captured using a low cost camera. The high frequency content of the super-resolved image is learned in form of DCT coefficients using a database of low resolution images and their high resolution versions. The suggested learning technique yields better close approximation to the solution as compared to that obtained using wavelet based learning technique. The proposed

Image 1

Image 2

Image 3

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

Figure 5.5: Perceptual comparison of the results. (a) The observation, (b) image expanded using bicubic interpolation, (c) super-resolved image using homogeneous AR model and (d) super-resolved image using nonhomogeneous AR model.

learning algorithm can be straightforwardly extended to incorporate learning for images compressed using DCT. The LR observation is represented using linear model and non-homogeneous AR model is used as prior for regularization. The observed image is divided into several homogeneous regions and the model parameters for each of the homogeneous regions are estimated from the close approximation obtained using DCT based learning technique. The cost function consisting of a data fitting term and a linear prior term is optimized using simple gradient descent technique. The proposed method yields better results considering both smoother regions as well as texture regions. It may be concluded that with learning the additional knowledge using the DCT, we have been able to add some structure to our solution and thereby bring more information.

Chapter 6

A Fast Approach to Learning Based SR Using Zoom cue

In chapter 3, we addressed the problem of super-resolving an image using zoomed observations captured by varying the zoom setting of a camera. In this problem, we modeled the super-resolution image as a MRF field and the MRF model parameters were adjusted by trial and error approach. We used MAP estimation and arrived at a convex cost function. The super-resolution was recovered by optimizing the cost function using simple and fast gradient descent technique. Here, MRF prior served as smoothness constraints to regularize the solution. However, the imposition of the smoothness constraints flattens the entity to be super-resolved causing distortion along the discontinuities and finally leads to undesired smoothed solutions. This approach cannot provide details of pixels representing lines, edges, corners, and texture regions. Since, the pixels with significant change in intensities carry important information, it is necessary to prevent them being smoothed while regularization. In order to prevent the discontinuities, MRF prior with line fields can used [1]. The resulting non convex cost function needs to be optimized using global optimization technique.

In this chapter, we solve the problem formulated in section 3.3. We model the super-resolved image using MRF and incorporate line fields in order to preserve the discontinuities. We first obtain a close approximation to the super-resolved image by using a learning based approach and learn both the aliasing and the discontinuity preserving MRF prior parameters using this approximation. We derive a non convex cost function and minimize it using a computationally efficient particle swarm optimization technique.

Although there is some overlap with the work in [1], novelty of our work is as under.

- We obtain close HR approximation to the super-resolved image using the DCT based learning technique that uses the database of low resolution images and their high resolution versions. We use the close HR approximation as an initial guess while optimizing unlike using expanded image obtained by successive bilinear interpolation in [1].
- The authors in [1] select MRF model parameters on trial and error basis. Since, accurate parameters play important role in reconstructing the true HR image, we estimate the same from the close HR approximation. We expect that the estimated parameters are close to their true values.
- The decimation model used in [1] is based on the assumption that the aliased low resolution pixel intensity of an image point is always an equally weighted sum of the high resolution intensities. However, in practice, the observed intensity at a pixel captured due to low resolution sampling depends on various factors and hence it has to be estimated. We estimate decimation (aliasing) using the close HR approximation and low resolution observations.
- While using the discontinuity preserving MRF prior, we employ particle swarm optimization technique and thus avoid the use of computationally taxing optimization techniques such as simulated annealing.

6.1 Problem Formulation

Although the problem formulation is described in section 3.3, we present it here briefly for the sake of convenience. Let $\{Y_i\}_{i=1}^{p}$ be a set consisting of p low resolution observed images each of which is obtained by adjusting the optical zoom. The size of each observation is $M \times M$ pixels. The relationship between these low resolution images and the high resolution image is shown in Figure 6.1. Here the most zoomed image of the scene Y_3 has the highest resolution. The least zoomed image Y_1, corresponding to the entire scene, needs to be upsampled to the size of $(l_1 l_2 \cdots l_{p-1}) \times (M \times M) = (N \times N)$ pixels, where $l_1, l_2, \ldots, l_{p-1}$ are the zoom factors between observed images of scene $Y_1Y_2, Y_2Y_3, \ldots, Y_{(p-1)}Y_p$, respectively. Given Y_p, the remaining $(p-1)$ observed images can be modeled as aliased and

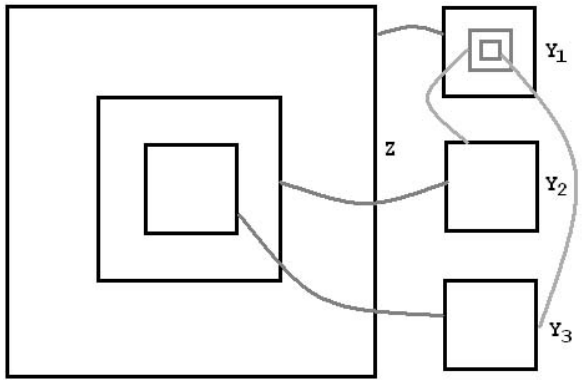


Figure 6.1: Illustration of observations at different zoom levels. Here Z is the super-resolved image, Y_1 and Y_3 are the least zoomed and the most zoomed observations, respectively.

noisy versions of this single high-resolution image (Z) of appropriate region. Note that although the observations are of same size, they are at different spatial resolutions. We write the forward model for the observed images as expressed in [1].

$$\mathbf{y}_m = D_m C_m \mathbf{z} + \mathbf{n}_m, \tag{6.1}$$

where $m = 1, 2, \ldots, p$. Here $\mathbf{y}$, $\mathbf{z}$ and $\mathbf{n}$ are vector representation of the observed image, super-resolved image and noise respectively. C is cropping operator and D is aliasing matrix. The noise $\mathbf{n}$ is assumed to be independent and identically distributed with zero mean and variance σ_n^2. Now the problem is to estimate $\mathbf{z}$ given $\mathbf{y}_m$s which is an ill-posed problem and requires proper regularization.

6.2 Proposed Approach

The proposed SR approach consists of two basic processes, namely a learning process and a regularization process. In the training process, we obtain a close approximation to final super-resolved image using the DCT based learning technique where the high frequency contents are learned from a database consisting of LR images and their HR versions. We learn decimation (aliasing) from this close SR approximation and the zoomed observations. For regularization, we represent the HR image by a discontinuity preserving MRF prior using line fields. The model parameters are estimated from the close approximation using homotopy continuation method [5]. We propose the use particle swarm optimiza-

tion technique for minimizing the non convex cost function. The particles required for optimization are obtained using existing image expansion techniques as well as using learning based techniques.

6.2.1 Learning the Close HR Approximation

We first obtain the close approximation to the super-resolved image using a learning based method. We use the DCT based learning method described in section 5.2.2. We have three observations Y_1, Y_2, and Y_3 of size $M \times M$ each. We aim to super-resolve the least zoomed observation Y_1 at the resolution of Y_3. The resolution of the observation Y_2 higher than that of Y_1 and lower than that of Y_3. In order to get the best result, we learn the HR estimate of size $4M \times 4M$ $(q = 4)$ for the least zoomed observation using the DCT based approach recursively. We then obtain the HR estimate of size $2M \times 2M$ $(q = 2)$ for Y_2 using same algorithm. We replace the centermost $M \times M$ region of this learned image by the most zoomed observation Y_3 as it is captured at the highest resolution. We then obtain the close approximation of size $4M \times 4M$ to the super-resolved image by inserting the resulting $2M \times 2M$ image of Y_2 into the centermost $2M \times 2M$ region of the HR estimate of Y_1.

Due to different configurations of camera hardware and different zoom settings, the aliasing differs from image to image. Hence the aliasing has to be estimated. Since the close approximation to the super-resolved image is now already available, we make use of the same to obtain the decimation matrix entries and thus learn the aliasing better. We use the least squares technique described in section 3.1.2 for learning the decimation (aliasing). We also estimate the MRF model parameters from the same close approximation.

6.2.2 Discontinuity Preserving MRF Prior Model

The MRF provides a convenient and consistent way of modeling spatially related features in an image. The MRF model prior is extensively used to represent the local dependencies among the pixels in images. Although this constraints helps to stabilize the optimization process, it pushes the reconstruction towards smooth entity and causes the distortions along discontinuities. In order to have a better reconstruction of natural images having

edges and varying textures, the simple MRF prior is to be amended with some mechanism to take care of these discontinuities. In order to incorporate provisions for detecting such discontinuities, Geman and Geman [164] introduced concept of line fields located on a dual lattice. One can define horizontal and vertical linefields on a lattice and choose binary variables over these line field in order to detect the discontinuities in horizontal and vertical directions, respectively. An on-state of the variable indicates that there exist a discontinuity in form of high gradient between neighboring sites in the lattice.

Given an $N \times N$ lattice Z of sites $\{(i,j)|1 < (i,j) \leq N\}$, one can define a horizontal line field L consisting of binary elements $l_{i,j}$ that connect the site (i,j) to its neighboring site $(i-1,j)$. The element $l_{i,j}$ is set to 1 (i.e., on-state) if there exists a horizontal discontinuity at the site $l_{i,j}$. Thus,

$$l_{i,j} = \begin{cases} 1 & \text{if } |z_{i,j} - z_{i-1,j}| > \theta_1, \\ 0 & \text{else.} \end{cases}$$

Here θ_1 is an appropriate threshold value that declares the horizontal discontinuity. Similarly, a binary elements $v_{i,j}$ of vertical line field V connecting the site (i,j) to its neighboring site $(i,j-1)$ is set to 1 (i.e., on-state) if there exists a vertical discontinuity at the site $l_{i,j}$. Thus,

$$v_{i,j} = \begin{cases} 1 & \text{if } |z_{i,j} - z_{i,j-1}| > \theta_2, \\ 0 & \text{else,} \end{cases}$$

where, θ_2 is a threshold value that declares the vertical discontinuity. In our experimentation, the thresholds for detecting the horizontal and vertical discontinuities are kept the same. Thus we have $\theta_1 = \theta_2 = \theta$. A trivial solution would declare a discontinuity at every location. In order to declare discontinuities at genuine locations, each turn on of a line field variable is penalized by a quantity γ so as to prevent spurious discontinuities. The energy function for Z with discontinuity fields L and V obtained by modifying equation

(3.13) is,

$$\begin{aligned} V(\mathbf{z}) &= \sum_{i,j} \mu \Big[(z_{i,j} - z_{i,j-1})^2 (1 - v_{i,j}) + (z_{i,j+1} - z_{i,j})^2 (1 - v_{i,j+1}) \\ &+ (z_{i,j} - z_{i-1,j})^2 (1 - l_{i,j}) + (z_{i+1,j} - z_{i,j})^2 (1 - l_{i+1,j}) \Big] \\ &+ \gamma \big[v_{i,j} + v_{i,j+1} + l_{i,j} + l_{i+1,j} \big]. \end{aligned} \tag{6.2}$$

This energy function consists of the smoothness term and the penalty term to prevent spurious discontinuities. Here parameters μ and γ correspond to the relative weights of the smoothness term and the penalty term necessary to prevent occurrence of spurious discontinuities. We estimate these parameters from the close HR approximation using homotopy continuation method proposed in [5]. Since, this energy function is non convex, there is a possibility of the steepest descent type of algorithms getting trapped in a local minima. Hence global optimization technique needs to be used for obtaining MAP estimates of the super-resolved image.

6.2.3 MAP-MRF Formulation

Using the data fitting term given by equation (6.1) and the energy function $V(\mathbf{z})$, the estimate of the super-resolved image using the MAP-MRF formulation can be obtained as,

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{argmin} \left[\sum_{m=1}^{p} \frac{\|\mathbf{y}_m - D_m C_m \mathbf{z}\|^2}{2\sigma_n^2} + V(\mathbf{z}) \right]. \tag{6.3}$$

The cost function in equation (6.3) balances two types of errors. The left term is minimized when a candidate $\mathbf{z}$, projected through the observation model described by equation (6.1), matches the observed data. The right term is a regularization term, which is necessary as directly minimizing the first term is an ill posed problem. The inclusion of the binary line fields makes the cost function non linear, which cannot differentiated. Hence, it cannot be minimized using simple differentiation based optimization techniques. Minimization of this cost function requires the use of global optimization techniques.

Global optimization techniques such as simulated annealing, graph-cuts are widely used for optimization of non-convex cost function. Although a good initial guess for

simulated annealing speed up the computation and lead to a global minima, it is computationally demanding and takes a very long time for convergence. The graph-cuts technique can be employed when the cost function is regular. Since the derived cost function is not regular, graph-cuts techniques can not be used. In order to speed up the convergence and reduce the computation time, we propose the use of particle swarm optimization technique. It is a method for solving global optimization problems [165].

6.3 Particle Swarm Optimization

Particle swarm optimization (PSO) is a new evolutionary computation technique proposed by Kennedy and Eberhart [166, 167]. The particle swarm concept was motivated from the simulation of social behavior. The original intent was to graphically simulate the graceful but unpredictable movement of bird flocking. The PSO algorithm mimics the behavior of flying birds and their means of information exchange to solve optimization problems. Each potential solution is seen as a particle with a certain velocity, and "flies" through the problem space. Each particle adjusts its flight according to its own flying experience and its companions' flying experience. The particle swarms find optimal regions of complex search spaces through the interaction of individuals in a population of particles. PSO has been successfully applied to a large number of difficult combinatorial optimization problems. Studies show that it often outperforms Genetic Algorithms [168]. It has been seen that PSO has a strong search capability in the problem space and can discover optimal solutions quickly. The advantages of the PSO technique are easy implementation, no need of derivatives, very few parameters and ease of parallelization for concurrent processing.

Recently, researchers have explored the use of PSO to solve variety of problems in image processing and pattern recognition [169]. PSO is initialized with a population of random solutions, called 'particles'. The algorithm then searches for the best solution through an iterative process. At every iteration, the fitness of each particle is evaluated using the fitness function. If it is the best value the particle has achieved so far, the particle stores that value as 'personal best'. The best fitness value achieved by any particle during current iteration is stored as 'global best'. The pseudocode of the particle swarm optimization technique is given by Algorithm is 6.1.

Algorithm 6.1: Particle swarm optimization

```
Data: swarm of particles
Result: best solution
foreach particle in the swarm do
    initialize the particle position;
    initialize the particle velocity;
    set personal best and global best;
end
repeat
    foreach particle in the swarm do
        evaluate the fitness of the particle;
        update velocity of the particle;
        update position of the particle;
        find the personal best;
    end
    find the global best;
until convergence or maximum iterations ;
```

Let $S = \{Z_b | b = 1, 2, \ldots, B\}$ be a swarm initialized with initial HR images as the particles. The fitness of the particle Z_b at n^{th} iteration is denoted by $F^n_{Z_b}$. The fitness function in our case is the cost function that has to be minimized. Let Z^n_{bp} be the personal best of particle Z_b and Z^n_g be the group best at the n^{th} iteration. The fitness values of Z^n_{bp} and Z^n_g are denoted by $F_{Z^n_{bp}}$ and $F_{Z^n_g}$, respectively. While using the PSO the velocity V_b and position Z_b after n^{th} iteration are updated according to the following two equations [170],

$$V^{n+1}_b = wV^n_b + c_1 r_1 (F_{Z^n_{bp}} - F^n_{Z_b}) + c_2 r_2 (F_{Z^n_g} - F^n_{Z_b}), \tag{6.4}$$

and

$$Z^{n+1}_b = Z^n_b + V^n_b, \tag{6.5}$$

where r_1 and r_2 are random numbers uniformly distributed in $[0, 1]$. The equation (6.4) consists of three terms. The first term provides the 'flying particles' with a degree of memory capability allowing the exploration of new search space areas. The second term is the 'cognition' term, which represents the private thinking of the particle itself. The third term is the 'social' term, which represents the collaboration among the particles. The inertia weight w serves as a memory of previous velocities [171]. A large value favors exploration, while a small inertia favors exploitation. The constants c_1 and c_2 are cognitive and social parameters reflecting the weighting of stochastic acceleration

terms that pull each particle toward personal best and global best positions, respectively. Equation (6.4) is used to calculate the particle's new velocity according to its previous velocity and the distances of its current position from its own best experience (position) and the group's best experience. The particle flies toward a new position according to equation (6.5). The performance of each particle is measured according to a pre-defined fitness function. The iterative process is repeated until the number of iteration exceeds the maximum number or stopping criteria is met. The 'global best' particle at the end of iterative process represents the solution. The flow-chart of the PSO algorithm is shown in Figure 6.2.

6.4 Experimental Results

We conducted experiments on real images captured by adjusting the zoom setting of a camera. The observed images Y_1, Y_2 and Y_3 are shown in Figure 6.3(a)-(c). The zoom factor between Y_1 and Y_2 is 2 that between Y_1 and Y_3 is 4. All observations are of size 64×64 pixels. Our results are compared with other approach using the qualitative as well as quantitative measures. In this chapter, the quality of a super-resolution image is defined as the similarity of the super-resolved image with the original high-resolution image. We use MSE and Structural Similarity index to measure the quality of results.

MSE is the most widely used metric for quantitative comparison of images in the areas of compression, restoration, super-resolution etc. MSE have clear physical meaning and have mathematically convenient expressions for comparing the performance of different approaches. However there are some implicit assumptions while using this measure [172]. These include: the signal fidelity is independent of spatial relationship between the samples of original signal and it is independent of any relationship between original signal and error signal. The faithfulness with which the signal is reconstructed is also independent of the signs of error signal samples giving equal importance to all signal samples. Because of these limitations, MSE is not an accurate measure for comparison. Motivated by the fact that the human visual system is highly sensitive to structural distortion, the authors in [173] introduced an alternative measure for quality assessment termed as structural similarity (SSIM). This new measure evaluates the structural changes between two complex structured signals directly (note that we use signal term in the above explanation,

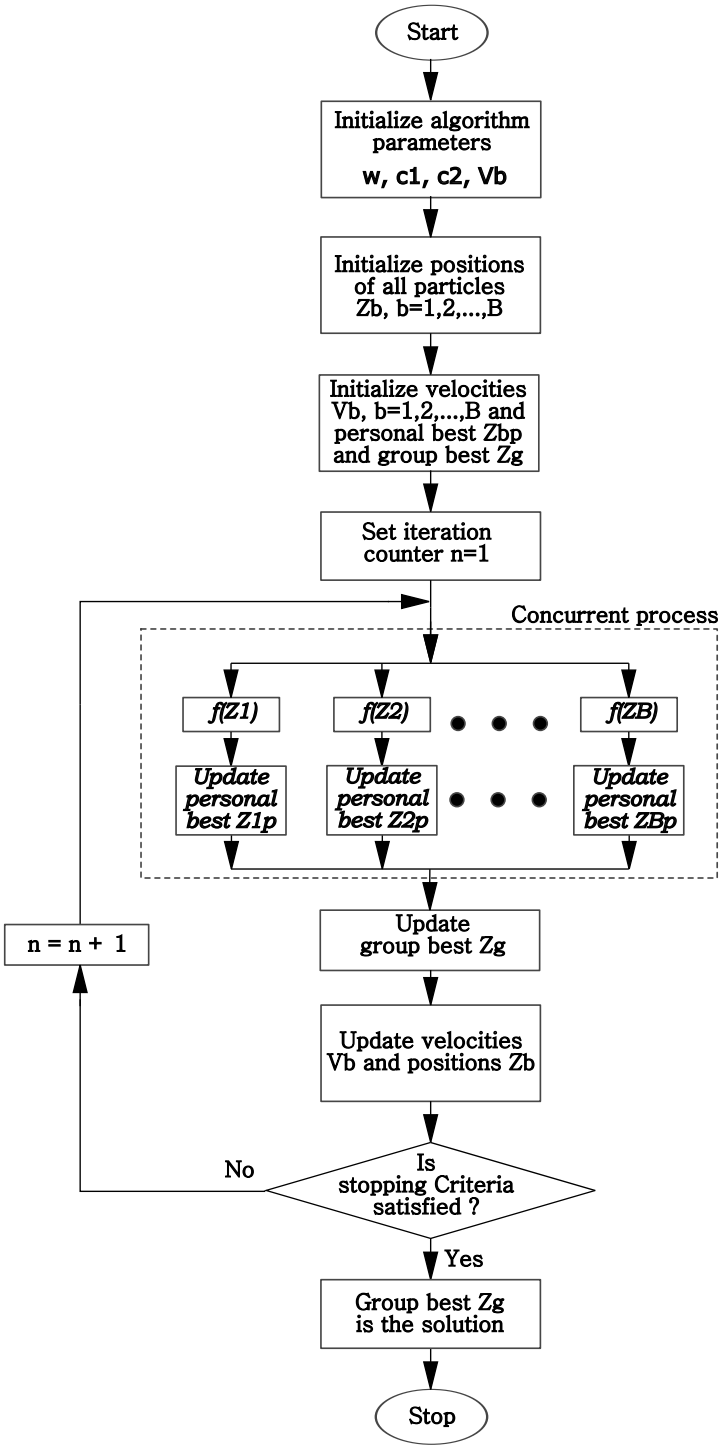


Figure 6.2: Flow-chart of the particle swarm optimization.

it applies to images as well). SSIM can be viewed as a quality measure of one of the images being compared, while the other image is regarded as of perfect quality. The implementation of SSIM takes into account contrast, luminance, and structure to determine similarity between the local patches in the images under comparison.

The SSIM between two images is computed as follows: Let the original image and the super-revolved image be divided into several patches. Let $\mathbf{x}$ and $\mathbf{y}$ be the image patches taken from the same locations of the two images under comparison. Then the local similarity $S(\mathbf{x}, \mathbf{y})$ can be calculated using,

$$S(\mathbf{x}, \mathbf{y}) = \Big(\frac{2\mu_x\mu_y + K_1}{\mu_x^2 + \mu_y^2 + K_1}\Big)\Big(\frac{2\sigma_x\sigma_y + K_2}{\sigma_x^2 + \sigma_y^2 + K_2}\Big)\Big(\frac{2\sigma_{xy} + K_3}{\sigma_x\sigma_y + K_3}\Big), \tag{6.6}$$

where μ_x and μ_y are local sample means of image patches $\mathbf{x}$ and $\mathbf{y}$, respectively. Similarly σ_x and σ_y are the local sample standard deviations of $\mathbf{x}$ and $\mathbf{y}$, respectively. Here K_1, K_2 and K_3 are small positive constants preventing the numerical instability that may arise because of division with small denominators and σ_{xy} is the sample cross correlation of $\mathbf{x}$ and $\mathbf{y}$ after removing their means. It is obtained using,

$$\sigma_{xy} = \frac{1}{T-1}\sum_{i=1}^{T}(x_i - \mu_x)(y_i - \mu_y). \tag{6.7}$$

Here x_i and y_i represent pixel intensities in images patches $\mathbf{x}$ and $\mathbf{y}$, respectively and T is the total number of pixels in each image patch $\mathbf{x}$ and $\mathbf{y}$. The SSIM score of the entire image is computed by averaging the SSIM values of the patches across the image. The value of SIMM ranges between 0 and 1, where a higher value means a higher structural similarity and hence better image quality and the highest value 1 means excellent quality indicating perfect similarity. For more details see [172]. In our case, the patch size is set at 5×5 and parameters K_1, K_2 and K_3 are set to 0.05. PSO parameters are chosen as $c_1 = 1$, $c_2 = 0.5$ and $w = 1$. We expand the least zoomed images using existing interpolation techniques and use them as particles in PSO. In our experiment, the swarm consists of 10 particles. Three particles are obtained using bilinear, bicubic and lanczose interpolation techniques. The initial SR approximation is used as a particle. The remaining particles are obtained by manipulating these four particles. The MRF model parameters μ and γ estimated from the SR approximation using homotopy continuation method proposed in

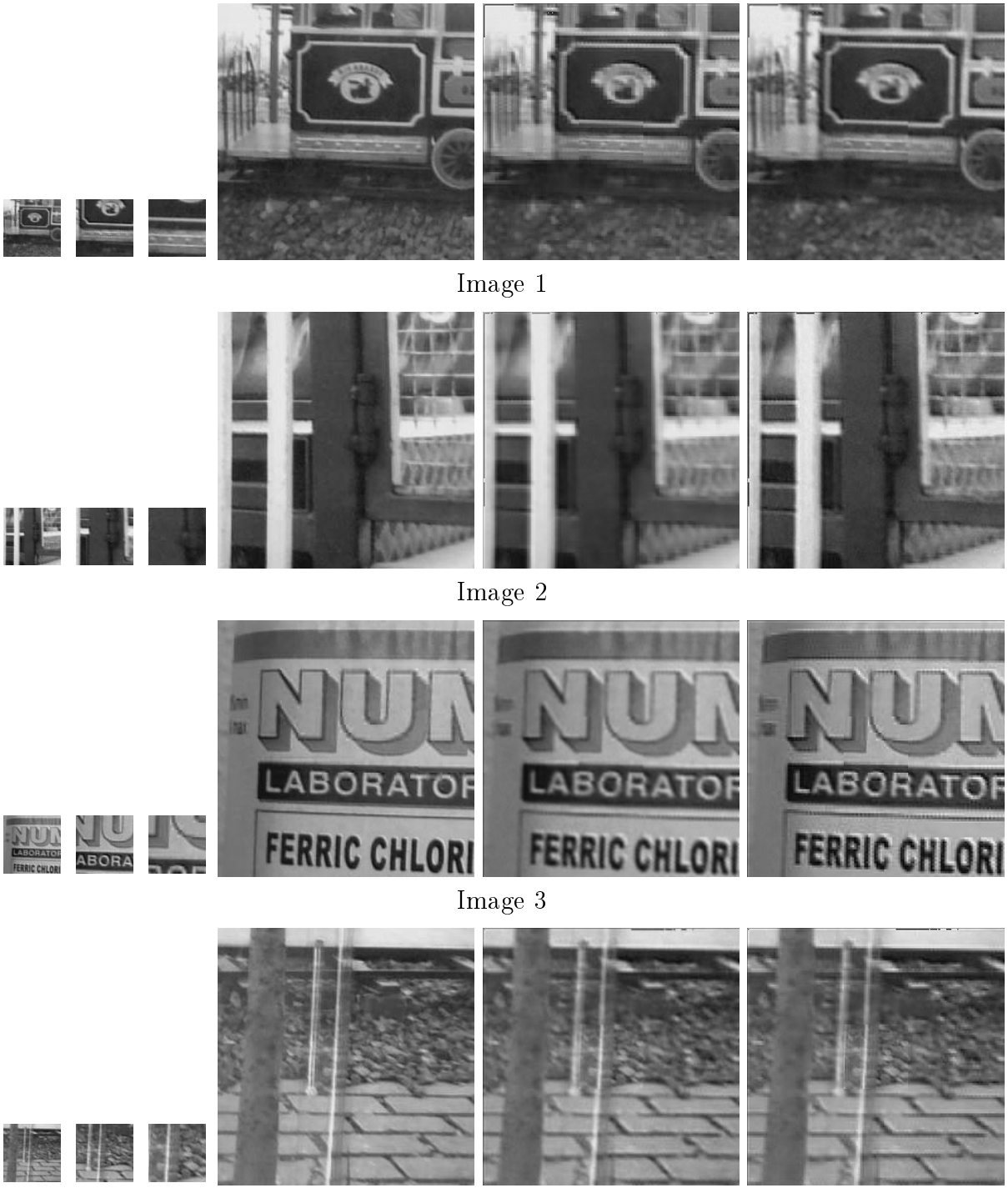


Figure 6.3: Performance comparison. (a)-(c) Observations: (a) is the least zoomed observation and (c) is the most zoomed one, (d) ground-truth image for observation in (a), (e) image super-resolved using the approach in [1], and (f) image super-resolved using proposed approach.

Image	μ	γ
Image 1	0.02057	4.2863
Image 2	0.03023	5.9252
Image 3	0.01337	2.7917
Image 4	0.00913	3.1809

Table 6.1: Model parameters estimated using homotopy continuation method [5].

[5] are shown in Table 6.1.

We show the effectiveness of the proposed approach by conducting experiment on real world images having different textures. The results of our experiments for decimation factor $q = 4$ are shown in Figure 6.3. Figure 6.3(d) show the ground-truth images or high resolution images for the observations in Figure 6.3(a). Figure 6.3(e) show the images super-resolved using the algorithm proposed in [1]. Images super-resolved using the proposed approach using PSO are displayed in Figure 6.3(f). Image 1 contains variety of objects: a rectangle with an oval shaped logo inside it, a wheel with spokes, irregularly shaped stones at bottom, and some smooth regions. Image 1 in Figure 6.3(f) exhibits better texture at bottom than that in Figure 6.3(e). The rectangle in the center and logo inside it look pleasant and appear structurally similar to those in the ground-truth image. Image 2 contains smooth regions along with a large number of horizontal and vertical edges. The Image 2 super-resolved using the proposed approach looks more similar to the ground-truth as compared to that displayed in Figure 6.3(e). Image 3 consists of English text in different shape and size. Here also characters in the text look sharper in the image super-resolved using the proposed approach. Image 4 is a combination of textured regions, smooth regions and edges in different directions. Image in Figure 6.3(f) shows marginal improvement over that in Figure 6.3(e). The comparison shows that the SR reconstruction using the proposed approach is visually better, filling the reconstructed image with realistic fine spatial details, such as sharp lines and edges. The close examination of the super-resolved images in Figure 6.3(e) reveals that the reconstruction at the peripheral region of the image is inferior than that at the central region. Visual assessments suggest that a the proposed method outperform the method in [1]. Possible reasons are: 1) The authors in [1] use the MRF model parameters that are adjusted on trial and error basis; 2) Very little information is available for SR reconstruction at the peripheral region of an image; 3) Initial estimate is obtained using successive bilinear

Image	Optimization Time (in seconds)	
	SA	PSO
Image 1	16730.94	10.586
Image 2	14641.36	10.529
Image 3	15419.78	10.285
Image 4	14827.32	10.549

Table 6.2: Optimization time comparison.

Image	MSE		SSIM	
	MRF with SA	Proposed approach	MRF with SA	Proposed approach
Image 1	0.026226	0.026181	0.9881	0.9789
Image 2	0.023515	0.022756	0.7876	0.8218
Image 3	0.020392	0.019700	0.6800	0.7912
Image 4	0.018696	0.017865	0.6767	0.7909

Table 6.3: Performance comparison between the proposed approach and the approach in [1] using two image quality measures. Lower values of MSE and higher values of SSIM represent better image quality.

interpolation.

Table 6.3 shows the performance improvement over the other method using MSE and SSIM. It show that our approach can attain lower MSE than the other method in all cases. The SSIM measure which is more sensitive to structural distortion show better enhancements with the proposed approach. Table 6.2 show computational advantage of using PSO for optimization over using simulated annealing. It can be seen that the optimization time for SA is almost around 4 hours, whereas that for the proposed method is less than 11 seconds. When comparing these two approaches, the proposed approach outperforms the other approach in terms of both, the processing time and reconstruction quality.

6.5 Conclusions

In this chapter, we have addressed the problem of capturing the right amount of scene information from the perspective of zoom based super-resolution. We have presented a fast approach for super-resolving the least zoomed observation at the resolution of the most zoomed observation. The solution is proposed in a MAP-MRF framework. Although,

MRF conveniently allows modeling of contextual constraints, it produces smooth solution. We incorporated line fields for preserving edges and arrived at non convex cost function. A learning based approach is proposed for obtaining an initial HR approximation. The aliasing matrix entries and the discontinuity preserving MRF parameters were estimated using the learned SR approximation. Visual comparison between the ground-truth images and the SR images obtained using proposed approach finds significant improvements in terms of details. Better enhancements are found at the edges and boundaries of objects. Low computational speed is one of the key issues while optimizing a non convex cost function. We have proposed use of particle swarm optimization. With the use of PSO significant speed up has been achieved. Our results show that the proposed method outperforms conventional zoom based image super-resolution approach in terms of both, reconstruction quality and processing speed.

Chapter 7

Conclusions and Future Research Work

7.1 Conclusions

In this thesis, we have addressed the problem of motion-free super-resolution. Motion based super-resolution techniques reconstruct super-resolution from multiple observations with sub-pixel shift among them. Accuracy of the registration of these observations plays an important role in the quality of the final solution. Since the registration is a computationally complex process, it has to be avoided. The motion-free super-resolution techniques use the observations that need not be registered. In addition, there may be situations where more than one observations are not available. Since, it is impossible to extract additional non-redundant information from the single observation, we needed to seek other sources for deriving additional information. Further, the fact that the richness of the natural images is difficult to capture analytically, motivated us to construct a database of low resolution images and its high resolution versions and use it to get non-redundant information for the super-resolved image. We have proposed learning based techniques which attempt to derive the details of high resolution image from the database by analyzing the spatial relationship between the image features across the scales. We have considered the problems of zoom based super-resolution and single frame super-resolution. We solved these inverse problems using regularization approach. We represented the forward process of image formation using a linear model. In zoom based

super-resolution problems we estimated the aliasing using the most zoomed observation and corresponding part in the least zoomed image. We represented the high resolution image field using different model and considered smoothness prior as well as edge preserving priors. The resulting cost functions were minimized using either simple gradient descent technique or global optimization technique depending on the nature of the cost function.

We began with the zoom based super-resolution. Given the observations of a scene captured by varying the zoom settings of a camera, we obtained the super-resolution of the least zoomed image at the resolution of the most zoomed observation. Normally, the intensity of the aliased pixel in the low resolution image obtained by averaging the corresponding pixels in the high resolution images. However in practice, the aliasing depends on several factors and has to be estimated. The availability of a portion of the scene at different zoom factor, motivated us to estimate the aliasing. We have proposed the technique of estimating the aliasing (decimation). In order to make the ill-posed problem a better posed, we have employed smoothness prior to constrain the space of solutions. We represented the high resolution image field using MRF field and adjusted the model parameters on trial and error basis. We have shown the efficacy of the proposed approach by conducting experiments on real world images. We have also applied the proposed approach to super-resolve the multispectral (MS) images in remote sensing. In this case we have used the high resolution panchromatic image to estimate aliasing on the MS images. A better solution could have been obtained if the MRF model parameters were estimated from the observations themselves. However, the estimation of these parameter is computationally taxing as it requires the computation of partition function. This motivated us to consider a different prior model that do not put computational burden. We assumed that the intensity of a pixel in an image is the linear combinations of the intensities of its neighboring pixels. We have considered the autoregressive prior model to represent high resolution image field and have obtained the super-resolution. It has been shown that proposed approach yields better results for both the super-resolution and multi-resolution fusion in remotely sensed images.

In order to capture the correspondence between low resolution image and their high resolution versions effectively, we have prepared a database of images captured at different resolutions by changing the zoom setting of a camera and used it to learn the high

frequency details of the super-resolved image. We have proposed a new learning based technique to obtain the close approximation to the super-resolved image. The accurate estimation of the prior model parameters requires the high resolution image. However in super-resolution problems the high resolution images are not available. We obtained better estimates of the parameters from the close approximation to the super-resolution image and use them in the minimization. In earlier experiments, we used homogeneous prior models. Since, the natural images consists of edges, and wide variety of textured regions, it cannot be considered as homogeneous fields. We have proposed here an inhomogeneous Gaussian Markov random field prior model that can adapt to the local structure in an image. Since it is adaptive to the local structures there is no need for separate prior for preserving edges as in previous experiment. We have applied the algorithm to color images and shown the results for real world images.

Although wavelets provide computational advantage, it fails to learn the edges in arbitrary directions other than horizontal, vertical and diagonal directions. In order to alleviate this difficulty, we have proposed a new learning technique based on discreet cosine transform (DCT). We use the database consisting of low resolution images and their high resolution versions and learn the DCT coefficients corresponding to the high resolution content. The learned image is considered as a close approximation to the super-resolved image and is used as the initial estimate while minimizing. As discussed earlier, the natural images cannot be represented using homogeneous models, we have employed non-homogeneous AR model prior and obtained the prior model parameters from the close approximation. We obtained super-resolution by minimizing the cost function using gradient descent technique.

We have proposed a fast approach to the zoom based super-resolution problem. We obtain close approximation to the super-resolved image using learning technique and use the same to estimate the aliasing and prior model parameters. We have solved the problem using MAP-MRF formulation. We have introduced discontinuity preserving MRF model prior and arrived at non convex cost function. The introduction of the discontinuity preserving prior give rise to a computationally expensive optimization process which makes the super-resolution reconstruction very slow. We have optimized the cost function using particle swarm optimization. The initial particles were obtained using conventional interpolation techniques and the learning approaches based on DCT and DWT. The

results shows that the proposed approach offers significant computational advantage over the other methods.

7.2 Future Research Work

Super-resolution is an ill-posed inverse problem. Solution of this problem requires additional information which may be available from either multiple observations or a huge database consisting of high resolution images. Most of the super-resolution techniques suffer from one or more disadvantages such as high computational complexity, need for huge training database, need for registration and poor visual quality. Therefore several of these issues are to be appropriately addressed in future. In this section, we discuss the directions for future research work to produce good practical super-resolution.

- The LR and HR images captured by varying the resolution setting of a camera truly represent the spatial features relationship between low resolution image and its high resolution versions. We have proposed wavelet transform based learning technique using the database consisting of such pairs. However the wavelets can capture only limited directional information. In addition 2-D wavelets are not good at capturing edges along counters. The natural images contain intrinsic geometric structures with edges located along smooth curves as well. The contourlet transform of an image consists of basis images oriented at various directions in multiple scales with flexible aspect ratios. Since the contourlet transforms are capable of capturing edges at arbitrary edges, one can consider contourlet transform for learning the fine details from the database of LR-HR images.

- In image formation process model we have assumed space invariant blur. This enables us the simple formulation for the data fitting term. However, in practice, the blur need not be space invariant hence it has to be estimated. One has to go down the resolution pyramid of the training images to locate the best match if the input images are blurred. This is a difficult problem as the scale is unknown. Further there may not exactly match the given levels of decomposition scales. A simple approach to handle the space variant blur is to divide the entire image into small blocks and super-resolve each of them individually.

- In our experiment we have employed a simple linear model for image formation process. While estimating decimation, a pixel in LR image is related with corresponding pixels in the HR image. Since there are limited number of the HR pixels for an LR pixel (depending on the decimation factor), the accurate estimation of the decimation may not be possible. One can consider the splines to model the image formation process and obtain better representation of the process by considering more HR pixels in larger neighborhood of an LR pixel.

- The prior for regularization is obtained by modeling the HR image as an MRF that can be expressed as a joint probability density function using the Gibbs distribution. The marginal distribution prior can be obtained by passing the HR image or its close approximation through different filters and computing the histograms of the filtered outputs. One can construct a bank of different types of filters such as Laplacian of Gaussian and Gabor filters with different parameters and arrive at cost function which can be optimized using simpler global optimization techniques such as particle swarm optimization or graph-cut.

- The proposed techniques for image super-resolution can be extended for video super-resolution. The video super-resolution requires the super-resolution reconstruction in both the spatial domain and temporal domain. One can develop algorithm for reconstructing additional temporal frames using motion fields. With the spatial super-resolution of all the frames and the temporal super-resolution of the video sequence, one can obtain super-resolved video by concatenating the resulting SR frames.

- Most of the super-resolution algorithms are based on iterative process. This limits the use of the algorithms for real-times application such as video super-resolution. It will be interesting to see the development towards the computationally efficient super-resolution algorithm that may use a strong but differentiable prior such that the optimization converges fast and lead to good quality output image.

Bibliography

[1] M. V. Joshi, S. Chaudhuri, and P. Rajkiran, "Super-resolution imaging:use of zoom as a cue," *Image and Vision Computing*, vol. 14, no. 22, pp. 1185–1196, 2004.

[2] P. P. Gajjar and M. V. Joshi, "New learning based super-resolution: Use of DWT and IGMRF prior," *IEEE Transaction on Image Processing*, vol. 19, no. 5, pp. 1201 – 1213, 2010.

[3] C. V. Jiji, M. V. Joshi, and S. Chaudhuri, "Single-frame image superresolution using learned wavelet coefficients," *International Journal of Imaging Systems and Technology*, vol. 14, no. 3, pp. 105–112, 2004.

[4] K. I. Kim and Y. Kwon, "Example-based learning for single-image super-resolution," *Thirtieth annual symposium of the Deutsche Arbeitsgemeinschaft fűr Mustererkennung*, pp. 456–465, 2008.

[5] P. K. Nanda, U. B. Desai, and P. G. Poonacha, "A homotopy continuation method for parameter estimation in MRF models and image restoration," *IEEE symposium on Circuits and Systems*, vol. 3, pp. 273–276, 1994.

[6] J. E. Estes and D. S. Simonett, *Manual of remote sensing.* Bethesda, MD: American society for photogrammetry and remote sensing, 1975, ch. Fundamentals of Image Interpretation, pp. 869–1076.

[7] 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 *Proceedings of Inst. Elect. Eng.*, vol. 140, no. 1, 1993, pp. 19 – 25.

[8] H. Stark and P. Oskui, "High resolution image recovery from image-plane arrays using convex projections," *Journal of the Optical Society of America - A*, vol. 6, no. 11, pp. 1715 – 1726, 1989.

[9] K. Aizawa, T. Komatsu, and T. Saito, "A scheme for acquiring very high resolution images using multiple cameras," in *Proceedings of International Conference on Aucostics, Speech, Signal Processing*, 1992, pp. 289 – 292.

[10] X. Li, "Adaptive markov random fields for example-based super-resolution of faces," *EURASIP Journal of Appllied Signal Processing*, pp. 1 – 12, 2006.

[11] S. Chaudhuri and D. R. Taur, "High resolution slow motion sequencing," *IEEE Signal Processing Magazine*, pp. 16 – 24, 2005.

[12] S. Zhang, "Application of super-resolution image reconstruction to digital holography," *EURASIP Journal of Appllied Signal Processing*, vol. 2006, pp. 1 – 7, 2006.

[13] M. S. Lee, M. Y. Shen, and C. C. J. Kuo, "Techniques for flexible image/video resolution conversion with heterogeneous terminals," *IEEE Communications Magazine*, pp. 61 – 67, 2007.

[14] D. Barreto, L. Alvarez, R. Molina, A. Katsaggelos, and G. Callico, "Region-based super-resolution for compression," *Multidimensional Systems and Signal Processing*, vol. 18, no. 2-3, pp. 59–81, 2007.

[15] T. Taxt and R. Jirik, "Superresolution of ultrasound images using the first and second harmonic signal ultrasonics," *IEEE Transactions on Ferroelectrics and Frequency Control*, vol. 51, no. 2, pp. 163 – 175, 2004.

[16] G. Clement, J. Huttunen, and K. Hynynen, "Superresolution ultrasound imaging using back-projected reconstruction," *The Journal of the Acoustical Society of America*, vol. 118, no. 6, pp. 3953 – 3960, 2005.

[17] H. Li and D. Doermann, "Superresolution-based enhancement of text in digital video," in *Proceedings on the International Conference on Pattern Recognition*, vol. 1, 2000, pp. 847 – 850.

[18] D. Capel and A. Zisserman, "Super-resolution enhancement of text image sequences," in *Proceedings of International Conference on Pattern Recognition*, vol. I, 2000, pp. 600 – 605.

[19] G. Dalley, B. Freeman, and J. Marks, "Single-frame text super-resolution: A Bayesian approach," in *Proceedings of International Conference on Image Processing*, vol. V, 2004, pp. 3295 – 3298.

[20] Y. Altunbasak, A. Patti, and R. Mersereau, "Super-resolution still and video reconstruction from mpeg-coded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 4, pp. 217 – 226, 2002.

[21] K. Donaldson and G. Myers, "Bayesian super-resolution of text in video with a text-specific bimodal prior," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 1188 – 1195.

[22] L. Teodosio and W. Bender, "Salient stills," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 1, no. 1, pp. 16–36, 2005.

[23] C. Rubert, L. Fonseca, and L. Velho, "Learning based super-resolution using yuv model for remote sensing images," in *Proceedings of WTDCGPI*, 2005.

[24] H. Tao, X. Tang, J. Liu, and J. Tian, "Superresolution remote sensing image processing algorithm based on wavelet transform and interpolation," in *Proceedings of SPIE: Image Processing and Pattern Recognition in Remote Sensing*, vol. 4898, 2003, pp. 259 – 263.

[25] N. K. Bose and K. J. Boo, "High-resolution image reconstruction with multisensors," *International Journal of Imaging Systems and Technology*, vol. 9, no. 4, pp. 294–304, 1998.

[26] M. Elad and A. Feuer, "Super-resolution restoration of an image sequence: adaptive filtering approach," *IEEE Transactions on Image Processing*, vol. 8, no. 3, pp. 387–395, 1999.

[27] J. C. Gillette, T. M. Stadtmiller, and R. C. Hardie, "Aliasing reduction in staring infrared imagers utilizing subpixel techniques," *Optical Engineering*, vol. 34, no. 11, pp. 3130–3137, 1995.

[28] G. Jacquemod, C. Odet, and R. Goutte, "Image resolution enhancement using subpixel camera displacement," *Signal Processing*, vol. 26, no. 1, pp. 139–146, 1992.

[29] T. Komatsu, K. Aizawa, T. Igarashi, and T. Saito, "Signal processing based method for acquiring very high resolution images with multiple cameras and its theoretical analysis," in *IEE Proceedings. I, Communications, Speech and Vision*, vol. 140, no. 1, 1993, pp. 19–24.

[30] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," *Advances in Computer Vision and Image Processsing*, pp. 317–339, 1984.

[31] S. Borman and R. Stevenson, "Spatial resolution enhancement of low-resolution image sequences: A comprehensive review with directions for future research," in *Technical Report*, 1998.

[32] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47 – 57, 2004.

[33] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 21 – 36, 2003.

[34] H. Ur and D. Gross, "Improved resolution from sub-pixel shifted pictures," *CVGIP: Graphical models and Image Processing*, vol. 54, pp. 181 – 186, 1992.

[35] M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda, "Infrared registration and high resolution reconstruction using multiple translationally shifted aliased video frames," *IEEE Transactions on Instruments and Measuremets*, vol. 49, pp. 915 – 923, 2000.

[36] N. Nguyen and P. Milanfar, "An efficient wavelet-based algorithm for image super-resolution," in *Proceedings of International Conference on Image Processing*, vol. 2, 2000, pp. 351–354.

[37] V. N. Dvorchenko, "Boounds (on deterministic) correlation functions with application to registration," *IEEE Transactions on Pattern Analalysis and Machine Intelligence*, vol. 5, no. 2, pp. 206 – 213, 1983.

[38] Q. Tian and M. N. Huhns, "Algorithm for subpixel registration," *Computer Vision, Graphics, Image Processing*, vol. 35, pp. 220 – 233, 1986.

[39] L. G. Brown, "A survey of image registration techniques," *ACM Computer Surveys*, vol. 24, no. 4, pp. 325 – 376, 1992.

[40] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, 1990.

[41] S. P. Kim and W. Y. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 534–539, 1993.

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

[43] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive implementation of total least squares algorithm for image construction from noisy, undersampled multiframes," in *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing*, vol. 5, 1993, pp. 269 – 272.

[44] M. C. Hong, M. G. Kang, and A. K. Katsaggelos, "An iterative weighted regularized algorithm for improving the resolution of video sequences," in *Proceedings of International Conference on Image Processing*, vol. 2, 1997, pp. 474 – 477.

[45] M. G. Kang, "Generalized multichannel image deconvolution approach and its applications," *Optical Engineering*, vol. 37, no. 11, pp. 2953 – 2964, 1998.

[46] R. C. Hardie, K. J. Barnard, J. G. Bognar, E. E. Armstrong, and E. A. Watson, "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," *Optical Engineering*, vol. 37, no. 1, pp. 247 – 260, 1998.

[47] N. K. Bose, S. Lertrattanapanich, and J. Koo, "Advances in superresolution using L-curve," in *Proceedings of International Symposium on Circuits and Systems*, vol. 2, 2001, pp. 433 – 436.

[48] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Transactions on Image Processing*, vol. 3, no. 3, pp. 233–242, 1994.

[49] 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 Transactions on Image Processing*, vol. 12, no. 6, pp. 1621–1633, 1997.

[50] M. Elad and A. Feuer, "Restoration of a single super-resolution image from several blurred, noisy and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 12, no. 6, pp. 1646–1658, 1997.

[51] D. Capel and A. Zisserman, "Automated mosaicing with super-resolution zoom," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 1998, pp. 885–891.

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

[53] B. K. Gunturk and M. Gevrekci, "High-resolution image reconstruction from multiple differently exposed images," *IEEE Signal Precessing Letters*, vol. 13, no. 4, pp. 197–200, 2006.

[54] K. V. Suresh, G. M. Kumar, and A. N. Rajagopalan, "Superresolution of license plates in real traffic videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 321–331, 2007.

[55] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challanges in super-resolution," *International Journal of Imaging Systems and Technology*, no. 14, pp. 47–57, 2004.

[56] S. Farsiu, M. Elad, and P. Milanfar, "Multi-frame demosaicing and super-resolution of color images," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 141–159, 2006.

[57] D. Rajan and S. Chaudhuri, "Generalized interpolation and its application in superresolution imaging," *Image and Vision Computing*, vol. 19, no. 13, pp. 957 – 969, 2001.

[58] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitary sampling lattices and non zero aperture time," *IEEE Transactions on Image Processing*, vol. 8, no. 6, pp. 1064–1076, 1997.

[59] P. E. Eren, M. I. Sezan, and A. Tekalp, "Robust, object-based high-resolution image reconstruction from low-resolution video," *IEEE Transactions Image Processing*, vol. 6, no. 10, p. 14461451, 1997.

[60] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 231–239, 1991.

[61] S. Park, M. K. Park, and M. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, no. 20, pp. 21–36, 2003.

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

[63] D. Robinson, S. Farsiu, and P. Milanfar, "Optimal registration of aliased images using variable projection with applications to super-resolution," *The Computer Journal*, April 2007.

[64] P. Vandewalle, S. Susstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP Journal of Appllied Signal Processing*, vol. 2006, 2006.

[65] W. Y. Zhao, "Super-resolution with significant illumination change," in *Proceedings of International Conference on Image Processing*, vol. 3, 2004, pp. 1771 – 1774.

[66] B. Zitova and J. Flusser, "Image registration methods : A survey," *Image and Vision Computing*, vol. 21, pp. 977 – 1000, 2003.

[67] H. Arora, A. M. Namboodiri, and C. V. Jawahar, "Robust image registration with illumination, blur and noise variations for super-resolution," in *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 1301 – 1304.

[68] W. Y. Zhao and H. S. Sawhney, "Is super-resolution with optical flow feasible?" in *Proceedings of European Conference on Computer Vision*, vol. 1, 2002, pp. 599 – 613.

[69] K. V. Suresh and A. N. Rajagopalan, "Robust and computationally efficient super-resolution algorithm," *Journal of the Optical Society of America - A*, vol. 24, no. 4, pp. 984 – 992, 2007.

[70] 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, 2003.

[71] M. Ng, H. Shen, S. Chaudhuri, and A. Yau, "A zoom based super-resolution reconstruction approach using total variation prior," *Optical Engineering*, vol. 46, no. 127003, 2007.

[72] D. Rajan and S. Chaudhuri, "Generation of super-resolution images from blurred observations using an MRF model," *Journal of Mathematical Imaging and Vision*, no. 16, pp. 5–15, 2002.

[73] ——, "Simultaneous estimation of super-resolved intensity and depth maps from low resolution defocussed observations of a scene," in *Proceedings of IEEE Conference on Computer Vision*, 2001, pp. 113 – 118.

[74] M. V. Joshi and S. Chaudhuri, "Simultaneous estimation of super-resolved depth map and intensity field using photometric cue," in *Computer Vision and Image understanding*, vol. 101, 2006, pp. 31 – 44.

[75] S. Sharma and M. V. Joshi, "A practical approach for super-resolution using photometric stereo and graph cuts," in *Proceedings of British Machine Vision Conference*, 2007.

[76] M. V. Joshi, S. Chaudhuri, and R. Panuganti, "A learning based method for image super-resolution from zoomed observations," *IEEE Transactions Systems, Man and Cybernetics, Part B, Special Issue on learning in computer vision and pattern recognition*, vol. 35, no. 3, pp. 527–537, 2005.

[77] R. R. Sahay and A. N. Rajagopalan, "Extension of the shape from focus method for reconstruction of high-resolution images," *Journal of the Optical Society of America - A*, vol. 24, no. 11, pp. 3649 – 3657, 2007.

[78] M. F. Tappen, B. Russell, and W. T. Freeman, "Exploiting the sparse derivative prior for super-resolution and image demosaicing," in *IEEE Workshop on Statistical and Computational Theories of Vision*, 2003.

[79] D. Tschumperle and R. Deriche, "Vector-valued image regularization with PDEs: a common framework for different applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 506 – 517, 2005.

[80] S. Y. Dai, M. Han, W. Wu, Y. Wu, and Y. H. Gong, "Soft edge smoothness prior for alpha channel super resolution," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1 – 8.

[81] R. Fattal, "Image upsampling via imposed edge statistics," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 95:1 – 95:8, 2007.

[82] J. Sun, Z. Xu, and H. Shum, "Image super-resolution using gradient profile prior," in *Proceedings of Compuer Vision and Pattern Recognition*, 2008.

[83] S. Dai, M. Han, Y. Wu, and Y. Gong, "Bilateral back-projection for single image super resolution," in *IEEE International Conference on Multimedia and Expo*, 2007, pp. 1039 – 1042.

[84] 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.

[85] 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, 2002.

[86] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image analogies," in *Computer Graphics (Proceedings of SIGGRAPH 2001)*, 2001, pp. 327–340.

[87] K. I. Kim, M. O. Franz, and B. Scholkopf, "Iterative kernel principal component analysis for image modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1351 – 1366, 2005.

[88] L. C. Pickup, S. J. Roberts, and A. Zissermann, *Advances in Neural Information Processing Systems.* Cambridge, MA: MIT Press, 2004, ch. A sampled texture prior for image super-resolution.

[89] C. V. Jiji and S. Chaudhuri, "Single-frame image super-resolution through contourlet learning," *EURASIP Journal of Appllied Signal Processing*, vol. 2006, no. 2, pp. 1 – 11, 2006.

[90] F. Malgouyres and F. Guichard, "Edge direction preserving image zooming: A mathematical and numerical analysis," *SIAM Journal on Numerical Analysis*, vol. 39, pp. 1 – 37, 2002.

[91] F. Guichard and F. Malgouyres, "Total variation based interpolation," in *Proceedings of the European Signal Processing Conference*, vol. 3, 1998, pp. 1741 – 1744.

[92] H. Aly and E. Dubois, "Image up-sampling using total-variation regularization," *IEEE Transaction on Image Processing*, vol. 14, pp. 1646 – 1659, 2005.

[93] ——, "Specification of the observation model for regularized image up-sampling," *IEEE Transaction on Image Processing*, vol. 14, pp. 567 – 576, 2005.

[94] S. Farsiu, M. Elad, and P. Milanfar, "A practical approach to super-resolution," in *Proceedings of the SPIE: Visual Communications and Image Processing*, vol. 6077, 2006, pp. 24 – 38.

[95] C. Jiji, P. Neethu, and S. Chaudhuri, "Alias-free interpolation," in *Proceedings of ECCV - Part IV. Lecture Notes in Computer Science*, 2006, pp. 255 – 266.

[96] P. Vandewalle, S. Susstrunk, and M. Vetterli, "Double resolution from a set of aliased images," in *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V, Proceedings of SPIE*, vol. 5301, 2004, pp. 374 – 382.

[97] ——, "A frequency domain approach to super-resolution imaging from aliased low resolution images," *Technical Journal, Department of Electrical Engineering and Computer Science, University of California, Berkeley, USA*, pp. 1–21, 2004.

[98] P. Vandewalle, L. Sbaiz, M. Vetterli, and S. Susstrunk, "Superresolution from highly undersampled images," in *Proceedings of International Conference on Image Processing*, vol. 1, 2005, pp. 889 – 892.

[99] P. Vandewalle, S. Susstrunk, and M. Vetterll, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP Journal of Applied Signal Processing*, vol. 2006, pp. 1 – 14, 2006.

[100] S. Rajaram, M. D. Gupta, N. Petrovic, and T. S. Huang, "Learning based nonparametric image super-resolution," *EURASIP Journal of Appllied Signal Processing*, vol. 2006, no. 2, pp. 1 – 11, 2006.

[101] Q. Wang, X. Tang, and H. Shum, "Patch based blind image superresolution," in *Proceedings of IEEE Conference on Computer Vision*, vol. 1, 2005, pp. 709 – 716.

[102] T. A. Stephenson and T. Chen, "Adaptive markov random fields for example-based super-resolution of faces," *EURASIP Journal of Appllied Signal Processing*, vol. 2006, no. 2, pp. 1 – 11, 2006.

[103] L. C. Pickup, S. J. Roberts, and A. Zisserman, "A sampled texture prior for image super-resolution," in *Proceedings of Neural Information Processing Systems*, 2004, pp. 1587 – 1594.

[104] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 627–634, 2001.

[105] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[106] R. Shyamsundar, M. D. Gupta, N. Petrovic, and T. S. Huang, "Learning-based nonparametric image super-resolution," *EURASIP Journal on Applied Signal Processing*, pp. 1–11, 2006.

[107] J. Sun, N. Zheng, H. Tao, and H. Shum, "Image hallucination with primal sketch priors," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, vol. II, 2003, pp. 729–736.

[108] A. Chakrabarti, A. N. Rajagopalan, and R. Chellappa, "Super-resolution of face images using kernel PCA-based prior," *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 888–892, 2007.

[109] K. Ni and T. Q. Nguyen, "Image superresolution using support vector regression," *IEEE Transactions on Image Processing*, vol. 16, no. 6, pp. 1596–1610, 2007.

[110] F. Brandi, R. de Queiroz, and D. Mukherjee, "Super resolution of video using key frames," in *Proceedings IEEE International Symposium on Circuits and Systems*, 2008, pp. 1608–1611.

[111] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2004, pp. 275 – 282.

[112] T. M. Chan and J. Zhang, "An improved super-resolution with manifold learning and histogram matching," in *Proceedings of IAPR International Conference on Biometric*, vol. 1, 2005, pp. 756 – 762.

[113] K. Jia and S. G. Gong, "Generalized face super-resolution," *IEEE Transactions on Image Processing*, vol. 17, no. 6, pp. 873 – 886, 2008.

[114] K. S. Ni and T. Q. Nguyen, "Image superresolution using support vector regression," *IEEE Transactions on Image Processing*, vol. 16, no. 6, pp. 1596 – 1610, 2007.

[115] B. S. Morse and D. Schwartzwald, "Image magnification using level set reconstruction," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, 2001.

[116] Y. W. Tai, W. S. Tong, and C. K. Tang, "Perceptually-inspired and edge-directed color image super-resolution," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, 2006.

[117] V. Rabaud and S. Belongie, "Big little icons," in *1st IEEE Workshop on Computer Vision Applications for the Visually Impaired*, 2005.

[118] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos, "Softcuts: A soft edge smoothness prior for color image super-resolution," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 969 – 981, 2009.

[119] F. Wei and D. T. Yeung, "Image hallucination using neighbor embedding over visual primitive manifolds," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1 – 7.

[120] T. M. Chan, J. P. Zhang, J. Pu, and H. Huang, "Neighbor embedding based super-resolution algorithm through edge detection and feature selection," *Pattern Recognition Letters*, vol. 30, no. 5, pp. 494–502, 2009.

[121] Y. W. Tai, W. S. Tong, and C. K. Tang, "Perceptually-inspired and edge-directed color image super-resolution," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1948–1955.

[122] C. Liu, H. Shum, and C. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local nonparametric model," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2001, pp. 192–198.

[123] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[124] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Total variation super resolution using a variational approach," in *International Conference on Image Processing*, 2008.

[125] A. Chakrabarti, A. N. Rajagopalan, and R. Chellappa, "Super-resolution of face images using kernel pca-based prior," *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 888 – 892, 2007.

[126] C. V. Jiji and S. Chaudhuri, "PCA based generalised interpolation for image super-resolution," in *Proceedings of Indian Conference on Computer Graphics, Vision and Image Processing*, 2004.

[127] M. Ng, E. Lam, and C. S. Tong, Eds., *Multidimensional Systems and Signal Processing: Special issue on Superresolution imaging: Theory, Algorithms and Applications.* Springer, Netherlands, 2007, no. 2-3.

[128] M. Ng, T. Chan, M. G. Kang, and P. Milanfar, Eds., *EURASIP Journal on Advances in Signal Processing: Special Issue on Super-resolution Imaging: Analysis, Algorithms, and Applications.* Hindawi, 2006, vol. 2006.

[129] R. C. Hardie, R. R. Schultz, and K. E. Barner, Eds., *EURASIP Journal on Advances in Signal Processing: Special Issue on Super-resolution Enhancement of Digital Video.* Hindawi, 2006, vol. 2006.

[130] N. K. Bose, R. H. Chan, and M. K. Ng, Eds., *International Journal of Imaging Systems and Technology: Special Issue on High-Resolution Image Reconstruction-I*, 2004, vol. 14.

[131] ——, *International Journal of Imaging Systems and Technology: Special Issue on High-Resolution Image Reconstruction-II*, 2004, vol. 14.

[132] M. G. Kang and S. Chaudhuri, Eds., *IEEE Signal Pfrocessing Magazine: Special Issue on Super-Resolution Image Reconstruction*, 2003, vol. 20, no. 3.

[133] R. Molina and A. Katsaggelos, Eds., *The Computer Journal: Special Issue on Super-resolution in Imaging and Video.* Oxford University Press, 2007.

[134] S. Chaudhuri, Ed., *Super-resolution Imaging.* Kluwer, 2001.

[135] D. Capel, *Image Mosaicing and Super-Resolution.* Springer, 2004.

[136] S. Chaudhuri and M. V. Joshi, *Motion-free Super-resolution.* Springer, 2005.

[137] P. P. Gajjar, M. V. Joshi, A. Banerjee, and S. Mitra, "Decimation estimation and super-resolution using zoomed observations," in *Proceedings of Indian Conference on Computer Graphics, Vision and Image Processing*, no. 1, 2006, pp. 45 – 57.

[138] R. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial-interaction models of images," *IEEE Transactions on Information Theory*, vol. 1, no. IT-29, pp. 60–72, 1983.

[139] S. Z. Li, *Markov Random Field Modelling in Computer Vision.* Springer-Verlag, Tokyo, 1995.

[140] X. Wu, K. U. Barthel, and W. Zhang, "Piecewise 2d autoregression for predictive image coding," in *Proceedings of IEEE International Conference on Image Processing, ICIP*, 1998, pp. 901 –904.

[141] J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern recognition*, vol. 25, no. 2, pp. 173–188, 1992.

[142] M. V. Joshi, L. Bruzzone, and S. Chaudhuri, "Model-based approach to multiresolution fusion in remotely sensed images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2549–2562, 2006.

[143] G. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control.* HoldenDay, 1976.

[144] A. H. J. M. Pellemans, R. W. L. Jorddans, and R. Allewiijin, "Merging multispectral and panchromatic spot images with respect to the radimetric properties of the sensor," *Photogrammetric Engineering and Remote Sensing*, vol. 59, no. 1, p. 8187, 1993.

[145] J. Vrabel, "Multispectral imagery band sharpening study," *Photogrammetric Engineering and Remote Sensing*, vol. 62, no. 9, pp. 1075–1083, 1996.

[146] P. S. Chavez(Jr), S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogrammetric Engineering and Remote Sensing*, vol. 57, no. 3, pp. 295–303, 1991.

[147] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images," *Photogrammetric Engineering and Remote Sensing*, vol. 63, no. 6, pp. 691–699, 1997.

[148] H. Li, B. S.Manjunath, and S. K.Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models and Image Processing*, vol. 27, no. 3, pp. 235–244, 1995.

[149] E. Yu and R. Wang, "Fusion and enhancement of the multispectral image with wavelet transform," *Computer Engineering and Science*, vol. 23, no. 1, pp. 47–50, 2001.

[150] J. Mateos, M. Vega, R. Molina, and A. Katsaggelos, "Pansharpening of multispectral images using a tv-based super-resolution algorithm," in *Seventh International Workshop on Information Optics (WIO-08), Journal of Physics: Conference Series*, vol. 139. Annecy (France), June 2008, pp. 012–022.

[151] C. Mariz, D. Gianelle, L. Bruzzone, and L. Vescovo, "Fusion of multi-spectral spot-5 images and very high resolution texture information extracted from digital orthophotos for automatic classification of complex alpine areas," *International Journal of Remote Sensing*, vol. 30, no. 11, pp. 2859–2873, 2009.

[152] T. Zaveri and M. A. Zaveri, "A novel hybrid pansharpening method using contourlet transform," in *Third International Conference on Pattern Recognition and Machine Intelligence*. IIT Delhi (India), December 2009, pp. 363–368.

[153] R. G. Aykroyd, "Bayesian estimation for homogeneous and inhomogeneous gaussian random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 533 – 539, 1998.

[154] H. M. Shapiro, "Embedded image coding," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[155] S. G. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, pp. 617–643, 1992.

[156] S. G. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, 1992.

[157] A. Jalobeanu, L. Blanc-Féruad, and J. Zerubia, "An adaptive gaussian model for satellite image blurring," *IEEE Transactions on Image Processing*, vol. 4, no. 13, pp. 613–621, 2004.

[158] A. Jalobeanu, L. Blanc-Fraud, and J. Zerubia, "Adaptive parameter estimation for satellite image deconvolution," in *Rep. 3956*, 2000.

[159] N. Kaulgud and U. B. Desia, *Super-resolution imaging.* Kluwer, 2001, ch. Image zooming: use of wavelets, pp. 21–44.

[160] K. Rudraraju and M. V. Joshi, "Nonhomogenous AR model based prior for multiresolution fusion," in *Proceedings of IEEE international Geoscience and Remote Sensing Symposium*, 2009.

[161] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. C-23, pp. 90 – 93, 1974.

[162] N. Ahmed and K. R. Rao, *Orthogonal transforms for digital signal processing.* Springer Verlag, 1975.

[163] K. R. Rao and P. Yip, *Discrete cosine transform: algorithms, advantages, applications.* Academic Press, 1990.

[164] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[165] J. Kennedy and R. Eberhart, *Swarm intelligence.* Morgan Kaufmann, 2001.

[166] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proceedinds of IEEE International Conference On Neural Networks*, 1995, pp. 1942 –1948.

[167] ——, "A new optimizer using particle swarm theory," in *Sixth International Symposium on Micro Machine and Human Science*, 1995, pp. 39–43.

[168] J. Kennedy and W. M. Spears, "Matching algorithms to problems: An experimental test of the particle swarm and some genetic algorithms on the multimodal problem generator," in *Proceedings of the IEEE International Conference on Evolutionary Computation*, 1998, pp. 39–43.

[169] M. Omran, P. A. Engelbrecht, and A. Salman, "Particle swarm optimization for image clustering," *International Journal of Pattern Recognition and Artifficial Intelligence*, vol. 19, no. 3, pp. 297–321, 2005.

[170] M. Vrahatis and K. Parsopoulos, *Natural Computing*. Kluwer, 2002.

[171] Y. Shi and R. C. Eberhart, "A modified particle swarm optimizer," in *Proceedings of IEEE International Conference On Evolutionary Computation*, 1998, pp. 69–73.

[172] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it?" *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 98–117, 2009.

[173] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

List of Publications

Journal

- Prakash P. Gajjar and Manjunath V. Joshi, "New Learning based Super-Resolution: Use of DWT and IGMRF Prior", IEEE Transactions on Image Processing, vol. 19, no. 5, pp. 1201-1213, 2010.

Conference

- Prakash P. Gajjar, Manjunath V. Joshi, Asim Banerjee, and Suman Mitra, "Decimation Estimation and Super-Resolution Using Zoomed Observations" in Proc. Indian Conf. on Computer Vision, Graphics and Image processing, 2006, pp. 45-57.

- Prakash P. Gajjar, Manjunath V. Joshi, Asim Banerjee, and Suman Mitra, "Decimation Estimation and Linear Model-Based Super-Resolution Using Zoomed Observations" in Proc. Iberian Conf. on Pattern Recognition and Image Analysis, 2007, pp. 314-321.

- P. P. Gajjar and M. V. Joshi, "A Fast Approach to the Learning-based Super-Resolution Using Autoregressive Model Prior and Wavelet Prior" in Proc. National Conf. on Communications, IIT, Bombay, 2008, pp. 229-233.

- P. P. Gajjar and M. V. Joshi, "Single frame super-resolution: A new learning based approach and use of IGMRF prior" in Proc. Indian Conf. on Computer Vision, Graphics and Image Processing, 2008, pp. 636-643.

- M. V. Joshi, P. P. Gajjar, S. Ravishankar, and K. V. V. Murthy, "Multiresolution Fusion in Remotely Sensed Images: Use of Gibbs Prior and PSO Optimization" IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Honolulu, July, 2010.

- Prakash Gajjar and Manjunath Joshi, "Zoom Based Super-resolution: A Fast Approach Using Particle Swarm Optimizatio" International Conference on Image and Signal Processing (ICISP), June, 2010, pp. 63-70.

Publications Under Review

- Kishor Upla, M. V. Joshi, P. P. Gajjar, Asim Banerjee, and Vineet Pratap Singh, "A New Learning Based Super-resolution Using Contourlet Transform" submitted to Indian Conference on Vision, Graphics and Image Processing (ICVGIP), Chennai, 2010.