Image Super-resolution and Inpainting from a Single Observation

M.Tech. Dissertation

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology

by

Priyam Chatterjee

Roll No: 03307911

under the guidance of

Prof. Subhasis Chaudhuri

Department of Electrical Engineering

Indian Institute of Technology, Bombay

Mumbai

July, 2006
Acknowledgment

From the very nascent stage of this project till its maturity into this dissertation, there has been a constant source of guidance and inspiration from my supervisor, Prof. Subhasis Chaudhuri. Words are not enough to express my gratitude for his help, encouragement and the freedom to explore the topic on my own terms. I just cannot thank him enough.

Vision and Image Processing (VIP) lab-mates have perhaps maintained the most perfect balance between work and fun. It has been an enriching experience, academically and personally, being a part of this group. I take this opportunity to thank C.V. Jiji for sharing his thoughts and inputs in the domain of super-resolution and Krishnan Iyer for discussions, both academic and otherwise. Hitesh Shah has been an amazing lab-mate and room-mate. I am constantly gifted by his little inputs that make work easier. And last, but definitely not the least, I express my heartfelt gratitude to Vinay P. Namboodiri. He is the biggest student encyclopedia that I know in the field of Computer Vision and Graphics. It has been an amazing experience working and being with him. At times, he has been more than a friend, philosopher and guide.

I thank my friends in the hostel and sporting activity center for making sure that my stay in IIT was anything but dull. I would also like to express my gratitude to my relatives in Mumbai, specially my much younger cousins, who often went out of their way to make my stay in Mumbai so very memorable.

Finally, I dedicate this work to my parents and my brother whose constant support and belief in me remain inexplicable at times.

Priyam Chatterjee
Abstract

In this work, we deal with the concepts of image super-resolution and inpainting from only a single input image. A study of the Papoulis-Gerchberg method is presented and its applications to these domains of computer vision are demonstrated. We address the issue of its inability to deal with blurred low resolution images and show results that prove the effectiveness of our method when the low resolution image formation model is known. Another learning based approach is presented where we try to do away with such an assumption by trying to learn a filter that bridges the gap between the reconstructed and the desired high resolution image. The PG method can also be extended to the domain of image inpainting where filling-in is to be done for thin scratches. We propose alterations which lead to better inpainting in specific cases. A faster method for inpainting within this framework is also proposed by making use of different orthogonal basis functions.

We then explore the use of regularization based techniques for image super-resolution. The total variation is used as a regularizer to take advantage of its edge preserving nature. However, a direct application results in loss of textural details in the super-resolved image. We propose the use of additional data fidelity constraints to perform texture preserving super-resolution when the degradation model is available. In order to make the process robust in the presence of noise, this process is further extended by using spatially varying weighing terms in the objective function. We demonstrate encouraging results obtained using our methods that strengthen our theory and pave a way for future research.
Contents

1 Introduction and Overview .................................................. 1
   1.1 Introduction ......................................................... 1
   1.2 Contribution ....................................................... 2
   1.3 Organization ....................................................... 4

2 Super-Resolution and Inpainting .......................................... 5
   2.1 Super-resolution .................................................. 5
      2.1.1 What is Super-resolution ? ................................. 5
      2.1.2 Related work ................................................. 6
      2.1.3 Observation Model ......................................... 8
   2.2 Inpainting .......................................................... 11
      2.2.1 What is Image Inpainting ? ............................... 11
      2.2.2 Related Work ............................................... 11

3 Image Restoration using Papoulis-Gerchberg Method ............... 13
   3.1 The Papoulis-Gerchberg Method ................................ 13
   3.2 Application of PG Method to Super-resolution ................. 16
   3.3 Modifications to the PG Method for SR ....................... 18
      3.3.1 PG Method with deblurring ............................. 18
      3.3.2 Learning based approach ................................ 20
   3.4 Inpainting using PG method .................................... 22
   3.5 Modifications to the PG method for Inpainting ............... 23
3.5.1 Filtering with a smooth cut-off ........................................ 23
3.5.2 Directional filtering ...................................................... 23
3.5.3 Extension to different orthogonal bases ............................. 24
3.6 Results .............................................................................. 25
  3.6.1 Super-resolution ............................................................ 25
  3.6.2 Inpainting ................................................................. 28

4 Super-resolution using TV Approach ..................................... 33
  4.1 Total Variation ............................................................... 33
    4.1.1 Definition .............................................................. 33
    4.1.2 Optimization with TV .............................................. 34
  4.2 TV for Image Restoration .................................................. 35
  4.3 TV based Super-resolution ............................................... 39
  4.4 Results ............................................................................ 41

5 Conclusion and Future Work .............................................. 48
List of Figures

2.1 Observation Model relating LR images to HR images ................. 9
2.2 Generic scheme in super resolution .................................. 9
2.3 Registration-interpolation based super resolution ................... 10

3.1 Illustration of iterative extrapolation of a signal using the PG method . 15
3.2 Iterative method of SR using PG method ............................. 17
3.3 SR using PG method in presence of blur and noise ................... 19
3.4 SR of Lena image using PG method with deblurring ................. 26
3.5 SR of fruits image using PG method with deblurring ............... 27
3.6 SR using learning based PG method ................................. 29
3.7 Inpainting using the PG method ..................................... 30
3.8 Inpainting using modified PG approaches ............................ 32

4.1 SR using modified TV approach with scalar weights ............... 43
4.2 SR using modified TV approach with scalar weights ............... 44
4.3 SR using modified TV approach with scalar weights for noisy case 45
4.4 SR using modified TV approach with spatially varying weights .... 47
Chapter 1

Introduction and Overview

1.1 Introduction

Signal interpolation and extrapolation methods are subjects of immense research. In digital form, a signal is stored and represented by a finite number of samples. Interpolation of such a signal is the method of finding out samples between those that are known. This is a highly ill-posed problem since there are infinitely many signals that may have the same samples. In most cases, signals are assumed to be bandlimited. Interpolation then becomes possible if the sampling process satisfies the Nyquist criteria. However, this may not always be true. In such cases, interpolation of signals becomes a non-trivial problem.

Signal interpolation is a requirement for zooming into signals, where a denser representation of the signal may be required. Interpolation techniques are used for image super-resolution where the spatial density of an image is required to be increased. In essence, given some sampled version of an image, we need to find out more samples of the image so that we get a more detailed spatial description. Another application of signal interpolation is in image inpainting. Image inpainting uses non-uniform signal interpolation to find sample values over a region where no signal samples are provided. Like super-resolution, the inpainting problem too is numerically ill-posed and the complexity increases as we attempt to recover sample values over a larger area.
Signal extrapolation, on the other hand, is the method of finding out a signal when only a subset of the signal is available. It is thus extending a signal beyond the range in which the signal is available to us. Signal extrapolation theory has been applied to perform image super-resolution by extrapolating the spectrum beyond the diffraction limit of a finite object [1]. In [2] Slepian uses the prolate spheroidal function-based extrapolation technique making an assumption on the bandlimitedness of the signal. An approach of analytic continuation of a signal from only a known segment was proposed by Harris in [3]. He establishes that, given a finite extent of an object and a continuous but finite portion of the spectrum of the object, the entire spectrum can be generated uniquely using the principle of analytic continuation. This leads to an exact and complete reconstruction of the object spectrum if the measurements are noise free. An iterative method to signal extrapolation was demonstrated independently by Papoulis [4] and Gerchberg [5]. However, these global methods becomes highly unreliable even if a small amount of noise or measurement error is present in the given portion of the spectrum.

For such ill-posed problems certain a priori information or assumptions are often required to converge to one of the multiple possible solutions. Regularization theory is used in such cases to obtain a suitable solution which conforms to the assumptions made on the desired signal. The choice of the regularizer is made such that the space of solutions have properties that are needed to be present in the solution. Regularization theory is thus a trade-off between fitting the available data and minimizing a choice of norm. The use of regularization is widespread in the field of inverse problems like image restoration. Since the problems we consider in this dissertation are highly ill-posed we make use of regularization to perform super-resolution.

1.2 Contribution of this dissertation

Super-resolution and image inpainting are popular areas of research where a multitude of literature is available addressing the related issues. Many researchers perform super-resolution where multiple low resolution observations are available. Recently techniques
have been proposed to perform super-resolution from a single observation using a database of related or unrelated images where multiple low and corresponding high quality images are present. We restrict ourselves chiefly to the case where only a single image is provided and attempt to super-resolve such an image without the presence of any image database. This kind of image restoration has received less attention in comparison to the other cases. We point out here that we differentiate the super-resolution method from data fitting interpolation methods where the aim of reconstruction is a visually pleasing result. In our dissertation we initially work with the Papoulis-Gerchberg framework of image super-resolution and demonstrate its limitations. We propose modifications to the method which enable us to get a sharper result through deblurring. This allows the method to be used in cases where the image is degraded by blurring in decimation process. We assume that the low resolution image formation model is known to us. Another method is provided where we do not assume anything about the image formation model. We apply a learning based method to obtain a higher resolution image. We then show that the theory developed by Papoulis and Gerchberg can be extended very easily to perform simple image inpainting for removal of thin scratches. We extend this theory to be able to achieve similar results faster and better results in the PSNR sense.

Since the PG method is sensitive to the presence of noise, we then experiment with a regularization based method for super-resolution. We propose a new cost function for the super-resolution case which iteratively checks for conformity with the available low resolution image under the known downsampling process. Using additional constraints on the data fidelity in the image spectrum, we introduce a texture preserving total variational approach to super-resolution. We propose a modified total variation based approach where we use spatially varying multipliers to develop a robust iterative process that produces a sharp output while handling the presence of noise.
1.3 Organization of the dissertation

The dissertation is organized as follows: We provide a brief overview of the issues of image super-resolution and inpainting in Chapter 2. Here we also provide a review of the work that has been done in the related areas. In Chapter 3 we present the Papoulis-Gerchberg method and its application to image super-resolution and inpainting. We then introduce the theory of total variation and make use of this regularization scheme for super-resolution in Chapter 4. We finally conclude with a word on the scope for further research in Chapter 5.
Chapter 2

Super-Resolution and Inpainting

2.1 Super-resolution

2.1.1 What is Super-resolution?

In its simplest form, the word image resolution is defined as the smallest discernible or measurable detail in a visual presentation [6]. It refers to the spacing of pixels in an image. The higher the spatial resolution the greater is the number of pixels in the image. In most imaging applications, images with a high resolution is desired. This can be obtained by more sophisticated image acquisition hardware. Since most digital imaging devices use charge-coupled devices (CCD) and CMOS image sensors, a higher resolution image can be obtained by reducing the pixel size by sensor manufacturing techniques [7]. As the pixel size decreases, the amount of light available also decreases resulting in shot noise which degrades the image. Hence, the pixel size can be limited to a certain extent. Current sensor technology has, at present, almost reached this level. Increasing chip size, so as to accommodate larger number of pixels is also not viable due to an increase in capacitance with chip size and hence slower charge transfer rates.

Hence we need some sort of post processing that will enable us to produce images of higher resolution than what the physical imaging system can produce. Signal processing techniques can be used to obtain a high resolution (HR) image from one or more observed
low-resolution (LR) images. Super-resolution (SR) is thus the problem of generating a high-resolution image from one or more low-resolution images [8]. Reconstruction methods increase pixel resolution beyond that of the physical imaging system by estimating values on a finer grid. Restoration methods increase fidelity by correcting for acquisition artifacts such as blurring, aliasing and noise. Super-resolution reconstruction and restoration can be combined to produce images with greater resolution and higher fidelity than the lower resolution image(s) [9]. The super resolution process includes three main tasks: an alias-free upsampling of the image, thereby increasing the maximum spatial frequency, and removing degradations that arise during the image capture, viz., blur and noise. In effect, the super-resolution process tries to generate the missing high frequency components and minimize aliasing, blur and noise. This has been a field of intense research and many different methods have been proposed to obtain high resolution images.

Super-resolution image reconstruction techniques have proved useful in cases where greater clarity in images are required. Some of the applications are in medical imaging such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), surveillance systems with CCTV, satellite imaging application such as remote sensing and LANDSAT, and conversion of NTSC video signal to High Definition TV (HDTV) signals.

2.1.2 Related work

The methods used to super-resolve images differ widely in their choice of domain (spatial, spectral or even both), in the choice of the observation model used, in the method of capturing low resolution images, in the choice of number of low resolution images required and also in the choice of cues that are used to extract information from images. Resolution enhancement using a multiple-aperture camera system has been carried out by Komatsu et al.[10]. In this the authors use different aperture systems to acquire multiple LR images. An iterative algorithm is used for alternate registration and reconstruction. Capel et al. [11] use different methods of image registration and then go on to estimate the HR image using maximum likelihood estimation (MLE), Maximum a-priori (MAP) estimation with
priors like Gaussian Markov random field (MRF) and Huber MRFs. Rajan et al. [12, 13] use blur as a cue for super-resolution. They model the HR image and the blur processes as separate independent MRFs. The HR space variant blur and SR image are estimated using a MAP estimator. They also show how MRF models are well suited for modeling the structural information in the form of surface normal. The shape from shading problem has also been used for restoration of an HR image in [12]. In [14, 15], Joshi et al. make use of zoom as a cue to perform super-resolution. They consider linear dependency of pixels in a neighborhood of the HR image and model this using a simultaneous auto-regressive (SAR) model. This SAR model is then used as a prior to perform super-resolution. A combination of bilateral filter and the TV norm has been effectively used by Farsiu et al. in [16, 17] to perform super-resolution under a regularization framework. The use of TV is also done by Chan et al. in [18] to perform simultaneous deblurring and interpolation from multiple LR images.

Researchers have also attempted to solve the super resolution problem by using learning based techniques. These methods try to recognize local features of a low resolution image and then try to retrieve the most likely high frequency information from the given training samples. They try to re-create the HR image from just a single SR image by making use of a image database that is used as a training set. Freeman et al. [19] describe image interpolation algorithms which use a database of training images to create plausible high frequency details in zoomed images. In [20], Baker and Kanade develop a super-resolution algorithm by modifying the prior term in the cost function to include the results of a set of recognition decisions, and call it as recognition based SR or “hallucination”. Their prior enforces the condition that the gradient of the HR image should be equal to the gradient of the best matching training image. In [21], Candocia et al. address the ill-posedness of the SR problem by assuming that the correlated neighbors remain similar across scales, and this prior information is learnt locally from the available image samples across scales. When a new image is presented, a kernel that best reconstructs each local region is selected automatically and the super-resolved image is reconstructed by simple convolution operation. In [22], Jiji et al. propose a single frame, learning based SR
restoration technique by using the wavelet domain to define a constraint on the solution. Wavelet coefficients at finer scales of the unknown HR image are learnt from a set of HR images that are used as a training database. This learnt image is further used for regularization while super-resolving the image. An appropriate smoothness prior is used with discontinuity preservation in addition to the wavelet based constraint to estimates the HR image. This method was further improved by the use of contourlets in [23] to capture edges better and hence results in better reconstruction of images.

A third kind of approach is where researchers have tried to perform image super-resolution from a single low resolution observation but without the aid of any image database. Jiji et al. in [24] separates a bandlimited LR image into its unaliased and aliased part and uses this information to perform image interpolation under a regularization framework. The method is able to handle aliasing due to improper downsampling rate but has the disadvantage of being memory intensive and the results depend on the accurate separation of aliased and unaliased parts. In [25] the TV norm has been used to perform image regularization for minimization of gradient energy of the super-resolved image. Aly et al. in [26] modify the data fidelity term of the regularization term using the knowledge of the downsampling process. Surveys of various SR techniques have been presented in [6, 7, 27, 28, 29].

2.1.3 Observation Model

For any method of super-resolution process it is important to first formulate an observation model which will determine how the observed LR image(s) are related to the HR image. In literature many different observation models are used. In this section we will deal with one of the generic observation model used for still images.

As shown in Figure 2.1, the observed LR images can be mathematically written as

\[ y_k = DB_k M_k x + n_k \]  

(2.1)

where \( y_k \) is the \( k^{th} \) observed image, ordered lexicographically, \( M_k \) is a warp matrix, \( B_k \) is the blur matrix, \( D \) is the subsampling matrix and \( n_k \) is the lexicographically ordered
Figure 2.1: Observation Model relating LR images to HR images

noise vector. As illustrated in the Figure 2.1, the continuous scene is first sampled and discretized. This results in the HR image that we want to obtain. However, the observed image is a degraded version of the HR image as it undergoes warping (rotation, translation) due to camera movement or scene change. Further reduction in quality occurs due to optical blur which may be caused by the optical system (out of focus, diffraction limit, etc.), relative motion between the imaging system and the original scene, and the point spread function (PSF) of the LR sensor [7].

From this observation it becomes clear that any SR technique will essentially consist of three parts - Registration (the method of calibrating one low-resolution image with respect to the others), Reconstruction (interpolation on a higher resolution grid, with information from low resolution images) and Restoration (removal of artifacts due to blur, aliasing etc.) as shown in Figure 2.2. The registration process is a very important step

Figure 2.2: Generic scheme in super resolution
for the success of any SR method with takes as input multiple LR images. Registration between images are done to estimate motion (rotation and translation), zooming and other parameter estimation which will relate one LR image with another. A comprehensive study of different image registration techniques have been done by Zitova et al.[30] and Brown [29]. The reconstruction process essentially consists of projecting the pixels from the LR images on the HR grid. The first LR image is first projected in the HR grid, taking into account the resolution of the LR and HR grid. The other registered images are then suitably placed in the HR grid so as to conserve the inter-pixel shifts. The grid is then filled up with a suitable interpolation technique as shown in Figure 2.3.

Figure 2.3: Registration-interpolation based super resolution
2.2 Inpainting

2.2.1 What is Image Inpainting?

Image inpainting is the art of touching up images which has been around for a long time among museum restoration artists. It is an image interpolation technique where a part of the image is unknown and has to be guessed from the surrounding known areas. Digital Inpainting is thus the process of predicting the unknown values of pixels so as to remove artifacts like scratches or unwanted objects from an image. In the latter case it is sometimes known as object disocclusion. Unlike in the case of super-resolution the reference grid is at the same resolution as the observed image.

Image inpainting finds use in image and film restoration, texture synthesis, disocclusion and in the entertainment industry in special effects. Inpainting methods are also used, in a variety of ways, to perform video inpainting.

2.2.2 Related Work

A lot of work has been done to address this problem since Bertalmio et al. introduced this term in [31]. Inpainting algorithms based on level lines were proposed by Masnou and Morel [32, 33]. Other authors have taken a variational approach like Ballester et al. in [34] and Chen and Shen in [35]. PDE based approaches have also been used by authors in [31, 35]. Oliveira et al. [36] propose a fast digital inpainting method using an isotropic diffusion kernel for convolution. They make use of anisotropic kernels to preserve edges. However, their method requires manual intervention to identify pixels where the scratches cross edges in the image. Chan et al. in [37] make use of the total variation norm to perform image inpainting as well as deblurring. The method makes use of blind deconvolution for deblurring and inpainting for missing pixels using an alternate minimization (AM) algorithm. In [38, 39] the authors make use of the Ginzburg-Landau equation to perform image inpainting.

A considerable amount of literature is also available where researchers have used projec-
tions onto convex sets (POCS) to perform image restoration of this kind. In [40] Hirani and Totsuka present a dual domain iterative method to restore parts of an image. The method requires the users to provide a noise mask and a corresponding mask (repair subimage) and a sample subimage. Their method makes use of the spectrum of the sample subimage to do scratch removal in the repair subimage. Patwardhan and Sapiro [41] demonstrated a similar method in which they make use of the wavelet transform for inpainting. They do away with the requirement of human provided sample subimage, instead assume that the corrupted portion of the image is as smooth as its neighborhood. They learn the wavelet coefficients from the neighborhood and impose spectral restriction on the noisy sections to restore images. In [42] Chan et al. have proposed a framelet-based method. They try to determine the missing framelet coefficients of the unknown pixels from those of the neighborhood. Analysis of some methods for sample recovery in band-limited images and application of adaptive weights conjugate gradient Toeplitz method (ACT) to perform fast image inpainting is presented in [43, 44]. These methods involve projecting the image or a subimage from one domain to another.
Chapter 3

Image Restoration using
Papoulis-Gerchberg Method

In this chapter we introduce the Papoulis-Gerchberg method for signal extrapolation and show how it can be used in a modified form for digital image super-resolution and inpainting. We then identify certain demerits of the method and propose modifications that deliver case specific improvements. Results using the modified approach are then shown in Section 3.6.

3.1 The Papoulis-Gerchberg Method

This method of super-resolution is based on the work done independently by Papoulis [4] and Gerchberg [5]. While Gerchberg proposed a method to perform signal reconstruction given the diffraction limit of the signal and a part of the spectrum, the motivation for Papoulis’ work was extrapolation of a bandlimited signal from only a part of the original signal, i.e., determination of the transform

\[ F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t}dt \]  

(3.1)
of a signal \( f(t) \) given a finite segment

\[ g(t) = f(t)p_T(t), \quad \text{where} \quad p_T(t) = \begin{cases} 1, & |t| \leq T \\ 0, & |t| > T \end{cases} \]  

(3.2)
This is shown in Figure 3.1(c) which is a truncated version of Figure 3.1(a). The signal extrapolation is carried out by the method of alternate projections [45], iterating alternately between time and spectral domains. The signal $g(t)$ is first low-pass filtered by truncating its Fourier transform outside the interval $[-\sigma, \sigma]$, assuming $\sigma$ to be the signal bandwidth of $f(t)$. This is illustrated in the formation of $F_1(\omega)$ in Figure 3.1(f) from $G(\omega) = G_0(\omega)$ shown in Figure 3.1(d). In the $n^{th}$ iteration this can be expressed as

$$F_n(\omega) = G_{n-1}(\omega) p_\sigma(\omega), \quad p_\sigma(\omega) = \begin{cases} 1, & |\omega| \leq \sigma \\ 0, & |\omega| > \sigma \end{cases} \quad (3.3)$$

The inverse function of $F_n(\omega)$ is then computed as $f_n(t)$ (Figure 3.1(e)). This results in a reduction of the error signal $|f(t) - f_n(t)|^2$ outside the known segment of the signal. This follows from Parseval’s theorem. However, the signal $f_n(t)$ does not match the observed signal $g(t)$ in the region $[-T, T]$. This part of the signal is then restored to the original known segment forming the function $g_n(t)$ for the next iteration as shown in Figure 3.1(g). The resultant change in the spectral domain due to introduction of higher frequency components can be seen in Figure 3.1(h).

$$g_n(t) = f_n(t) + [f(t) - f_n(t)] p_T(t) = \begin{cases} g(t), & |t| \leq T \\ f_n(t), & |t| > T \end{cases} \quad (3.4)$$

This process is then iterated with the new $g_n(t)$. In each iteration the mean square error of the extrapolated signal is reduced [4]. Hence with successive iterations the generated extrapolated signal approaches the desired signal $f(t)$. Convergence of the method is guaranteed and is shown in [4]. However, the process requires an infinite number of iterations. If we stop after $r$ iterations, the reconstructed signal is given by $f_r(t)$ instead of $f(t)$. Also, in practice, the measured data $g(t) = g_0(t)$ will contain error. The propagation of this measurement error can be controlled by early termination of the iterative process [5, 46]. The process also assumes the signal $f(t)$ to be bandlimited, but it is found that the method works reasonably well for signals with sufficiently low energy in their higher frequency components.
Figure 3.1: Illustration of iterative extrapolation of $g(t)$ using the Papoulis-Gerchberg method. (a) The signal that we want to recover, (b) partial spectrum of the signal, (c) the time-limited available signal, (d) spectrum of (c), (e) low-pass filtered spectrum of (d), (f) time domain signal from (e), (g) portion of (c) reinstated in (f), (h) spectrum of (g).
3.2 Application of PG Method to Super-resolution

The Papoulis-Gerchberg algorithm has been used in a modified form by Vanderwalle et al. in [47] to super-resolve images when multiple low-resolution registered images are available. We restrict ourselves to the case when only a single LR image is available. The initial image is thus a higher dimensional grid where the values of some pixels are known and some are unknown. The unknown pixel values are initially set to zero. In the next step, the image is taken to its frequency domain and the higher frequencies (frequencies larger than the maximal allowed frequency $\sigma$) are taken to zero. This is effectively low-pass filtering the image. After this step, the unknown pixels will have some values. But, the known pixel values have changed as a result of the filtering. In the next step, these values are set back to their original values, creating high frequency components. The whole process is repeated. Figure 3.2 shows the reconstructed image after one such iteration of the method. Figure 3.3 compares the output of the method after 50 iterations with that of the standard bicubic interpolation.

Though the method is quite fast it has some drawbacks. Due to the steep cut-off in the frequency domain, the resultant HR image has ringing artifacts near the edges. Also, it relies heavily on the fact that the measured (known) pixel values are the values that are to be obtained in the reconstructed high resolution image. In other words, it assumes that the low resolution images are downsampled versions of the expected high resolution image without any low-pass (averaging) filtering. Moreover, they should also be totally free from noise perturbations. Hence, it is not able to compensate effectively for blur and noisy measurement of data. Figure 3.3 demonstrates these drawbacks. In Figure 3.3(a) we have a fairly good quality LR image, and Figure 3.3(b & c) show corresponding results of interpolation using bicubic and PG methods, respectively. One may notice that the PG method offers a better result. This can be seen as sharper features near the nose, eyes and freckles. In Figure 3.3(d) we show a poor quality LR image. The input image is quite blurred. Hence the corresponding HR reconstructions in Figure 3.3 (e & f) are poor. The PG method offers no help as it assumes correct pixel values on the LR grid. Similarly

16
Figure 3.2: (a) The starting of iteration 1 with the HR grid partly filled with the given LR data, (b) image after low-pass filtering with 40\% of the higher frequencies set to zero, (c) image after known pixel values are reset to original values. Spectra in log scale (d) of (a), (e) of (b), and (f) of (c). Notice that the spurious peaks at the higher frequency regions in (d) reduce in (e) due to the spatial filling-in process. This increases again in (f) and the process continues.
Figure 3.3(g) shows a noisy LR input image and the corresponding HR reconstructions are shown in Figure 3.3(h & i). For the same reason, the PG method does not offer a quality reconstruction.

3.3 Modifications to the PG Method for SR

3.3.1 PG Method with deblurring

As mentioned in Section 3.2, the PG method cannot deal with images that are blurred due to averaging during the downsampling process. We now propose an extension which allows the method to be applied to super-resolution of images for which the downsampling process is known. In our experimentation we assume that the low resolution image is formed by averaging and subsampling the high resolution image that we aim to recover. This means that for $2 \times$ zooming of an image we assume that every pixel in the low resolution input image is the average of a corresponding $2 \times 2$ pixel block of the high resolution image. The classical PG method will not deliver a good result in such a case as shown in Figure 3.3. This is because we enforce the available pixel values in the LR image to be pixel values in HR image. This results in an overall blurred HR image. We overcome this drawback by introducing a different constraint enforcing part. We want to enforce that the super-resolved image obtained after every iteration of the method conforms to the input LR image. To do this we introduce a back projection part within the PG method. Iterative back projection has been used in image super-resolution by Irani and Peleg [48]. But as opposed to their case, we deal with the case where only a single low resolution image is available.

In our method, after every iteration of the PG method we calculate the error between the LR image and the simulated LR image formed by applying the known decimation model to the obtained SR image. Mathematically, the error $\epsilon_i$ can be expressed as

$$\epsilon_i = y - Dz_i$$

(3.5)

where $y$ is the input LR image, $D$ is the decimation matrix that we assume to be known
Figure 3.3: (a) A LR face image, (b) 2× zoom of (a) using bicubic interpolation, (c) SR of (a) using PG method, (d) blurred LR image, (e) 2× zoom of (d) using bicubic interpolation, (f) SR of (d) using PG method, (g) noisy LR image, (h) 2× zoom of (g) using bicubic interpolation, (i) SR of (g) using PG method.
to us and $z_i$ is the obtained SR image after the $i^{th}$ iteration of the method. We then compensate for this error in the obtained SR image by adding back the error for each pixel of the simulated LR image to the corresponding block of pixels in the obtained SR image. The compensation is thus done such that the SR image best approximates the desired HR image, under the assumed downsampling process. We then proceed with the next iteration. It may be noted here that due to this error compensation in blocks a certain blockiness will be introduced to the SR image obtained at this point. However, this is then taken care of in the low-pass filtering part of the next iteration. The algorithm terminates when the error is small enough. This method may thus seem as a process where we attempt to reverse the process of averaging. Since this is an ill-posed problem, we can only hope to achieve one of the possible solutions. The combination of low-pass filtering and back projection of the error $\epsilon_i$ in every iteration ensures that the resultant SR image is smooth and approximates one of the possible SR images, under the specified downsampling process. This method can easily be modified to handle any general downsampling process. The proposed process thus generalizes the PG method by making it applicable to any class of images where the downsampling process is known.

We have found that the method performs better than the standard PG method and the results obtained through bicubic interpolation. The results obtained through different methods are compared in Section 3.6.1.

### 3.3.2 Learning based approach

In the previous section we have discussed a method of improving the PG method assuming that the decimation process in the forming of the LR image is known to us. However, there may be cases where one does not have such knowledge about the downsampling process. We propose the use of a learning based method using a pseudo-inverse filter to achieve better spectral reconstruction of the HR image. In this method we make use of a training LR-HR image pair to learn the spectral difference between the HR image and a super-resolved image obtained by the PG method. We learn the difference in the form of
a pseudo-inverse filter. We then apply this learnt information to perform super-resolution of other images. We assume the pseudo-inverse filter to be inverse Gaussian and thereby model it accordingly. This method can be easily extended to perform this learning from multiple training data pairs. When multiple training images are used, it is important that images with similar spectral distribution be used. The reason for this will be clear once we explain the technique that we use.

In this method we assume that we are given a training HR-LR pair of images. The learning phase of the method consists of first generating the upsampled super-resolved image from the training LR image using the PG method. We now try to learn the difference between the obtained SR image and the original HR image. This difference is learnt as a pseudo-inverse filter which will have the transfer function $H(u, v)$ calculated as

$$H(u, v) = \begin{cases} 
F_{hr}(u,v)/F_{sr}(u,v), & F_{sr}(u,v) \neq 0 \text{ and } u,v \leq \sigma \\
1, & \text{otherwise}
\end{cases}$$

(3.6)

where $F_{hr}$ and $F_{sr}$ denote the DFT of the HR and SR image respectively. This pseudo-inverse filter is assumed to be inverse Gaussian. We calculate its variance, $\nu$ as

$$\nu = \frac{1}{A} \int_{u} \int_{v} \frac{-2}{u^2 + v^2} \log H(u, v) dudv$$

(3.7)

where $A$ is the area of the region in the $(u, v)$ space not containing the singularities (if any). Using $\nu$ as the variance and mean as zero, we next model a new transfer function as inverse Gaussian filter. To this filter we use a $\beta$ ($> \nu$) factor to suppress higher frequencies. The $\beta$ factor comes into play only at frequencies higher than a pre-determined frequency $(u_0, v_0)$. As a result frequencies greater than this particular frequency get suppressed. So the new transfer function becomes

$$H_{\text{new}}(u, v) = \begin{cases} 
e^{-\beta^2(u^2+v^2)} e^{-\beta^2(u-u_0)^2+(v-v_0)^2)}, & u \leq u_0 \text{ and } v \leq v_0 \\
e^{2(u^2+v^2)}, & \text{otherwise}
\end{cases}$$

(3.8)

The filter, $H_{\text{new}}(u, v)$ learns the amount of boosting or suppression to be done on different frequency components to bridge the difference between the desired HR image and the SR image obtained by the PG method. The next phase is to apply the learning coefficient to super-resolve a test LR image. One way of using the learnt function would
be to perform super-resolution using the classical PG method and then apply \( H_{new}(u, v) \) to the resultant image to obtain a better SR image. The final DFT of the SR image \( F_{res} \) would hence be

\[
F_{res}(u, v) = F_{sr}(u, v)H_{new}(u, v)
\]  

(3.9)

We propose to use the learnt function \( H_{new}(u, v) \) within the PG iterative method itself. In the iterative steps, we introduce the effect of \( H_{new}(u, v) \) after truncation of frequencies higher than the assumed bandwidth of the signal. So equation (3.3), for \( 2 - D \) case, now becomes

\[
F_n(u, v) = G_{n-1}(u, v)p_\sigma(u, v)H_{new}(u, v)
\]  

(3.10)

where \( p_\sigma(u, v) \) is a lowpass filter of bandwidth \( \sigma \) having a sufficiently smooth transition band. This has the effect of trying to restore some of the frequencies but also amplifies noise at each iteration. But this is followed by smoothening, due to low-pass filtering, in the next iteration. To suppress the amplification of noise, we introduce the effect of \( \beta \) only in the higher frequency region. There is thus a kind of optimality for which we get a better reconstruction without amplification of noise. The result of super-resolving using the learning co-efficients is shown in Figure 3.6.

### 3.4 Inpainting using PG method

As opposed to super-resolution discussed in the previous section, image inpainting requires filling-in of lesser number of pixels but in many cases the corrupt pixels are contiguous in nature, rendering simple smoothing insufficient. We have applied the Papoulis-Gerchberg algorithm directly to address the issue of image inpainting. To implement this we assume that the pixels to be inpainted are known to us. As is common in inpainting, these are user specified and these pixels are marked as the unknown pixels. Next we fill in values for the unknown pixels using the technique suggested in Section 3.2. We find that the method works considerably well for areas where the region to be inpainted in not very wide. The performance of this method varies with the thickness of the contiguous unknown pixels,
even when the number of unknown points remain the same. The problem of inpainting becomes more complex when we encounter larger areas to be inpainted. The PG method, as the results show, is capable of providing a fast and acceptable quality of inpainting when the width of the scratches or areas to be inpainted is relatively thin (say, about 1 to 4 pixels wide).

3.5 Modifications to the PG method for Inpainting

For thicker edges the method does not generate a good result. In such cases the inpainted region displays a ringing effect such as the one experienced in the super-resolution case. We suggest suitable modifications to address this issue.

3.5.1 Filtering with a smooth cut-off

The ringing of the inpainted region is a result of the steep cut-off in the filtering process of the PG method. We introduce a smooth transition band for the cut-off region. The amount of smoothing is controlled by the width of this transition band. A wide transition band would considerably reduce the effect of ringing but it also translates to a poorer low-pass filtering and hence a poorer reconstruction. A trade-off condition is hence to be reached in deciding how smooth we would want our cut-off to be. For our implementation we have used a filter where the slope falls off gently as opposed to the steep cut-off for a ideal filter. This modification does away with the ringing effect to some extent.

3.5.2 Directional filtering

We observe that scratches in images are often highly directional. Hence, we investigate the possibility of using a directional filtering approach to solve the inpainting problem for scratches which have some sort of a directional nature such as the image shown in Figure 3.8(a). A scratch in one direction introduces higher frequency components in a direction perpendicular to the direction of the scratch. This effect can be seen when we compare the
spectrum of the unscratched original image with that of the corrupt image. We modified
the filtering process to take advantage of this directional nature of the image corruption.
Rather than taking a uniform filtering mask (which was circular in nature for all the above
methods), we use a directional filter with two different cut-offs - one for the direction where
higher frequency components have been introduced due to the scratch, and another for all
other directions. We set a lower bandwidth for the filter in the direction perpendicular to
that of the scratches. This has a greater smoothing effect across the edges of the scratch
and allows faster flow of information across the edges into the region to be inpainted.

3.5.3 Extension to different orthogonal bases

Till now our discussion has been restricted to various modifications to the method of fil-
tering in the frequency domain using the discrete Fourier transform (DFT) as the bases.
However, it may be noted here that the method will work under any orthogonal transfor-
mation maintaining sequency ordering. This encourages us to try out PG method with the
DCT bases instead of DFT. DCT implementation using integer arithmetic is also faster
than the DFT computation. The choice of DCT also stems from its high energy comp-
paction property. In DCT most of the information about the signal is captured in a few
lower frequencies. This leads us to expect that the filtering part in the PG method can be
achieved faster as DCT computation is quite fast and we have to deal with only a few DCT
coefficients. Taking advantage of the information packing capacity of DCT, in the low-pass
filtering process we simply retain the coefficients near the origin of the transformed image.
However, like the filtering in the DFT based PG method, a steep cut-off in the DCT do-
main also leads to ringing effect due to Gibb’s phenomenon. The DCT based method in
conjunction with a smooth cut-off filtering process in the DCT domain reduces the ringing
effect.
3.6 Results

3.6.1 Super-resolution

We applied the algorithm discussed in Section 3.3.1 on a Lena image of size 128 to perform 2× zooming. We found that the method performs better than the standard PG method. The results obtained are shown in Figure 3.4.

The image obtained using bicubic interpolation shown in Figure 3.4(b) is quite blurred with respect to that obtained using the standard PG method shown in Figure 3.4(c). We show that the result is further improved using the PG method with the deblurring step. This can be seen in Figure 3.4(d) which is sharper than the other two images shown even though its has a lower PSNR. Though PSNR is used as a measure of data fidelity, it is well known that it is not a good measure for comparing super-resolution schemes. Note the increased details in the texture of the hat, the sharpness of the lips and nose in the figure with respect to those obtained in Figures 3.4(b) and (c). Results with 3× magnification are shown in Figure 3.5. The bicubic image in Figure 3.5(b) is much blurred compared to the output of the standard PG method result shown in Figure 3.5(c). The result obtained using the deblurring step, shown in Figure 3.5(d), can be seen to be sharper than the other two images, specifically the scales and leaves of the pineapple.

Then we deal with the case where the downsampling process is not known to us. For our experiment we used the peppers image as the LR-HR image pair from which we try to learn the higher frequency components. First we applied the classic PG algorithm on the LR peppers image to perform 2× zoom. We then use the spectrum of the obtain peppers SR image and the HR image to calculate the filter with transfer function $H(u, v)$. Once this is done we now estimate the filter using equation (3.8). We choose a suitable $\beta$ parameter to be much higher than the estimated $\upsilon$. Using the new transfer function we iterate with the modified PG method to obtain Figure 3.6. However, we see little improvement using this method. Possible reasons for this are that our assumption of inverse-Gaussian filter may not hold true. Further investigation is required in this domain.
Figure 3.4: SR of Lena image using PG method with deblurring: (a) Low resolution Lena image formed by averaging and downsampling of a high resolution image, (b) $2 \times$ zoomed image using bicubic interpolation, (c) super-resolved image formed using the standard PG method, (d) super-resolved image formed using the PG method with deblurring.
Figure 3.5: SR of fruits image using PG method with deblurring: (a) Low resolution image formed by $3 \times 3$ pixel averaging and downsampling of a high resolution image, (b) $3 \times$ zoomed image using bicubic interpolation, (c) super-resolved image formed using the standard PG method, (d) super-resolved image formed using the PG method with deblurring.
Though this method promises to be able to handle blur due to the downsampling process, it still remains sensitive to the presence of noise.

### 3.6.2 Inpainting

In our experiment we first took the Lena image and scratched it arbitrarily with a variable width of 3 to 5 pixels. Subtraction of this image with the original image provided us with an accurate mask of the scratch. This mask is thus the region which is to be inpainted. Then taking all pixels outside this mask as known pixel values we perform the standard PG method to obtain an inpainted image. Similar experiment was carried out with overlaid text of variable font size. The results of the method are shown in Figure 3.7. Note that while the thinner areas are inpainted without any visual artifacts, this is not the case for the thicker regions.

We then performed inpainting with the proposed modifications. This time we used thicker edges of width 5 to 7 pixels. The scratches used this time were loosely directional in nature as shown in Figure 3.8(a). We obtain the scratch mask similar to the method explained above. The resultant inpainted image using the standard PG method with a low-pass bandwidth of 20% is shown in Figure 3.8(b). Note the ringing effect in the inpainted regions. Next we tried to better this method by using a smooth filter by doing away with the ideal low-pass filter and replacing it with a gradual cut-off. The result obtained shows lesser ringing as pointed out in Figure 3.8(c). This results in a 0.19dB gain over the standard method, as can be seen in Table 3.1. Figure 3.8(d) illustrates the result using DCT instead of the DFT, in order to take advantage of the greater information packing density of the DCT. We see that this produces nearly similar results, in terms of quality and PSNR, as compared to using DFT. The result obtained using a smooth cut-off on DCT of the image in the filtering process is shown in Figure 3.8(e). It shows a 0.29dB increase in PSNR from that obtained using a simple DCT based filtering. We then show the result obtained using the directional filtering method to take advantage of the directionality of the scratches. In this experiment the lowpass filtering bandwidth was kept to 10% of

28
Figure 3.6: SR using learning based PG method. (a) LR Peppers image for learning, (b) Original HR Peppers image for learning, (c) HR Peppers image using PG method, (d) LR Fruits image, (e) HR Fruits image using PG method without learning, (f) HR Fruits image using PG method with learning.
Figure 3.7: Inpainting using the PG method. (a) Lena image with different widths of scratching, (b) Inpainted image of (a), (c) Lena image with varying text sizes, (d) Inpainted image of (c).
<table>
<thead>
<tr>
<th>METHOD</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scratched Image</td>
<td>20.16</td>
</tr>
<tr>
<td>PG Method</td>
<td>37.67</td>
</tr>
<tr>
<td>PG with smooth cut-off</td>
<td>37.86</td>
</tr>
<tr>
<td>PG with directional filtering</td>
<td>37.72</td>
</tr>
<tr>
<td>PG using DCT</td>
<td>37.65</td>
</tr>
<tr>
<td>PG using DCT with smooth cut-off</td>
<td>37.94</td>
</tr>
</tbody>
</table>

Table 3.1: PSNR values of results from different inpainting methods as shown in Figure 3.8.

the dimension of the DFT in the direction perpendicular to that of the scratch and 30% elsewhere. The result obtained is shown in Figure 3.8(f). Notice that the ringing along the smoother areas have reduced. We see that in a PSNR sense all these modifications result in better inpainted images than that obtained using the standard PG method.
Figure 3.8: (a) Example of a scratched image requiring inpainting. Inpainting using (b) the PG method setting 80% of higher frequencies to zero, (c) a smooth filter with a smoothening band of 30 pixels, (d) the DCT based PG method, (e) DCT based PG with smooth cut-off and (f) directional filtering approach.
Chapter 4

Super-resolution using TV Approach

The theory of total variation (TV) was first introduced by Rudin et al. in [49]. Consequently, a lot of researchers have used this approach for image denoising, deblurring and restoration. In this chapter we give a short introduction and mathematical analysis of total variation. An in depth analysis of the theory of total variation has been discussed by Dibos and Koepfle in [50]. We briefly review the application of TV based regularization schemes for image restoration. We then discuss how the properties of TV have been used for the process of super-resolution and propose modification to get better results. While researchers have worked on image super-resolution from multiple low resolution observations, we restrict our work to the case when only one LR image is available.

4.1 Total Variation

4.1.1 Definition

Total variation (TV) for a given continuous and differentiable function $u$ is defined as [51]

$$T(u) \equiv \int_{\Omega} |\nabla u(t)| dt,$$

(4.1)

where $\Omega$ is the domain and

$$|\nabla u(t)| = \left( \left( \frac{\partial u}{\partial t_1} \right)^2 + \ldots + \left( \frac{\partial u}{\partial t_n} \right)^2 \right)^{\frac{1}{2}}$$
From the definition presented above it is clear that the TV is an integration of all the length of all gradients at any point in the domain $\Omega$. The gradient at a point is a measure of the variation of the function at the point, so an integration over the entire domain results in the total variation.

### 4.1.2 Optimization with TV

We define a constrained minimization problem as

$$
\min T(u) = \int_{\Omega} |\nabla u| \, dt
$$

subject to some constraint

$$
Au = b
$$

We convert the problem to an unconstrained minimization problem using the regularization parameter $\lambda$ to form the objective function

$$
\min_u J = T(u) + \frac{1}{2} \lambda \|Au - b\|^2
$$

The objective function in equation (4.3) is a sum of two terms: one regularization term, also known as the TV term, and the other fidelity term. Hence, we can calculate the gradient of the two terms separately. From the standard least-square calculations it is known that

$$
\nabla \left( \frac{1}{2} \|Au - b\|_2^2 \right) = A^T(Au - b)
$$

The gradient of the TV term is

$$
\nabla (T(u)) = -\nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right)
$$

The Euler-Lagrange equation for equation (4.3) is then

$$
-\nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) + \lambda A^T(Au - b) = 0
$$

It is to be noted here that the gradient for the TV term in equation (4.6) degenerates at regions where $|\nabla u| \approx 0$. To overcome this problem a small positive $\beta$ is introduced in the
cost function. Equation (4.3) then takes the form
\[
\min_u J = \int_\Omega \sqrt{\|\nabla u\|^2 + \beta^2} \, dt + \frac{1}{2} \lambda \|Au - b\|^2
\] \hspace{1cm} (4.7)

Numerical solutions to the constrained minimization problem have been proposed by solving the associated Euler-Lagrange equation. The term \(\beta\) plays an important role in the nature of the solution as has been established in [51]. In general \(\beta\) has a smoothening effect on the solution. If the value of \(\beta\) is large in comparison to the derivative term \(|\nabla u|\) we will get an unregularized solution, which is the least-squares solution. Too small a value, on the other hand, results in numerical instability due to division by zero. To overcome such a computational problem duality-based methods have been proposed. Such methods have been discussed in [51, 52].

### 4.2 TV for Image Restoration

In this section we will discuss the total variation approach in the 2 dimensional case. Since an image can be thought of as a discrete representation of a \(2 - D\) signal, we need to adapt the continuous TV introduced in Section 4.1.1 for images. TV is a non-linear edge preserving method. It allows discontinuities (i.e. sharp edges) in its solution and the resultant image tends to be piece-wise constant in nature. We have provided a brief survey of how it has been used for image restoration, chiefly in areas of image denoising, deblurring and inpainting. The TV was first used for image denoising by Rudin \(\text{et al.}\) in [49]. In their model (referred to as the ROF model), the authors solve the constrained minimization problem
\[
\min \int_\Omega \sqrt{u_x^2 + u_y^2} \, dx \, dy \quad \text{(4.8)}
\]
subject to the constraints
\[
\int_\Omega u \, dx \, dy = \int_\Omega u_0 \, dx \, dy
\]
and
\[
\frac{1}{2} \int_\Omega (u - u_0)^2 \, dx \, dy = \sigma^2
\]
where $u(x, y)$ is the denoised image and $u_0(x, y)$ is the noisy input image. The noise is assumed to be zero mean white noise with a known standard deviation of $\sigma$. The solution procedure uses a parabolic equation with time as the evolution parameter for the gradient descent method. The authors thus solve

$$u_t = \frac{\partial}{\partial x} \left( \frac{u_x}{\sqrt{u_x^2 + u_y^2}} \right) + \frac{\partial}{\partial y} \left( \frac{u_y}{\sqrt{u_x^2 + u_y^2}} \right) - \lambda(u - u_0), \quad t > 0, \; x, y \in \Omega \quad (4.9)$$

where in the discrete case $u_x = u(i, j) - u(i - 1, j)$ and $u_y = u(i, j) - u(i, j - 1)$ is known as the forward difference. This representation results in an iterative algorithm for solving the optimization problem using the equation

$$u^{k+1} = u^k + \Delta t \left( \frac{u_x (\beta^2 + u_y^2) - 2u_x u_y u_{xy} + u_{yy} (\beta^2 + u_x^2)}{\sqrt{(\beta^2 + u_x^2 + u_y^2)^3}} + \lambda(u_0 - u) \right)^k \quad (4.10)$$

where $\beta$ is a small positive value introduced to avoid division by zero in the smooth regions of the image where $|\nabla u| \approx 0$ and $\Delta t$ is the time step for gradient descent method. The $k$ in the superscript implies that all calculation are done for the $k^{th}$ iteration.

The authors use this model for denoising of images. They have also presented an iterative method for choosing $\lambda$ to obtain a favorable solution. The TV method has been found to be effective for images which can be classified as piece-wise constant. While the method preserves edges, it does not work well with texture detail in images. Gilboa et al. in [53] propose a spatially varying fidelity term which preserves the finer details of an image. Based on the local variance measures the extent of denoising is controlled in a spatially varying nature. A modified TV regularization functional has also been applied for restoration of blurred and noisy images by Vogel and Oman in [54]. They combine a fixed point iteration to handle non-linearity in TV with a preconditioned conjugate gradient method for linear systems.

Total variation has also been used for image deblurring when the PSF of the camera system (i.e. the blurring kernel) is known. In that case the objective function is taken as

$$J(u) = \int_{\Omega} |\nabla u| \; dx \; dy + \frac{1}{2} \lambda \int_{\Omega} (u * h - u_0)^2 \; dx \; dy$$

(4.11)
where $h$ is the blurring kernel and $*$ denotes convolution. The process of deblurring is thus deconvolution under the total variation sense. Chan \textit{et al} in [55] present an alternate minimization (AM) algorithm to perform blind deconvolution on blurred images in cases where the blurring kernel is not known. The objective function in this case is taken as

$$
\min_{u, h} J(u, h) = \min_{u, h} \alpha_1 \int_\Omega |\nabla u| \, dx \, dy + \alpha_2 \int_\Omega |\nabla h| \, dx \, dy + \frac{1}{2} \|u \ast h - u_0\|^2
$$

(4.12)

where $\alpha_1$ and $\alpha_2$ are the Lagrange multipliers. In this case both $u$ and $h$ are unknown. Using an iterative process which alternately minimizes one of the two unknowns in each iteration while keeping the other fixed, the authors are able to perform deconvolution as well as predict the blurring kernel. The authors in [37] have shown that such a method is a good candidate for performing inpainting on images which have been scratched and blurred. They show that under such cases it is not possible to perform inpainting and deblurring separately. A simultaneous TV based inpainting is performed along with blind deconvolution in places where the blurring kernel is defined.

Some researchers have also applied different methods to address the issue of zooming into an image. The main challenge of this issue is how to preserve edges that are present in any natural image. A variety of linear and non-linear tools are available which try to address this issue. A detailed mathematical analysis of such schemes have been provided by Malgouyres and Guichard in [56]. The authors in [25] 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 has also been demonstrated by Aly and Dubois in [26]. In their method they modify the data fidelity term to closely model the image acquisition method. The iterative algorithm in their case is

$$
u^{(k+1)} = \left[u + \Delta T \left(\lambda u_k - H^T (Hu - u_0)\right)\right]^{(k)}
$$

(4.13)

where $u$ is a vectorized form of the image, $u_k = \frac{u_{xx} u_y^2 - 2u_{xx} u_x u_y u_{yy} + u_{yy} u_x^2}{(u_x^2 + u_y^2)^{3/2}}$ similar to the form in equation (4.10) and $H$ is the matrix that captures the blurring and downsampling of the
LR image formation process. Then $H^T$ becomes the upsampling process applied to the residue in a the low dimensional grid. In that sense the data fidelity term becomes similar to the case of iterative back projection of Irani and Peleg [48] with a regularization term. 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 [57].

While all the methods presented above deal with enhancement of spatial resolution of images from a single low resolution image, a lot of work has also been done where total variational approaches have been used to super-resolve images from multiple frames. Such methods, as previously discussed, require a robust registration process and have to deal with non-uniformly spaced samples in the higher resolution grid. One such approach has been used by Chan et al. in [18]. In this the authors make use of the algorithm developed for their work in [37]. They obtain multiple blurred, noisy LR frames from adjacent frames of a video and attempt to form a high resolution image. A total variation based regularization method is then used to perform simultaneous inpainting and deblurring to obtain a super-resolved image. A combination of TV and bilateral filter named Bilateral TV (BTV) has been used as a regularizing term for super-resolution by Farsiu et al. in [16, 17, 28]. As opposed to the $L_2$ norm used as a data fidelity term in most cases, the authors use the $L_1$ norm. They present a two step algorithm where they first use the median filter to build a high resolution grid from multiple LR images. Regularization is then done to perform an iterative interpolation to deblur as well as inpaint missing pixels in the HR grid. Another algorithm is presented in [17] where the authors make use of the temporal information between frames of a low resolution video. SR is then performed under a control theoretic approach using an approximation of the Kalman Filter along with the previous framework.
4.3 TV based Super-resolution

Since we restrict ourselves to the case of super-resolution when only one low resolution observation is available our system of equations is always under-determined and multiple solutions exist which form the LR image under any given deblurring and downsampling process. A regularization based system is thus a necessary choice and a tool which does not hamper the sharpness of edges is preferred. Our motivation for the use of TV based regularization is stemmed from its edge preserving property which is vital for super-resolution. However, the disadvantage in using TV as the regularizer is that it results in an image where finer details like texture are smoothened out. It is for this over-smoothening effect that it becomes important to select a proper data fidelity term for the objective function. In our work we assume that the process of formation of the LR image is known to us. This means that we know the blurring and the downsampling process of the LR image formation model. For our work we take the blurring effect to be simple averaging in a rectangular window, though the methodology should hold for any general blurring kernel. The LR image is assumed to be formed by simple subsampling of this blurred HR image.

In the objective functions used in literature, it is found that when the data fidelity term is chosen as the $L_2$ norm of the error energy, we penalize the error from data fidelity more than the derivative term due to the former’s quadratic nature. A $L_1$ norm, as taken in [17], penalizes the error equally at all places. Ideally, for a sharper reconstruction, we would like to penalize the error more when the estimated LR image deviates from the input LR image at the edges than at the smoother regions. Consequently, we propose an extra data fidelity term where such a differential treatment will be possible. For this purpose, we split the LR image and the estimated LR image (formed by applying the decimation model to the evolved SR image at each iteration) into bands using bandpass filters having non-overlapping pass bands. In essence we split up the images into a number of subband component images and enforce data fidelity differently over such subbands. This is used as additional constraints in the commonly used total variation regularization scheme. The
objective function we use is then

\[ J(u) = \int |\nabla u| \, dx \, dy + \frac{1}{2} \alpha \int (u * h - u_0)^2 \, dx \, dy + \frac{1}{2} \sum_k \lambda_k \int (\tilde{U}_k - U_{0k})^2 \, dv \, d\omega \]  

(4.14)

Here \( u \) denotes the restored HR image, \( h \) is the known blur kernel, \( u_0 \) is the starting image formed by interpolation of the LR image, \( \tilde{U}_k \) denotes the \( k^{th} \) sub-band of the estimated LR image formed under the known decimation model, \( U_{0k} \) is a similar decomposition for the input LR image. Under this framework, it then becomes possible to assign different weights \((\lambda_k)\) to errors in different bands. As opposed to the other schemes discussed before, the additional error term in this case is calculated and weighed in the spectral domain. It may be argued here that it follows from Parseval’s theorem that calculating error power in the spatial domain and the frequency domain should be equivalent. However, the operation in the spectral domain makes it easy to split an image into separable components based on spectral contribution. Under the absence of noise, we use a higher weight factor for the higher spectral bands which capture the finer details and the edges of the image. Using such a model it is then possible for us to enforce that more importance is given to data fidelity at the edges. This should ensure that the image that is formed is a sharper reconstruction of the available LR image under the known image formation model. On the other hand, noise in an image can be expected to be captured in the higher frequency bands, which necessitates the use of smaller weights for the higher bands when the input image is noisy. The first data fidelity enforcing term in this case is the same as that of the ROF model shown in equation (4.8). An appropriate choice of \( \lambda_k \)s would ensure a proper trade-off between the sharpness of the super-resolved image with the accentuation of the noise present. We have experimented with both noisy and noiseless cases and the results are shown in Section 4.4.

As with all cases of regularization, the solution differs a lot with the choice of the weighing terms \( \lambda_k \). Too low a value for these terms tends to over-smoothen the image and too high a value will tend to the least mean square solution of the under-determined system. So it becomes important to choose the parameters of the system properly. Unfortunately, in our case, as we split the image into multiple bands, the number of controlling parameters
increase linearly. Hence it becomes difficult to tune all the parameters to obtain the “best” solution. This leads us to investigate the possibility of using the available images itself to ascertain the weights for the bands using the iterative nature of the algorithm. We propose to use a spatially varying $\lambda_k$ as the scaling term as opposed to the scalar terms used in the previous case. In our system, we first split each of the formed SR images and the LR image into $k$ images, each consisting of contribution from one of the $k$ spectral component bands only. We calculate the weighing parameter matrix $\lambda_k$ as

$$\lambda_k = (\tilde{u}_k - \tilde{u}_{k,\text{mean}})(u_{0_k} - u_{0_k,\text{mean}})$$

where $\tilde{u}_k$ is the estimated LR image formed using the decimation model with contribution from only the $k^{th}$ spectral band. The advantage of using such a $\lambda$ matrix is that the weighing will vary spatially in each of the $k$ images. For instance one would expect that for the highest band which captures the sharpest edges, error penalization is done mainly at the edges. Our model achieves specifically that. In this case we do not apply the weighing term to the spectral components of the image as in equation (4.14). The objective function now used is

$$J(u) = \int |\nabla u| \, dx \, dy + \frac{1}{2} \sum_k \int \lambda_k (\tilde{u}_k - \tilde{u}_{0_k})^2 \, dx \, dy$$

where $\tilde{u}_k$ denotes the estimated LR image formed with contributions from the $k^{th}$ spectral band only and $\tilde{u}_{0_k}$ is the same for the input LR image. This model can also be expected to work better in cases where the system is corrupted by additive noise. Assuming that noise remains restricted to the higher spectral bands, the weight of the band should automatically adjust accordingly and yet not sacrifice the quality of the edges. This is because of the spatial variance of the $\lambda$ value in each band. We present results obtained using this modified approach of TV based regularization with spatially varying multipliers in the next section.

### 4.4 Results

We first performed experiments based on the scalar value of $\lambda_k$. The images were decomposed into two bands and based on the theory presented above we apply a higher weight
to higher frequency component of the image. For the results shown here we use the values \( \lambda_1 = 0.6 \) and \( \lambda_2 = 0.8 \) where a higher index of \( \lambda \) value implies a higher frequency band. The results obtained using these parameter values are shown in Figure 4.1 and Figure 4.2.

In Figure 4.1 we can see that the total variational deblurring performed on the bicubic reconstruction sharpens the image at edges but at the cost of loss of the texture. This is not the case for Figure 4.1(d). This can be noted from the presence of texture in the hat and the hair, even though the overall reconstruction remains sharp. This is specifically what we wanted to achieve by our method. A similar effect can be seen in Figure 4.2 where the result from our method yields a better texture than that of TV based deblurring. This is visible at the terrain and finer details on the tank. This proves that band splitting and differential weighing of the bands indeed performs better as far as restoration of texture is concerned.

We attempt to try our method where the image is corrupted by zero mean additive noise. We use a Gaussian noise of variance of 25 ([0, 255] being the range of pixel values in the image). In our reconstruction process we do not make use of any information about the nature of noise. Our theory builds on the assumption that this noise remains limited to higher frequency bands of the image. Hence we use a lower scaling factor for the higher frequency band. The results obtained using parameter values of \( \alpha = 0.7, \lambda_1 = 0.8 \) and \( \lambda_2 = 0.6 \) are shown in Figure 4.3.

As can be seen from Figure 4.3 the bicubic interpolation does not perform any denoising, as expected. Total variation based regularization performed on the bicubic interpolated image reduces a lot of noise but the resultant image is smooth. The modified approach yields a sharper reconstruction but the presence of noise is clearly visible. Apart from an enhancement of details, an improvement in the PSNR is also observed with the proposed method. More denoising will lead to a smoother reconstruction, as can be expected from the method. The reason for this is that spectral bands which capture sharp edges and texture will also contain most of the noise and hence it becomes difficult to distinguish between noise and texture using the present method.
Figure 4.1: SR using modified TV approach with scalar weights for $2 \times$ zoom: (a) Input LR image, (b) bicubic interpolated image, (c) TV based deblurring of (b), (d) SR using modified TV based approach using only 2 bands.
Figure 4.2: SR using modified TV approach with scalar weights for $2 \times$ zoom: (a) Input LR image, (b) Bicubic interpolated image, (c) TV based deblurring of (b), (d) SR using modified TV based approach using only 2 bands.
Figure 4.3: SR using modified TV approach with scalar weights for 2× zoom for noisy case: (a) input LR image, (b) bicubic interpolated image, (c) TV based denoising of (b), (d) SR using modified TV based approach using only 2 bands.
We now move on to the second part of the experiment where we work with a spatially varying scaling parameter. We expect this model to capture the spatial as well as the spectral contribution of different parts of the image. Unlike the previous case we now apply the scaling parameter in the spatial domain itself. However, each of the $k$ images formed are from contribution of the corresponding $k^{th}$ band only. In our experiments we decompose the image into three bands. However, it can be easily extended to multiple such bands. We take the lower 40\% of the spectrum to be the lower band, the next 30\% to be the middle band and the rest as the higher frequency band. Using this configuration we perform 14 iterations of our method. Since calculation of $\lambda_k$ values is time consuming, we perform updation of the multiplier matrix in every 5\textsuperscript{th} iteration of the restoration process. The results thus obtained are shown in Figure 4.4. For comparison we use the bicubic image which has been restored using simple TV based regularization. In Figure 4.4(b) we find that our new method does introduce certain amount of texture visible in the hat. Also, the TV reconstruction shown in Figure 4.4(a) appears to be flatter in comparison. In Figure 4.4(c) we show the case of TV denoising for a noisy LR image where the variance of the Gaussian additive zero mean noise is 25 (pixel values ranging from $[0 - 255]$). In Figure 4.4(d) we show the output of the modified TV based restoration with spatially varying multipliers.

In the case of spatially varying multipliers we do not make use of any deblurring or denoising term, as the case may be, apart from what is done using the image bands. As can be seen from the figure, the result is not comprehensively better than that obtained using scalar multipliers. However, it is an improvement from the normal TV method in the PSNR sense. More investigation is required to determine a better method for calculation of the $\lambda_k$ matrices so that they can mathematically capture the essence of the method.
Figure 4.4: SR using modified TV approach with spatially varying $\lambda_k$ for 2x zoom: (a) TV deblurring of bicubic image, (b) SR using modified TV based restoration, (c) TV based denoising of noisy bicubic image, (d) SR using modified TV based approach in presence of noise.
Chapter 5

Conclusion and Future Work

In this dissertation we discuss some techniques which deal with image restoration when only a single degraded observation is available. We make use of an old Papoulis-Gerchberg technique of signal extrapolation to perform image super-resolution and inpainting and combine it with newer techniques to achieve better results and allow the method to be applicable to a larger class of images. We modify the method to better handle blurring effects introduced in the decimation process. For this we need to have knowledge about the decimation process, specifically the blurring kernel involved in the process. This limitation motivates us to investigate the case when such information is not available. We propose the use of a LR and HR image pair to infer a filter which tries to bridge the gap between the quality of the image obtained by the PG method and the actual desired HR image. However, we do not get very encouraging results in our endeavors in this direction. A separate approach may be to model the filter as a weighted average filter and try to learn the weighing coefficients. From a single LR image this method is ill-posed in nature and the possible use of regularization to model a filter will be an interesting exercise.

The PG method as used for image super-resolution was easily adapted for the image inpainting case. In fact, we could make use of the method without having to perform deblurring or denoising. The difficulty in inpainting is that since missing pixels tend to be clustered together, the problem becomes more ill-posed and the PG method is not
suited to handle interpolation for large areas. We propose methods that produce better results, in terms of either a faster scheme or lesser ringing, to make the method applicable to slightly larger scratch areas. We are aware of the fact that there exist quite a few advanced techniques in literature which outperform the method when it comes to inpainting over large regions. However, the very simplicity of the process is its main advantage and attraction. The possibility of using a PDE based methodology in the PG framework to perform inpainting over larger regions remains to be investigated.

Even though the PG method in all its simplicity performs quite well for image super-resolution, it remains sensitive to measurement errors and presence of noise in the image. We attempt to try and solve ill-posed problems of this nature with a regularization scheme. As edge preservation is one of the major requirements of super-resolution we choose the TV norm for the regularization term. A severe drawback of the process in SR terms is its inability to differentiate between texture and noise, leading to a overly smooth reconstruction. Methods in literature address this issue for image restoration by decomposition into cartoon and texture components. But we propose something simpler which can perform texture preserving super-resolution. We propose further extensions by way of a spatially varying fidelity term which promises to be able to perform denoising as well as deblurring - an important requirement for super-resolution. We need to further experiment with a mathematical method to calculate this fidelity term.

In our work on super-resolution we have come across a multitude of methods, each addressing certain specific problem. However, an ideal method should address all the three issues of blur, noise and aliasing in the LR image. A method targeting all three has remained elusive till now. Another goal of research should be finding computationally inexpensive algorithms for super-resolution and inpainting. The PG method is fast but still falls far short of being used in real-time applications. The regularization approach using TV is much slower in comparison. Even though the computing speed of computers and other hardware keep increasing, a real-time implementable method is still unavailable. Finally, the research community needs a mathematical measure for super-resolution. The PSNR or the MSE measure is a poor indicator of data fidelity as far as super-resolution
is concerned. A possible measure would be using only higher frequency components to
calculate these measures since we are more interested in the addition of higher frequency
components in super-resolution. The fields of super-resolution and inpainting still pose a
lot of challenges with each new solution throwing up many more problems. A lot of work
thus remains to be done in these areas of research.

“This is not the end. It is not the beginning of the end. But it is perhaps the
end of the beginning” – Sir William Churchill.
References


