SUPER-RESOLUTION WITH BETTER EDGE ENHANCEMENT

by

GAURAV HANSDA

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

MASTERS OF SCIENCE IN ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

November 2012

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. K. R. Rao for supervising me on this thesis and helping me in every step of my work. His patience, motivation and guidance helped me immensely during my research and the writing of this thesis. I am grateful to Dr. Jonathan Bredow and Dr. Howard Russell for their invaluable advice and interest in my research and for taking time to serve on my thesis committee. I would like to thank all the members of the Multimedia Processing Lab for creating a positive and enjoyable work environment and being available at any time to discuss problems. I would also like to thank my roommates Subhasish Dutta Chowdhuri and Ankit Upadhyaya for being there with me in my difficult times and supporting me throughout.

Finally, I thank my parents for their constant encouragement and unconditional support and my sister for her consistent motivation.

November 19, 2012

ABSTRACT

SUPER RESOLUTION WITH BETTER EDGE ENHANCEMENT

Gaurav Hansda, M.S.

The University of Texas at Arlington, 2012

Supervising Professor: Prof. K. R. Rao

Super resolution image reconstruction is a promising technique of digital imaging which attempts to generate a raster image with a higher resolution than its source. The source can consist of one or more images or frames. This thesis focuses on single-frame super-resolution i.e. the source is a single raster image. Changing the resolution directly refers to resampling the image. Specific algorithms have been designed for scaling up, scaling down and rotating images that deliver high quality results. Process of scaling up or super-resolution is the primary focus of this thesis, which in contrast scaling down or rotating deals with the problem of increasing the amount of pixels or, more general, data. In comparison to various other image enhancement techniques, super-resolution image reconstruction technique not only improves the quality of under-sampled, low-resolution images by increasing their spatial resolution but also attempts to filter out distortions.

The Iterative back-projection (IBP) is a classical super-resolution method with low computational complexity that can be applied in real time applications. However, it often produces many artifacts especially along the strong edges. To reduce these jaggy artifacts, this thesis proposes a novel approach to single image super-resolution image

reconstruction. First, an image up-sampling scheme is implemented which takes the advantage of bilateral filtering as well as mean shift image segmentation. This produces a better smoothed image. Then, a complex shock filter is used to enhance strong edges in the initial up-sampling result and obtain an intermediate high-resolution image. Complex shock filters are less sensitive to noise than other shock filters. After that, a reconstruction constraint on the high-resolution image is applied, so that fine details can be suppressed in the back projection step. IBP is the state of art method for this. Finally, as a post-processing step on the reconstructed image, the structural content of low resolution image and the correlation among the pixels with similar structure are learned and employed on the reconstructed image to further stop across-edge propagation. The experimental results proved that this proposed method can remove the artifacts subjectively and obtain sharp edges in visual perception. Comparing this algorithm with several other state-of-the-art image super-resolution algorithms shows that this approach performs better in terms of both qualitatively and quantitatively (peak signal-to-noise ratio and mean square error).

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

LIST OF ACRONYMS

AWGN - Additive White Gaussian Noise

CAT- Computed Axial Tomography

CC – Cross Correlation

CCD – Charge-Coupled Device

CMOS – Complementary Metal-Oxide-Semiconductor

CT – Computed Tomography

DDT – Data-Dependant Triangulation

DVD – Digital Versatile Disc

FR - Full reference

FSIM – Feature SIMilarity

GM – Gradient Magnitude

HDTV – High-Definition TeleVision

HR – High Resolution

HVS – Human Visual System

IBP – Iterative Back Projection

IQA – Image Quality Assessment

LAZA - Locally Adaptive Zooming Algorithm

LR – Low Resolution

MAD – Mean Absolute Differences

MP – Mega-Pixel

MRI – Magnetic Resonance Imaging

MSE – Mean Square Error

NLIBP – Non-Local Iterative Back Projection

NOAD - New American Oxford Dictionary

NTSC – National Television System Committee

PC – Phase Congruency

POCS – Projection Onto Convex Set

PSF – Point Spread Function

PSNR – Peak Signal-to-Noise Ratio

ROI – Region Of Interest

SIAD - Smart Interpolation by Anisotropic Diffusion

SR – Super-Resolution

SSIM – Structural SIMilarity

TVP – Total Variation Preserving

UAVs – Unmanned Aerial Vehicles

VISTA – Vision by Image/Scene TrAining

CHAPTER 1

INTRODUCTION TO SUPER-RESOLUTION

1.1 Introduction

The world has seen an immense global advancement in the technology, both in hardware and software. Digital pictures today are all around us- on the web, on digital versatile discs (DVDs), on satellite systems; they are everywhere. Having these pictures in digital format allows us to manipulate them the way one want them. Digital image processing helps in enhancing the features of interest and extracting useful information about the scene from the enhanced image. Initial ideas on image processing were used in 1920s for just cable transmission of pictures [1] [5]. Since majority of the information received by a human being is visual, integrating the ability to process visual information into a system would certainly enhance the overall utility. Work on using computer techniques for improving the quality of images obtained from space probe began at the Jet Propulsion Laboratory in 1964 when pictures of the moon transmitted by Ranger 7 were processed by a computer which also corrects various types of the image distortions inherent in the on-board television camera [1]. The field of image processing has grown considerably during the past few decades with improvements in size, speed, and cost effectiveness of the digital computers. The camera sensor manufacturing units also have advanced in their manufacturing techniques to produce good quality high-resolution (HR) digital cameras. Although, HR digital cameras are available, many computer vision applications such as satellite imaging, target detection, medical imaging, and many more still have a strong requisition for higher resolution imagery which very often exceed the capabilities of these HR digital cameras. To cope up with strong demand of higher

-resolution imagery, these applications have approached image-processing techniques for a solution to generate good quality HR imagery.

The topic of super resolution (SR) first appeared in the early 1980s, with one of the first papers in the signal processing community, the paper by Tsai and Huang [2]. Super-resolution image reconstruction is a promising technique of digital imaging which attempts to reconstruct HR imagery by fusing the partial information contained within a number of under sampled low-resolution (LR) images of that scene during the image reconstruction process. Super-resolution image reconstruction involves up-sampling of under-sampled images thereby filtering out distortions such as noise and blur. In comparison to various image enhancement techniques, super-resolution image reconstruction technique not only improves the quality of under-sampled, low-resolution images by increasing their spatial resolution but also attempts to filter out distortions.

1.2 Image Resolution

Optical resolution is a measure of the ability of a camera system, or a component of a camera system, to depict the picture detail [3]. On the other hand, image resolution is defined as the fineness of detail that can be clearly distinguished in an image. Resolution is a fundamental issue in judging the quality of various image acquisitions or processing systems. In its simplest form, *image resolution* is defined as the smallest discernible or measurable detail in a visual presentation. Both the definitions apply to digital and analog camera systems and images. However, in this research, the term resolution will only relate to digital camera systems and digital images. Researchers in digital image processing and computer vision classify resolution into three different types [4].

- *Spatial Resolution:* An image is made up of small picture elements called pixels. Spatial resolution refers to the spacing of the pixels in an image and is measured in pixels per unit length. The higher

the spatial resolution, the more are the pixels in an image. High spatial resolution allows a clear perception of sharp details and subtle color transitions in an image. In case an image with high levels of details is not represented by a spatially dense set of pixels, the image is said to suffer from aliasing artifacts. For an output device such as a printer the spatial resolution is expressed in dots per inch (dpi) [4].

- *Brightness Resolution:* Also known as gray-level resolution, it refers to the number of brightness levels or gray-levels used to represent a pixel. The brightness resolution increases with the number of quantization levels used. A monochrome image is usually quantized using 256 levels with each level represented by 8 bits. For a color image, at least 24 bits are used to represent one brightness level, i.e., 8 bits per color plane (red, green, blue). It should be noted that the number of gray value quantization levels is also intrinsically related to the spatial sampling rate. If the camera sensor has fewer quantization levels, it should have a much higher spatial sampling rate to capture the scene intensity. This idea is quite similar to that of delta modulation used in communication systems and to that of dithering used in half-tone printing [4].
- *Temporal Resolution:* It represents the frame rate or the number of frames captured per second. Higher the temporal resolution, lesser is the flicker observed. The lower limit on the temporal resolution is proportional to the amount of motion that occurs between two consecutive frames. The typical frame rate for a pleasing view is about 25 frames per second or above [4].

Another kind of resolution of interest is the spectral resolution and it refers to the frequency or spectral resolving power of a sensor that gives the bandwidth of the light (or electro-magnetic wave) frequencies captured by the sensor. It is defined as the smallest resolvable wavelength difference by the sensor. The spectral resolution plays an important role in satellite imaging.

In this research, the term resolution refers to the spatial resolution unless otherwise mentioned. As stated earlier, spatial resolution essentially describes the total number of pixels in an image, horizontally and vertically. For instance, a digital image 300 pixel (wide) x 300 pixel (high) consists of a total of 90,000 pixels or is nearly 0.1 megapixel (MP) image. If this image is quadrupled, the dimensions will be 1,200 pixels (wide) x 1,200 pixels (high) with a total of 1,440,000 pixels or nearly 1.5 MP. Clearly, the detail carrying capacity of an image is directly proportional to the number of pixels in an image. Higher the number of pixels, higher is the detail representation of the image.

1.3 Image Degradation Factors

The acquired image usually represents the scene usually in an unsatisfactory manner. Since, real imaging systems as well as imaging conditions are imperfect: an observed image represents only a degraded version of the original scene. These degradations in the images are caused due to various factors such as blur, noise and aliasing. Figure 1.1 shows an example of the original scene and a corrupted image. Such distortions may get introduced into an imaging system due to the following reasons:

- Motion between the camera sensor and the scene or subject.
- Camera optics and lenses.
- Atmosphere.
- Insufficient sampling.

Factors like motion of the scene, wrong focus, atmospheric turbulence and optical point spread function can introduce degradations in an image known as blur, during the imaging process. Removing the effect of blurring in an image is known as de-blurring which is a well known image enhancement technique. If the conditions at the time of acquiring an image are known, it is much easier to de-blur the image accurately. Figure 1.2 shows an example of a blurred image and the original scene.

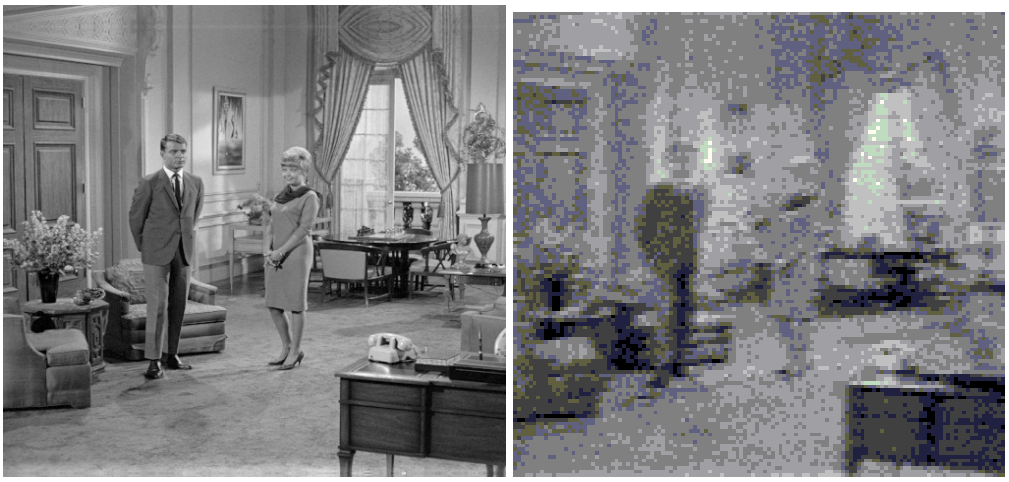


Figure 1.1 Original image (left panel) and its noisy, blurred and under-sampled image (right-panel).

Noise is a random background event and is certainly not a part of the ideal scene/signal and may be caused by a wider range of sources such as variations in detector sensitivity, optical imperfections and environmental changes. Although many noise models exist in literature, only Gaussian white noise is considered as it provides a good model for noise in most of the imaging systems. The noise is also assumed to be spatially uncorrelated with respect to the image, i.e., there is no correlation between the

image pixel values and the noise components. Figure 1.3 shows an example of a noise-corrupted image.

Figure 1.2 Original image (left panel) and its blurred image (right-panel).

Figure 1.3 Example of the noise-corrupted image (left panel) of an original scene (right-panel)

Another factor affecting image resolution is due to the insufficient spatial sampling of images. As per the Shanon-Nyquist sampling theorem [5] [6] [7], the sampling rate should be greater than twice the highest frequency. If the sampling frequency is less than twice the highest frequency, then all frequency components higher than the half the sampling frequency are reflected as lower frequencies in the reconstructed signal. This is referred to as under-sampling of images which occurs in many imaging sensors. Because of under-sampling, the high frequency component overlap with the low frequency components and get introduced into the reconstructed image/signal causing degradation of the image. Such degradation is known as aliasing which consequently causes partial loss of scene information. The aliasing may also give rise to artifacts thereby corrupting the reconstructed image. In order to reduce these artifacts, anti-aliasing techniques are implemented. Figure 1.4 shows as example of an aliased, under-sampled image and the original scene.

Figure 1.4 Example of an under-sampled image (left panel) of an original scene (right-panel)

1.4 Significance of Super-resolution

As defined earlier, spatial resolution refers to the spacing of pixels in a digital image. Therefore, more the number of pixels, more detailed are the information contained within the image. So a fundamental question arises as to why in the first place the super-resolution algorithms are required. To answer this question first one should know about the image sensor.

An image sensor or camera is a device which converts optical energy into an electrical signal. Modern imaging sensors are based on the charge-coupled device (CCD) technology or a complementary metal-oxide-semiconductor (CMOS) active-pixel sensor. This essentially consists of an array of photo-detector elements or pixels that have a voltage output proportional to the incident light [8]. The sensor size or equivalently the number of sensor elements per unit area in the first place decides the spatial resolution of the image to capture. Higher the number of detector elements, more is the resolution. An imaging system with inadequate detectors will produce a low resolution image, with blocky effect. This is because when a scene is photographed from a low resolution camera, it is sampled at a low spatial sampling frequency, causing aliasing effect. Now based on this, there can be two possible ways to increase the resolution.

One is to reduce the size of the photo-detector elements. Thereby increasing the density and hence the sampling rate. But as pixel size decreases, the amount of light incident on each pixel also decreases, and this causes shot noise [9] [10], which degrades the image quality. Increasing the pixel density increases the resolution but also causes the shot noise. Thus there exists a limitation on the size of a pixel in a sensor and the optimal size is estimated to be about $40\mu m^2$ [11]. The current image sensor technology has almost reached this level.

Another approach to increase the resolution is to increase the wafer size which leads to increase in the capacitance [12]. But such an increase in capacitance will certainly cause a decrease in charge transfer rate. This limitation causes the image of a point light source to be blurred. Also there is distortion due to aliasing because of a low sampling rate for a low resolution sensor. In certain defense computer vision applications, unmanned aerial vehicles (UAVs) are used for acquiring images. All UAVs have a payload carrying capacity and is usually half the UAVs take off or launch weight [13]. It is therefore not feasible to mount heavy HR cameras with image stabilization equipment to counteract the vibrations of UAVs. The ever demanding need for high resolution imagery stimulated research and development of super-resolution techniques.

While the image sensors limit the spatial resolution of the image, the image details (high-frequency bands) are also limited by the optics, due to lens blurs (associated with sensor point spread function (PSF)), lens aberrations effects, aperture diffractions, and optical blurring due to motion [14]. Constructing imaging chips and optical components to capture very high resolution images is prohibitively expensive and not practical in most real applications, e.g., widely used surveillance and cell phone built-in cameras. Besides the cost, the resolution of a surveillance camera is also limited in camera storage and hardware storage. Thus there is a need for developing post acquisition signal processing techniques to enhance the resolution. These techniques being post processing methods applied on low resolution images, they offer flexibility as well as cost benefit since there is no additional hardware cost involved. However, the increased computational cost may be the burden that any user has to bear.

Hence to summarize, the more appropriate way of addressing the previously mentioned problems is to accept the image degradations and use signal processing to post-process the captured images and to trade off computational cost with the hardware

cost. These techniques are specifically referred to as super-resolution (SR) reconstruction.

1.5 Applications of Super-resolution

The field of super-resolution has a vast area of application. Although the concept of super-resolution remains the same, the techniques of achieving HR imagery may or may not be same for each and every application. This is because in certain applications such as real time video surveillance or target detection, computational time is of great importance and hence requires a super-resolution technique with high accuracy and low computational cost. On the other hand, for certain applications such as astronomical imaging or text recognition, computational cost is not a constraint and therefore such application can implement super-resolution techniques with high accuracy and a higher computational cost. Few of the applications are:

1. Surveillance video [15] [16]: frame freeze and zoom region of interest (ROI) in video for human perception (e.g., look at the license plate in the video), resolution enhancement for automatic target recognition (e.g., try to recognize a criminal's face).
2. Remote sensing [17]: several images of the same area are provided, and an improved resolution can be sought.
3. Medical imaging (CT, MRI, ultrasound, etc.) [18] [19] [20] [21]: several images limited in resolution quality can be acquired, and SR technique can be applied to enhance the resolution.
4. Video standards conversion, e.g., from NTSC video signal to HDTV signal.
5. Astronomical imaging, [22], [23].
6. Target detection and recognition [24].

1.6 Research Outline

In this research, a novel approach to super-resolution image reconstruction is proposed. The problem of super-resolution is treated as an inverse problem, where it is assumed that LR frames are degraded versions of a HR image. One criteria of solving this inverse problem is minimizing the reconstruction error. In other words, the result which can produce the same low resolution image as the input one is preferred. The iterative back-projection (IBP) [25] algorithm is a classical and efficient method to obtain the HR image by minimizing the reconstruction error. It is state of the art method when magnification factor is 2. The original IBP method was designed to reconstruct the HR image from multiple LR inputs, but this research focuses on only one input image. It can minimize the reconstruction errors efficiently by an iterative process. However, it can produce many jaggy artifacts along the edges. Hence, focus is on improving the quality of the reconstructed image generated by the original IBP method and proposes a new super-resolution framework based on edge is preserved.

The quality of the initial interpolation image has a very critical impact on the final results, especially along the edges. Then this initial image is modified using bilateral filtering [26] to preserve true edges, since bilateral filtering can achieve edge-preserving image smoothing. The across-edge error propagation can be further reduced by learning non-local similarity by the structure of the pixels in LR image. This method is compared with bicubic interpolation [27], IBP and Non-local IBP. The reconstructed images can be evaluated using perceptual image quality assessment metrics (FSIM [28], SSIM [29]), PSNR and MSE.

1.7 Summary

This chapter gives a brief yet comprehensive overview of what super-resolution is. It also makes familiarize with some digital image processing concept mainly with the factors that induces degradation in images. It also gives a clear idea on why one requires super-resolution. This in turn produces the events that demand super-resolution. Hence some essential applications of super-resolution are also studied in this chapter. The next chapter will give a clearer idea about image super-resolution.

CHAPTER 2

OVERVIEW OF SUPER-RESOLUTION TECHNIQUES

2.1 Introduction

The idea of image super-resolution was first introduced by Tsai and Huang [2] in 1984. Super-resolution image reconstruction has widely been researched in the last two decades. Most of the researches has been carried out for combining multiple LR images of the same scene to reconstruct a single or more HR image(s) [2] [30] [31] [32] [33]. The basic principle underlying most of the aforementioned techniques is to take multiple images from the same object under similar lighting conditions but from slightly different sensor locations or orientations. When two images provide different views of the same object or landscape, the motion or motion vector field (a set of displacements of the pixel grid points between the images) is used to keep the grid points tied to their corresponding fixed locations on the viewed surface. The true motion between the images is not known, and must therefore be approximated that is a complicated and difficult task [34] [35]. Furthermore, most of the aforementioned SR techniques are computationally expensive. Another major limitation of these techniques is perhaps the requirement of multiple images that are often costly to procure and difficult to acquire at the same temporal instant, especially for remote sensing. Commonly, the improvement of spatial resolution of multi-frame SR algorithm always has to sacrifice the temporal resolution [2]. Other limitations include non-suitability of LR images for SR reconstruction; the application of SR algorithms is possible only if the images are sub-pixel shifted.

Alternatively, many researchers tackled the image fusion problem of reconstructing an LR image using a HR image. A typical example is the use of

panchromatic image for sharpening multi/hyper-spectral images [36] [37] [38] [39] [40] [41]. However, SR reconstruction and image fusion are little different. Image fusion combines one or several LR images with one or more HR images in order to obtain a useful final image with better spatial resolution than LR image. Therefore, fusion methods require the use of at least one HR image and the spatial resolution of their results is limited by that HR pixel-size. In contrast, SR algorithms do not use any HR image; they only depend on LR image(s).

Another category of super-resolution is called single image (frame) super-resolution where only one low resolution input is used to produce the high resolution image. However, most of the researchers of the existing literatures believe that the quality of a single LR image is limited; and interpolation based on an under-sampled image does not allow recovering the lost high-frequency information. Hence single LR image cannot be used for SR reconstruction and multiple observations of the same scene are needed. Typical single-frame SR construction techniques have been criticized widely as 'image enhancement' by means of image scaling, interpolation, zooming and enlargement [42]. Despite the criticisms, these approaches are preferred where multi-frame techniques are not applicable or affordable.

2.2 Image Interpolation

Literally the word interpolation means – "insert (an intermediate value or term) into a series by estimating or calculating it from surrounding known values," in New American Oxford Dictionary (NOAD). In signal processing, image interpolation refers to the technique of recovering a continuous signal by estimating image data from a set of discrete image data samples. Hence, it acts as a bridge between the continuous and the discrete domains. Image interpolation forms a fundamental base in image processing and is the heart of many computer vision applications such as medical imaging [18] [19],

target detection and recognition [24], and astronomical imaging [22]. Almost every image processing software implements some interpolation technique for transformations, rotations and many other manipulations performed on an image. It is very important that the interpolation techniques have a very low computational cost in terms of both, time and memory utilization since they are usually implemented at some intermediate step in a system. At the same time, it is necessary for the technique to yield good and accurate results, or else it could jeopardize the final solution. For instance, in the field of medical imaging, computed tomography (CT) or computed axial tomography (CAT) and magnetic resonance imaging (MRI) scan employ interpolation techniques during the registration process [19]. A slight error in the interpolated data could cause mis-registration thereby significantly affecting the accuracy of reconstruction of the final image which may lead to wrong diagnosis of a patient. It is therefore very important to choose a correct type of interpolation technique, depending upon the nature of its application, which provides the best trade-off between accuracy and computational cost.

2.2.1. Interpolation Properties

As per the Shannon-Nyquist sampling theorem [43] [2-7] [38] [39], a continuous signal (band-limited) can be completely recovered from its samples, if the sampling frequency is at least twice the highest frequency (Nyquist frequency) in the original signal. For a 1D case, let $f(x)$ be the continuous signal to be reconstructed from its samples $f_k(m)$, where k = 1, 2, …, p. The interpolation process in terms of convolution in the spatial domain can then be given as,

$$f(x) = f_k(m) * h(x) \quad \text{for k=1, 2, ..p} \tag{2.1}$$

where $h(x)$ is the interpolation reconstruction kernel. For image resampling, the interpolation step must reconstruct a two-dimensional (2-D) continuous signal $s(x, y)$ from its discrete samples $s(k, l)$ with $s, x, y \in \mathbb{R}$ and $k, l \in \mathbb{N}^0$. Thus, the amplitude at the

position (x, y) must be estimated from its discrete neighbors. This can be described formally as the convolution of the discrete image samples with the continuous 2-D impulse response ${}_{2D}h(x, y)$ of a 2-D reconstruction filter

$$s(x,y) = \sum_k \sum_l s(k,l) \cdot h(x-k, y-l) \tag{2.2}$$

In order for the interpolation technique to reconstruct the continuous signal from its discrete data samples, the kernel should be,

- Symmetric: $h(x) = h(-x)$ (2.3)
- Zero for all non-zero integers and one if its argument is zero. This rule ensures that the interpolation coefficients become the sampled data points.

$$h(x) = \begin{cases} 1, & if\ x = 0 \\ 0, & if\ x \neq 0 \end{cases} \tag{2.4}$$

- Separable in order to reduce computational cost. For example, a 2D interpolation kernel is given by $h(x, y)$, then they are separable as:

$$h(x, y) = h(x) \bullet h(y) \tag{2.5}$$

Therefore, using eq. 3.5 for 2D, eq. 3.1 can be rewritten as:

$$f(x,y) = f_k(m,n) * h(x,y) \qquad for\ k=1, 2, \ldots p \tag{2.6}$$

$$= ((f_k(m,n) *_x h(x)) *_y h(y))$$

where $*_x$ and $*_y$ denotes convolution in x and y direction respectively.

It is a fact that the type, size and shape of the kernel chosen for interpolation are major factors contributing to the reconstruction quality of the final image or signal. Generally, the size of the interpolation kernel is very crucial since it determines the computational cost of the system.

Figure 2.1 illustrates the interpolation of the point *(x, y)* in a 4 x 4 neighborhood. Interpolation is performed in the x direction first. The small grey intermediate points in Fig. 4 are generated by four one-dimensional (1-D) interpolations. They are used for the final 1-D interpolation in the y direction.

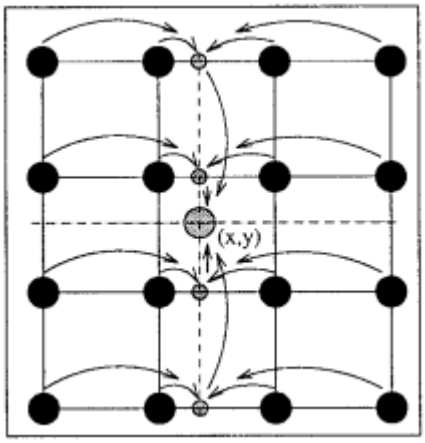


Figure 2.1 One-dimensional decomposition of the 2-D N x N interpolation of the point *(x,y)* [44].

The reason for resampling from a smaller to a larger matrix size is often to make an image more pleasing to a human viewer. In this case, the properties of the human visual system must be taken into account. Certain types of distortions will be much better tolerated by the observer than other distortions. For example, noise which is correlated with an image is much more noticeable than noise which is uncorrelated with the image. The property which is sought in the final image is not necessarily its mathematical similarity to the original scene, but rather the appearance of similarity, i.e., verisimilitude (the quality of seeming to be true). Considerable work has been done on interpolation for a human observer (for example,see Ratzel [45]). Often, the resampled images are produced for further processing by a computer. In this case, verisimilitude is not necessarily the best property. Rather, mathematical similarity is more desirable. The form of the mathematical similarity will depend on the processing which is to be performed.

Keys [27] emphasized similarity of the Taylor series expansion of the two signals. Alternately, it may be desirable for the interpolating function to have a flat frequency response.

2.2.2. Ideal Interpolation

Shanon sampling theorem [6] can be formally described as: for a uniformly sampled DSP system, an analog signal can be perfectly recovered as long as the sampling rate is at least twice as large as the highest-frequency component of the analog signal to be sampled. In other words, following two conditions must be true in order to get perfect reconstruction:

1. The signal must be bandlimited. This avoids spectra with in finite extent that are impossible to replicate without overlap.
2. The sampling frequency f_s must be greater than twice the maximum frequency f_{max} present in the signal. This minimum sampling frequency, known as the Nyquist rate, is the minimum distance between the spectra copies, each with bandwidth f_{max}.

The first condition merely ensures that a sufficiently large sampling frequency exists that can be used to separate replicated spectra from each other. Since all imaging systems impose a bandlimiting filter in the form of a point spread function, this condition is always satisfied for images captured through an optical system. Note that this does not apply to synthetic images, e.g., computer-generated imagery.

The second condition proves to be the most revealing statement about reconstruction. It answers the problem regarding the sufficiency of the data samples to exactly reconstruct the continuous input signal. It states that exact reconstruction is possible only when $f_s > f_{Nyquist}$ where $f_{Nyquist} = 2f_{max}$. Collectively, these two conclusions

about reconstruction form the central message of sampling theory, as pioneered by Claude Shannon in his landmark papers on the subject [6].

If the original function *f(x)* was discretized in accordance with the sampling theorem, then *f(x)* must have been "band limited"—it could not contain any signal components with frequencies higher than half the sampling frequency ω_s. This means that the reconstructed signal can only contain a limited set of frequencies and thus its trajectory between the discrete

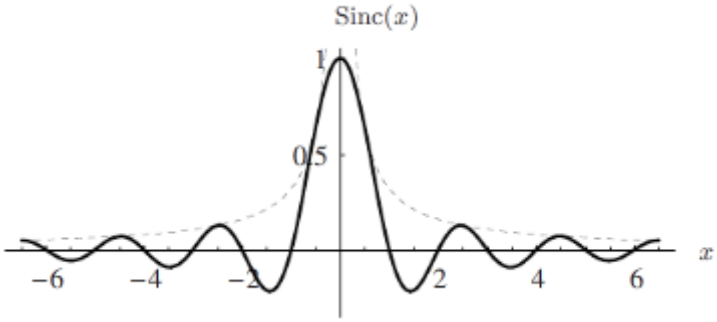


Figure 2.2 Sinc function in 1D.The function Sinc(x) has the value 1 at the origin and zero values at all integer positions. The dashed line plots the amplitude $\left|\frac{1}{\pi x}\right|$ of the underlying sine function [46].

sample values is not arbitrary but naturally constrained. In this context, absolute units of measure are of no concern since in a digital signal all frequencies relate to the sampling frequency. In particular, if $\tau_s = 1$ as the (unitless) sampling interval, the resulting sampling frequency is $\omega_s = 2\pi$ and thus the maximum signal frequency is $\omega_{max} = \omega_s/2 = \pi$. To isolate the frequency range $-\omega_{max} \ldots \omega_{max}$ in the corresponding (periodic) Fourier spectrum, multiply the spectrum $G(\omega)$ by a square windowing function $H_\pi(\omega)$ of width $\pm\omega_{max} = \pm\pi$,

$$\hat{G}(\omega) = G(\omega) \cdot H_\pi(\omega) \tag{2.7}$$

$$where\ H_{\pi}(\omega) = \begin{cases} 1 & for\ -\pi \leq \omega \leq \pi \\ 0 & otherwise \end{cases} \tag{2.8}$$

This is called an ideal low-pass filter, which cuts-off all signal components with frequencies greater than π and keeps all lower-frequency components un-changed. In the signal domain, the operation in Eqn. (2.7) corresponds to a linear convolution with the inverse Fourier transform of the windowing function $H_{\pi}(\omega)$, which is the Sinc function, defined as

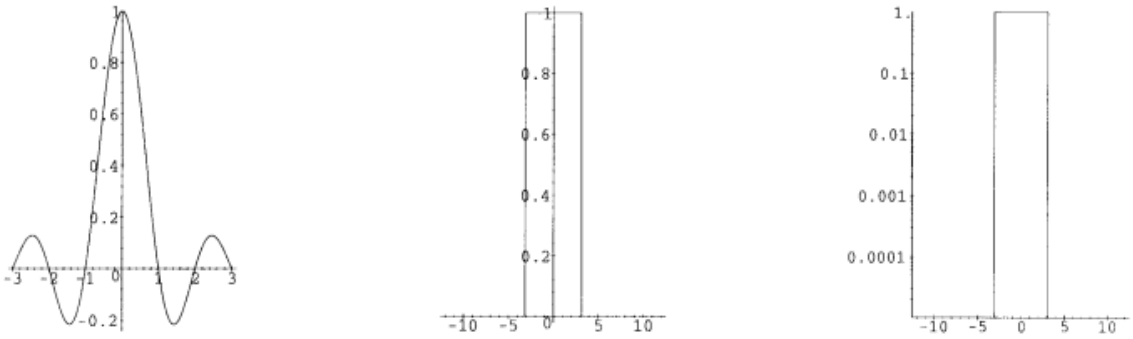
$$Sinc(x) = \begin{cases} 1 & for\ |x| = 0 \\ \dfrac{\sin(\pi x)}{\pi x} & for\ |x| > 0 \end{cases} \tag{2.9}$$

The sinc function as shown in fig. 2.2 is one instance of a large class of functions known as cardinal splines, which are interpolating functions de fined to pass through zero at all but one data sample, where they have a value of one. This allows them to compute a continuous function that passes through the uniformly spaced data samples.

Since multiplication in the frequency domain is identical to convolution in the spatial domain, $\mathrm{sinc}(x)$ represents the convolution kernel used to evaluate any point x on the continuous input curve g given only the sampled data g_s :

$$g(x) = sinc(x) * g_s(x) = \int_{-\infty}^{\infty} sinc(\lambda) g_s(x - \lambda) d\lambda \tag{2.10}$$

Equation 2.3 highlights an important impediment to the practical use of the ideal low-pass filter. The filter requires an in finite number of neighboring samples (i.e., an infinite filter support) in order to precisely compute the output points. The *Sinc* function is also known as the ideal reconstruction filter. The function is symmetric, *Sinc(x) = Sinc (-x)*. Also, the *Sinc* function is zero for all integer values of its argument except for zero.



(a) (b) (c)

Figure 2.3 Ideal interpolation [44]. (a) Kernel plotted for |x| < 3. (b) Magnitude of Fourier transform. (c) Logarithmic plot of magnitude.

2.3 Kernel Interpolation

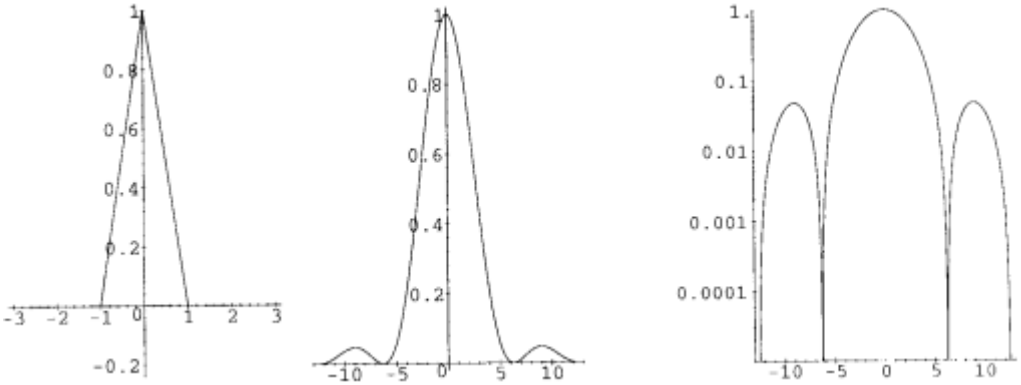
Simple single image resolution enhancement methods on smoothing and interpolation techniques for noise reduction have been commonly used in image processing. Smoothing can be achieved by applying various spatial filters such as Gaussian, Wiener and median filters. Commonly used interpolation kernels are linear, bicubic interpolation [27], cubic spline interpolation [47], and Lanczos [48]. Although the implementation of this approach is easy, the shortcomings of the kernel-based approach are apparent and can be attributed to the underlying model of the interpolated signal. Some of the well-known interpolation techniques that can be utilized as convolution kernel are described.

2.3.1. Linear Interpolation

This is one of the most popular interpolation techniques. Low complexity hence became very popular. The general expression can be given as:

$$f(x) = \begin{cases} 1 - |x| & for\ 0 \leq |x| < 1 \\ 0 & elsewhere \end{cases} \quad (2.11)$$

The values of both direct neighbors are weighted by their distance to the opposite point of interpolation. The triangular function $f(x)$ corresponds to modest low-pass filter in frequency domain as shown in the figure 2.4. As one can notice that the sidelobes in the stopband are below 10%, which still is considerable. Therefore, the main disadvantages of linear interpolation are both the attenuation of high frequency components and the aliasing of the data beyond the cut-off point into the low frequencies [49].



(a) (b) (c)

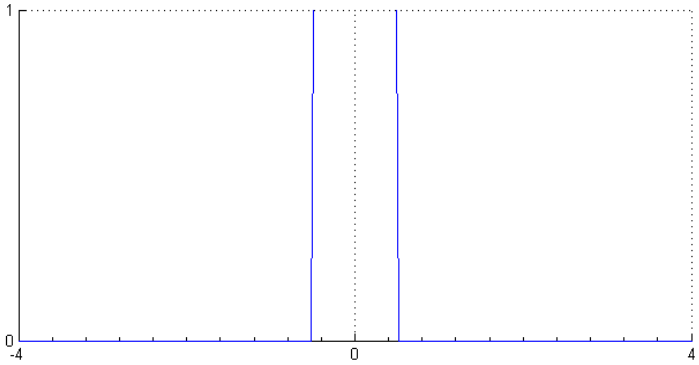
Figure 2.4 Linear interpolation [44]. (a) Kernel. (b) Magnitude of Fourier transform. (c) Logarithmic plot of magnitude.

2.3.2. Nearest-Neighbor Interpolation

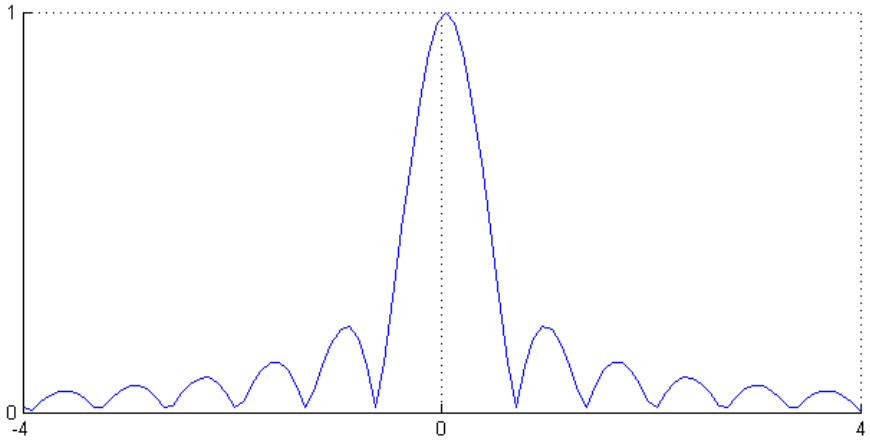
Nearest-neighbor technique simply uses the value from the nearest pixel. In terms of a convolution kernel this is a rectangular function, with width of one pixel. The general expression is given as

$$h_2(x) = \begin{cases} 1 & for\ -0.5 \leq |x| < 0.5 \\ 0 & elsewhere \end{cases} \quad (2.12)$$

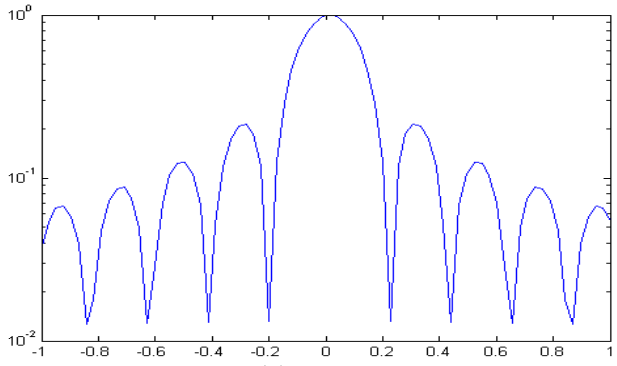
Fig. 2.5(b) shows that the Fourier spectrum of the nearest neighbor kernel equals the sinc function (expressed in the frequency domain). The logarithmical scale



(a)



(b)



(c)

Figure 2.5 Nearest-neighbor. (a) Kernel. (b) Magnitude of Fourier transform. (c) Logarithmic plot of magnitude.

shows prominent sidelobes in those regions of the frequency domain where the repetitions of caused by scanning should be suppressed [Fig. 2.5(c)]. The gain in the passband rapidly falls off to 64% at the cutoff point, and the amplitude of the side maxima is more than 20% [44]. Therefore, strong aliasing and blurring effects are associated with the nearest neighbor method for image interpolation. This is one of the most popular interpolation techniques. Low complexity hence became very popular.

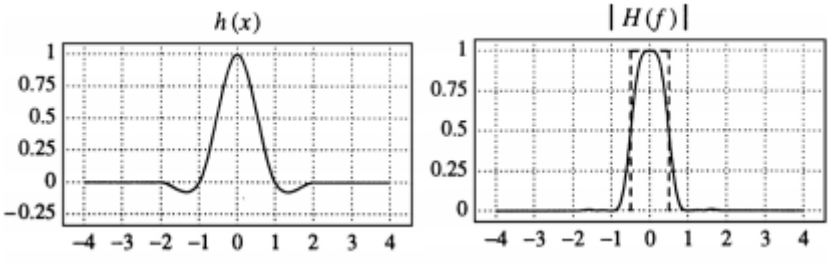
2.3.3. Cubic Interpolation

The cubic convolution kernel is a third-degree approximation to the sinc function [50]. It is symmetric, space-invariant, and composed of piecewise cubic polynomials:

$$h(x) = \begin{cases} (a+2)|x|^3 - (a+3)|x|^2 + 1, & 0 \le |x| < 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a, & 1 \le |x| < 2 \\ 0, & 2 \le |x| \end{cases} \quad (2.13)$$

where−3 < a < 0 is used to make h resemble the sinc function.

Of all the choices for a, the value −1 is preferable if visually enhanced results are desired. That is, the image is sharpened, making visual detail perceived more readily. However, the results are not mathematically precise, where precision is measured by the order of the Taylor series. To maximize this order, the value a =−0.5 is preferable [27]. A cubic convolution kernel with a =−0.5 is shown in figure 2.6.



(a) (b)

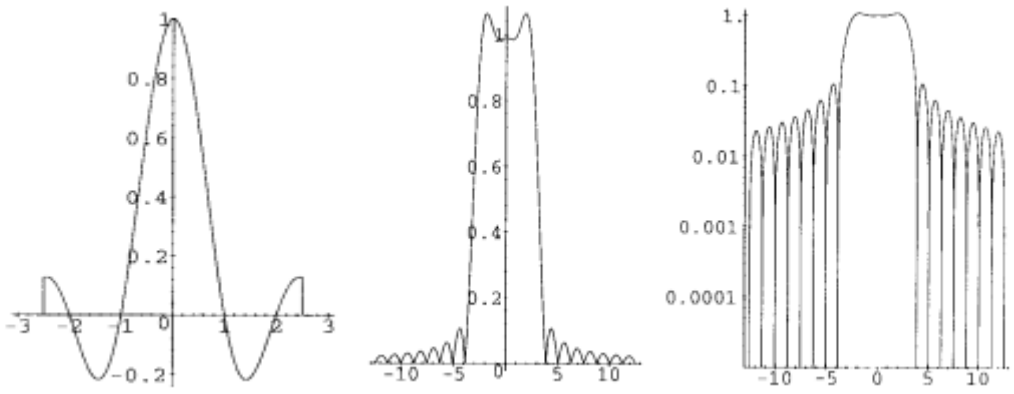
Figure 2.6 Cubic convolution [46]: (a) Kernel (a=-0.5). (b) Fourier transform

2.3.4. Windowed Sinc Interpolation

Sampling theory establishes that the sinc function is the ideal interpolation kernel. Since the Sinc filter is spatially unlimited and has an infinite impulse response (IIR) filter defined by a slowly converging infinite sum, the function is impracticable. To solve this problem, the Sinc function can be multiplied by a function which is non-zero in a finite range. This function is referred to as the Window Function $w(x)$ and the Sinc is then known as Windowed Sinc Function.

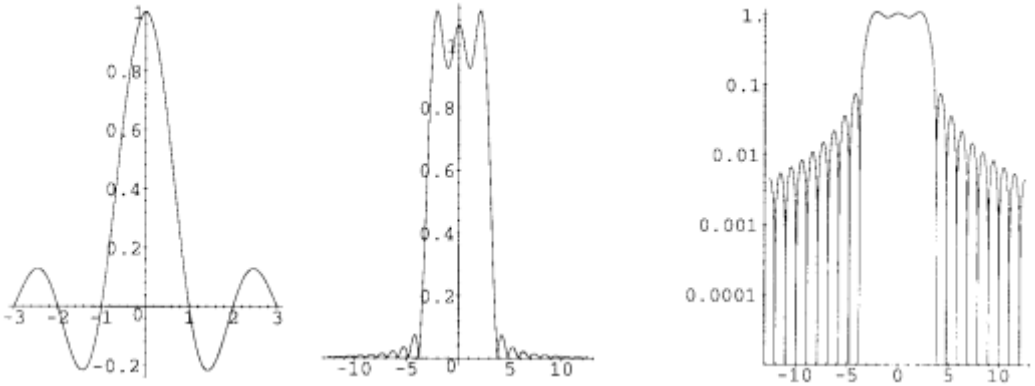
$$Sinc_{h_N} = \begin{cases} h_{ideal}(x) \cdot w(x), & 0 \leq |x| < N/2 \\ 0, & elsewhere \end{cases} \tag{2.14}$$

The results of this operation are predicted by sampling theory, which demonstrates that truncation in one domain leads to ringing in the other domain. Truncation is equivalent to the multiplication of with a rectangular function in the spatial domain, which is tantamount to a convolution with a sinc function in the frequency domain. Therefore, truncations of the ideal interpolator produce ringing effects in the frequency domain because a considerable amount of energy is discarded. Figures 2.7 and 2.8 demonstrate this effect, which also is referred to as the Gibbs's phenomenon [47], produced by a truncated sinc function with $N=5$ and $N=6$ supporting points, respectively.



(a) (b) (c)

Figure 2.7 Truncated sinc interpolation, $N = 5$ [44]. (a) Kernel. (b) Magnitude of Fourier transform. (c) Logarithmic plot of magnitude



(a) (b) (c)

Figure 2.8 Truncated sinc interpolation, $N = 6$ [44]. (a) Kernel. (b) Magnitude of Fourier transform. (c) Logarithmic plot of magnitude

2.3.5. Drawbacks of Kernel Interpolation

These three shortcomings are identified next and will have a central role in the subjective evaluation.

A common problem with kernel super-resolution is that it smoothes the image data in discontinuous regions, producing a larger image which appears rather blurred. Kernel filters typically perform very well in smooth areas, but not in edge areas [51]. Figure 2.9 shows a graphical image at the top left and at the bottom left the same image after decimation and super-resolution by one octave (a factor of two) using a linear kernel. The intensities of a 10 pixel part are displayed in the graph on the right of the image. The darker line represents the original image, while the lighter line represents the super-resolved version. It is clear that the originally steep edge has become less steep in the super-resolved image, which is visible as edge blurring.

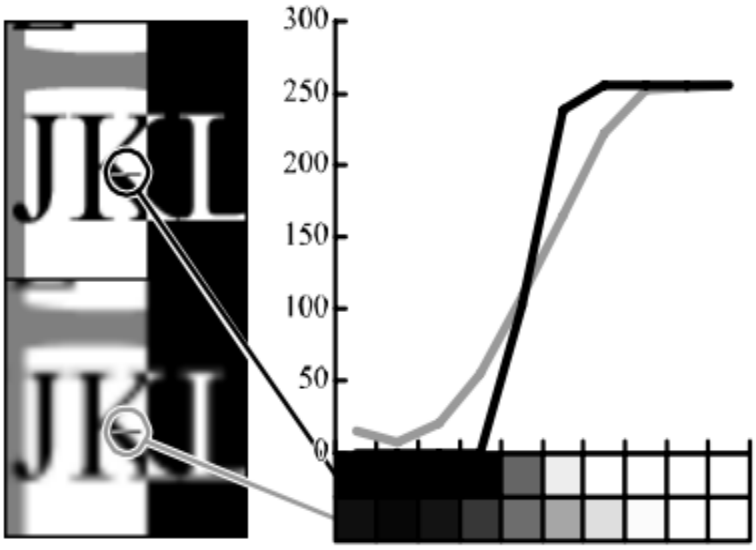
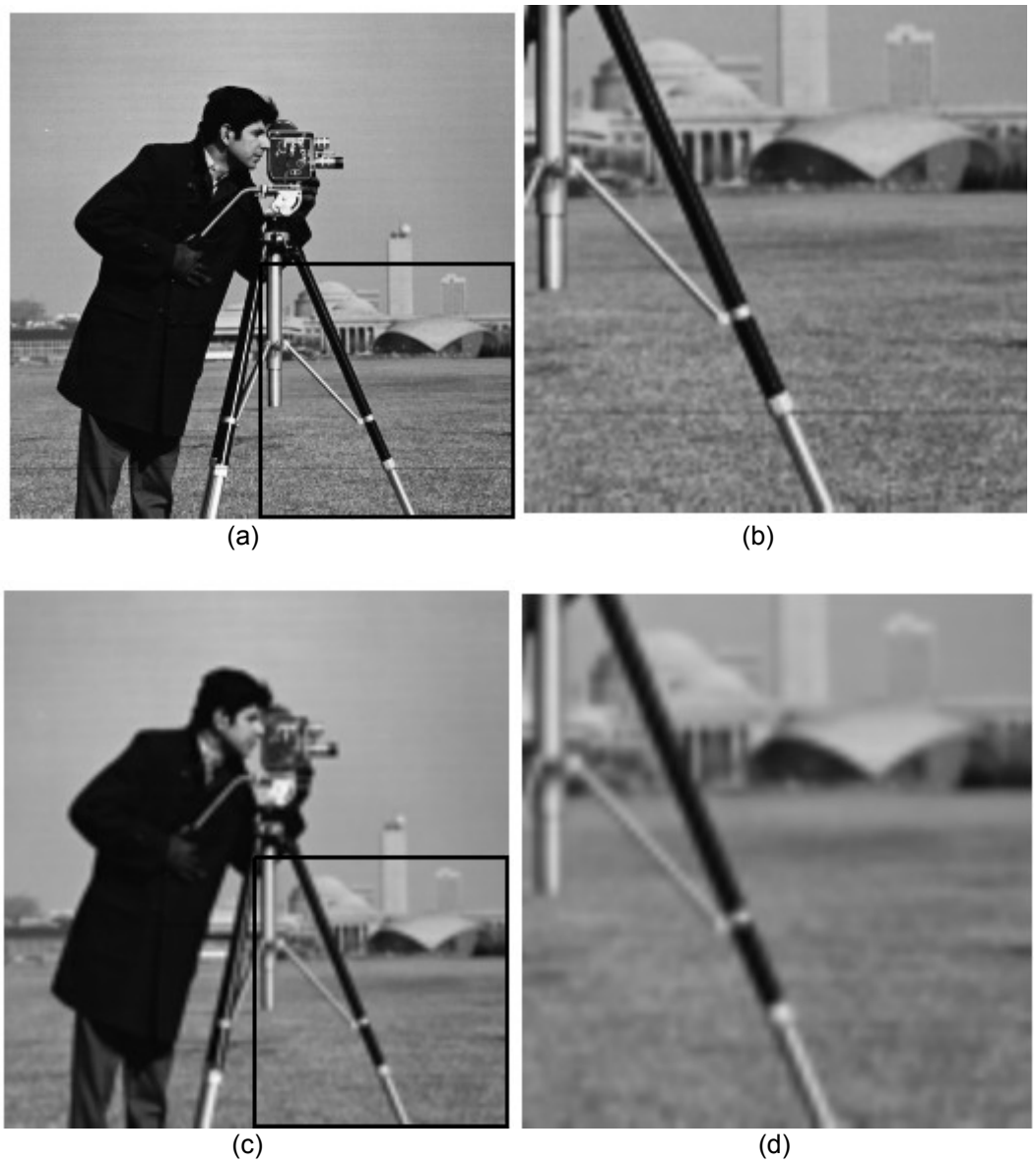


Figure 2.9 Original image and result (in form of graph) after decimation and bilinear super-resolution by one octave (factor of two) [51].

Cubic spline interpolators also tend to overshoot sharp discontinuities, producing a ringing effect on the edges. Hence it introduces blocking artifacts in diagonal edges or lines as shown in fig. 2.10. The diagonal line in this image not only looks blurry, but a staircase pattern is also emerging in the bilinearly super-resolved image. This staircase pattern or blocking artifacts is caused by the horizontal and vertical orientation of the

resampling kernels. Kernel super-resolution is unable to recognize or follow diagonal lines, which causes blocking.



(a) (b)

(c) (d)

Figure 2.10 Example of blocking artifacts in images. (a) Original image (b) zoomed up portion of image. (c) Result after decimation and bilinear super-resolution by one octave. (d) Result after decimation and bilinear super-resolution by one octave of the portion of image.

The third problem is the inability to generate high frequency components or fine detail. This is needed to make the super-resolved image look more plausible. The original and super-resolved images are taken through low pass and high pass filters to show the higher deterioration in the high pass signal compared to the low pass signal.

2.4 Single Image SR Algorithms

In this section some of the works specifically on single image super-resolution is reviewed, which is the topic of this thesis. Early efforts on super-resolution deal with the property of analytic continuation of a signal. Essentially, these techniques derive the missing high frequency components from a portion of the entire spectrum. This process sometimes also referred as spectral extension. Harris [52] established 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 free of noise. Furthermore, if even a small amount of noise is present this entire method becomes highly unreliable. A new view of the problem of continuing a given segment of the spectrum of a finite object is presented in [53]. Here the signal extrapolation is carried out by the method of alternate projections [54], iterating alternately between time and frequency domains. This method relies on the notion of reducing the 'error energy'. A dual of same problem is solved by Papoulis [55] where in the spectrum of the bandlimited object is recovered from a finite segment of the object using an iterative procedure.

Recently, many learning-based methods have been proposed for single image super-resolution reconstruction. Most of these techniques make use of a set of training images to learn the best features for the given low resolution image in order to produce a high resolution image. Freeman et al. approached super-resolution from a low level vision

learning perspective. An approach to low level vision tasks using belief propagation is presented in [56]. In this a learning framework called VISTA – Vision by Image/Scene TrAining was proposed. By blurring and down-sampling sharply defined images, a training set of sharp and corresponding blurred images was constructed. This scene underlying the supplied image data is estimated using a Markov network. To make the estimation feasible, both the image data and the scene are separated into patches. This approach to low level vision is specifically applied to super-resolution in [57]. It uses the high frequency part of the low resolution image as ground truth and the high frequency part of the high resolution image as scene to be estimated. A faster version of the algorithm that only uses a single pass is introduced in [58]. But these methods are somewhat dependent on the training set and hence the results are not stable and sometimes produce artifacts in real applications.

Atkins et al. [59] approached super-resolution using pixel classification. Pixel classification aims to sort pixels into classes like horizontal edges and smooth areas. In this paper, a tree-based classification approach to super-resolution is introduced. This algorithm builds a decision tree with a linear interpolator at each leaf of the tree. Each non-leaf node in the tree represents a binary choice. In [59] a similar algorithm is introduced, which assigns a pixel to one or more classes in a single step instead of using a set of binary choices.

Battiato et al. [60] have published papers on several super-resolution algorithms. In this rule based super-resolution approach called locally adaptive zooming algorithm (LAZA) is described. In LAZA, the authors use simple rules and configurable thresholds to detect edges and update the interpolation process accordingly. In [61] the same authors introduce an algorithm that incorporates anisotropic diffusion like the smart interpolation by anisotropic diffusion (SIAD) to sharpen edges.

Muresan and Parks published several papers [62], [63], [64] on super-resolution based on the optimal recovery principle. The authors model the image as belonging to a certain ellipsoidal signal class. Together with Kinebuchi [65] a wavelet-based algorithm using hidden Markov trees was introduced. It uses lower frequency wavelet coefficients to predict the highest frequency coefficients. By applying an inverse wavelet transform after prediction, a one octave super-resolved image results.

Su and Willis [66] present super-resolution by triangulation on pixel level. Yu et al. [67] present super-resolution by data-dependant triangulation (DDT). These methods use linear interpolation that is not generally aligned with the coordinate axes to reduce visible artifacts caused by this alignment. It is shown by the last authors that results can be further improved by improving the algorithm that searches for the optimal triangulation of the source image.

Another approach that is explicitly directed at maintaining sharp edges is the use of subpixel edge localization by Jensen and Anastassiou [68]. This approach detects the most prominent edge in the local window with subpixel precision and uses the resulting edge template to obtain sharper edges in super-resolved images.

There are also a range of commercial products available that rely on an algorithm more advanced than kernel-based resampling. Some worth mentioning are PhotoZoom Professional by BenVista [69], Imagener by Kneson Software [70], Qimage by Digital Domain [71], and SmartScale by Extensis [72]. The Pictura software makes use of a modified version of the algorithm presented in [62] by Muresan and Parks. The PhotoZoom Professional software was previously known as S-Spline by Shortcut, but has gone through some changes in name and company.

2.5 Summary

In this chapter, the basics of interpolation techniques were discussed. It was also learnt that why this simple and robust kernel interpolation do not produce desirable images. In addition discussed are some of the earlier super-resolution techniques to some of the newer methods. In the next chapter the integral components on this proposed framework for super-resolution are explained in detail.

CHAPTER 3

INTEGRAL PARTS

3.1 Bilateral Filtering

3.1.1. Introduction

Filtering is perhaps the most fundamental operation of image processing and computer vision. In the broadest sense of the term "filtering", the value of the filtered image at a given location is a function of the values of the input image in a small neighborhood of the same location. For example, Gaussian low-pass filtering computes a weighted average of pixel values in the neighborhood, in which the weights decrease with distance from the neighborhood center [1]. Although formal and quantitative explanations of this weight fall-off can be given, the intuition is that images typically vary slowly over space, so near pixels are likely to have similar values, and it is therefore appropriate to average them together. The noise values that corrupt these nearby pixels are mutually less correlated than the signal values, so noise is averaged away while signal is preserved. The assumption of slow spatial variations fails at edges, which are consequently blurred by linear low-pass filtering. How can one prevent averaging across edges, while still averaging within the smooth regions? Many efforts have been devoted to reducing this undesired effect [73] [74] [75] [76] [77] [78] [79]. Bilateral filtering is a simple, non-iterative scheme for edge-preserving smoothing.

3.1.2. Concept

The basic idea underlying bilateral filtering is to apply the range of an image what traditional filters implement in its domain. Two pixels can be close to one another, that is, occupy nearby spatial location, or they can be similar to one another, that is, have nearby

values, possibly in a perceptually meaningful fashion. Closeness refers to vicinity in the domain, similarity to vicinity in the range. Traditional filtering is domain filtering, and enforces closeness by weighing pixel values with coefficients that fall off with distance. Similarly, range filtering can be viewed as a filtering method, which averages image values with weights that decay with dissimilarity [26]. Range filters are nonlinear because their weights depend on image intensity or color. Computationally, they are no more complex than standard non-separable filters. Most importantly, they preserve edges.

Spatial locality is still a predominant factor. In fact, range filter by itself merely distorts an image's color map. Hence, the combination of range and domain filtering is worth noticing. This combined filtering is denoted as bilateral filtering. Consider a shift-invariant low-pass domain filter applied to an image $\mathbf{f}(x)$ [26]:

$$\boldsymbol{h}(\boldsymbol{x}) = k_d^{-1}(x) \iint_{-\infty}^{\infty} \boldsymbol{f}(\xi) c(\xi - \boldsymbol{x}) d\xi \tag{3.1}$$

where $c(\xi, x)$ measures the geometric closeness between the neighborhood center $\mathbf{x}$ and a nearby point $\boldsymbol{\xi}$. The bold font for **f** and **h** emphasizes the fact that both input and output images may be multi-band. In order to preserve the DC component, k_d must be

$$k_d(x) = \iint_{-\infty}^{\infty} c(\xi, x) d\xi \tag{3.2}$$

If the filter is shift-invariant, $c(\xi, x)$ is only a function of vector difference $\xi - x$, and k_d is constant.

Range filtering is similarly defined:

$$h(x) = k_r^{-1}(x) \iint_{-\infty}^{\infty} f(\xi) s\big(f(\xi) - f(x)\big) d\xi \tag{3.3}$$

except that now $s(f(\xi), f(x))$ measures the photometric similarity between the pixel at the neighborhood center x and that of a nearby point ξ. Thus, the similarity function s operates in the range of the image function f, while the closeness function c operates in the domain of f. The normalization constant k_r in this case is

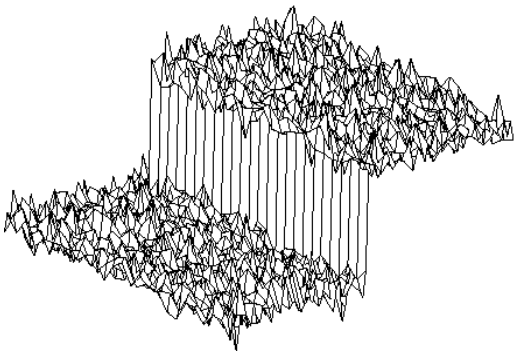
$$k_r(x) = \int_{-\infty}^{\infty} s\big(f(\xi) - f(x)\big) d\xi \tag{3.4}$$

The spatial distribution of image intensities plays no role in range filtering taken by itself. Combining intensities from the entire image, however, makes little sense, since the distribution of image values far away from **x** ought not to affect the final value at **x**. In addition, one can show that range filtering without domain filtering merely changes the color map of an image, and is therefore of little use. The appropriate solution is to combine domain and range filtering, thereby enforcing both geometric and photometric locality. Combined filtering can be described as follows [26]:

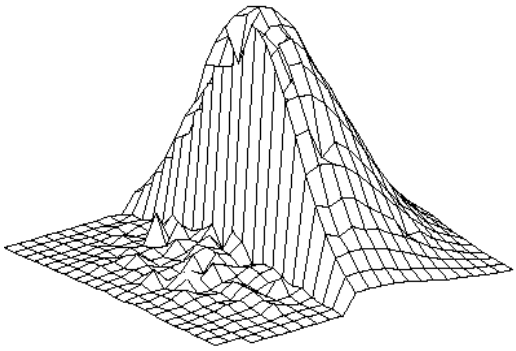
$$h(x) = k^{-1}(x) \iint_{-\infty}^{\infty} f(\xi) c(\xi - x) s\big(f(\xi) - f(x)\big) d\xi \tag{3.5}$$

with the normalization

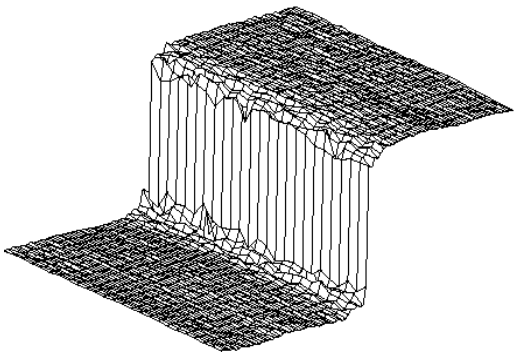
$$k(x) = \int_{-\infty}^{\infty} c(\xi - x) s\big(f(\xi) - f(x)\big) d\xi \tag{3.6}$$



(a)



(b)



(c)

Figure 3.1 (a) A 100-gray-level step perturbed by Gaussian noise with σ = 10 gray levels. (b) Combined similarity weights $c(\xi, x)s(f(\xi), f(x))$ for a 23 x 23 neighborhood centered two pixels to the right of the step in (a). The range component effectively suppresses the pixels on the dark side. (c) The step in (a) after bilateral filtering with σ_r = 50 gray levels and σ_d = 5 pixels [26].

Combined domain and range filtering is denoted as *bilateral filtering*. It replaces the pixel value at **x** with an average of similar and nearby pixel values. In smooth regions, pixel values in a small neighborhood are similar to one another, and the bilateral filter acts essentially as a standard domain filter, averaging away the small, weakly correlated differences between pixel values caused by noise [26]. Consider now a sharp boundary between a dark and a bright region, as in figure 3.1(a).

When the bilateral filter is centered, say, on a pixel on the bright side of the boundary, the similarity function *s* assumes values close to one for pixels on the same side, and values close to zero for pixels on the dark side. The similarity function is shown in figure 3.1(b) for a 23x23 filter support centered two pixels to the right of the step in figure 3.1(a). The normalization term ***k(x)*** ensures that the weights for all the pixels add up to one. As a result, the filter replaces the bright pixel at the center by an average of the bright pixels in its vicinity, and essentially ignores the dark pixels. Conversely, when the filter is centered on a dark pixel, the bright pixels are ignored instead. Thus, as shown in figure 3.1(c) with σ_r = 50 and σ_d = 5, good filtering behavior is achieved at the boundaries, thanks to the domain component of the filter, and crisp edges are preserved at the same time, thanks to the range component.

A simple and important case of bilateral filtering is shift-invariant Gaussian filtering, in which both the closeness function *c* and the similarity function *s* are Gaussian functions of the Euclidean distance between their arguments. More specifically, *c* is radially symmetric [26]:

$$c\left(\xi - \mathbf{x}\right) = e^{-\frac{1}{2}\left(\frac{d\left(\xi - \mathbf{x}\right)}{\sigma_d}\right)^2} \tag{3.7}$$

where

$$d(\xi - \mathbf{x}) = \|\xi - \mathbf{x}\| \tag{3.8}$$

is the Euclidean distance [80]. The similarity function *s* is perfectly analogous to *c* :

$$s(\xi - \mathbf{x}) = e^{-\frac{1}{2}\left(\frac{\delta(\mathbf{f}(\xi) - \mathbf{f}(\mathbf{x}))}{\sigma_r}\right)^2} \tag{3.9}$$

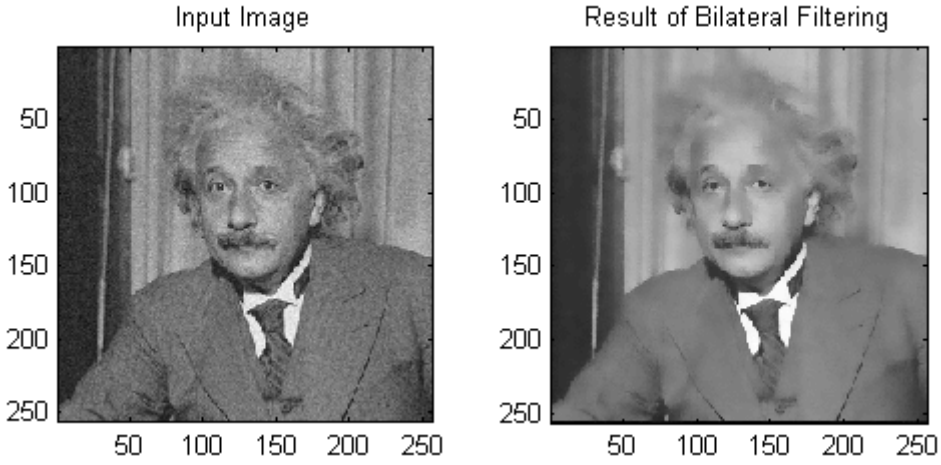
where

$$\delta(\mathbf{f}(\xi) - \mathbf{f}(\mathbf{x})) = |\mathbf{f}(\xi) - \mathbf{f}(\mathbf{x})| \tag{3.10}$$

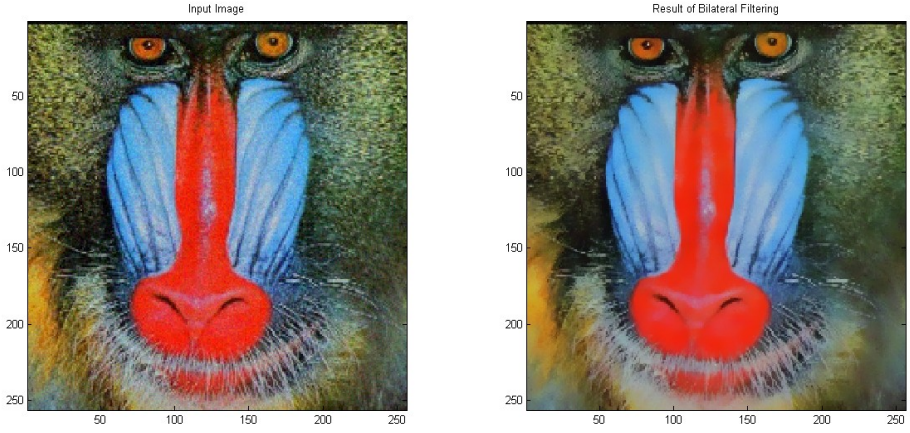
is a suitable measure of distance in the intensity space. In the scalar case, this may be simply the absolute difference of the pixel difference or, since noise increases with image intensity, an intensity-dependent version of it. Just as this form of domain filtering is shift-invariant, the Gaussian range filter introduced above is insensitive to overall additive changes of image intensity. Of course, the range filter is shift-invariant as well.

The geometric spread σ_d in the domain is chosen based on the desired amount of low-pass filtering. A large σ_d blurs more, that is, it combines values from more distant image locations. Also, if an image is scaled up or down, σ_d must be adjusted accordingly in order to obtain equivalent results. Similarly, the photometric spread σ_r in the image range is set to achieve the desired amount of combination of pixel values. Loosely speaking, pixels with values much closer to each other than σ_r are mixed together and values much more distant than σ_r are not. If the image is amplified or attenuated, σ_r must be adjusted accordingly in order to leave the results unchanged [26].

Fig 3.2 (a) and Fig 3.2 (b) show the effectivenesss of bilateral filtering on grayscale images and color images by introducing Additive White Gaussian Noise (AWGN) to the input images. An interesting application of bilateral filtering is shown in Fig. 3.3 as given in [81].



(a)



(b)

Figure 3.2 Gray-scale original image with AWGN (left panel) and bilateral filtered image(right panel) with σ_r = 0.1 gray levels and σ_d = 3 pixels (b) (b) Color original image with AWGN (left panel) and bilateral filtered image(right panel) with σ_r = 0.1 gray levels and σ_d = 3 pixels

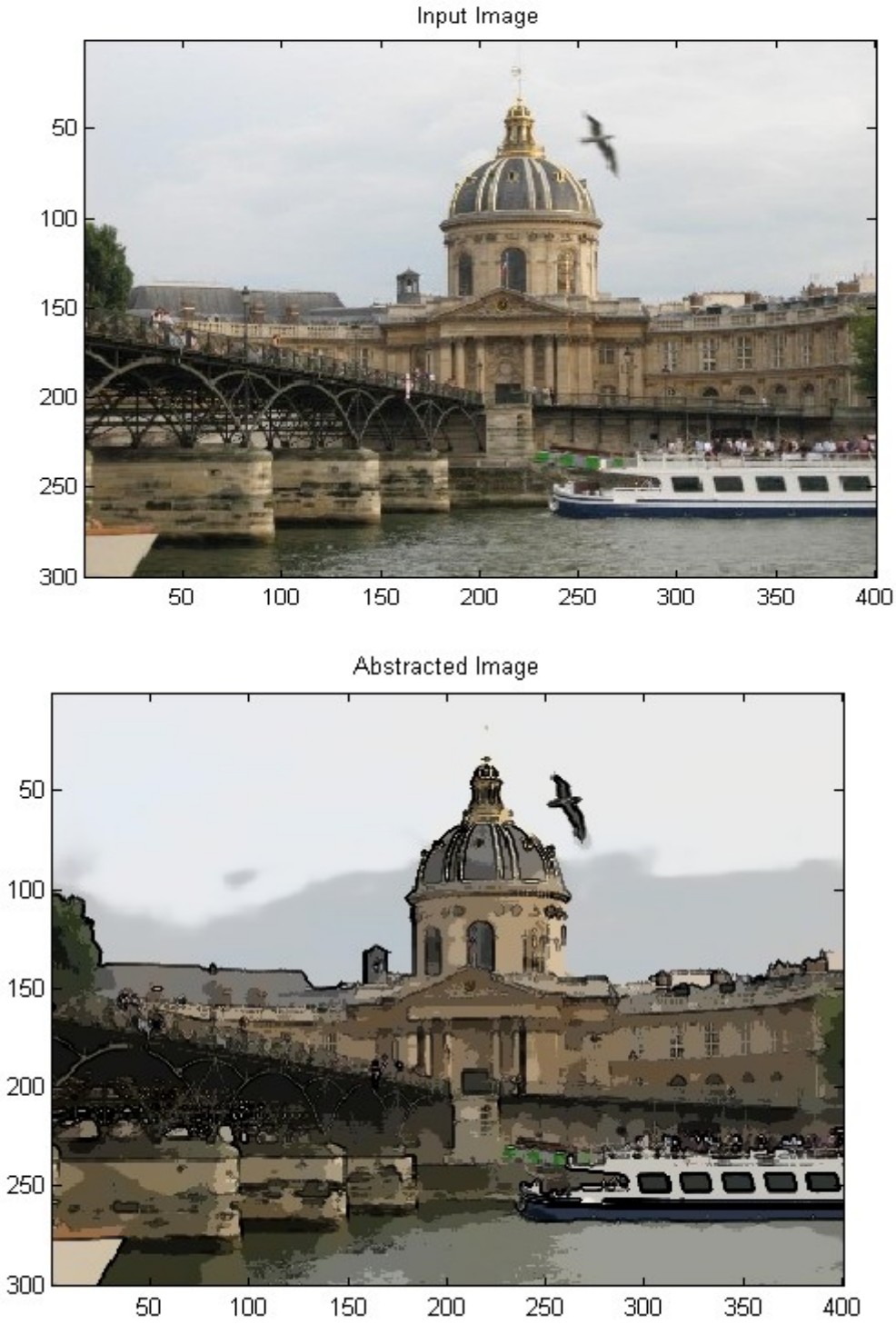


Figure 3.3 CARTOON Image abstraction using bilateral filtering [81].

3.2 Mean Shift Image Segmentation

3.2.1. Introduction

Image segmentation plays a crucial role in various image processing applications in several domains, including industrial as well as medical applications [82]. It describes the task of partitioning an image into several segments or regions. In other words, it is the decomposition of a gray level or color image into homogeneous tiles. This is arguably most important low-level vision task. Common segmentation approaches include simple thresholding techniques [83], graph-based methods [84], and level set techniques [85] among others. They have been applied to images from different imaging modalities in typically two or three dimensions, e.g., gray/color images, high dynamic range images, CT/MR datasets, and multispectral images. In general, these methods are adapted to the specific application [86] [87] [88] [89].

Mean shift is an unsupervised clustering algorithm [90], which estimates the gradient of a probability density function to detect modes in an iterative fashion. Hence, image segmentations that take color/intensity-similarity as well as local connectivity into account, can be obtained by applying this algorithm to the combined spatial-range domain [91]. Mean shift segmentation has been successfully applied to several applications [91] [92] [93].

The basic processing of mean shift image segmentation relies on the joint domain. The description for the joint domain is as follows:

An image is typically represented as a two-dimensional lattice of p-dimensional vectors (pixels), where p=1 in the gray-level case, three for color images, and $p > 3$ in the multispectral case. The space of the lattice is known as spatial domain, while the gray level, color or spectral information is represented in range domain. The concatenation of location vectors and range vectors gives rise to joint spatial-range domain of dimension

$d=p+2$. Now this concatenation must be compensated by a proper normalization. Thus, the multivariate kernel is defined as the product of two symmetric kernels and the Euclidean metric allows a single bandwidth parameter for each domain [94]:

$$K_{h_s h_r}(x) = \frac{C}{h_s^2 h_r^2} v\left(\left\|\frac{x^s}{h_s}\right\|^2\right) v\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \tag{3.11}$$

where x^s is the spatial part, x^r is the range part of the feature vector, $v(x)$ the common profile in both the two domains, h_s and h_r the employed kernel bandwidths, and C the corresponding normalization constant. In practice, an Epanechnikov [95] or a (truncated) normal kernel always provides satisfactory performance, so the user only has to set the bandwidth parameter $\mathbf{h} = (h_s, h_r)$, which, by controlling the size of the kernel, determines the resolution of the mode detection [94].

3.2.2. Edge preserving smoothing

Smoothing through replacing the pixel in the window by the (weighted) average of the pixels in the window indiscriminately blurs the image, removing not only the noise but also salient information. Discontinuity preserving smoothing techniques, on the other hand, adaptively reduce the amount of smoothing near abrupt changes in the local structure, i.e. edges.

Many approaches have been implemented to achieve this goal, form adaptive Wiener filtering [31], to implementing isotropic [50] and anisotropic [44] local diffusion processes. The diffusion based techniques, however, do not have a straightforward stopping criterion and, after a sufficiently large number of iterations, the processed image

collapses into a flat surface. The connection between anisotropic diffusion and M-estimators is analyzed in [5].

The bilateral filters also work in the joint spatial-range domain. The data is independently weighted in the two domains and the center pixel is computed as the weighted average of the window. The fundamental difference between the bilateral filtering and the mean shift-based smoothing algorithm is in the use of local information [94].

3.2.3. Mean Shift Filtering

Mean shift filtering is a data clustering algorithm commonly used in computer vision and image processing. For each pixel of an image (having a spatial location and a particular color), the set of neighboring pixels (within a spatial radius and a defined color distance) is determined. For this set of neighbor pixels, the new spatial center (spatial mean) and the new color mean value are calculated. These calculated mean values will serve as the new center for the next iteration. The described procedure will be iterated until the spatial and the color (or grayscale) mean stops changing. At the end of the iteration, the final mean color will be assigned to the starting position of that iteration.

The kernel (window) in the mean shift procedure moves in the direction of the maximum increase in the joint density gradient, while the bilateral filtering uses a fixed, static window. In the image smoothed by mean shift filtering, information beyond the individual windows is also taken into account [94].

Mean shift filtering with uniform kernel having (hs, hr) = (8,4) has been applied to the often used 256x256 gray-level cameraman image (Fig. 3.4(a)), the result being shown in Fig. 3.4(b). The regions containing the grass field have been almost completely

smoothed, while details such as the tripod and the buildings in the background are preserved.

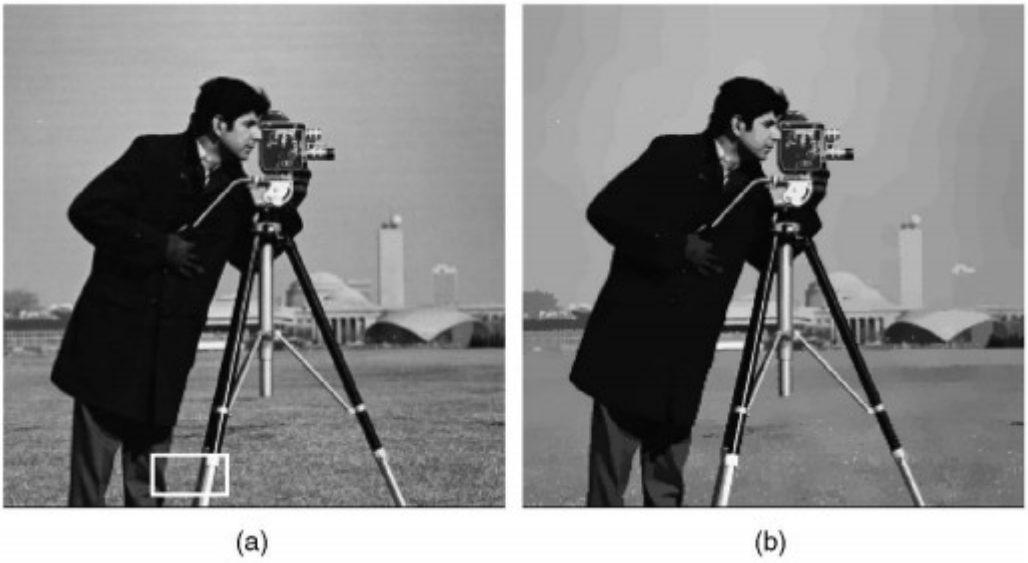


Figure 3.4 Cameraman image. (a) Original (b) Mean shift filtered $(h_s,h_r) = (8,4)$.

3.3 Shock filter

3.3.1. Introduction

In the past decade there has been a growing amount of research concerning partial differential equations in the fields of computer vision and image processing [96]. Applications, supported by rigorous theory, were developed for purposes such as image denoising and enhancement, segmentation, object tracking and many more [97]. The research is focused mostly on linear and nonlinear parabolic schemes of diffusion-type processes. Osher and Rudin [98] proposed a hyperbolic equation called shock filter that can serve as a stable deblurring algorithm approximating deconvolution.

The basic idea behind shock filters is the process of applying either erosion or dilation in a much localized manner, in order to create a “shock” between two influence

zones, one belonging to a maximum and the other to a minimum of the signal. By iterating this process (modeled using a PDE framework) according to a small time increment dt, one can ultimately obtain a piecewise constant segmentation of the input image, thus leading to a deblurred output.

The use of shock filters as a means of image enhancement is recommended by the advantages this particular method offers: they create strong discontinuities at image edges and furthermore, the filtered signal within a region delineated by those edges becomes flat. In other words, shock filters create segmentation. Due to their discrete mathematical definition they are inherently unstable, meaning that they require special discretization schemes in order to preserve the total variation of the signal. Another property of shock filters underlined in [99] is that they satisfy the maximum-minimum principle which states that the range of the filtered image remains within the range of the original image. Another advantage of shock filters over other image enhancement methods, such as Fourier or wavelet-based ones is that phenomena like the Gibbs phenomenon cannot appear [100].

3.3.2. Concept

The first definition of the shock filter can be traced back to 1975 when Kramer and Bruckner have defined the first concepts regarding shock filter theory [101]. The formulation of the shock filter equation is:

$$I_t = -|I_x| F(I_{xx}), \tag{3.12}$$

where F should satisfy F(0)=0, F(s)sign(s) ≥0. Choosing F(s) = sign(s) gives the classical shock filter equation as per Kramer and Bruckner definition [102]:

$$I_t = -\text{sign}(I_{xx})|I_x| \tag{3.13}$$

where $\mathbf{I}$ represents the image and $\mathbf{I}_x$ and $\mathbf{I}_{xx}$ represent the first, respectively the second directional derivatives of the image $\mathbf{I}$. Equation 3.13 represents a generic definition since the direction x is not properly defined. The first term of the equation represents the edge detector (in this case the Canny edge detector [103]) used for shock filter steering [100]. The actual term of shock filter was introduced in 1990 by Osher and Rudin [98] when they proposed this new class of filters based on PDEs and defined the minmod numerical scheme for successfully avoiding any instabilities of the algorithm, since the shock filter theory is defined only on a discrete domain. The minmod function can be given as:

$$m(x,y) \doteq \begin{cases} (\text{sign}x)\min(|x|,|y|) & \text{if } xy > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3.14}$$

The shock filter main properties are [104]:

- Shocks develop at inflection points (second derivative zero-crossings).
- Local extrema remain unchanged in time. No new local extrema are created. The scheme is total variation preserving (TVP).
- The steady state (weak) solution is piece-wise constant (with discontinuities at the inflection points of I_0).
- The process approximates deconvolution.

As noted in [98], any noise in the blurred signal will also be enhanced. As a matter of fact this process is extremely sensitive to noise. Theoretically, in the continuous domain, any white noise added to the signal may add an infinite number of inflection points, disrupting the process completely. The performance of the shock filter with and without noise can be compared in Fig. 3.5. Clearly the signal in the noisy case is not enhanced and the process results mainly in noise amplification.

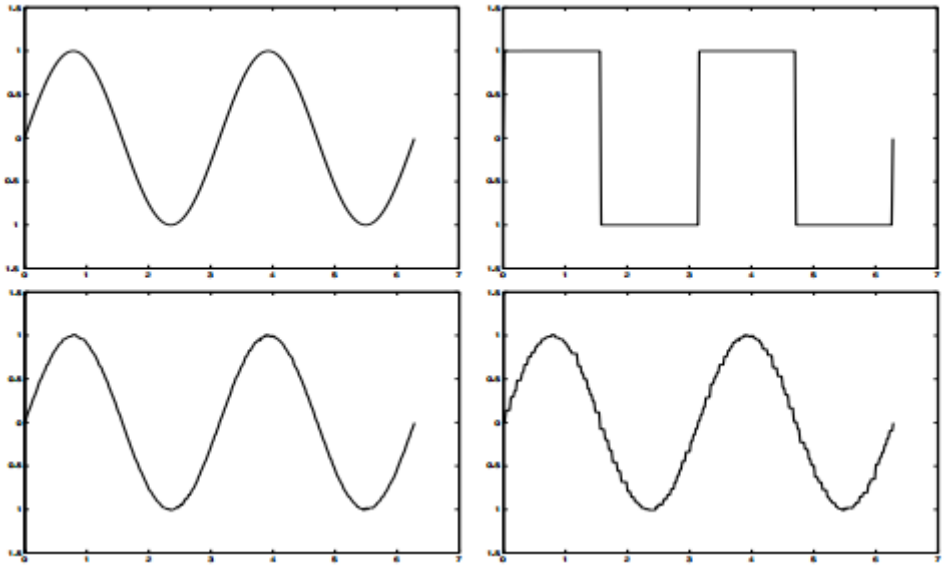


Figure 3.5 Signal (sine wave) and its steady state shock filter solution without noise (top) and with very low additive white Gaussian noise, SNR=40dB (bottom) [104]

3.3.3. Previous Works

The noise sensitivity problem is critical and unless properly solved - might prevent most practical uses of shock filters. Previous studies addressed the issue suggesting several solutions. The common way seen in literature to increase robustness ([105] - [108]) is to convolve the signal's second derivative with a lowpass filter, such as a Gaussian:

$$I_t = \text{-sign}(G_\sigma * I_{xx})|I_x| \tag{3.15}$$

where G_σ is a Gaussian of standard deviation σ.

This is generally not sufficient to overcome the noise problem: convolving the signal with a Gaussian of moderate width will in many cases not cancel the inflection points produced by the noise. Their magnitude will be considerably lower, but there will still be a change of sign at these points, which will lead the flow to go in opposite direction at each side. For very wide (large scale) Gaussians - most inflection points produced by

the noise are diminished, but at a cost: the locations of the signal's inflection points are less accurate. Moreover, the effective Gaussian's width σ is in many cases larger than the length of the signal, thus causing the boundary conditions imposed on the process to strongly affect the solution. Lastly, from a computational point of view, the convolution process at each iteration is costly [104].

If the issue is addressed as an enhancing-denoising problem, one can devise an approach in which smoother parts are denoised keeping edges sharpened. The main idea is to add some sort of anisotropic diffusion term with an adaptive weight between the shock and the diffusion processes. Alvarez and Mazorra were the first to couple shock and diffusion in [1] proposing an equation of the form:

$$I_t = -sign\left(G_\sigma * I_{\eta\eta}\right)|\nabla I| + cI_{\xi\xi} \tag{3.16}$$

where c is a positive constant and ξ is the direction perpendicular to the gradient ∇I.

In order to account for the magnitude of the second derivative controlling the flow - from the original shock filter formulation of equation 3.12 and choose $F(s) = (2/\pi)\arctan(a\ s)$. This function is a "soft" sign, where a is a parameter that controls the sharpness of the slope near zero. The equation is therefore:

$$I_t = -\frac{2}{\pi}\arctan(aI_{xx})|I_x| + \lambda I_{xx}. \tag{3.17}$$

where $\lambda > 0$ is a constant weight parameter. In this way the inflection points are not of equal weight anymore; regions near edges, with large magnitude of the second derivative near the zero crossing, will be sharpened much faster than relatively smooth regions.

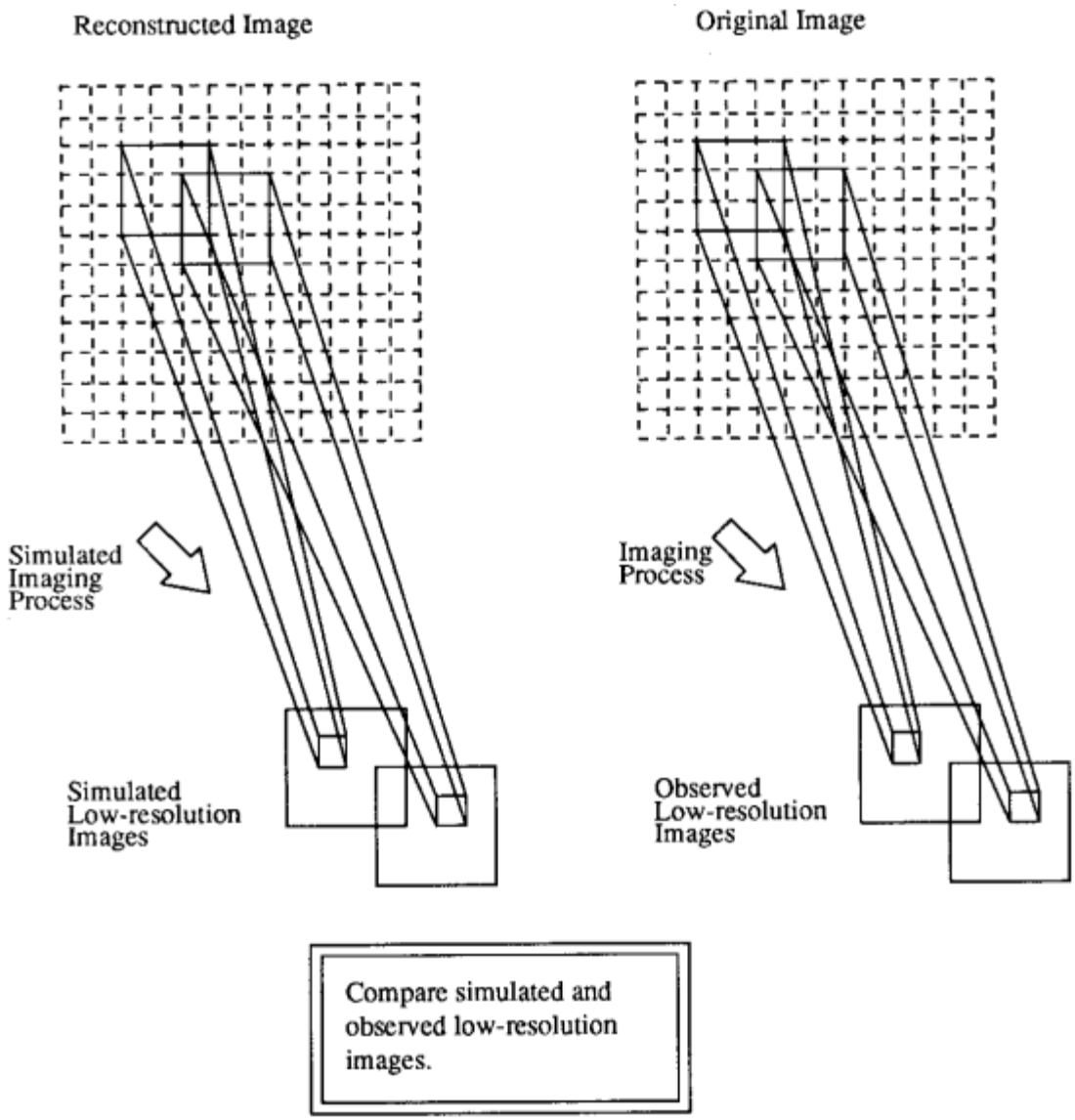


Figure 3.6 Schematic diagram of the super-resolution algorithm. A high-resolution reconstructed image (left) is sought, which gives simulated low-resolution images that are as close as possible to the observed low-resolution images [109].

3.4 Iterative Back-Projection

The iterative back-projection (IBP) technique [25] can accomplish the HR image interpolation and de-blurring simultaneously. Its underlying idea is that the reconstructed HR image from the degraded LR image should produce the same observed LR image if

passing it through the same blurring and downsampling process. The IBP technique can minimize the reconstruction error by iteratively back-projecting the reconstruction error into the reconstructed image.

IBP is an efficient algorithm to acquire the HR image by minimizing the norm of the reconstruction error. Given an estimate of the reconstructed HR image and a model of the imaging process, a set of simulated LR images can then be generated. Each simulated LR image is compared with the actual version and then the error can be used for correcting the estimated image. In fact, it is difficult to restore a HR image in a one-shot manner. Hence, an iterative procedure is needed. This simulate/correct process is iterated until some stopping condition is achieved. Generally, the minimization of some error criterion between the simulated and observed LR images is adopted as the stopping condition [110].

Super-resolution of monochrome and color low resolution image sequences was considered by Irani and Peleg [25]. They derived an iterative back-projection algorithm based on computer aided tomography. The schematic diagram of super-resolution algorithm is given in Fig. 3.6. The algorithm starts with an initial guess ($\mathbf{X}^0$) for the output high-resolution image ($\mathbf{X}^1$) and the imaging process ($\mathbf{A}$) is simulated to generate low-resolution images ($\mathbf{b}^{sim}$) based on the initial guess. These simulated low-resolution images are then compared with the observed ones ($\mathbf{b}$) and the error generated between them is back-projected onto the initial guess via back-projection operator ($\mathbf{A}^{bp}$), thereby minimizing the error iteratively.

$$X^1 = X^0 + A^{bp}(b - b^{sim}) \tag{3.18}$$

The algorithm considers translational and rotational motion but the authors claim that the same concept can be applied to other motions also. They considered multiple motion analysis in [25] including occlusion and transparency. The algorithm successfully solves the issue of blur and noise, however due to the ill-posed nature of super-resolution; the technique is unable to generate a unique solution.

3.5 Similar structure learning

Because of the constant back-projection kernel, the reconstruction error is back projected into the reconstructed image isotropically. This produces artifacts across the edges. These artifacts can be removed by the post-processing step. The structure of low resolution (LR) image is employed to guide the high resolution (HR) image. Hence it is better to first learn the structural content of each HR pixel, from the LR input pixels. Then find the correlation among the pixels with the similar structure. At the end, one can employ this correlation to correct the reconstructed image.

The matching process involves computation of the similarity measure for each disparity value, followed by an aggregation and optimization step. Since these steps consume a lot of processing power, there are significant speed-performance advantages to be had in optimizing the matching algorithm.

One way to learn the similar structure can be described as follows: First learn the structure of HR pixel from the LR image. The image signal is not a static signal, but local image signal can be considered as stationary signal. Consider small squared windows ($\mathbf{B}$ with n^2 pixels) as basic structure elements. The criterions that can be used to find the pixels with similar structure are the zero-mean normalized cross correlation (CC) and the mean absolute differences (MAD) [111].

$$E_{cc}(x,x') = \frac{\sum_{x\in\Omega}(B(x)-\bar{B})\left(B_{ref}(x')-\overline{B_{ref}}\right)}{\sqrt{\sum_{x\in\Omega}(B(x)-\bar{B})^2 \sum_{x\in\Omega}\left(B_{ref}(x')-\overline{B_{ref}}\right)^2}} \quad (3.19)$$

$$E_{MAD}(x,x') = \frac{1}{n^2}\sum_{x\in\Omega}\left|B(x) - B_{ref}(x')\right| \quad (3.20)$$

where Ω contains all the pixels of the reference window $\mathbf{B}_{\text{ref}}$; x and x' are the given pixel and reference pixel respectively; $\mathbf{B}$ and $\overline{B_{ref}}$ are denoted as the mean values of respectively **B** and $\mathbf{B}_{\text{ref}}$. The transformation of the coordinates is characterized by the mapping function m. To simplify the registration problem and particularly to save computation time, pure translation motions of **B** are assumed. The main motive to use these two criterions is because they are somewhat complementary: CC emphasizes the similarity of the structural or geometrical content of the windows, while MAD underlines the similarity of the luminance (and color) information. A matched window is found if the two measures E_{CC} and E_{MAD} satisfy the respective thresholds τ_{CC} and τ_{MAD}, more specifically: $E_{CC} > \tau_{CC}$ and $E_{MAD} < \tau_{MAD}$ [111]. This implementation uses an exhaustive search in order to find the matching windows, but more intelligent (pattern-based) search algorithms can reduce the computation time enormously.

3.6 Summary

In this chapter some of the integral components of this thesis are introduced. It started with bilateral filtering which is a type of edge preserving smoothing filter. To further preserve the edges in the images mean shift image segmentation can be used. Shock filters can help to clearly discriminate between the homogenous patterns. Iterative back projection algorithm reduces the errors present in the reconstructed image to a

minimum. Finally, by learning similar structure in the low resolution image, one can incorporate in the final high resolution image to get better perceptual quality. In the next chapter the proposed flow for the image super-resolution using these techniques are given.

CHAPTER 4

PROPOSED FRAMEWORK FOR SUPER-RESOLUTION

4.1 Introduction

This thesis proposes a novel approach to single image super-resolution. The goal of single image super-resolution is to recover a high-resolution image from a low-resolution image. Although much work has been done, super-resolution has not been solved very well yet. As pointed out in [112], the generation process of a low-resolution image can be modeled as smoothing and down-sampling a high-resolution image. Single image super-resolution is a challenging problem because there is inevitable information loss in down-sampling and the number of unknowns in the recovery exceeds the number of observed data.

Almost all image super-resolution methods [[56], [113], [114], [115], [116], [117]] need an initial high-resolution image for latter steps via up-sampling the input low-resolution image. Previous approaches often use interpolation methods (e.g., bicubic interpolation [27]) to obtain the initial high resolution results. Such a simple method severely degrades the final results. In this thesis, a novel up-sampling method is implemented to obtain the initial high-resolution result. The primary focus is on improving the quality of the reconstructed image generated by the original Iterative back projection (IBP) [25] method and proposes a new SR framework based on edge preservation.

4.2 Problem Statement

The iterative back-projection (IBP) algorithm [25] is a state of art method when the magnification factor is 2 [118]. This is an efficient method to obtain the HR image by

minimizing the reconstruction error. The original IBP method was designed to reconstruct the HR image from multiple LR inputs. Here, the focus is just on one input image only.

In theory, the degradation process from HR image to LR image can be modeled by a combination of the blur effect (due to the atmosphere, the object/camera motion, and the sensor), and the down-sampling operations [119]. This can be formulated as follows:

$$\mathrm{I^L} = (\mathrm{I^H} * \mathrm{g}) \downarrow d \tag{4.1}$$

where $\mathrm{I^H}$ and $\mathrm{I^L}$ are the HR and LR images respectively, g is the point spread function (PSF), $*$ is the convolution operator, and $\downarrow d$ is the down-sampling operator with scaling factor d. In iterative procedure, there are two steps:

(1) Compute the reconstruction error as follows:

$$\mathrm{e_r(I)} = \mathrm{I^L} - (\mathrm{I} * \mathrm{g}) \downarrow \mathrm{d} \tag{4.2}$$

The reconstruction error of an HR image I can be defined as the difference between the LR input image $\mathrm{I^L}$, and the synthesized LR image by I.

(2) Update the HR image with the reconstruction error as follows

$$I_{t+1}^{H} = I_t^H + e_r(I_t^H) \uparrow d * p \tag{4.3}$$

where I_t^H is the HR image at the $\mathrm{t^{th}}$ iteration, $\uparrow$ is the up-sampling operator, p is a constant back-projection kernel.

It is shown in [25] that for d=1(the problem of SR is equivalent to the problem of deblurring in this case) and for multiple LR input images, the back-projection algorithm [25] can converge to the desired deblurred image, which satisfies Eqn. 4.1 for all LR inputs under their corresponding geometry transform, with an exponential rate under certain conditions.

Theorem 1 By updating the HR image with the back-projection iteration I_t^H will converge to desired image $\mathrm{I^c}$, which satisfies Eqn. 4.1, with an exponential rate for all $\mathrm{d} \geq 1$, given $||\delta - \mathrm{g} * \mathrm{p} \downarrow \mathrm{d}||_1 < 1$.

The proof of Theorem 1 is presented in [119]. It means that by applying the back-projection method iteratively, the reconstruction error can be minimized efficiently for any positive integer scaling factor d, with one LR input image, when $||\delta - \mathrm{g} * \mathrm{p} \downarrow \mathrm{d}||_1 < 1$. Similar to the discussion in [25], back-projection filter corresponding to smaller value of $||\delta - \mathrm{g} * \mathrm{p} \downarrow \mathrm{d}||_1$ will have faster converging speed, while it may produce numerically unstable results [119].

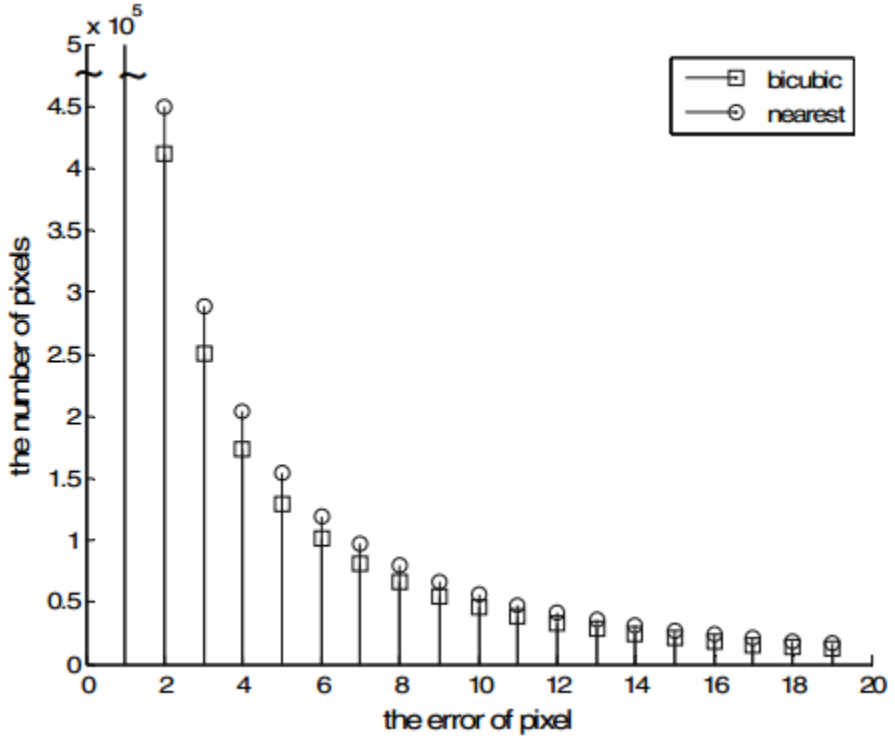
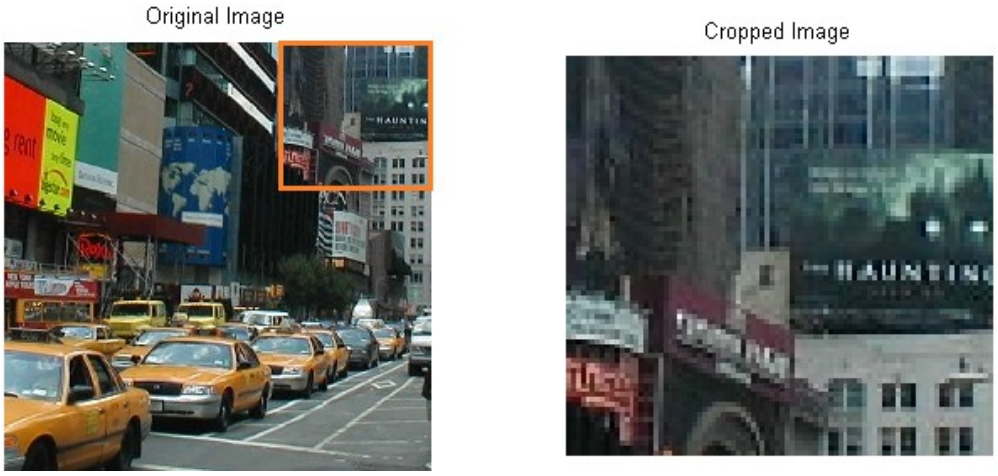


Figure 4.1 The distribution of the error using different interpolation algorithms [118].

I_t^H converges to the original image with an exponential rate, which was proved in [25]. However, it can be easily observed that I_t^H may converge to an HR image with

jaggy artifacts especially along the strong edges. There are two reasons for this result. One is the initial interpolation cannot preserve the edge. The other is that the reconstruction error is back projected into the reconstructed image without considering the structural content of the pixel.



(a)

(b)

Figure 4.2 (a) Original image on left panel and its top right cropped image on the right panel. (b)Interpolation of the cropped image using nearest neighbor (left panel) and bicubic (right panel).

The extent with which the initial interpolation affects the quality of the HR image generated by the IBP method can be shown experimentally. Consider two interpolation algorithms, bicubic interpolation and nearest neighbor interpolation, as the initial image estimate. As in [118], 50 natural gray images are tested whose size is $M \; x \; N$. Hence in total $50xMxN$ pixels are examined. Define $f(x)$ is the value of the original pixel, $\hat{f}(x)$ is the value of the reconstructed pixel. The error of pixel can be defined as the difference between the original pixel and the output pixel as follows:

$$e = \left| f(x) - \hat{f}(x) \right| \tag{4.4}$$

Figure 4.1 is the distribution diagram of the error of $50 \; x \; M \; x \; N$ pixels. When the error of pixel is zero, the number of pixels using bicubic interpolation as the initial estimate is more than that using nearest neighbor. It can be seen that the total number of errors of pixels using bicubic interpolation as the initial estimate is fewer than the nearest neighbor as the initial estimate.

Figure 4.2 shows the HR images generated by using the bicubic interpolation method and using the nearest neighbor interpolation. It is observed that with bicubic interpolation, there is better edge preservation and fewer jaggy artifacts than nearest neighbor interpolation (see the letters in the fig. 4.2(b)). So an initial estimate with better edge preservation will lead to promising final results.

4.3 The proposed SR framework

In this thesis, a new framework is proposed to improve the HR image reconstruction quality. The proposed SR framework is shown in Fig. 4.3. It includes a pre-processing procedure and a post-processing procedure. The pre-processing step is to

strengthen the true edge. The post-processing step can remove the jaggy artifacts produced by the back projected reconstruction error.

A novel up-sampling method is proposed to obtain the initial high-resolution result. Our scheme incorporates a soft-edge and a hard-edge constraint in up-sampling. The soft-edge constraint is enforced via bilateral filtering [26] which smoothes the image while preserving edges. The hard-edge constraint is enforced after applying the mean shift image segmentation algorithm [94].

After obtaining the initial result via up-sampling, the complex shock filter [104] is used to enhance strong edges in the high-resolution image, instead of imposing prior knowledge on the high-resolution image which is used in previous methods [[114], [115], [120], [116]].

Then, a reconstruction constraint on the high-resolution image is enforced and the final result is solved by back projection [119]. The results obtained by the original back projection [119] often suffer from ringing artifacts. Due to the combination of up-sampling scheme and the complex shock filter, this algorithm obtains results without noticeable ringing artifacts.

Finally, in order to further stop the across-edge error propagation, non-local similarity by the structure of the pixels in LR image is learnt and this similar correlation is employed to correct the reconstructed HR image [111]. The experimental results show that the proposed method can improve the output results both in visual perception and in objective estimation.

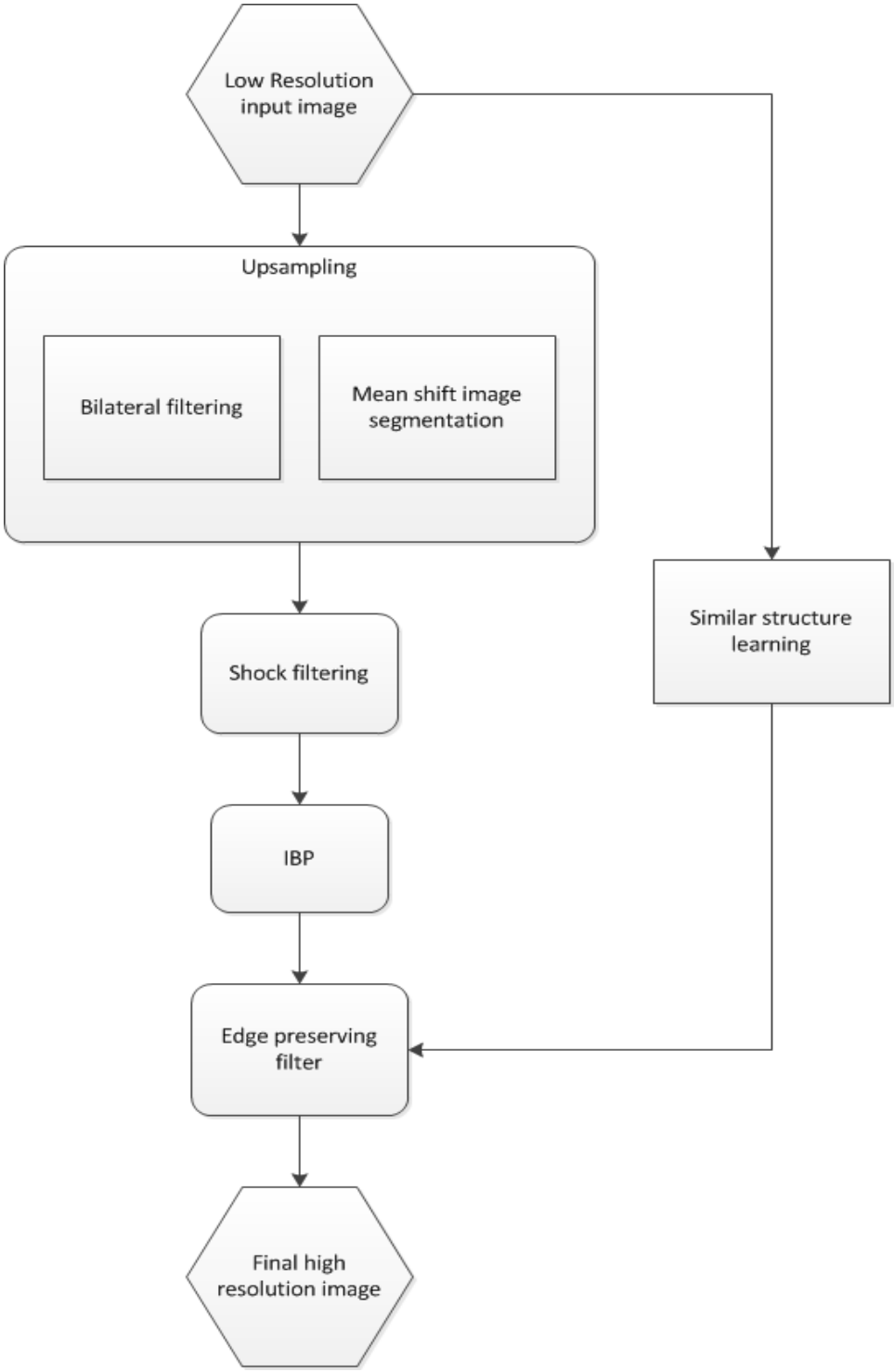


Figure 4.3 The proposed SR framework.

4.3.1. Upsampling

In previous image up-sampling algorithms [60], the filter is constructed considering only spatial information with useful intensity information discarded. Motivated by bilateral filtering in [26], a novel up-sampling scheme considering both spatial and intensity information is implemented. Bilateral filtering can be defined as:

$$B(p) = \frac{1}{W(p)} \sum_{q \in \Omega} I(q) G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(\|I(p) - I(q)\|), \tag{4.5}$$

where I is the input image, and Ω is the set of all pixels of the image, p and q denote pixel locations over the image, and $\mathrm{W}(p)$ is the normalization constant at p, and σ_s and σ_r are standard deviations of two Gaussian functions G_{σ_s} and G_{σ_r}, respectively.

As explained in section 3.1, the intuition of bilateral filtering is to smooth the image using pixels which are close both in the spatial domain and in the intensity domain. This can be called as soft-edge constraint. In this algorithm, to interpolate a target pixel, the nearby pixels with similar colors are chosen to interpolate a target pixel. In order to preserve edges even better, a hard-edge constraint can be utilized [121]. In this method, first segmentation is performed on the input low resolution image by mean shift algorithm [94]. Then for each pixel, only the pixels in the same segment [121] are used.

Mean shift can be easily viewed as a nonparametric estimator of density gradient. It is employed in the joint, spatial-range (value) domain of gray level and color images for discontinuity preserving filtering and image segmentation. As stated by Comaniciu et al. [94], "an image is typically represented as 2-dimensional lattices of r-dimensional vectors (pixels), where r is1 in the gray level case, 3 for color images, or $\mathrm{r} > 3$ in the multispectral case. The space of the lattice is known as the spatial domain while

the gray level, color, or spectral information is represented in the range domain. However, after a proper normalization with σ_s and σ_r global parameters in the spatial and range domains, the location and range vectors can be concatenated to obtain a spatial-range domain of dimension $d = r+2$."

To explain this clearly, consider an image as set of vectors in the spatial-range domain:

$$I = \{ x_j \} \text{ where } j=1,\ldots,n \tag{4.6}$$

Each vector x_j has two parts, a spatial part and a range part, where the range part may be written as a function of the spatial part:

$$x_j = \left(x_j^s, x_j^r\right) = \left(x_j^s, I(x_j^s)\right) \tag{4.7}$$

The superscripts s and r denote the spatial and range parts of the vectors, respectively. Hence, the image vectors can be referred as pixels, and name the spatial and the range part, pixel location and pixel value, respectively.

The mean shift procedure [94] is applied for all the data points in the joint spatial-range domain. Each data point becomes associated with a point of convergence which represents the local mode of the density in the d-dimensional space. The process, having the parameters σ_s and σ_r, takes into account simultaneously both the spatial and range information.

The output of the mean shift filter for an image pixel is defined as the range information carried by the point of the convergence. This process achieves a high quality discontinuity preserving spatial filtering. For the segmentation task, the convergence points sufficiently close in the joint domains are fused to obtain the homogeneous regions in the images.

Let I_M be the result of mean shift image segmentation. By combining both the soft-edge constraint and the hard-edge constraint, the up-sampling is formulated as:

$$I_u(p_h) = \frac{1}{W(p_l) + \beta}\left(\sum_{q\epsilon\Omega} BF(p_l, q)M(p_l, q)I_l(q) + \beta I_L(p_l)\right) \quad (4.8)$$

$$BF\ (p_l, q) = G_{\sigma_s}(\|p_l - q\|)G_{\sigma_r}(\|I_L(p_l) - I_L(q)\|) \quad (4.9)$$

$$M(p_l, q) = \begin{cases} 1 & if\ I_M(p_l) = I_M(q) \\ 0 & if\ I_M(p_l) \neq I_M(q) \end{cases} \quad (4.10)$$

where β is a balance parameter, p_h is pixel position in the high-resolution image, and p_l is the position in the low-resolution image corresponding to p_h. If the coordinates of p_l are not integers, $I_L(p_l)$ and $I_M(p_l)$ are obtained by nearest interpolation on the low-resolution image. The term $\beta I_L(p_l)$ in equation 4.8 is used to emphasize the importance of the filter center p_l.

4.3.2. Shock Filtering

From section 3.3 it is known that the Shock filters are based on the idea to apply locally either dilation or erosion process, depending on whether the pixel belongs to the influence zone of a maximum or a minimum. The decision between dilation and erosion is made using the sign function (also called the signum) s in the set {-1, 0, +1} based on the Laplace operator [101]. Applying this procedure produces a sharp discontinuity called shock at the borderline between two influence zones.

Another desirable goal is the ability to change the process behavior with time in a controlled manner. The basic idea is that processes controlled by the gradient magnitude

have large errors in estimating gradients at the initial stages, where the signal is still very noisy. Therefore a preliminary phase of mainly noise removal can be advantageous [122]. It is suggested to use two processes with continuous transition in time, beginning with linear diffusion at time zero (strong denoising), and advancing towards high nonlinearity (strong edge-preserving properties) [78].

Similar ideas can be applied here. It is desired to decrease the shock effects of the process at the beginning (when estimating the signal's inflection points is difficult) - allowing the diffusion process to smooth out the noise. As the evolution advances, false inflection points produced by the noise are greatly reduced and the enhancing shock part can gain dominance. A simple way to do that is to multiply the second derivative of the shock part by time t. Hence, equation 3.17 for time t can be rewritten as:

$$I_t = -\frac{2}{\pi}\arctan(aI_{xx}t)|I_x| + \lambda I_{xx} \tag{4.11}$$

From equation 4.11 and approximating the solution for small angle, one can derive the complex shock filter formulation. The complex filter is an elegant way to avoid the need of convolving the signal in each iteration and still get smoothed estimations. The time dependency of the process is inherent, without the need to explicitly use the evolution time t. Moreover, the imaginary value receives feedback - it is smoothed by the diffusion and enhanced at sharp transitions by the shock, thus can serve better for controlling the process than a simple second derivative.

The complex shock filter generalized to 2D is:

$$I_t = -\frac{2}{\pi}\arctan(a\mathrm{Im}(\frac{I}{\theta}))|\nabla I| + \lambda I_{\eta\eta} + \tilde{\lambda} I_{\xi\xi}, \tag{4.12}$$

where, $I_{\xi\xi} = I_{xx}I_y^2 - 2I_{xy}I_xI_y + I_{yy}I_x^2$,

$\lambda = re^{i\theta}$ is a complex scalar and $\tilde{\lambda}$ is a real scalar, the parameter a controls the sharpness of the slope near zero, θ is the phase angle of the complex part, and $Im\left(\frac{I}{\theta}\right)$ denotes the complex part of $\left(\frac{I}{\theta}\right)$. More details can be found in [104]. The final output after shock filtering is obtained by the following operation:

$$I_s^{t+1} = I_s^t + I_t * d_t \tag{4.13}$$

where d_t is a iteration step. In this experiment d_t=0.1 and the initial input in the iteration is I_u.

4.3.3. Reconstruction refinement

The obtained intermediate result I_S often does not satisfy the reconstruction constraint. Reconstruction error is defined as

$$E(I_L, I_H) = I_L - (I_H * g)\downarrow d \tag{4.14}$$

where $*$ is the convolution operator, g is a spatial filter, and ↓d denotes the down-sampling operator with factor d. The reconstruction constraint requires that the high-resolution image after smoothing and down-sampling should be as close as possible to the input low-resolution image.

As in many image super-resolution methods [113-116], the reconstruction error can be minimized by the back projection algorithm [25] which is an iterative gradient-based minimization method. In this thesis also iterative back projection is incorporated to obtain the high resolution image by minimizing the reconstruction error as described in section 3.4.

Hence, the high resolution image can be continually updated as:

$$I_H^{t+1} = I_H^t + (((I_t^H * g) - I_L \downarrow d) * p \quad (4.15)$$

where I_H^t is the high-resolution image at iteration *t,* and g and p are the spatial filter and Gaussian "back projection" filter, respectively. The intermediate result I_S from the equation 4.13, is used as the initial I_H at the first iteration. In this experiment, the variances of g and p are set to $d/2$.

4.3.4. Similar structure learning

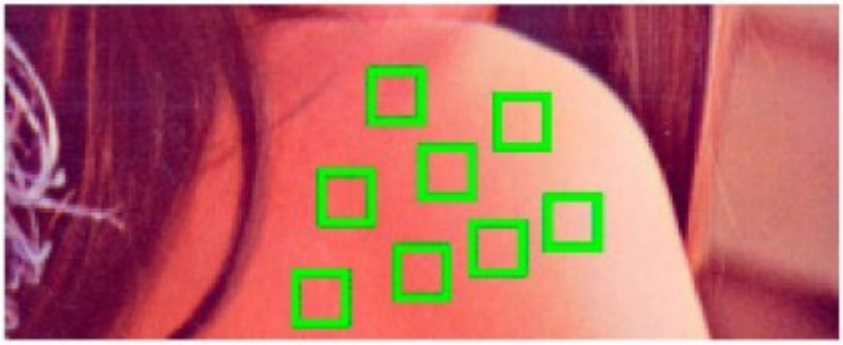
Because of the constant back-projection kernel "p" in equation 4.15, the reconstruction error is back projected into the reconstructed image isotropically. This produces some artifacts across the edges. In addition to this, it is often assumed that true motion is needed for super-resolution, however many registration methods do not yield true motion: their results are optimal to some proposed cost criterion, which are not necessarily equal to true motion. With this in mind, one can hypothetically assume that repetitive structures can serve as multiple noisy observations of the same structure (after proper registration). Besides, repetitivity in texture, one can also find this recurrent property in other parts of the image. Some examples are shown in Fig. 4.4.

Hence with the help of post-processing step, one can reduce the artifacts produced by back projection kernel to minimum. The structure of low resolution input image is employed to guide the high resolution image. This can be implemented as:

1. Learn the structural content of the low resolution input pixels for each high resolution pixel. This can be done using two criterions, namely zero-mean normalized cross correlation (CC) and the mean absolute differences as explained in section 3.5.

(a)

(b)



(c)

Figure 4.4 (a) Repetition in different objects. (b) Repetition along edges. (c) Repetition in uniform areas [111].

2. Find the correlation among the pixels with the similar structure.

3. Employ the correlation to correct the reconstructed image.

In this thesis, search window is limited to 20x20. One can choose this to be as large as the whole image, but it will result in very large computation. Continuing the discussion from section 3.5, subpixel registration in the spatial domain can be achieved either by interpolating image data or by interpolating the correlation data. In order to save computation time, resample only the reference window B_{ref}, on a higher resolution. The subpixel shifts can be estimated using the criterion of equation 3.15. After the registration, the pixel values of B are mapped onto the high resolution (HR) grid. Most existing techniques [57, 116] use linear methods to upscale B_{ref}. However, these interpolation methods typically suffer from blurring, staircasing and/or ringing. These artifacts not only degrade the visual quality but also affect the registration accuracy. That is why a fast non-linear restoration-based interpolation based on level curve mapping [123] is adopted.

After extracting the pixels with the similar structure, the correlation of the similar pixels can be learned, and use this information to correct the reconstructed HR image. Data fusion in is similar to that in the non-local IBP [124] method. The reconstructed image can be updated as the weighted average of all elements in S(x'), considering the initial interpolated image to be $\widehat{H}$:

$$\tilde{H}(x) = \sum_{x' \epsilon S(x')} w(x, x') \hat{H}(x') \quad (4.16)$$

where $S(x')$ is a set which contains the pixels with the similar structure, and weight $w(x,x')$ shows the correlation between the given pixel and the similar structure pixel. $w(x,x')$ can be obtained as follows:

$$w(x,x') = \frac{1}{C(x)} exp\left(\frac{E_{MAD}(x,x')}{t_1}\right) \quad (4.17)$$

$$C(x) = \sum_{x' \epsilon S(x')} w(x,x') \quad (4.18)$$

where $C(x)$ is the normalization constant and t_1 is a parameter to control the decaying speed.

4.4 Summary

In this chapter the proposed methodology is discussed. Firstly, the input low resolution image is upsampled using bilateral filtering and mean shift image segmentation. Secondly, this upsampled image is processed by shock filter with which one can get better smoothed filter. Thirdly, Iterative back projection is applied to this image as a reconstruction refinement of the shock filtered image. Lastly, to further stop the across-edge the similar structure in the low resolution image is learnt and based on zero mean-mean normalized cross correlation (CC) and the mean absolute differences (MAD). This data is then fused with the IBP high resolution image. In the next chapter, the results of this experiment are shown. Also the comparisons of this proposed method with other established super-resolution methods are discussed.

CHAPTER 5

RESULTS AND CONCLUSION

5.1 Image quality assessment

Any processing applied to an image may cause an important loss of information or quality. Image quality evaluation methods can be subdivided into objective and subjective methods [125], [126]. Subjective methods are based on human judgment and operate without reference to explicit criteria [127]. Objective methods are based on comparisons using explicit numerical criteria [128], [129], and several references are possible such as the ground truth or prior knowledge expressed in terms of statistical parameters and tests [130-132].

Image quality is a characteristic of an image that measures the perceived image degradation (typically, compared to an ideal or perfect image). Imaging systems may introduce some amounts of distortion or artifacts in the signal, so the quality assessment is an important problem. The metrics that has been used in this thesis for the image quality assessment (IQA) are discussed in the subsequent sections

5.1.1. PSNR

The peak-signal-to-noise ratio (PSNR) is the most common measure of picture quality. Another more popularly used measure is Mean-Squared Error (MSE). The mean-squared error (MSE) between two images $g(\mathrm{n,m})$ and $\hat{g}(\mathrm{n,m})$ can be given by:

$$E_{MSE} = \frac{1}{MN}\sum_{n=1}^{N}\sum_{m=1}^{M}[\hat{g}(n,m) - g(n,m)]^2 \qquad (4.1)$$

One problem with mean-squared error is that it depends strongly on the image intensity scaling. A mean-squared error of 100.0 for an 8-bit image (with pixel values in

the range 0-255) looks dreadful; but a MSE of 100.0 for a 10-bit image (pixel values in [0, 1023]) is barely noticeable. PSNR avoids this problem by scaling the MSE according to the image range:

$$PSNR = -10\log_{10}\left(\frac{E_{MSE}}{S^2}\right) \tag{4.2}$$

where S is the maximum pixel value. PSNR is measured in decibels (dB). These are appealing because they are simple to calculate, have clear physical meanings, and are mathematically convenient in the context of optimization. But they are not very well matched to perceived visual quality [133-138]. Its main drawback is that the signal strength is estimated as S^2, rather than the actual signal strength for the image. Hence, PSNR is not as adequate as perceptually meaningful measures [139].

5.1.2. SSIM

The Structural SIMilarity (SSIM) index [29] is a method for measuring the similarity between two images. The SSIM index can be viewed as a quality measure of one of the images being compared provided the other image is regarded as of perfect quality. The SSIM index is a full reference (FR) metric, in other words, the measuring of image quality based on an initial uncompressed or distortion-free image as reference. SSIM is designed to improve on traditional methods like peak signal-to-noise ratio and mean squared error, which have proved to be inconsistent with human eye perception.

The difference with respect to other techniques mentioned previously such as MSE or PSNR, is that these approaches estimate perceived errors on the other hand SSIM considers image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies

especially when they are spatially close. These dependencies carry important information about the structure of the objects in the visual scene [29].

At a high level, SSIM attempts to measure the change in luminance, contrast, and structure in an image. Luminance is modeled as average pixel intensity, contrast by the variance between the reference and distorted image, and structure by the cross-correlation between the 2 images. The resulting values are combined (using exponents referred to as alpha, beta, and gamma) and averaged to generate a final SSIM index value.

The general form of SSIM can be given as:

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{4.3}$$

where $\alpha>0$, $\beta>0$, and $\gamma>0$ are parameters used to adjust the relative importance of the three components, and x, y are image patches, and

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{4.4}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{4.5}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{4.6}$$

l(x,y) in equation 4.4 is luminance comparison, c(x,y) in equation 4.5 is contrast comparison and s(x,y) in equation 4.6 is structural comparison. C_1, C_2 and C_3 are

constants to avoid instability. μ_x and μ_y are mean average of x and y, and σ_x^2 , and σ_y^2 are variances of x and y. σ_{xy} is covariance of x and y. In order to simplify the equation 4.3, set $\alpha = \beta = \gamma = 1$ and

$$C_1 = (K_1 L)^2 \tag{4.7}$$

$$C_2 = (K_2 L)^2 \tag{4.8}$$

$$C_3 = C_2/2 \tag{4.9}$$

Where L is the dynamic range of the pixel values (255 for 8-bit grayscale images), and $K_1, K_2 << 1$ is a small constant. Therefore, equation 4.3 can be re-written as:

$$\mathrm{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4.10}$$

5.1.3. FSIM

The great success of SSIM and its extensions owes to the fact that human visual system (HVS) is adapted to the structural information in images. The visual information in an image, however, is often very redundant, while the HVS understands an image mainly based on its low-level features, such as edges and zero crossings [140-142]. In other words, the salient low-level features convey crucial information for the HVS to interpret the scene. Accordingly, perceptible image degradations will lead to perceptible changes in image low-level features, and hence, a good IQA metric could be devised by comparing the low-level feature sets between the reference image and the distorted image. Based on the aforementioned analysis, a novel low-level feature similarity (FSIM) induced FR IQA metric, namely, FSIM has been evolved.

Consider calculating the similarity between images f_1 and f_2. Denote by PC_1 and PC_2 the phase congruency (PC) maps extracted from f_1 and f_2, respectively, and G_1 and G_2 the gradient magnitude (GM) maps extracted from them. FSIM index from [28] can be defined as follows:

$$\mathrm{FSIM} = \frac{\sum_{\mathbf{x}\in\Omega} S_L(\mathbf{x}) \cdot PC_m(\mathbf{x})}{\sum_{\mathbf{x}\in\Omega} PC_m(\mathbf{x})} \tag{4.11}$$

where Ω is the whole image, and

$$S_L(\mathbf{x}) = [S_{PC}(\mathbf{x})]^{\alpha} \cdot [S_G(\mathbf{x})]^{\beta} \tag{4.12}$$

$$S_G(\mathbf{x}) = \frac{2G_1(\mathbf{x}) \cdot G_2(\mathbf{x}) + T_2}{G_1^2(\mathbf{x}) + G_2^2(\mathbf{x}) + T_2} \tag{4.13}$$

$$S_{PC}(\mathbf{x}) = \frac{2PC_1(\mathbf{x}) \cdot PC_2(\mathbf{x}) + T_1}{PC_1^2(\mathbf{x}) + PC_2^2(\mathbf{x}) + T_1} \tag{4.14}$$

For simplicity, $\mathrm{PC_m(x)} = \max\{\mathrm{PC_1(x)}, \mathrm{PC_2(x)}\}$, and $\alpha = \beta = 1$, and $\mathrm{T_1}$ and $\mathrm{T_2}$ denotes positive constants to increase the stability. Detailed explanation of this metric can be found on [28].

5.2 Results

The proposed algorithm is compared with some already established super-resolution methods. The methods which are used to compare are bicubic interpolation [27], Projection onto Convex Set (POCS) Technique [55], non-local iterative back projection (NLIBP) [111] and 2D auto regressive model based image interpolation [143]. Test images used in this thesis are shown in fig. 5.1.

Figure 5.1. Test image: estatua

Figure 5.2. Test image: lena

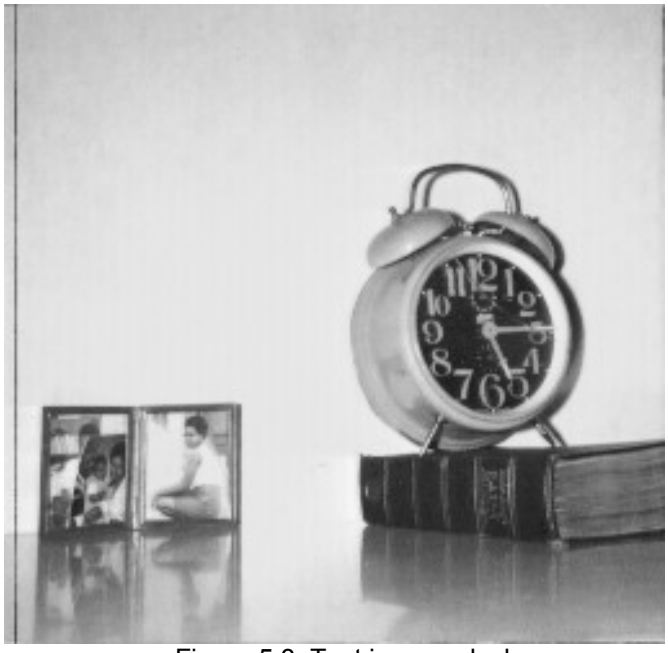


Figure 5.3. Test image: clock

Figure 5.4. Test image : portofino

Figure 5.5. Test image: barche

Table 5.1 Comparison using Lena image

	PSNR (dB)	SSIM	FSIM
Proposed algorithm	28.91	0.9106	0.9518
Bicubic interpolation	28.27	0.8631	0.9418
POCS	27.93	0.9457	0.9467
NLIBP	28.69	0.8573	0.9259
2D auto regressive model	29.05	0.8754	0.9487

Table 5.2 Comparison using Barche image

	PSNR (dB)	SSIM	FSIM
Proposed algorithm	29.24	0.8913	0.9243
Bicubic interpolation	26.98	0.9331	0.9077
POCS	21.65	0.6372	0.7567
NLIBP	27.46	0.8226	0.8845
2D auto regressive model	27.37	0.8403	0.9161

Table 5.3 Comparison using Estatua image

	PSNR (dB)	SSIM	FSIM
Proposed algorithm	30.19	0.7933	0.9412
Bicubic interpolation	29.81	0.7760	0.9204
POCS	24.51	0.6187	0.8101
NLIBP	30.66	0.7785	0.9033
2D auto regressive model	30.08	0.7840	0.9226

Table 5.4 Comparison using Portofino image

	PSNR (dB)	SSIM	FSIM
Proposed algorithm	26.97	0.7995	0.9217
Bicubic interpolation	25.64	0.7728	0.8904
POCS	22.46	0.5880	0.7539
NLIBP	26.61	0.7710	0.8673
2D auto regressive model	25.71	0.7743	0.8924

Table 5.5 Comparison using Clock image

	PSNR (dB)	SSIM	FSIM
Proposed algorithm	29.19	0.9290	0.9268
Bicubic interpolation	28.63	0.9331	0.9305
POCS	21.74	0.8201	0.8049
NLIBP	29.19	0.9292	0.9215
2D auto regressive model	28.98	0.9357	0.9379

Table 5.6 Overall comparisons for all images

	PSNR (dB)	SSIM	FSIM
Proposed algorithm	28.9	0.8674	0.93316
Bicubic interpolation	28.866	0.8556	0.9186
POCS	23.658	0.72194	0.81446
NLIBP	28.522	0.83172	0.9005
2D auto regressive model	28.238	0.84194	0.92354

5.3 Conclusions

In this thesis a novel image interpolation method is proposed with special attention to edges present in the images. Tables 5.1-5.5 clearly show that the proposed algorithm performs better than its counterparts. There has been 3-6% increase in the FSIM index. Also worth noting is that the proposed algorithm out-perform the other existing techniques when the image under consideration has more details (complex) in it. The given algorithm is more suitable for applications where computational cost is not a

constraint. Using mean-shift image segmentation definitely preserves edges during the upsampling. In addition to this, through post processing on the reconstructed images improves the image quality.

This new framework improves the quality of the HR image reconstructed by the IBP method. The upsampling process which is combination of bilateral filtering and mean-shift image segmentation is used in the initial step to strengthen the true edge. In order to further reduce the across-edge error propagation, post processing on the reconstructed HR image is employed, which takes the advantage of the correlation of pixels with the similar structure to guide the reconstructed image. The proposed algorithm can improve the results in both subjective visual quality and PSNR measure.

5.4 Future Work

The algorithm can be optimized for better performance. One can tweak this algorithm and have a check on the error convergence, which is important for several other applications. Other than that the super-resolution problem is not yet solved. There are always other frameworks yet to be discovered towards optimal super-resolution.

REFERENCES

[1] R. C. Gonzalez and R. E. Woods, "Digital Image Processing". New Jersey: Prentice Hall, 2008.

[2] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in Advances in Computer Vision and Image Processing. (1st ed.), T. S. Huang, Ed. London: JAI Press Inc., pp. 101-106, 1984.

[3] B. S. I. Staff, "Photography. electronic still picture cameras. resolution measurements", B S I Standards, Feb. 2000. Available: http://books.google.com/books?id=aQbFAAAACAAJ.

[4] S. Chaudhuri and M. V. Joshi. Research on image super-resolution. pp. 15-31. 2005.

[5] H. Nyquist , "Certain topics in telegraph transmission theory," Transactions of the American Institute of Electrical Engineers, vol. 47, no.2, pp. 617-644, Apr.1928.

[6] C. E. Shanon, “A mathematical theory of communication," Bell System Technical Journal, vol. 27, no.4, pp. 623 - 656, Oct. 1948.

[7] C. E. Shannon, "Communication in the presence of noise,” Proceedings of the IEEE, vol. 86, no.2, pp. 447-457. Feb.1998.

[8] R. J. Schalkoff. "Digital image processing and computer vision ", New York, Wiley, 1989. Available: http://books.google.com/books?id=-vtRAAAAMAAJ.

[9] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections,” Jour. of Optical Society of America, vol. 6, no.11, pp. 1715-1726. Nov.1989. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-6-11-1715.

[10] K. Aizawa, T. Komatsu and T. Saito, "A scheme for acquiring very high resolution images using multiple cameras," IEEE International Conference on Acoustics, Speech, and Signal Processing,San Francisco, CA, vol.3, pp. 289-292,Mar. 1992.

[11] E.-.Eid, "Study of limitations on pixel size of very high resolution image sensors," Proceedings of the Eighteenth National Radio Science Conference,vol.1, pp. 15-28, 2001.

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

[13] V. Bannore and L. Swierkowski, "An iterative approach to image super-resolution." in Intelligent Information Processing III, vol. 228, pp. 473-482, Springer, Boston, 2007.

[14] P. Milanfar. "Super-resolution imaging ", Taylor & Francis Group, 2010. Available: http://books.google.com/books?id=fjTUbMnvOkgC.

[15] M. Cristani, D. S. Cheng, V. Murino and D. Pannullo, "Distilling information with super-resolution for video surveillance," Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks, New York, NY, USA, pp. 2-11, 2004. Available: http://doi.acm.org/10.1145/1026799.1026803.

[16] F. C. Lin, C. B. Fookes, V. Chandran and S. Sridharan, "Investigation into optical flow super-resolution for surveillance applications," APRS Workshop on Digital Image Computing: Pattern Recognition and Imaging for Medical Applications, Brisbane, Australia, pp. 73-78, Feb. 2005. Available: http://eprints.qut.edu.au/17945/.

[17] F. Li, X. Jia and D. Fraser, "Universal HMT based super resolution for remote sensing images," IEEE International Conference on Image Processing, pp. 333-336, Oct. 2008.

[18] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," Med. Image Anal. vol. 2, no.1, pp. 1-36, Oct.1997.

[19] S. Peled and Y. Yeshurun, "Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging," Magn. Reson. Med., vol. 45, no.1, pp. 29-35, Jan. 2001.

[20] J. A. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom and H. Azhari, "Super-resolution in PET imaging," IEEE Transactions on Medical Imaging, vol. 25, no.2, pp. 137-147, Feb. 2006.

[21] K. Malczewski and R. Stasinski, "Toeplitz-based iterative image fusion scheme for MRI," 15th IEEE International Conference on Image Processing, San Diego, CA, pp. 341-344, Oct. 2008.

[22] T. Bauer, "Super-resolution imaging: The use case of optical astronomy. " Proceedings of the IADIS International Conference Computer Graphics, Visualization, Computer Vision and Image Processing,Rome, Italy, pp. 49-59, 2011.

[23] K. G. Puschmann and F. Kneer, "On super-resolution in astronomical imaging," Astronomy and Astrophysics, vol. 436, no.1, pp. 373-378, June 2005.

[24] L. M. Novak, G. J. Owirka and A. L. Weaver, "Automatic target recognition using enhanced resolution SAR data," IEEE Transactions on Aerospace and Electronic Systems, vol. 35, no.1, pp. 157-175, Jan.1999.

[25] M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion, and transparency," Journal of Visual Communication and Image Representation, vol. 4, no.4, pp. 324-335, Dec.1993.

[26] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," Proceedings of the Sixth International Conference on Computer Vision, Bombay, India, pp. 839-846, Jan.1998. Available: http://dl.acm.org/citation.cfm?id=938978.939190.

[27] R. Keys, "Cubic convolution interpolation for digital image processing," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 29, no.6, pp. 1153-1160, Dec. 1981.

[28] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Transactions on Image Processing, vol. 20, no.8, pp. 2378-2386, Aug. 2011.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing vol. 13, no.4, pp. 600-612, Apr. 2004.

[30] 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, Jun.1990.

[31] S. P. Kim and W.-Su, "Recursive high-resolution reconstruction of blurred multiframe images," IEEE Transactions on Image Processing, vol. 2, no.4, pp. 534-53, Oct.1993.

[32] S. Rhee and M. G. Kang, "DCT-based regularized algorithm for high-resolution image reconstruction," International Conference on Image Processing, vol.3, pp. 184-187, Oct.1999.

[33] R. H. Chan, T. F. Chan, L. Shen and Z. Shen, "Wavelet algorithms for high-resolution image reconstruction," SIAM J.Sci.Comput. vol. 24, no.4, pp. 1408-1432, Apr. 2002.

[34] B. K. Horn. "Robot Vision ", 1st ed., Cambridge, MA: McGraw-Hill Higher Education, 1986.

[35] P. Packalén, T. Tokola, J. Saastamoinen and M. Maltamo, "Use of a super - resolution method in interpretation of forests from multiple NOAA/AVHRR images," International Journal of Remote Sensing, vol. 27, no.24, pp. 5341-5357, Dec. 2006.

[36] Z. Wang, D. Ziou, C. Armenakis, D. Li and Q. Li, "A comparative analysis of image fusion methods," IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no.6, pp. 1391-1402, June 2005.

[37] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti and L. Wald, "Image fusion—the ARSIS concept and some successful implementation schemes," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 58, no.1–2, pp. 4-18, June 2003.

[38] M. Gonzalez-Audicana, X. Otazu, O. Fors and J. Alvarez-Mozos, "A low computational-cost method to fuse IKONOS images using the spectral response function of its sensors," IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no.6, pp. 1683-1691, June 2006.

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

[40] J. H. Park and M. G. Kang , "Spatially adaptive multi-resolution multispectral image fusion," International Journal of Remote Sensing, vol. 25, no.23, pp. 5491-5508, Dec. 2004.

[41] B. Bhatta. "Remote Sensing and GIS ", Oxford University Press, 2008.

[42] J. C. Chan, J. Ma and F. Canters, "A comparison of superresolution reconstruction methods for multi-angle CHRIS/Proba images," Proc. SPIE Image and Signal Processing for Remote Sensing XIV, vol.7109, no. 1, Oct. 2008.

[43] A. J. Jerri, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," Proceedings of the IEEE vol. 65, no.11, pp. 1565-1596, Nov.1977.

[44] T. M. Lehmann, C. Gonner and K. Spitzer, "Survey: interpolation methods in medical image processing," IEEE Transactions on Medical Imaging, vol. 18, no.11, pp. 1049-1075, Nov.1999.

[45] J. N. Ratzel, "The discrete representation of spatially continuous images," Thesis (Ph. D.) MIT, 1980.

[46] W. Burger and M. J. Burge. "Principles of digital image processing: core algorithms ", Springer, 2009. Available: http://books.google.com/books?id=s5CBZLBakawC.

[47] H. Hou and H. Andrews, "Cubic splines for image interpolation and digital filtering," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 26, no.6, pp. 508-517, Dec.1978.

[48] K. Turkowski, "Filters for common resampling task," in Graphics Gems, A. S. Glassner, Ed. San Diego, CA, USA: Academic Press Professional, Inc., pp. 147-165,1990.

[49] J. A. Parker, R. V. Kenyon and D. E. Troxel, "Comparison of interpolating methods for image resampling," IEEE Transactions on Medical Imaging, vol. 2, no.1, pp. 31-39. Mar.1983.

[50] G. Wolberg, "Sampling, reconstruction, and antialiasing," in Computer Science Handbook (2nd ed.), A. B. Tucker, Ed. CRC Press, pp. 39-1 - 39-25, 2004.

[51] J. D. van Ouwerkerk , "Image super-resolution survey," Image and Vision Computing, vol. 24, no.10, pp. 1039-1052, Feb.2006.

[52] J. L. Harris, "Diffraction and Resolving Power," J. Opt. Soc. Am. vol. 54, no.7, pp. 931-933, July1964. Available: http://www.opticsinfobase.org/abstract.cfm?URI=josa-54-7-931.

[53] R. Gerchberg, "Super-resolution through error energy reduction." Optica Acta: International Journal of Optics vol.21, no.9, pp. 709, Jan.1974.

[54] A. K. Jain, "Fundamentals of Digital Image Processing ", Upper Saddle River, NJ, USA: Prentice-Hall, Inc, 1989.

[55] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," IEEE Transactions on Circuits and Systems, vol. 22, no.9, pp. 735-742. Sept. 1975.

[56] W. T. Freeman, E. C. Pasztor and O. T. Carmichael, "Learning low-level vision". IJCV vol.40, no.1, pp. 25-47, Jan. 2000.

[57] W. T. Freeman and E. C. Pasztor, "Markov networks for super-resolution," Proceedings of 34th Annual Conference on Information Sciences and Systems,NJ, Mar. 2000.

[58] 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, Mar. 2002.

[59] C. B. Atkins, C. A. Bouman and J. P. Allebach, "Optimal image scaling using pixel classification," International Conference on Image Processing, vol.3, pp. 864-867, Oct. 2001.

[60] S. Battiato, G. Gallo and F. Stanco, "A locally adaptive zooming algorithm for digital images," Image Vision Comput. vol. 20, no.11, pp. 805-812. Sept.2002. Available: http://www.sciencedirect.com/science/article/pii/S0262885602000896.

[61] S. Battiato, G. Gallo and F. Stanco," Smart interpolation by anisotropic diffusion". 12th International Conference on Image Analysis and Processing, pp. 572-577, Sept. 2003.

[62] D. D. Muresan and T. W. Parks, "Adaptive, optimal-recovery image interpolation," IEEE International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, vol.3, pp. 1949-1952, May 2001.

[63] D. D. Muresan and T. W. Parks, "Adaptively quadratic (AQua) image interpolation," IEEE Transactions on Image Processing, vol. 13, no.5, pp. 690-698, May 2004.

[64] D. D. Muresan and T. W. Parks, "Prediction of image detail," International Conference on Image Processing, Vancouver, BC, Canada, vol.2, pp. 323-326, Sept. 2000.

[65] K. Kinebuchi, D. D. Muresan and T. W. Parks, "Image interpolation using wavelet based hidden Markov trees," IEEE International Conference on Acoustics, Speech, and Signal Processing, ,Salt Lake City, UT, vol.3, pp. 1957-1960, May 2001.

[66] D. Su and P. Willis, "Image interpolation by pixel-level data-dependent triangulation," Computer Graphics Forum vol. 23, no.2, pp. 189-201, Jul. 2004. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2004.00752.x/abstract.

[67] Xiaohua Yu, B. S. Bryan and T. W. Sederberg, "Image reconstruction using data-dependent triangulation," IEEE Computer Graphics and Applications, vol. 21, no.3, pp. 62-68, May 2001.

[68] K. Jensen and D. Anastassiou, "Subpixel edge localization and the interpolation of still images," IEEE Transactions on Image Processing, vol. 4, no.3, pp. 285-295. Mar.1995.

[69] http://www.benvista.com.; Commercial tool for photo enhancement

[70] http://www.imagener.com/; Commercial tool for photo enhancement..

[71] http://www.ddisoftware.com/qimage/plugins/index.html; Commercial tool for photo enhancement.

[72] http://www.extensis.com; Commercial tool for photo enhancement..

[73] L. Davis and A. Rosenfeld. , "Noise cleaning by iterated local averaging," IEEE Transactions on Systems, Man and Cybernetics, vol. 8, no.9, pp. 705-710, Sept. 1978.

[74] N. Himayat and S. A. Kassam. , "Approximate performance analysis of edge preserving filters," IEEE Transactions on Signal Processing, vol. 41, no.9, pp. 2764-2777, Sept.1993.

[75] T. Huang, G. Yang and G. Tang. , "A fast two-dimensional median filtering algorithm," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 27, no.1, pp. 13-18. Feb.1979.

[76] J. Lee, "Digital image enhancement and noise filtering by use of local statistics," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-2, no.2, pp. 165-168. Mar1980.

[77] P. M. Narendra, "A separable median filter for image noise smoothing," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-3, no.1, pp. 20-29. Jan.1981.

[78] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," Pattern Analysis and Machine Intelligence, IEEE Transactions on vol. 12, no.7, pp. 629-639. Jul1990.

[79] L. Yin, R. Yang, M. Gabbouj and Y. Neuvo, "Weighted median filters: a tutorial," IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 43, no.3, pp. 157-192, Mar. 1996.

[80] L. Wang, Y. Zhang and J. Feng, "On the Euclidean distance of images," IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 27, no.8, pp. 1334-1339, Aug.2005.

[81] H. Winnemoller, S. C. Olsen and B. Gooch, "Real-time video abstraction," ACM Trans.Graph. vol. 25, no.3, pp. 1221-1226, Jul. 2006. Available: http://doi.acm.org/10.1145/1141911.1142018.

[82] Y. J. Zhang, "Advances in image and video segmentation ", USA: IRM Press, 2006Available: http://books.google.com/books?id=VMWrAh2Di2QC.

[83] Q. Hu, Z. Hou and W. L. Nowinski, "Supervised range-constrained thresholding," IEEE Transactions on Image Processing, vol. 15, no.1, pp. 228-240, Jan.2006.

[84] L. Grady, "Random walks for image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no.11, pp. 1768-1783, Nov. 2006.

[85] J. A. Sethian, "Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science ", Cambridge University Press, 1999. Available: http://books.google.com/books?id=ErpOoynE4dIC.

[86] S. Chabrier, B. Emile, H. Laurent, C. Rosenberger and P. Marche, "Unsupervised evaluation of image segmentation application to multi-spectral images," Proceedings of the 17th International Conference on Pattern Recognition,vol.1, pp. 576-579, Aug. 2004.

[87] X. Artaechevarria, A. Munoz-Barrutia and C. Ortiz-de-Solorzano, "Combination strategies in multi-atlas image segmentation: application to brain MR data," IEEE Transactions on Medical Imaging, vol. 28, no.8, pp. 1266-1277. Aug. 2009.

[88] A. Roy, S. Kumar Parui, A. Paul and U. Roy, "A color based image segmentation and its application to text segmentation," Sixth Indian Conference on Computer Vision, Graphics & Image Processing,Bhubaneswar, India, pp. 313-319, Dec. 2008.

[89] G. Huilin, D. Lihua, C. Wenjie and X. Gang, "The applications of image segmentation techniques in medical CT images," 30th Chinese Control Conference (CCC),Yantai, pp. 3296-3299, July 2011.

[90] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function with applications in pattern recognition," IEEE Transactions on Information Theory, vol. 21, no.1, pp. 32-40, Jan. 1975.

[91] D. Comaniciu and P. Meer, "Robust analysis of feature spaces: Color image segmentation," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, pp. 750-755, Jun 1997.

[92] J. S. Suri, S. K. Setarehdan and S. Singh, "Advanced algorithmic approaches to medical image segmentation: state-of-the-art application in cardiology, neurology, mammography and pathology ", 1st ed., New York, NY, USA: Springer-Verlag New York, Inc, 2002.

[93] A. A. Bell, J. N. Kaftan, T. Aach, D. M.-Ebrecht and A. Bocking, "High dynamic range images as a basis for detection of argyrophilic nucleolar organizer regions under varying stain intensities," IEEE International Conference on Image Processing, Atlanta, GA, pp. 2541-2544, Oct. 2006.

[94] D. Comaniciu and P. Meer. , "Mean shift: a robust approach toward feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no.5, pp. 603-619, May 2002.

[95] Q. Guo, X. Chang and H. Chu. "Mean-shift of variable window based on the epanechnikov kernel," International Conference on Mechatronics and Automation, Harbin, pp. 2314-2319, Aug. 2007.

[96] M. Kerckhove, “Scale-space and morphology in computer vision. scale-space and morphology in computer vision,” Vancouver, Canada: Springer-Verlag, 2001. Available: http://books.google.com/books?id=DlRvjAKY2WAC.

[97] B. M. Romeny, "Geometry-driven diffusion in computer vision ", Norwell, MA, USA: Kluwer Academic Publishers, 1994.

[98] S. Osher and L. I. Rudin, "Feature-oriented image enhancement using shock filters," SIAM J.Numer.Anal., vol. 27, no.4, pp. 919-940, Aug. 1990. Available: http://dx.doi.org/10.1137/0727053.

[99] J. Weickert, "Coherence-enhancing shock filters," in Pattern Recognition, Lecture Notes in Computer Science, B. Michaelis and G. Krell, Eds., Springer, Berlin, Heidelberg.

[100] L. Cosmin, L. Olivier, P. Sorin, T. Romulus and B. Monica, "Image enhancement using a new shock filter formalism," ACTA Technica Napocensis:Electronics and Telecommunications, vol. 50, no.3, pp. 27-30, Mar. 2009.

[101] H. P. Kramer and J. B. Bruckner, "Iterations of a non-linear transformation for enhancement of digital images," Pattern Recognition vol. 7, no.1-2, pp. 53-58, Jun.1975. Available: http://www.sciencedirect.com/science/article/pii/0031320375900138.

[102] F. Guichard and J. Morel, "A note on two classical enhancement filters and their associated PDE's," Int.J.Comput.Vision vol. 52, no.2-3, pp. 153-160, May 2003. Available: http://dx.doi.org/10.1023/A:1022904124348.

[103] O. Marques, "Edge detection," in Practical Image and Video Processing using MATLAB, 1st ed., New Jersey: Wiley-IEEE Press, pp. 335-363, 2011.

[104] G. Gilboa, N. A. Sochen and Y. Y. Zeevi, "Regularized shock filters and complex diffusion," Proceedings of the 7th European Conference on Computer Vision-Part I, pp. 399-413, 2002. Available: http://dl.acm.org/citation.cfm?id=645315.649481.

[105] L. Alvarez and L. Mazorra, "Signal and image restoration using shock filters and anisotropic diffusion," SIAM J. Numer. Anal. vol. 31, no.2, pp. 590-605, Feb.1994. Available: http://epubs.siam.org/doi/abs/10.1137/0731032.

[106] O. Coulon and S. R. Arridge, "Dual echo MR image processing using multi-spectral probabilistic diffusion coupled with shock filters." Proc. of the British Conference on Medical Image Understanding and Analysis, (MIUA 2000),London, pp. 141-144, 2000.

[107] P. Komprobst, R. Deriche and G. Aubert, "Image coupling, restoration and enhancement via PDE's," IEEE International Conference on Image Processing, Santa Barbara, CA, vol.2, pp. 458-461, Oct 1997.

[108] F. R. Nicolas and J. P. Francoise, "Controlled anisotropic diffusion," Proc. SPIE Conf. on Nonlinear Image Processing VI - IS&T / SPIE Symp. on Electronic Imaging, Science and Technology, San Jose, CA, vol.2424, pp. 329-340, Mar. 1995.

[109] M. Irani and S. Peleg, "Improving resolution by image registration," CVGIP: Graph.Models Image Process. vol. 53, no.3, pp. 231-239, Apr. 1991. Available: http://dx.doi.org/10.1016/1049-9652(91)90045-L.

[110] G. -. Lin and M. -. Lai, "Enhancing resolution using iterative back-projection technique for image sequences," Journal of Computers vol. 19, no.3, pp. 44-54. Oct. 2008.

[111] H. Luong, A. Ledda and W. Philips, "Non-local image interpolation," Proc. Int. Conf. Image Process., Atlanta, GA, United States, pp. 693-696, Oct. 2006.

[112] 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, Sept. 2002.

[113] S. Jian, Z. Nan-Ning, T. Hai and S. H.-Yeung, "Image hallucination with primal sketch priors," IEEE Computer Society Conference on Computer Vision and Pattern Recognition,vol.2, pp. II-729-36, Jun 2003.

[114] S. Dai, M. Han, W. Xu, Y. Wu and Y. Gong, "Soft edge smoothness prior for alpha channel super resolution," IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, pp. 1-8, June 2007.

[115] Q. Shan, Z. Li, J. Jia and C. Tang, "Fast image/video upsampling," ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH Asia, vol. 27, no.5, pp. 153:1-153:7, Dec. 2008.

[116] J. Sun, Z. Xu and H.-Y. Shum, "Image super-resolution using gradient profile prior," IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, pp. 1-8, Jun. 2008.

[117] J. Yang, J. Wright, T. Huang and Y. Ma, "Image super-resolution as sparse representation of raw image patches," IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, pp. 1-8, Jun 2008.

[118] Y. Fan, Z. Gan, Y. Qiu and X. Zhu, "Single image super resolution method based on edge preservation," Sixth International Conference on Image and Graphics (ICIG),Hefei, Anhui, pp. 394-399, Aug 2011.

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

[120] R. Fattal, "Image upsampling via imposed edge statistics," ACM Trans. Graph. vol. 26, no.3, pp. 95. Jul. 2007. Available: http://doi.acm.org/10.1145/1276377.1276496.

[121] Q. Zhou, S. Chen, J. Liu and X. Tang, "Edge-preserving single image super-resolution", Proceedings of the 19th ACM international conference on Multimedia, Arizona, USA, pp. 1037-1040. Nov. 2011.

[122] G. Gilboa, Y. Y. Zeevi and N. Sochen, "Image enhancement segmentation and denoising by time dependent nonlinear diffusion processes," IEEE International Conference on Image Processing, Thessaloniki, vol.3, pp. 134-137, Oct. 2001.

[123] H. Q. Luong, P. de Smet and W. Philips, "Image interpolation using constrained adaptive contrast enhancement techniques," IEEE International Conference on Image Processing.vol.2, pp. II-998-1001, Sept. 2005.

[124] W. Dong, L. Zhang, G. Shi and X. Wu, "Nonlocal back-projection for adaptive image enlargement," 16th IEEE International Conference on Image Processing (ICIP), Cairo, pp. 349-352, Nov. 2009.

[125] R. Kreis, "Issues of spectral quality in clinical 1H-magnetic resonance spectroscopy and a gallery of artifacts," NMR Biomed., vol. 17, no.6, pp. 361-381, Oct. 2004.

[126] I. Avcıbas¸, B. Sankur and K. Sayood, "Statistical evaluation of image quality measures," Journal of Electronic Imaging, vol. 11, no.2, pp. 206-223, April 2002.

[127] L. W. MacDonald and R. Luo, "Colour Imaging: Vision and Technology ", Wiley, 1999. Available: http://books.google.com/books?id=rO9RAAAAMAAJ.

[128] M. Cadik and P. Slavik, "Evaluation of two principal approaches to objective image quality assessment," Eighth International Conference on Information Visualisation, IEEE Computer Society Press, pp. 513-518, Jul 2004.

[129] T. B. Nguyen and D. Ziou, "Contextual and non-contextual performance evaluation of edge detectors," Pattern Recogn.Lett. vol. 21, no.8, pp. 805-816, Jul. 2000. Available: http://dl.acm.org/citation.cfm?id=351795.351800.

[130] O. Elbadawy, M. R. El-Sakka and M. S. Kamel, "An information theoretic image-quality measure," IEEE Canadian Conference on Electrical and Computer Engineering, Waterloo, Ont., vol.1, pp. 169-172, May 1998.

[131] A. Medda and V. DeBrunner, "Color image quality index based on the UIQI," IEEE Southwest Symposium on Image Analysis and Interpretation, Denver, CO, pp. 213-217, Jun 2006.

[132] R. Dosselmann and X.- D. Yang, "Existing and emerging image quality metrics," Canadian Conference on Electrical and Computer Engineering, Saskatoon, Sask., pp. 1906-1913, May 2005.

[133] B. Girod, "What's wrong with mean-squared error?" in Digital images and human vision, A. B. Watson, Ed. Cambridge, MA, USA: MIT Press, pp. 207-220, 1993. Available: http://dl.acm.org/citation.cfm?id=197765.197784.

[134] P. C. Teo and D. J. Heeger, "Perceptual image distortion," IEEE International Conference Image Processing, Austin, TX, vol.2, pp. 982-986, Nov. 1994.

[135] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," IEEE Transactions on Communications vol. 43, no.12, pp. 2959-2965, Dec.1995.

[136] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," Signal Process- Special Issue on Image and Video Quality Metrics, vol. 70, no.3, pp. 177, Nov.1998.

[137] Z. Wang and A. C. Bovik, "A universal image quality index," IEEE Signal Processing Letters, vol. 9, no.3, pp. 81-84, Mar. 2002.

[138] Z. Wang, A. C. Bovik and L. Lu, "Why is image quality assessment so difficult?" IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),Orlando, FL, USA, vol.4, pp. IV-3313-IV-3316, May 2002.

[139] B. Rani, R. K. Bansal and S. Bansal, "Comparative analysis of wavelet filters using objective quality measures," IEEE International Advance Computing Conference, Patiala, India, pp. 402-407, Mar. 2009.

[140] D. Marr, "Vision". New York: Freeman, 1980.

[141] D. Marr and E. Hildreth, "Theory of edge detection," Proc. R. Soc. Lond. B. Biol. Sci., vol. 207, no.1167, pp. 187-217, Feb. 1980.

[142] M. C. Morrone and D. C. Burr, "Feature detection in human vision: a phase-dependent energy model," Proc. R. Soc. Lond. B. Biol. Sci., vol. 235, no.1280, pp. 221-245, Dec. 1988.

[143] X. Zhang and X. Wu, "Image Interpolation by Adaptive 2-D Autoregressive Modeling and Soft-Decision Estimation," IEEE Transactions on Image Processing vol. 17, no.6, pp. 887-896, June 2008.

BIOGRAPHICAL INFORMATION

Gaurav Hansda was born on 22nd July, 1988 in Jabalpur, Madhya Pradesh, India. He is the elder son of Mr. Anoop Hansda and Mrs. Georgina Hansda. He received his Bachelor's Degree in Electronics Engineering from D. J. Sanghvi college of Engineering (Mumbai University), Mumbai in 2010. Immediately after getting his Bachelor's Degree, he decided to pursue his Masters Degree in Electrical Engineering at University of Texas at Arlington. During his study period in Arlington he was interested in Image and Video Processing and joined the Multimedia Group at UTA in Jan. 2011 under the guidance of Dr. K. R. Rao. He got an opportunity to intern at Qualcomm Technology Inc. from Aug. 2012 to Dec. 2012 in San Diego, California in LINUX Android camera team which broadened his understanding in these fields in depth. After his graduation, he intends to find a job in multimedia field where he can utilize his knowledge and experience practically.