Super-Resolution Image Reconstruction from Low-Resolution Images

By

Haidawati Mohamad Nasir

In the fulfilment of the requirement
for the degree of Doctor of Philosophy

Centre for excellence in Signal and Image Processing,
Department of Electronic and Electrical Engineering,
University of Strathclyde

© May 2012

The copyright of this thesis belongs to the author under the terms of the United Kingdom copyright Acts as qualified by University of Strathclyde Regulation 3.49. Due acknowledgment must always be made of the use of any material contained in, or derived from this thesis.

© Copyright 2012

Declaration

I declare that this thesis embodies my own research work and that it is composed by myself. Where appropriate, I have made acknowledgments to the work of others.

Haidawati Mohamad Nasir

Acknowledgment

This thesis would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, I would like to express my deepest gratitude to my PhD supervisors, Prof Stephen Marshall and Dr.Vladimir Stankovic. Throughout my study, both of them have provided constant support, guidance, encouragement, sound advice, good teaching and lots of good ideas. I would have been lost without them and without their help and support, this thesis would never been completed.

I am grateful to Malaysia Government under its agency Majlis Amanah Rakyat (MARA) and University Kuala Lumpur for giving me the opportunity and providing me financial support to pursue my PhD study in Strathclyde University.

My deepest and sincere thanks are always to my parents, parent-in-laws, brothers, sisters especially Niza and Husna, brother-in-laws and sister-in-laws for their constant support and prayers. I am so lucky and so proud to have such a great family.

Most importantly, I wish to thank my husband Kushsairy, my eldest girl, Adriana and my twins Arissa and Adam, without their understanding, endless patience and encouragement it would have been impossible for me to finish this work. To them I dedicate this thesis.

Abstract

The thesis addresses the problem of obtaining high-resolution image from a set of one or more low-resolution images. The thesis focused on three building blocks of super-resolution algorithms i.e., image registration for super-resolution, image fusion for super-resolution and super-resolution image reconstruction. These three parts are addressed separately and singular value decomposition-based fusion is introduced before performing interpolation or single-image super-resolution.

An accurate image registration is crucial for super-resolution. An image registration approach for super-resolution based on a combination of Scale Invariant Feature Transform (SIFT), Belief Propagation (BP) and Random Sampling Consensus (RANSAC) is described to automatically register the low-resolution images. The results have shown effective for the removal of the mismatched features in the image.

A novel SVD-based image fusion for super-resolution is developed for integrating the significant features from low-resolution images. The SVD-based image fusion is shown to enhance the super-resolution results.

The implementation of a novel interpolation method based on a linear combination of the bicubic interpolation and their first-order derivates and the use of first-order difference equation to extract the features from the low-resolution images are described and shown to improve the method of single image super-resolution using sparse representation. The proposed method has shown to reduces the computational time and enhance the prior estimation of the high-resolution image as well as the final super-resolution results.

The performance of the algorithms is evaluated using synthetic sequences and also on real sequences subjectively and objectively.

CONTENTS

List of Figures

List of Tables

List of Acronyms

BIQI	Blind image quality indices
BP	Belief Propagation
CS	Consensus set
DCT	Discrete cosine transform
DoG	Difference of Gaussian
DWT	Discrete wavelet transform
EDI	Edge-directed interpolation
FFT	Fast fourier transform
FRI	Finite rate of innovation
HMRF	Huber- Markov random field
HR	High-resolution
IBP	Iterative Backward Projection
ICM	Iterated conditional modes
K-SVD	K-means Singular value decomposition
LDA	Linear discriminant analysis
LLE	Locally linear embedding
LR	Low-resolution
LSI	Linear space-invariant
LSV	Linear space variant
MAP	Maximum a-posteriori estimator
MEDI	Modified edge-directed interpolation
ML	Maximum likelihood
MLP-PNN	Multi-layer perceptron-probabilistic neural network
MMSE	Minimum mean squared error
MOD	Method of optimal directions
MRF	Markov random fields
MSE	Mean square error
MSS	Minimal sample sets
NEDI	New edge-directed interpolation
NSS	Natural scene statistics
PC	Personal computer
PCA	Principal component analysis
PDE	Partial differential equations
PDF	Gaussian Probability Distribution function
POCS	Projection onto convex sets
PSF	Point spread function
PSNR	Peak signal-to-noise ratio
RANSAC	Random Sampling Consensus
RMS	Root mean square
SDP	Semi-definite programming
SIFT	Scale Invariant Feature Transform
SNR	Signal to noise ratio
SVD	Singular value decomposition
SVR	Support vector regression
uHMT	Universal hidden Markov tree

CHAPTER 1

1. INTRODUCTION

Recent years have seen a growing interest in the problem of obtaining high-resolution image from low-resolution images obtained from various sources including medical imaging devices, traffic cameras, CCTV etc (Park et al, 2003). The process to retrieve high-resolution data from a set of one or more low-resolution images is called *super-resolution*.

Two approaches to perform super-resolution image reconstruction are single-image (or single-frame) super-resolution and multi-image (or multi-frame) super-resolution. Single-frame super-resolution usually requires an interpolation of neighbouring pixels to estimate the high-resolution image from a single low-resolution image. Multi-frame super-resolution, on the other hand, uses available information from a low-resolution image sequence to reconstruct a high-resolution image. The super-resolution can be applied to an image sequence to improve the spatial resolution by incorporating into the final high-resolution result the additional new details revealed in each low-resolution image. In particular, camera and scene motion leads to frames in the image sequence containing similar, but not identical information. The additional information available in these frames makes it possible to reconstruct visually superior frames at a higher resolution than that of the original data. The major advantage of this approach is that it may cost less and the existing low-resolution imaging systems can be still utilised.

The super-resolution image reconstruction is proved to be useful in many practical applications where multiple frames of the same scene can be obtained, including medical imaging (Robinson et al., 2010), traffic control (Hao et al., 2009), video applications (Gunturk et al., 2004) and satellite imaging (Chang and Wu, 2007). Super-resolution can also be used in surveillance (Gehani et al., 2007) and forensics

(Lin et al., 2005) where it is often needed to magnify objects in the scene. One of the super-resolution applications is to enhance license plates as shown in Figure 1-1. In this example, the license plates in the low-resolution images are super-resolved using super-resolution method.

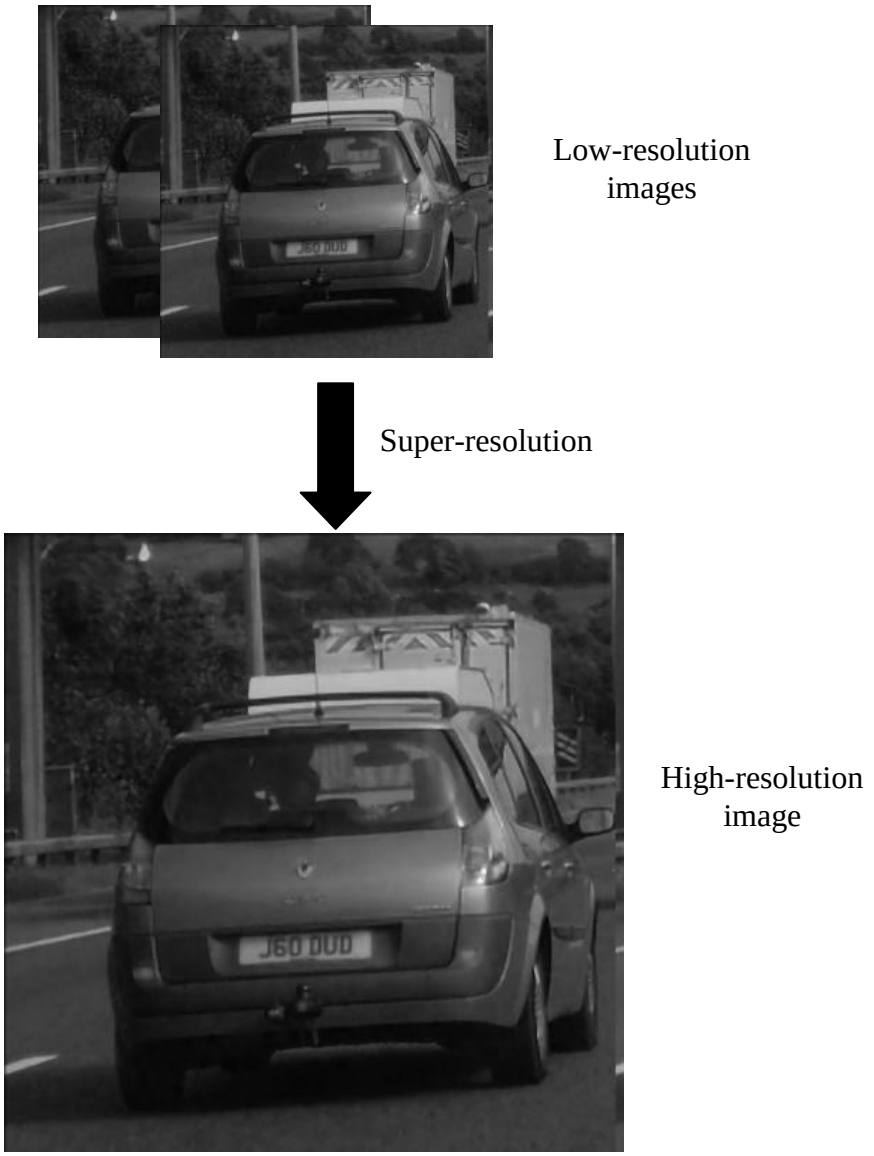


Figure 1-1: Super-resolution of license plates

1.1 Stages in super-resolution

Generally multi-image super-resolution involves image registration, image interpolation and image restoration. The stages can be performed successively or simultaneously. The implementation of these stages are depends on the method used to solve the problem. For example, the non-uniform interpolation based super-resolution performs the stages successively while super-resolution based on estimation method simultaneously performs image interpolation and restoration. The fusion based super-resolution introduces image fusion prior to interpolation stage to integrate the important information from the low-resolution images. Figure 1.2 shows the super-resolution stages.

Image registration is used to register the low-resolution image sequence by finding the disparity between the low-resolution images. The image registration will be discussed in more details in Section 3.5 of Chapter 3. Image fusion is used to integrate complementary information from the low-resolution images. The image fusion will be discussed in more details in Section 3.9 of Chapter 3. An interpolation step is used to improve the resolution by combining all pixel information from low-resolution images according to the estimated sub-pixel shifts. While image restoration is applied to enhance the quality of the low resolution images in the sequence by removal of system blur and noise.

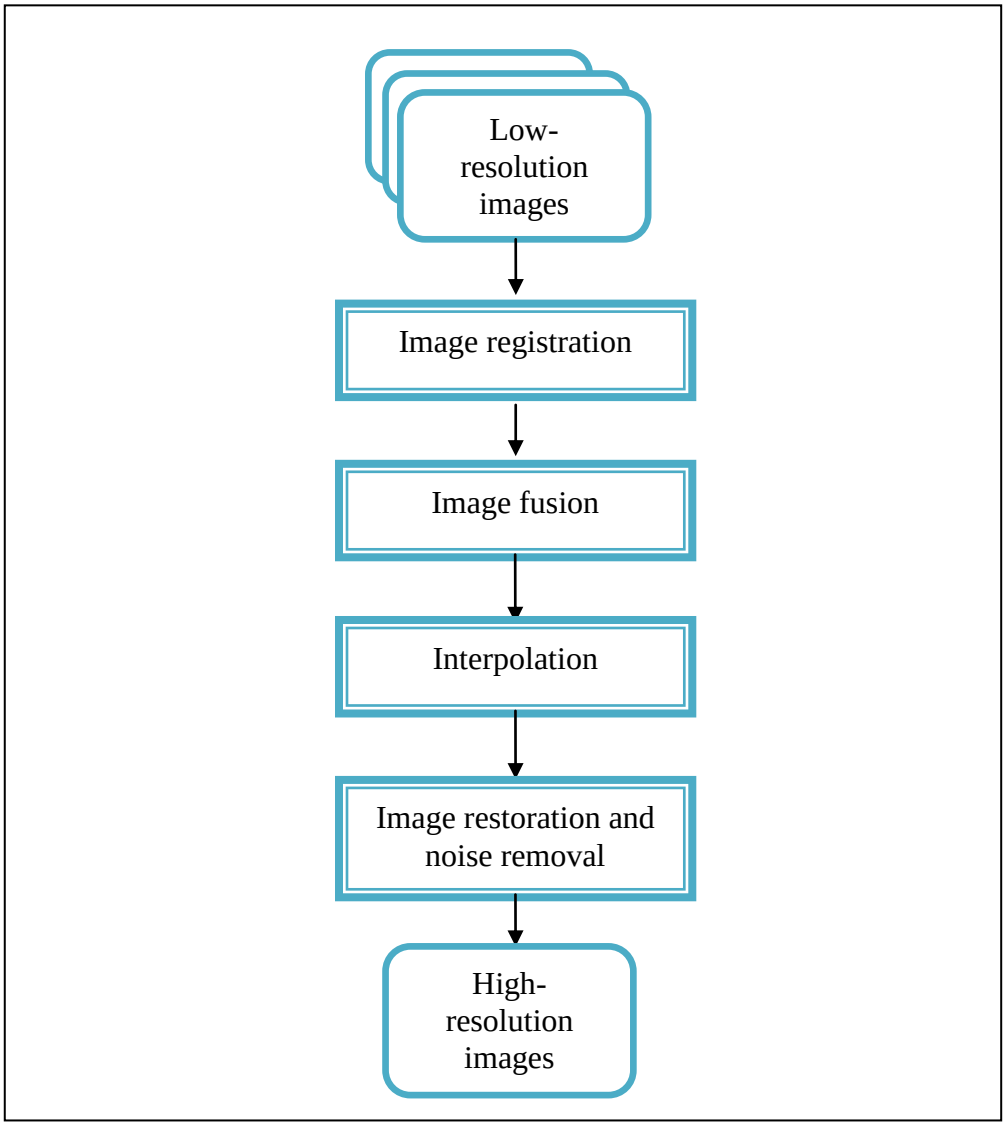


Figure 1-2: The super-resolution stages

This thesis focuses on three building blocks of super-resolution algorithms, i.e., image registration for super-resolution, image fusion for super-resolution and super-resolution image reconstruction. The proposed algorithms are tested on synthetic sequences and also on real sequences. In this thesis, the different parts of the super-resolution problem are addressed separately and introduced singular value decomposition (SVD)-based fusion before performing interpolation or single-image super-resolution.

1.2 Motivation

Images with high pixel density are desirable in many applications such as high-resolution medical images for medical diagnosis (Villanueva et al., 2010), high quality video conference, high definition television broadcasting (S. Mallat.), etc. Imaging devices have limited achievable resolution due to many theoretical and practical restrictions. Thus, software resolution enhancement techniques are very desirable for these applications. The performance of the simple interpolation method is poor due to the aliasing effect. Image super-resolution is therefore a solution to overcome the physical limitations of hardware capabilities (Park et al., 2003).

Super-resolution is important in addressing the current need for the higher resolution images in most applications, the improvement of existing super-resolution methods is important. There has been little work so far on incorporating image fusion and sparse-representation into the super-resolution procedure. The improvement on existing super-resolution methods by inclusion of image-fusion in existing literature is desirable. The new methods for single image super-resolution using sparse representation should also be expected to provide improved super-resolution results. Thus this motivates the work in the development of a super-resolution algorithm that enables high-resolution display of low-resolution images.

1.3 Original contributions of the work

In this work, the novel contributions are described below:

1. A novel image registration for super-resolution based on the combination of Scale Invariant Feature Transform (SIFT) with Belief Propagation (BP) and Random Sampling Consensus (RANSAC) is studied in Chapter 4. The technique incorporates the RANSAC algorithm after BP in SIFT matching procedure to match image features. Incorporating RANSAC after BP makes an extension to the SIFT for image registration which filter out the outliers introduced in the SIFT-BP method.
2. A novel image fusion for super-resolution using singular value decomposition (SVD) is introduced in Chapter 5. This method fuses the singular values of the reference image and the registered image. SVD-based image fusion which is introduced in this thesis in order to preserves the good contrast of the low-resolution image which is degraded in the registration process. The method enhances super-resolution results by integrating the significant features from low-resolution images.
3. A novel multi-image super-resolution scheme by incorporating SVD-based image fusion prior to interpolation. The technique improves the super-resolution result by registering the low-resolution images using SIFT-BP-RANSAC and integrating the important features from low-resolution images prior to interpolation using SVD-based image fusion.
4. A new image interpolation method using bicubic interpolation and first- order difference equations is carried out in Chapter 6. The method provides image with stronger edges by adding the first order difference of the interpolated image to the bicubic interpolated image. This linear combination enhances the important features of the image especially edge information.
5. An improvement of (Zeyde et al., 2010) methods of single-image super-resolution using sparse representation is studied by adapting the proposed new image interpolation for initial estimation of the high-resolution image and the use of the first-order difference equation to extract the features from

the low-resolution images. The proposed interpolation technique enhances the prior estimation of the high-resolution image and provides better features for the image patches. The use of the first-order filter for feature extraction reduces the computational time.

1.4 Organisation of the thesis

The organisation of the thesis is as follows:

Chapter 1 describes the objective and motivation for the research, as well as original contributions that are presented in this work.

Chapter 2 provides a review of super-resolution approaches which describes existing work on single image super-resolution and multi-frame super-resolution. This chapter also presents existing works on image registration and image fusion for super-resolution.

Chapter 3 presents an overview of super-resolution methods particularly the non-uniform interpolation approach. The chapter describes the SIFT, belief propagation and RANSAC algorithms for detecting and matching feature points in the image. This chapter also presents an overview of image fusion and interpolation techniques. The quality metric used for assessing the performance of the methods are explained at the end of the chapter.

Chapter 4 presents an image registration approach for super-resolution based on a combination of Scale Invariant Feature Transform (SIFT), Belief Propagation (BP) and Random Sampling Consensus (RANSAC). The chapter explains how the SIFT algorithm is used to detect and extract the local features in images, BP is used to match the features while RANSAC is adopted to filter out the mismatched points and then estimate the transformation matrix. The chapter provides experimental results by comparing the proposed method with traditional SIFT. This chapter also presents the result of using the proposed approach in the super-resolution application by compared to the traditional approach where SIFT is used for image registration step.

Chapter 5 presents a novel SVD image fusion approach for super-resolution. SVD image fusion is used to enhance the super-resolution results. The objective of using SVD is to integrate the important features from low resolution images into the method. The chapter describes how the proposed method converts the registered and reference image into the SVD domain and then the images' singular values are fused based on the fusion rule before performing the interpolation. The chapter also provides simulation results of applying SVD-fusion prior to interpolation by comparing to standard interpolation techniques and existing learning-based super-resolution approaches.

Chapter 6 presents an improvement of (Zeyde et al., 2010) methods of single-image super-resolution using sparse representation by proposing a novel image interpolation based on linear combination of the original patches with their first-order derivates. The chapter describes how the proposed method enhances the prior estimation and reduces the computation time and hence improves the super-resolution results. The chapter also provides the simulation results by comparing to methods of (Zeyde et al., 2010).

Chapter 7 presents the summary of the research and with some suggestions for future work.

All references can be found at the end of the thesis.

CHAPTER 2

2.LITERATURE REVIEW

2.1 Introduction

There is a great deal of research on super-resolution in the literature and a variety of techniques have been proposed so far. A review of earlier super-resolution techniques can be found in (Borman and Stevenson, 1998b). More recent reviews can be found in (Park et al., 2003) and (Tian and Ma, 2011). Super-resolution methods can be classified into several categories based on the method adopted to solve the super-resolution problem (Borman and Stevenson, 1998, Park et al., 2003, Tian and Ma, 2011). Super-resolution image reconstruction methods can also be categorised into single-image (or single frame) super-resolution and super-resolution from multiple images (or multi-frame super-resolution). Super-resolution can be carried out in two different domains i.e., either in frequency domain or spatial domain.

Extensive work has been published on single-image super-resolution (Yang et al., 2008, Glasner et al., 2009., Adler et al., 2010, Zeyde et al., 2010, Kim and Kwon, 2008) and multi-frame super-resolution (Yang et al., 2008, Glasner et al., 2009., Adler et al., 2010, Zeyde et al., 2010, Kim and Kwon, 2008). Single-image super-resolution methods usually require a large amount of training data to generate efficient learning models, while multi-image/frame super-resolution methods regularly involve an ill-posed inverse problem with matrices of large dimensions.

In this chapter, the existing work on super-resolution approaches is reviewed. The remainder of the chapter is organised as follows: The interpolation-based super-resolution method is presented in Section 2.2. A brief review of early multi-frame super-resolution methods is presented in Section 2.3. In Section 2.4, the existing

work on super-resolution based as an estimation problem is presented. Work on wavelet-based super-resolution is given in Section 2.5 while previous work on non-uniform interpolation based super-resolution is discussed in Section 2.6. Image fusion in super-resolution and learning-based super-resolution is presented in Section 2.7 and 2.8, respectively. Finally, chapter summary are presented in Section 2.9.

2.2 Interpolation-based method

The simplest method to reconstruct a high-resolution image from a single low-resolution image is an interpolation-based method such as bilinear and bicubic interpolation. The fine details in the reconstructed image usually suffer from blur and artefacts. More advanced interpolation methods are proposed in the literature to overcome the shortcomings. A number of edge-directed interpolation (EDI) methods that make use of the local statistical and geometrical properties to interpolate the unknown pixel values are able to obtain high visual quality interpolated images with the use of edge map (Allebach and Wong, 1996, Tam et al., 2010, Li and Orchard, 2001). Work in (Allebach and Wong, 1996) proposed an edge directed interpolation by assigning weights to different structures in the image. (Tam et al., 2010) presents an improvement of the NEDI method, namely the *Modified Edge-Directed Interpolation* (MEDI), which is an extension of work in (Li and Orchard, 2001). (Takeda et al., 2007) developed a kernel regression method by considering the underlying structural kernel regression. (Cha and Kim, 2007) developed an efficient Partial Differential Equations (PDE)-based interpolation algorithm for image super-resolution. A texture enhancement method is incorporated as a post-process of PDE-based interpolation to overcome the drawback of conventional PDE-based interpolation that tends to weaken fine structures. The algorithm is suitable for real-time processing for resolution enhancement of images and video clips that are downloaded through internet connections or wireless communications.

(Lertrattanapanich and Bose, 2002) developed a high-resolution image using Delaunay Triangulation. An algorithm based on spatial tessellation and approximation of each triangle patch in the Delaunay triangulation (with smoothness constraints) by a bivariate polynomial is used to construct a high-resolution image

from a set of low-resolution frames. The high-resolution algorithm is accompanied by a site-insertion algorithm for update of the initial high-resolution image with the availability of more low-resolution frames till the desired image quality is attained. The algorithm is followed by post filtering and is suitable for real-time image sequence processing because of the fast expected (average) time construction of Delaunay triangulation and the local update feature.

2.3 Early works on multi-frame super-resolution

The problem of image super-resolution from a set of low-resolution images received much attention in early 80s as an alternative to interpolation-based methods. (Huang and Tsay, 1984) were first to propose multi-frame image restoration in the frequency domain using the properties of Fourier transforms and disregarded the blur in the imaging process. The point spread function (PSF) was not considered and the sampling rate was assumed to be ideal across all frames. (Tekalp et al., 1992) extended Tsai and Huang formulation by addressing the problem of noise and blur during acquisition. Later, (Kim et al., 1990) extended this to blur and noisy low-resolution images using a recursive implementation based on the weighted least square theory, provided the noise had zero mean and the blur and noise were identical across all low-resolution images. They did not address any motion estimation. Later, Bose et al used total least squares (Bose et al., 1993).

Early methods tended to be implemented in the frequency domain. They were relatively simple but unable to accommodate general scene observation models including spatially varying degradations or blur, non-global relative scene motion or noise models (Borman and Stevenson, 1998a). However, improved performance and greater flexibility was obtained via spatial domain methods. Spatial domain methods improve the super-resolution formulations by incorporating spatially varying noise and blur processes as well as complex motion relationships between the low-resolution images. (Ur and Gross, 1992) suggested an alternative method carried out in the spatial domain based on (Papoulis, 1977) and (Brown Jr, 1981). An algorithm based on a minimum mean squared error (MMSE) approach for multiple image

restoration problems and interpolation of the restored images into a single image is proposed by (Srinivas and Srinath, 1990). The limitations of all the above methods were global uniform translational displacement between the measured images, linear space-invariant (LSI) blur and homogeneous additive noise.

(Irani and Peleg, 1991) introduced an iterative back-projection method, similar to that used in tomography. The main feature of the Irani and Peleg's method is that it iteratively updates the high-resolution estimate by back-projecting the difference between the observed and the simulated low-resolution images. The method allows arbitrary smooth motion flow, although the convergence of the proposed algorithm is demonstrated only for an affine geometric warp between the measured images.

A high-resolution image can also be reconstructed using a Projection Onto Convex Sets (POCS) algorithm (Stark and Oskoui, 1989), where the estimated reconstruction is successively obtained on different convex sets. POCS is one of the early approaches. POCS was first applied to super-resolution by Stark and Oskoui (Stark and Oskoui, 1989). The method of POCS converges to a point in the feasible region called as the feasible solution by successive projection of an initial estimate of the solution onto the convex constraint sets. POCS is an iterative method that employs a priori information about the degradation operator, the noise statistics and the actual high-resolution image distribution. With this prior information an estimate of the low-resolution observation is generated. The difference between the actual and the estimated low-resolution observation is called the residual of the imaging or observation model. The advantages of POCS are its simplicity and that prior information can be incorporated into the algorithm. The disadvantages are non-uniqueness of solution and slow convergence. The POCS formulation for super-resolution reconstruction was explored by (Stark and Oskoui, 1989) and their work was extended by (Tekalp et al., 1992) to include an observation noise. In this method an estimate of high-resolution of the reference image is determined iteratively starting from some arbitrary initialisation. Successive iterations are obtained by projecting the previous estimate onto the convex sets. (Patti et al., 1997) developed a POCS super-resolution technique to consider space varying blur and nonzero

aperture time. The POCS method works better than the Iterative Backward Projection (IBP) since it can easily be combined with the restoration process.

2.4 Super-resolution as an estimation problem

The super-resolution problem has also been tackled as an estimation problem that could be solved by means of statistical estimation, specifically, the maximum likelihood (ML) and the maximum *a-posteriori* estimator (MAP). The ML estimate of the high-resolution image reduces the mean square error (MSE) between the low-resolution images and the simulated ones. (Schultz and Stevenson, 1996) used a ML and MAP framework to solve super-resolution.

(Elad and Feuer, 1994) used the observation models of POCS, ML and MAP approaches to solve for single-image restoration from linear blur and additive noise. (Farsiu et al., 2004) propose an alternate approach using the l_1 norm instead of l_2 norm and robust regularisation based on a bilateral prior to deal with different data and noise models. The l_1 norm is more robust than l_2 norm and it improves the performance of super-resolution in terms of errors in motion and blurs estimation and results in images with sharp edges.

Following a Bayesian framework, (Martins et al., 2007) propose a procedure for super-resolution image reconstruction based on Markov random fields (MRF) (Kindermann, 1980), where a Potts-Strauss model is assumed for the a priori probability density function of the actual image. The first step is aligning all the low-resolution observations over a high-resolution grid and then improving the resolution through the Iterated Conditional Modes (ICM) algorithm (Besag, 1986). The method is analysed by considering a number of simulated low-resolution and globally translated observations.

There exists extensive work in the literature on super-resolution using estimation techniques. More details on the existing work in this area are not discussed here as the thesis is focused on non-uniform based interpolation and learning-based approaches.

2.5 Wavelet-based super-resolution

Wavelet domain methods have been employed to solve the super-resolution problem. This is motivated by the properties of wavelets which provide an efficient multi-scale representation of an image for recovering the high frequency information (Nguyen and Milanfar, 1999). (Nguyen and Milanfar, 1999) propose an efficient algorithm based on representing the low-resolution images using wavelet coefficients and relating these coefficients to the desired super-resolution image. (Bose and Chappalli, 2004) propose a super-resolution method based on second-generation wavelets. A Biorthogonal wavelet system is proposed in (Shen and Sun, 2004) for high-resolution image reconstruction.

(El-Khamy et al., 2005b) perform image registration in the wavelet domain, and then the registered low-registration wavelet coefficients are fused to obtain a single image. This single image is then interpolated to get a higher-resolution image. Interpolation in the Discrete Wavelet Transform (DWT) domain been introduced in (Demirel and Anbarjafari, 2011) to produce high-resolution images.

(Ji and Fermüller, 2006) propose a robust wavelet super-resolution approach which performs a wavelet-based denoising scheme in each iteration of super-resolution reconstruction. (Chappalli and Bose, 2005) develop a simultaneous denoising and super-resolution approach in the wavelet domain. An efficient denoising and adaptive interpolation in the wavelet domain are also used in (Liyakathunisa and Ananthashayana, 2009) for a robust wavelet based super-resolution reconstruction of low-resolution images. An image interpolation method which is combined with deblurring and denoising is proposed by (Li et al., 2007).

The MAP (Maximum a Posteriori) estimate is adopted to obtain a super-resolution image from a sub-sampled, blurred and contaminated image in the wavelet domain. The universal hidden Markov tree (uHMT) theory in the wavelet domain is applied to construct a prior model for the MAP estimate. The results of (Li et al., 2008) show that reconstructed images are much better and sharper than those recovered images by the Huber-Markov random field (HMRF) prior model for MAP in the space domain.

2.6 Non-uniform interpolation based super-resolution

The non-uniform interpolation approach usually consists of three (Park et al., 2003).
1) Registration: includes motion estimation and image registration to map the pixels from all available low-resolution frames onto a common reference frame.
2) Interpolation: this step is used to map the registered pixels (usually non-uniformly distributed) onto a rectangularly sampled super resolution grid.
3) Restoration: this step is needed to limit the effect of sensor blur and noise in the low resolution frames and to reconstruct sharper image details.

The three steps can be implemented separately or simultaneously according to the reconstruction approach adopted.

An image registration and image fusion for super-resolution particularly the application of SIFT in image registration and different fusion techniques are reviewed in the next section.

2.6.1 Image registration in super-resolution

Image registration is used to match two or more images of the same scene taken at different times, from different viewpoints, and / or from different sensors. It is the process of spatially matching two images, i.e., the reference and target images so that the corresponding coordinate points in the two images correspond to the same physical region of the scene being imaged.

Accurate image registration is a crucial step in the super-resolution process. In super-resolution, image registration is used to register low-resolution image frames. A subpixel-registered image sequence of the same scene potentially contains more information than any single view alone. Image registration enables sub-pixel shifts and hence combines useful information from multiple frames. There is a great deal of image registration research in the literature. The survey by (Zitova and Flusser, 2003) provides an overview of image registration methods. Reported image registration methods can be classified into two main approaches: intensity-based methods and feature-based methods. Intensity-based methods compare the intensity

patterns in images via correlation metrics, while feature-based methods find correspondence between image features.

In recent years, a number of different image registration techniques have been used for super-resolution. Optical flow image registration by (Lucas and Kanade, 1981) is used in optical flow super-resolution by (Baker and Kanade, 1999). Another approach can be found in (Farsiu et al., 2004) and (Vandewalle et al., 2006) where a frequency domain registration method that removes the aliased part of the spectrum of the low-resolution images before registering the images is proposed. (Fan et al., 2006) proposed an improve registration method of (Keren et al., 1988) and used the projection onto convex set (POCS) method to reconstruct high-resolution image from several low-resolution image sequences.

(Madhusudhan and Pais, 2007) used frequency domain registration by segmenting the video based on histogram for enhancing the resolution of video. Then bicubic interpolation is applied to the video to generate the super-resolution video frames. A computationally simple super-resolution algorithm using an adaptive Wiener filter is developed by (Hardie, 2007). The algorithm uses gradient based sub-pixel registration to position each low-resolution pixel value on a common spatial grid that is referenced to the average position of the input frames. Then used a weighted nearest neighbour interpolation approach to reconstruct high-resolution image and finally wiener filter is applied to reduce blur and noise.

(Robinson et al., 2009) use the method of variable projections that explore multi-frame registration under aliasing, while (He et al., 2007) focus on multi-frame registration under rotations and shifts. Using recent results from the sampling theory for signals with Finite Rate of Innovation (FRI), (Baboulaz and Dragotti, 2007) proposed a new technique for sub-pixel extraction from low-resolution images of local features like step edges and corners for image registration. By exploiting the knowledge of the sampling kernel, the techniques are able to locate exactly the step edges on synthetic images.

2.6.2 SIFT for image registration

In this section, an image registration approaches using SIFT and RANSAC are reviewed. The Scale Invariant Feature Transform (SIFT) is one of the most popular feature-based methods introduced by (Lowe, 2004). SIFT is able to detect and describe local features that are invariant to scaling and rotation. Various improvements have been made to the SIFT algorithm, and a recent advance reported in (Cheng et al., 2009) uses belief propagation to achieve better matching than is achieved with the minimum Euclidean distance method (Lowe, 2004). In (Yuan et al., 2008) Random Sampling Consensus (RANSAC) (Fischler and Bolles, 1981) is used to improve the mismatch points in the SIFT algorithm and then a support vector machine is adopted to estimate the transformation matrix. Geometrical information between the descriptors is used in (Brown and Lowe, 2007) where SIFT with RANSAC is used for robust homography estimation with a probabilistic model to verify the match. In, (Seong and Park, 2008), matching in the traditional SIFT algorithm is improved by using principal-component analysis (PCA) and RANSAC is used to estimate the homography matrix. In (Tang et al., 2008) the SIFT algorithm is used to register medical microscopic image sequences where the Gaussian weighting function is used to optimise the feature descriptor.

(Amintoosi et al., 2009a) used SIFT features for registration of images under a projective model in a super-resolution problem. The method has been improved in (Amintoosi et al., 2009b) by using displacement restriction criteria for removing the incorrect matches with the assumption of a Gaussian Probability Distribution Function (PDF) for the mentioned displacements.

Robust SIFT features are used for image registration in (Vrigkas et al., 2011). The image registration part is divided into two steps. First the low-resolution images are registered by finding the corresponding SIFT features then the registration parameter is fine tuned together with the estimation of the high-resolution image in an iterative procedure using the maximisation of the mutual information criterion.

SIFT features have also been used in (Zhu et al., 2010) for image registration in the mosaicing problem. The SIFT features and image mosaicing are integrated in order to construct a full high-resolution unmanned aerial vehicle (UAV) map.

2.7 Image fusion in super-resolution

In this section, the super-resolution methods that include fusion in the super-resolution steps are presented. Image fusion is needed to integrate useful information from the low-resolution images.

Various methods have been reported in the literature which incorporates image fusion in the super-resolution framework. Image fusion is the process of incorporating complementary information that exists in different observation images into a single unified image (Hill et al., 2002). It has been successfully used in many fields such as remote sensing, robotics, and medical applications. There are a number of different techniques for image fusion such as wavelet transform methods and (Li et al., 1994) and discrete cosine transform methods (Tang, 2004).

Singular value decomposition (SVD)-based fusion has been used in many papers such as (Pang et al., 2004, Jia and Zhang, 2009, Cernekova et al., 2005, Wang et al., 2007). In (Pang et al., 2004), a fusion of a linear discriminant analysis (LDA) and SVD is used for face recognition. LDA is a classical method for feature extraction and dimensionality reduction which has been widely used in several classifications. SVD is used as an enhancement of classical LDA by overcoming the singularity problem and to reduce dimensions of the scatter matrix. In (Jia and Zhang, 2009) the curvelet transform and SVD are used for facial feature extraction and recognition. The curvelet energy features of low-frequency and the high-frequency are first extracted and then singular value compression is performed and fused. SVD-based fusion is applied in (Repperger et al., 2009) for image registration. In (Cernekova et al., 2005) the features obtained by SVD are fused for the detection of shot boundaries in video sequences. In (Cernekova et al., 2005), SVD is used for its capabilities to derive a refined low-dimensional feature space from the high-dimensional raw feature space, where similar video patterns are placed together and can be easily clustered. In (Wang et al., 2007) the SVD-based fusion is applied to the multiple spectrum face recognition problem. In (Wang et al., 2007) the image is decomposed into three layers with different energy distribution and their eigenvalues are fused for reconstruction.

Image fusion has also been used in the super-resolution image reconstruction process because of its ability to preserve important features from multiple images. For example, in (El-Khamy et al., 2005a), the ill-posedness is removed by breaking the problem into four consecutive steps: registration, restoration, image fusion and interpolation. The low-resolution images are assumed to be correctly registered prior to the multi-channel restoration step. The image fusion step is introduced with the purpose of integrating the features obtained from each output of the multichannel restoration step. Then the fused image is interpolated using maximum entropy interpolation.

In recent years, a number of different fusion techniques have been used for super-resolution (Liyakathunisa and Ananthashayana, 2009, Liyakathunisa, 2009, Bhushan et al., 2010). A wavelet-based fusion approach is used for super-resolution in (El-Khamy et al., 2005a, Liyakathunisa and Ananthashayana, 2009, Liyakathunisa, 2009). In (El-Khamy et al., 2005a) a discrete wavelet transform (DWT) image fusion approach is used to integrate the outputs of the maximum entropy image restoration step, while in (Liyakathunisa and Ananthashayana, 2009) the images are registered using affine transformation and the aligned images are then fused in the DWT domain which simultaneously reduces the noise. In (Liyakathunisa, 2009) a wavelet lifting scheme is adopted as a fusion technique. The low-resolution images are registered using an FFT-based algorithm, then these images are decomposed and fused using a wavelet lifting scheme before performing wavelet interpolation. Recently, a framelet-based fusion is used for super-resolution blind reconstruction in (Bhushan et al., 2010).

(Liyakathunisa and Ananthashayana, 2009) proposed a super-resolution blind reconstruction technique for linearly degraded images. The algorithm is divided into three parts: image registration, wavelets based fusion and image restoration. The three low resolution images are considered which may be sub pixel shifted, rotated, blurred or noisy, the sub-pixel shifted images are registered using an affine transformation model. A wavelet based fusion is performed and the noise is removed using soft thresholding. The technique reduces blocking artifacts and also smoothes the edges and it is also able to restore high frequency details in an image.

2.8 Learning-based super resolution

Some super-resolution techniques use learning-based methods to tackle the super-resolution problem. Learning-based super-resolution requires a sufficient and representative database of patches pairs from high-resolution images and their corresponding low-resolution images or the coefficients of alternative representation. The database is used for training. For the reconstruction, an appropriate high-frequency or smoothing information needs to be added to the estimated high-resolution image. (Freeman et al., 2002) proposed a novel approach for interpolating high-frequency details from a training set. (Patil et al., 2008) proposed a better learning algorithm using a real image training set that enhances the high frequency information. The training set consists of preprocessed images which capture the structural correlation. The technique learns the fine details that correspond to different image structures seen at a low-resolution and then used those learned relationships to predict fine details in other images. (Candocia and Principe, 1999) proposed an adaptive super-resolution based on local correlations. The procedure locally projects image samples onto a family of kernels that are learned from image data. First, an unsupervised feature extraction is performed on local neighbourhood information from a training image. These features are then used to cluster the neighbourhoods into disjoint sets for which an optimal mapping relating homologous neighbourhoods across scales can be learned in a supervised manner. A super-resolved image is obtained through the convolution of a low-resolution test image with the established family of kernels.

A thorough investigation of the application of support vector regression (SVR) to the super-resolution problem is conducted by (Ni and Nguyen, 2007). Prior to their study, the SVR problem was enhanced by finding the optimal kernel. This was done by formulating the kernel learning problem in SVR form as a convex optimization problem, specifically a semi-definite programming (SDP) problem. The idea was improved upon by observing structural properties in the discrete cosine transform (DCT) domain to aid in learning the regression. Further improvement involved a combination of classification and SVR-based techniques, extending works in resolution synthesis.

(Miravet and Rodriguez, 2005) proposed a method based on the use of a hybrid multi-layer perceptron-probabilistic neural network (MLP-PNN) architecture, which incorporates the use of the local image models as the input for the neural processing stage. The data dimensionality was firstly reduced by application of PCA. An MLP, trained on synthetic sequences with various amounts of noise, estimates the high-resolution image data.

(Chang et al., 2004) used a locally linear embedding (LLE) method to generate the high-resolution image patch which depends on multiple nearest-neighbours in the training set. (Datsenko and Elad, 2007) proposed finding several candidates for high quality patches at each pixel position in the observed low resolution image, these candidates are found as the nearest-neighbours image database that contains pairs of corresponding low-resolution and high-resolution image patches. These patches are used as the prior image model and then merged onto an MAP cost function to arrive at the closed-form solution of the desired high-resolution image.

Recently various attempts have been made to regularise the ill-posed inverse problem of image super-resolution using sparse representations of low-resolution image patches (Yang et al., 2010, Zeyde et al., 2010). Sparse representation is applied in many fields in signal processing including applications such as image denoising, image restoration and classification (Elad and Aharon, 2006, Huang and Aviyente, 2007, Mairal et al., 2008). Image patches can be represented as a sparse linear combination of elements from an over-complete image patch dictionary (Wang et al., 2011, Kim and Kwon, 2008). The idea is to seek a sparse representation for each patch of the low-resolution input, followed by exploiting this representation to generate the high-resolution output. By jointly training two dictionaries for the low-resolution and high-resolution image patches, the sparse representation of a low resolution image patch can be applied with the high-resolution image patch dictionary to generate a high-resolution image. Yang et al proposed a sparse representation based super-resolution method (Yang et al., 2008), predicted on the assumption that the high-resolution and low-resolution patches have the same sparse representation coefficients with respect to a high-resolution dictionary and a

corresponding low-resolution dictionary. Recently the work has been extended and improved by (Zeyde et al., 2010).

(Sun et al., 2011) proposed a method to search for a mapping between a pair of low-resolution and high-resolution image patches in the gradient domain by learning a generic image database and the input image itself. Given a low-resolution image, the high-resolution image is reconstructed using sparse representation in the gradient domain and solving a Poisson equation. (Yin et al., 2012) used a texture constrained sparse representation for single image super-resolution. Firstly, the low-resolution observed image is segmented into different texture regions. Through pre-prepared texture databases, the low-resolution regions are classified into different texture categories using the designed texture classifier. Then, the high-resolution segments are reconstructed by sparse representation with relevant texture dictionaries. Integrating all segments, the high-resolution result is obtained.

(Jing et al., 2010) used both the human visual perception and image gradient features in the proposed super-resolution framework, the image total variation are decomposed into structural components and texture components. Based on the theory of sparse signal representation, the K-SVD method is used to generate an ultra-complete dictionary and to achieve the reconstruction of the texture component. The super-resolution reconstruction of the whole original low-resolution image is realised by fusing them with the bi-cubic interpolated image reconstruction of the structural components. The proposed method, without external image database support, brings in the whole image information while depending on the fixed neighbourhood.

(Ravishankar et al., 2011) used singular values as priors for regularizing the ill-posed nature of the single image super-resolution problem. Method of Optimal Directions algorithm (MOD) (Engan et al., 1999) has been used in the proposed algorithm for obtaining high-resolution and low-resolution dictionaries from training image patches. Using the two dictionaries, the given low-resolution input image is super-resolved.

(Zhang et al., 2011) proposed a method that speeds up the sparse-representation based super-resolution method by dual dictionary learning. The dual dictionary learning method is used and further extended to learn the dual filters together with

coupled dictionary training procedure for sparse representation based super-resolution. Yu et al proposed using contextual dictionaries for sparse-representation based super-resolution. The contextual dictionaries refer to the local relationships between patches rather than using the entire samples. This will lead to efficient discriminative features (Yu et al., 2011).

2.9 Summary

This chapter presents the review of super-resolution approaches. Existing work can be categorised into single image super-resolution and multi-image super-resolution.

The existing works on both methods have been presented in this chapter. For work on single image super-resolution, an interpolation-based method and learning-based method are described in this chapter. The interpolation-based method is the simplest method to reconstruct a high-resolution image from a single low-resolution image. More advanced interpolation-based methods and learning-based methods are proposed in the literature to overcome the drawback of these basic interpolation techniques.

An alternative to interpolation-based methods is multi-frame super-resolution where set of low-resolution images are used to reconstruct high-resolution images. There exist many variations on the proposed methods in the literature. This chapter presents the non-uniform interpolation approach which divides the super-resolution problem into three main stages i.e. image registration, interpolation and restoration. Existing work on image registration and image fusion for super-resolution is discussed and more details on image registration using SIFT and SVD-based fusion is described.

CHAPTER 3

3.KEY COMPONENTS OF SUPER-RESOLUTION

3.1 Introduction

Super-resolution is a set of techniques that in some way enhance the resolution of an imaging system, increasing image or video resolution. The term super-resolution is typically used in the literature to describe the process of obtaining a high-resolution image or a sequence of high-resolution images from a set of low-resolution observations. The super-resolution image reconstruction is useful in many applications including medical and satellite imaging, and video applications.

Multi-frame super-resolution uses multiple low-resolution images which provide different views of the same scene. The availability of multiple low-resolution images captured from the same scene makes super-resolution reconstruction possible. Super-resolution will be possible if the low-resolution images are shifted with subpixel precision and have different sub-pixels shifts. If low-resolution images are shifted by integer units, then each image contains the same information, and then there is no new information that can be used to reconstruct a high-resolution image so super-resolution serves no purpose. Individual frames can also be obtained from a video sequence. If these scene motions are known or can be estimated within sub-pixel accuracy, then super-resolution image reconstruction is possible by combining these new information contained in each low-resolution image**.**

This chapter presents key components of super-resolution and its stages. The remainder of the chapter is organised as follows. The sampling theorem is given in Section 3.2. In Section 3.3, the observation model of the super-resolution problem is presented. A non-uniform interpolation approach is given in Section 3.4. Section 3.5 outlines the image registration techniques followed by SIFT, Belief Propagation and RANSAC algorithm in Section 3.6, 3.7 and 3.8. An overview of image fusion is

presented in Section 3.9. Section 3.10 discusses interpolation techniques. The quality metric used for assessing the performance of the methods are discussed in Section 3.11. Finally, summary are presented in Section 3.12.

3.2 The Sampling theorem

The Nyquist–Shannon sampling theorem states that perfect reconstruction of a signal is possible when the sampling frequency (F_s) is greater than twice the maximum frequency $(F_{\max})$ of the signal being sampled, or equivalently, when the Nyquist frequency (half the sample rate) exceeds the highest frequency of the signal being sampled. If lower sampling rates are used, the original signal's information may not be completely recoverable from the sampled signal (Shannon, 1949).

For signal upsampling, consider a discrete signal $f(k)$ on a radian frequency digital frequency range. Let L denote the upsampling factor. The process starts by adding $L-1$ zeros between each sample in $f(k)$ or equivalently define as

$$g(k)=\begin{cases} f\left(\frac{k}{L}\right) & \text{if } \frac{k}{L} \text{ is an integer} \\ 0 & otherwise \end{cases} \tag{3-1}$$

$$F_{\max}=\frac{LF_s}{2}$$

Then perform low-pass filter with frequency cut off at $\frac{\pi}{L}$ to eliminate copies.

3.3 Observation Model

The observation model is a model that relates the high-resolution image to the observed low-resolution images (Park et al., 2003). Figure 3-1 from top to bottom shows how the high-resolution image is transformed into an observed low-resolution image. The reverse process shall be applied in order to reconstruct the high-resolution image from low-resolution images.

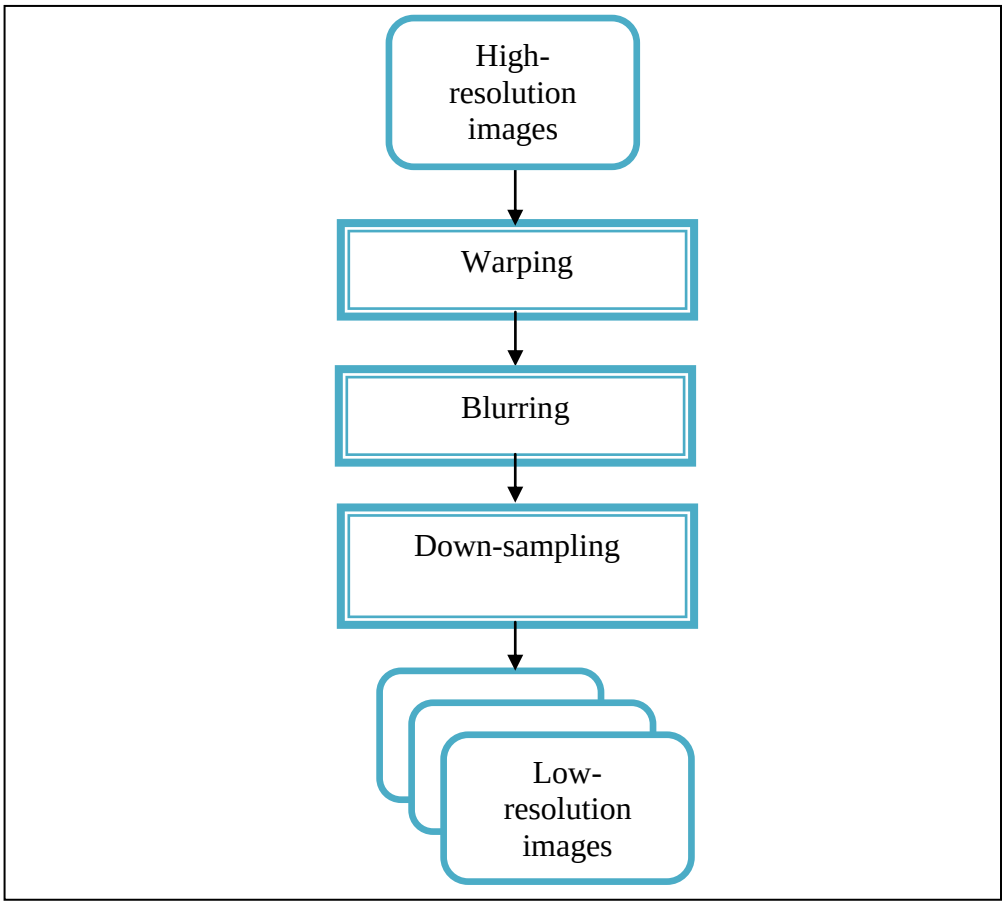


Figure 3-1: Observation model relating LR imaged to HR images (Park et al., 2003)

In super-resolution, the mathematical model is assumed as the *i*th low-resolution image y_i that has $N_1 \times N_1$ pixels and the original high-resolution image x is of size $L_1N_1 \times L_2N_2$ where L_1 and L_2 represent the down-sampling factors in the horizontal and vertical directions respectively. Generally, the *i*-th low-resolution image y_i can be obtained from the original high-resolution image x through blurring, warping and down-sampling. The mathematical model can be expressed as follows:

$$y_i = DH_iM_ix + n_i \tag{3-2}$$

where D is the down-sampling matrix and n_i represents the additive white Gaussian noise. M_i is the warping matrix which contains global or local translation, rotation etc and can be determined through motion estimation algorithms. H_i is the blurring matrix.

The motion that occurs during the image acquisition is represented by warp matrix M_i. It may contain global or local translation, rotation, and so on. Since this information is generally unknown, the scene motion for each frame with reference to one particular frame needs to be estimated. The warping process performed on high-resolution image x is actually defined in terms of low-resolution pixel spacing during the estimation process. Thus, this step requires interpolation when the fractional unit of motion is not equal to the high-resolution sensor grid.

Blurring may be caused by an optical system (e.g., out of focus, diffraction limit, aberration, etc.), relative motion between the imaging system and the original scene, and the point spread function (PSF) of the low-resolution sensor. It can be modeled as linear space invariant (LSI) or linear space variant (LSV), and its effects on high-resolution images are represented by the matrix H_i.

This blur is usually modeled as a spatial averaging operator as shown in Figure 3-2. In the use of super-resolution reconstruction methods, the characteristics of the blur are assumed to be known. However, if it is difficult to obtain this information, blur identification should be incorporated into the reconstruction procedure.

The subsampling matrix D may generate aliased low-resolution images from the warped and blurred HR image. Although the size of low-resolution images is the same here, in more general cases, the different size of low-resolution images can be addressed by using a different subsampling matrix (e.g., D_k for $1 \le k \le p$).

The next section will discuss in details one of the super-resolution approach i.e. non-uniform interpolation approach.

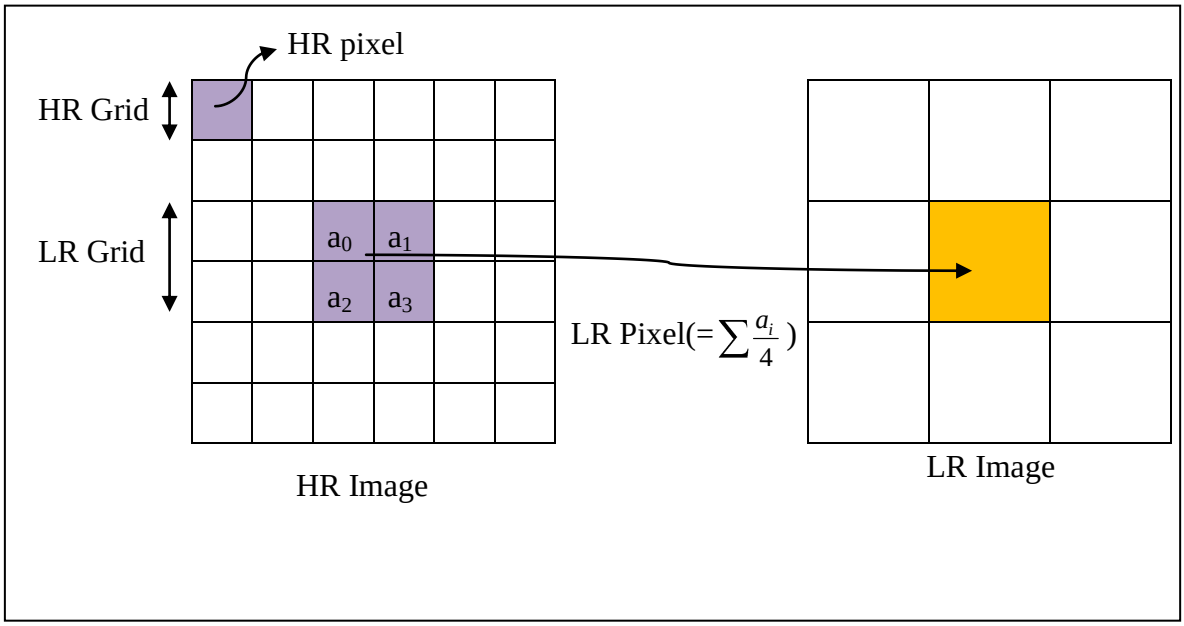


Figure 3-2: Low resolution sensor PSF (Park et al., 2003)

3.4 Non-uniform interpolation approach

There exist many approaches to solve the super-resolution problem as described in Chapter 2 in this thesis. This section focuses on the non-uniform interpolation method. The non-uniform interpolation method aims to fuse all information from low-resolution images as affectively as possible. Generally, three problems must be solved for non-uniform interpolation methods: image registration, interpolation and restoration. The three stages are performed successively as shown in Figure 3-3.

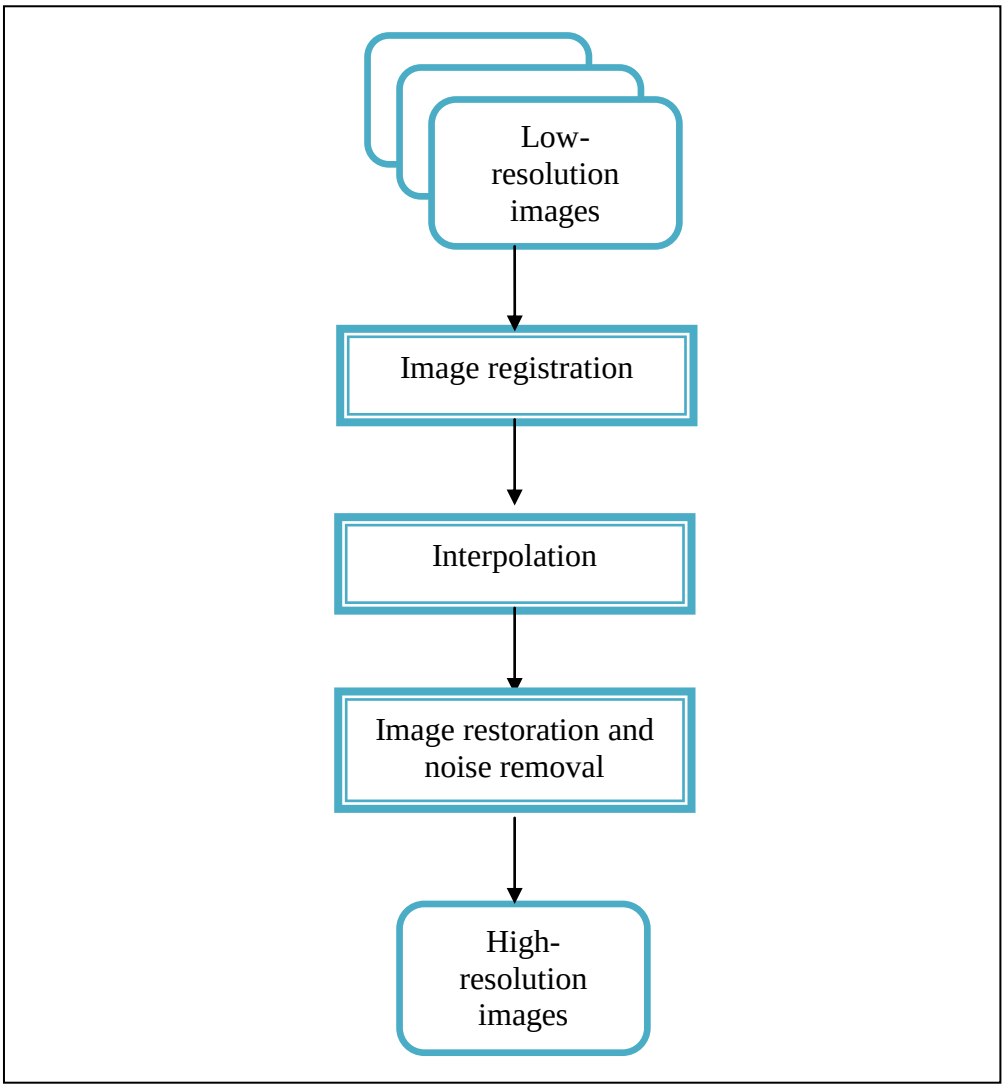


Figure 3-3: Three steps for non-uniform interpolation approach

The first step is to register the low-resolution image sequence. An image sequence of the same scene potentially contains more information than any single view alone. Image registration is used to determine the disparity between the low-resolution images and hence combines useful information from multiple frames. An accurate image registration has crucial implications on overall super-resolution performance. The image registration will be discussed in more details in Section 3.5.

The second step is to combine all the low-resolution frames according to the estimated sub-pixel shifts. Since the shifts between the low-resolution images are

arbitrary, the images will not always match up to a uniformly high-resolution grid. Thus, non-uniform interpolation is necessary to obtain a uniformly spaced high-resolution image from a non-uniformly spaced composite of low-resolution images. Non-uniform interpolation between low-resolution images is used to improve resolution.

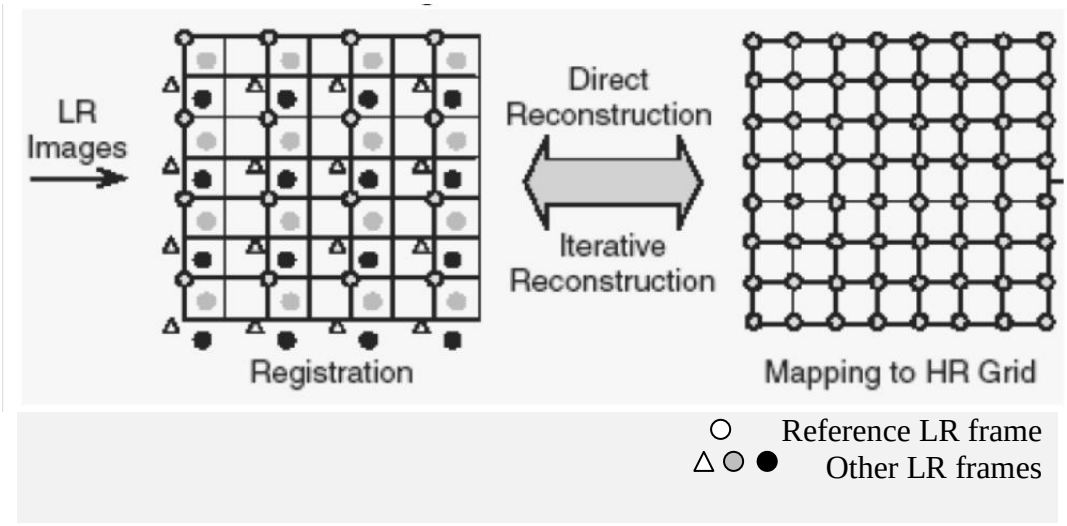


Figure 3-4: Mapping of non-uniform grid to uniform grid (Park et al., 2003)

The last step is to enhance the quality of the low resolution images in the sequence by removal of system blur and noise. In this step, any image deblurring and denoising techniques can be applied. Next, each of the steps will be explained in details.

3.5 Image registration

Image registration is the process of determining the optimal spatial transformation that maps one image to another. A spatial transformation must be found so that the points in one image can be related to their corresponding points in the other. The images to be registered have a mapping function to the reference image. Spatial transformations are applied to the images using the mapping functions to align images to a reference. This mapping is the transformation matrix and its matrix is called the homography matrix (Szeliski, 2005). Homography relates the pixel co-ordinates in the two images. It is a 3 by 3 matrix *H* which has eight degrees of freedom to represent spatial transformation in 2D space.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \qquad \begin{bmatrix} Sx' \\ Sy' \\ S \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3-3}$$

$$x' = \frac{xh_{11} + yh_{12} + h_{13}}{S}$$
$$y' = \frac{xh_{21} + yh_{22} + h_{23}}{S}$$
$$S = xh_{31} + yh_{32} + 1$$

where *x'* and *y'* represents the transformed coordinates and *x*, *y* are the original coordinates of the pixels.

There is a great deal of image registration research in the literature. In this section, background details on SIFT, belief propagation and RANSAC is discussed as the work on image registration of this thesis focus in this area.

3.6 SIFT algorithm

The Scale Invariant Feature Transform (SIFT) is one of the most popular feature-based methods introduced by (Lowe, 2004). The SIFT algorithm presents a method for extracting local features that are tolerant to changes in scale, illumination and rotation (Lowe, 2004). There are four main steps when extracting local features:

(i) **Keypoints detection**: First, a set of Difference of Gaussian (DoG) images covering the range of scales are generated using a Gaussian pyramid and then local minima and maxima are tracked through scale space by comparing each pixel with its 26 nearest neighbours. Each local minima and maxima form a candidate keypoint.

(ii) **Keypoints localisation**: The second step is to determine location and scale for each candidate keypoint. Points with low contrast and poorly localized edge points are rejected.

(iii) **Orientation assignment**: In the orientation assignment step, each keypoint is assigned a direction based on the local image gradient. Additional keypoints may be created if strong directions exist.

(iv) **Keypoints descriptor generation**: Lastly, the local neighbourhood of each keypoint is used to generate an array of SIFT descriptors.

The SIFT algorithm extracts features in an image that correspond to local extrema of difference-of-Gaussian filters at different scales. Detection of interest points in the image start by the convolution of the image with Gaussian filters at different scales and the generation of difference-of-Gaussian images from the difference of adjacent blurred images.

A Gaussian-blurred image described as the formula

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{3-4}$$

where L is a blurred image, I is an image, x and y are the location coordinate, σ is the scale parameter, the * is the convolution operator in x and y and G is Gaussian blur operator.

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{\frac{-(x^2+y^2)}{2\sigma^2}} \quad (3\text{-}5)$$

The blurred images are grouped by octave. An octave corresponds to doubling the value of σ. A fixed number of blurred images per octave are obtained by selecting the value of *k* (blur level). This will ensure that the same figure of difference-of-Gaussian images per octave is obtained.

Then the difference-of-Gaussian images are generated from the difference of adjacent blurred images which is a difference between the Gaussian-blurred images at scales σ and *k*σ.

$$D(x,y,\sigma)=L(x,y,k\sigma)-L(x,y,\sigma) \quad (3\text{-}6)$$

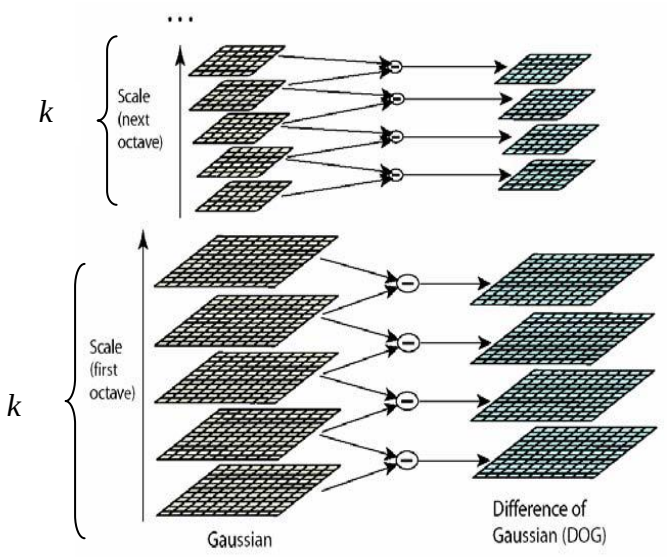


Figure 3-5: Diagram showing the blurred images at different scales, and the computation of the difference-of-Gaussian images (Lowe, 2004).

Interest points or keypoints are identified as local maxima or minima of the DoG images across scales. Each pixel in the DoG images is compared to its 8 neighbours at the same scale, plus the 9 corresponding neighbours at neighbouring scales. If the pixel is a local maximum or minimum, it is selected as a candidate keypoints.

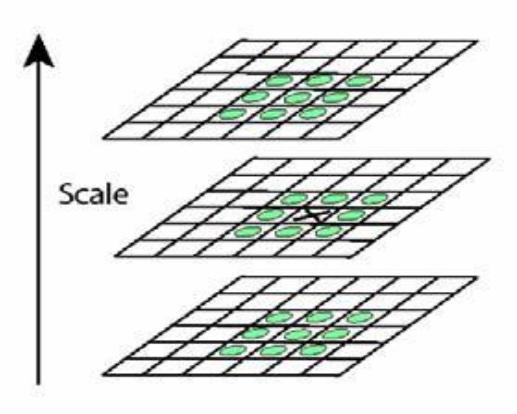


Figure 3-6**:** Local extrema detection, the pixel marked × is compared against its 26 neighbors in a 3 × 3 ×3 neighbourhood that spans adjacent DoG images (Lowe, 2004)

The position for each candidate keypoint is determined by interpolation of the nearby data. Then all the candidate keypoints with low contrast are removed and responses along edges are eliminated. Accurate keypoints' locations are computed by discarding points below a predetermined value.

$$D(\hat{X}) = D + \frac{1}{2}\frac{\partial D^T}{\partial x}\hat{X} \tag{3-7}$$

In (3-7) is calculated by setting the derivative $D(x, y, \sigma)$ to zero. The extremas of difference-of-Gaussian have large principal curvatures along edge, it can be reduced by checking

$$\frac{Tr(H)}{Det(H)} < \frac{(r+1)^2}{2} \tag{3-8}$$

H in (3-8) is a 2× 2 Hessian matrix, r is the ratio between the largest magnitude and the smallest one.

After that an orientation is assigned to the keypoint. The gradient orientation histogram is computed in the neighbourhood of the keypoint in order to determine the keypoint orientation. Keypoint orientation is used to collect the gradient directions and magnitudes around each keypoint.

To achieve invariance to rotation, the gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ are precomputed as the following equations.

$$m(x,y) = \sqrt{L(x+1,y) - L(x-1,y))^2 + L(x,y+1) - L(x,y-1))^2} \qquad (3\text{-}9)$$

$$\theta = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right)$$

The magnitude and orientation is calculated for all pixels around the keypoint. Then, a histogram is created for this. In this histogram, the 360 degrees of orientation are broken into 36 bins (each 10 degrees). Each sample added to the histogram is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a σ that is 1.5 times that of the scale of the keypoint. The histogram will have a peak at some point.

Peaks in the histogram correspond to dominant orientations. A separate keypoint will be created that relates to the direction corresponding to the histogram maximum, and any other direction within 80% of the maximum value. All the properties of the keypoint are measured relative to the keypoint orientation, which provides invariance to rotation.

The SIFT descriptor is generated by calculating orientations and magnitude of the pixel neighbourhood relative to the keypoint in question. Each descriptor consists of an area of 16 x 16 pixels and quantized into 8 bins. Each pixel contributes its magnitude to the bin closest to its orientation. Histograms contain 8 bins each, and each descriptor contains an array of 4 histograms around the keypoint. This leads to a SIFT feature vector with $4\times 4\times 8 = 128$ elements. This vector is normalized to ensure invariance to changes in illumination. More details on how SIFT descriptors are generated can be found in (Lowe, 2004).

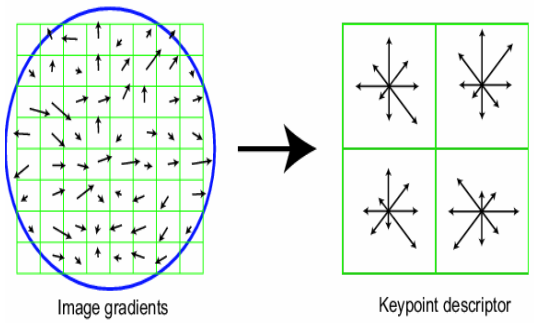


Figure 3-7: Keypoint magnitude is generated and weighted by a Gaussian window (circle) (Lowe, 2004).

3.7 Belief propagation (BP)

Belief propagation is a method to calculate the marginal distributions for random variables through passing messages in factor graphs (Kschischang et al., 2001, Yedidia et al., 2003, Loeliger, 2004). Suppose that $F(x_1,...,x_n)$ factors into a product of several local functions, each having some subset of $\{x_1,...,x_n\}$ as arguments i.e. suppose that

$$F(x_1, x_2, ..., x_n) = \prod_{j \in J} f_j(X_j) \qquad (3\text{-}10)$$

Where J is a discrete index set, X_j is a subset of $\{x_1,...,x_n\}$, and $f_j(X_j)$ is a function having the elements of X_j as arguments.

Factor graphs are graphical representations used to model the factorisation given in Figure 3-8. A factor graph consists of nodes and edges. Nodes in a factor graph can be listed in two categories: variable nodes representing independent variables x_i and factor nodes representing local functions f_j. In a factor graph, edges connect a factor variable node x_i to factor node f_j if and only if x_i is an argument of f_j. For example, the factor graph in Figure 3-8 represents the factorisation of the global function $F(x_1, x_2, x_3, x_4)$ given by

$$F(x_1,x_2,x_3,x_4)=f_A(x_1,x_2)f_B(x_3,x_4)f_C(x_3)f_D(x_3,x_4) \quad (3\text{-}11)$$

where the variable nodes are represented by circles and the factor nodes are represented by squares. $F(x_1,x_2,x_3,x_4)$ is a function of four variables, and *F* can be expressed as a product of four factors so that *J={A,B,C,D}*, $X_A=\{x_1,x_3\}$, $X_B=\{x_1,x_4\}$, $X_c=\{x_3\}$, $X_4=\{x_3,x_4\}$.

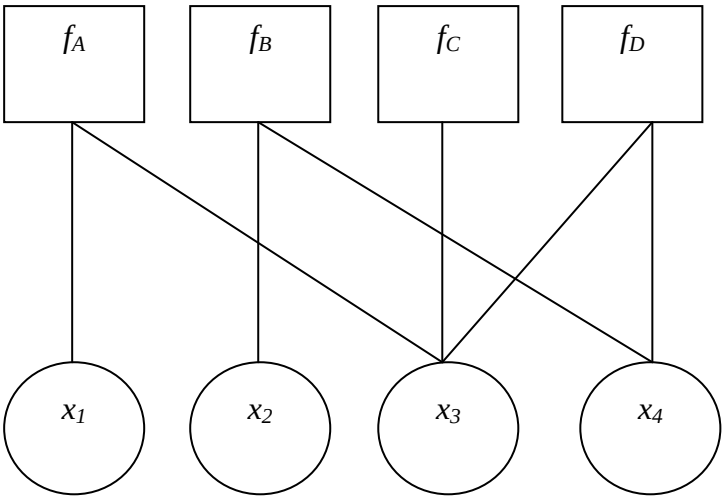


Figure 3-8: The factor graph for the global function factorisation in (3.11)

The BP algorithm is based on exchanging messages between variable and factor nodes in a factor graph. In every iteration, each function node receives a-prior information from the variable nodes connected to it through edges. It then calculates the a-posteriori probabilities for the connected variable nodes and passes the information back to the variable nodes. Upon receiving the updated information, the variable nodes calculate their new-prior information. Hence, two types of messages can be found in the BP algorithm. The first type is the message passing on the edge from the variable node *x* to the factor node *f* which is denoted by $\mu_{x\to f}(x)$. The second type is the message passing on the edge from the factor node *f* to the variable node *x* and is denoted by $\mu_{f\to x}(x)$. These messages are exchanged according to the following update rules (Kschischang et al., 2001).

The message from variable node x to function node f is updated as follows

$$\mu_{x\to f}(x)=\prod_{h\in n(x)\setminus\{f\}}\mu_{h\to x}(x) \tag{3-12}$$

where $X=n(x)$ is the set of arguments of the function f. Equation (3-12) implies that the message sent from variable node x to function node f is the product of all messages received by variable node x on all other edges.

The message from function node f to variable node x is updated as follows

$$\mu_{f\to x}(x)=\prod_{\sim\{x\}}(f(X)\prod_{y\in n(f)\setminus\{x\}}\mu_{y\to x}(y) \tag{3-13}$$

where $X=n(f)$ is the set of all variable nodes connected to f and $\sim\{x\}$c is the set of all variable nodes connected to f excluding x. In other words, the updated message sent from function node f on all other edges. The result is then marginalised to be a function of x only.

When the BP algorithm converges, the marginal distribution $f(x)$ for variable node x can be computed by taking the product of all messages received by x, that is

$$f(x)=\prod_{u\in n(x)}\mu_{u\to x}(x) \tag{3-14}$$

When the algorithm is implemented on a factor graph which is a tree, the algorithm is guaranteed to converge, and the marginal functions calculated in (3-14) are exact after a number of iterations equal to the depth of the tree. However, in the case that the factor graph contains cycles (a cycle is a loop that starts and ends at the same node), the BP algorithm does not have a natural termination and it needs to be stopped after sufficient number of iterations where improvements in the messages are minor. Although convergence is not guaranteed in the case of factor graphs containing cycles. Numerical results have shown that the algorithm achieves near-optimal results (Mooij and H. Kappen, 2007, Frey and Mackay, 1998).

3.8 RANSAC

This section introduces RANSAC and the steps of the general RANSAC algorithms. In this thesis, RANSAC algorithm is applied after BP as an extension of the SIFT image registration. RANSAC is used to filter out the outliers introduced in the SIFT-BP method and to estimate the transformation matrix. The algorithm will be able to estimate the model parameters more accurately with the present of less outlier in the dataset.

RANSAC is a robust estimator originally proposed by (Fischler and Bolles, 1981) where it was used to derive a usable model from a set of data. It is used to classify the matching features into inliers and outliers. Inliers are the data that adhere to the model while the outliers are the data that do not. The RANSAC algorithm consists of two steps that are repeated in an iterative fashion. The steps are "hypothesize" and "test-framework". In the hypothesize step, first minimal sample sets (MSSs) are randomly selected from the input dataset and the model parameters are computed using only the elements of the MSS. The cardinality of the MSS is the smallest sufficient to determine the model parameters. In the test step, RANSAC checks elements of the entire dataset are consistent with the model instantiated with the parameters estimated in the first step. The set of such element is called consensus set (CS) or inliers. RANSAC is terminated when the probability of finding a better ranked CS drops below certain threshold.

The steps of the general RANSAC algorithm are as follows (Fischler and Bolles, 1981)

1. Suppose there are *n* training data samples $X = x_1; x_2; ...; x_n$ which are fitted to a model determined by (at least) *m* samples $(m \leq n)$
2. Set an iteration counter $k = 1$.
3. Choose at random *m* items from *X* and compute a model.
4. For some tolerance ε, determine how many elements of *X* are within the derived model. If this number exceeds a threshold *t*, re-compute the model over this consensus set and stop.
5. Set $k \leftarrow k + 1$. If, $k < K$ for some predetermined *K*, go to 3. Otherwise, accept the model with the biggest consensus set so far, or fail.

3.9 Image fusion

Image fusion is the process of incorporating complementary information that exists in different observation images into a single unified image (Hill et al., 2002). Image fusion can be placed into three categories: pixel level, feature level and decision level (Lanir et al., 2006). In pixel level fusion, the arithmetic operations are used to fuse pixel by pixel information from the input images in time domain. Similarly frequency transformations are used in frequency domain. This is the lowest level fusion that fuses raw images pixel by pixel and it requires less processing time. The objective of pixel level fusion is to produce an output image with more useful information that the input images. Feature level fusion involves extraction of features from the input images such as edges. Then the fusion is based on the extracted features. While in decision level fusion, the fusion is perform after decisions on each input images are made. Decision level and feature level fusions are high level fusions that require more complex algorithms and involve high computational cost. This thesis focuses on pixel level fusion techniques. The pixel level fusion will be discussed further in the next section.

3.9.1 Pixel-level image fusion

Pixel-level image fusion methods can be categorised into two types: arithmetic fusion method and multiscale fusion method (Petrović et al 1999).

Arithmetic fusion algorithms produce the fused image pixel by pixel, as an arithmetic combination of the corresponding pixels in the input images. For pixel level fusion, typical arithmetic operations include weighted average and pixel comparison. Other methods include, fusion based on colour and intensity-hue saturation, Principal-Component-Analysis (PCA) based fusion and neural network.

Arithmetic fusion can be summarised by the expression given in Equation (3-15):

$$F(n,m) = k_A A(n,m) + k_B B(n,m) + C \tag{3-15}$$

where *A*, *B*, and *F* represent the input and the fused images respectively at location *(n,m).* Furthermore k_A, k_B and *C* are constants characterising the fusion method, with

k_A and k_B defining the relative influence of the individual inputs on the fused image and C the mean offset. For weighted average, the fusion is based on fixed or variable weights assigned to each original image.

Image averaging is the most commonly used arithmetic fusion method. In this case, the fused signal is evaluated as the average value between the inputs, i.e. k_A=½, k_B=½ and C=0. However, despite being significantly more computationally efficient than most other fusion systems, image averaging, like all other arithmetic fusion methods, does not achieve good performance. The main reason for this is the loss of contrast, as a result of destructive superposition when the input signals are added.

For comparison based fusion, the maximum or minimum intensity at each pixel location is selected from the input images to generate the output image as described in equation below.

$$F(n,m) = \max(A(n,m), B(n,m)) \qquad (3\text{-}16)$$
$$F(n,m) = \min(A(n,m), B(n,m))$$

For multi-scale image fusion, the input images are decomposed into a series of sub-bands signals (Zhang and Blum, 1999). Multi-scale fusion includes the pyramid transform and Discrete Wavelet Transform (DWT) (Lewis et al., 2007). In the pyramid transform, information from the input images are represented in different scales/pyramids. The pyramid is formed by iterative application of low-pass filtering, followed by subsampling with a factor of 2, a process also known as reduction. The Gaussian and Laplacian pyramids are the most commonly used pyramid transforms in image fusion (Lanir et al., 2006, Blum and Liu, 2005). For DWT, input signals are transformed using the wavelet decomposition process into the wavelet representation (Lewis et al., 2007).

The mutiscale based fusion such DWT and pyramid transform decompose an image into its multiscale edge representation. Based on the fact that the human visual system is primarily sensitive to local contrast changes (i.e. edges) the fusion can be performed as a combination of the multiresolution edges resulting from the DWT or image pyramid decomposition process. This results in a fused image which combines the most salient image features from all scale (Rockinger and Fechner, 1998).

3.10 Interpolation

Interpolation is the process of estimating the values of a continuous function from discrete samples. Image interpolation techniques are referred in the literature by many terminologies, such as image resizing, image re-sampling, digital zooming, image magnification, image enhancement, etc. Image interpolation applications include image magnification or reduction, sub-pixel image registration, image decompression and super-resolution. Basically, for super-resolution image interpolation algorithms convert or resize a digital image from one resolution (dimension) to another resolution without losing the visual content in the picture (Wittman, 2005).

Most of the interpolation techniques interpolate the pixels based on image characteristics such as edge information. Image interpolation techniques can be categorised into non-adaptive and adaptive techniques.

Adaptive interpolation algorithms mainly rely on the intrinsic image features or contents of the image and accordingly the computational logic is mostly dependent upon the intrinsic image features and contents of the input image. The non-adaptive algorithms do not rely upon the image features or its contents and the same computational logic is repeated in every pixel or group of local pixels irrespective of the image contents (Acharya and Tsai, 2007). The most common interpolation techniques are nearest neighbour, bilinear and bicubic interpolation.

The thesis focuses on one of the popular non-adaptive interpolation technique i.e., bicubic interpolation. A detailed review of image interpolation methods can be found in (Acharya and Tsai, 2007).

3.10.1 Bicubic interpolation

Bicubic interpolation uses a 4 by 4 neighbourhood to find the missing pixels in the high-resolution grid. It uses a polynomial passing through four pixels to make a decision. The function employs more complex cubic equations to calculate the output intensities. An advantage of using a bicubic interpolation is that it provides smoother

outputs. This is because bicubic interpolation uses the larger neighbourhood of pixels to estimate the unknown pixels.

In order to carry out a bicubic interpolation within a grid square, it is necessary to calculate the gradients (the first derivatives) in both the *x* and *y* directions and the cross derivative at each of the four corners of the square. This gives 16 equations that determine the 16 coefficients (a_{ij}), as explained in (Press et al., 2002).

3.11 Assessment Measures

This section describes the assessment index measures used in the thesis. In order to objectively assess the quality of the reconstructed high-resolution image, a peak signal-to-noise ratio (PSNR), sharpness index measure (Lee et al., 2009) and blind image quality indices (BIQI) (Moorthy and Bovik, 2010b) (implemented using BIQI software release in (Moorthy and Bovik, 2010a) were used. While for image registration and image fusion, the mutual information metric and the Root Mean Square (RMS) contrast metric were used to assess the results.

The well-known metrics peak signal-to-noise ratio (PSNR) is used to compare the reconstructed high-resolution image with the original high-resolution image. Sharpness index measure (Lee et al., 2009) and blind image quality indices (BIQI) (Moorthy and Bovik, 2010b) (implemented using BIQI software release in (Moorthy and Bovik, 2010a) were used when the reference or original image is not available.

PSNR is a derivation of mean square error (MSE). MSE calculates error between two images while PSNR measures the error variance against the maximum possible image variance. MSE and PSNR are defined as follows:

$$MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} \mathrm{Im}(x,y) - \mathrm{Im}'(x,y)^{2}}{M.N} \quad (3\text{-}17)$$

$$PSNR = 20 * \log\left[\frac{255}{\sqrt{MSE}}\right] \quad (3\text{-}18)$$

Where *Im(x,y)* is the original image, *Im'(x,y)* is the test image, *M* and *N* are the size of image in both horizontal and vertical axes.

The sharpness index (S-index) is used for sharpness evaluation. The second order derivative method was suggested in (Lee et al., 2009) to be used for evaluation of sharpness because it is highly sensitive to high frequency regions. To assess the sharpness for an image, a summation of the absolute values of all the coefficients in single frame is calculated. The sharper the image, the larger the summation value is.

The sharpness index *S* is calculated as:

$$S = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} |L(x,y)| \tag{3-19}$$

where *N* x *N* is the size of the image whose input is the pixel *(x,y)* and *L(x,y)* is the output of the Laplacian filter,. The sharper the image, the larger will be the summation value *M*.

The BIQI, on the other hand, calculates the quality scores of the image. Work in (Moorthy and Bovik, 2010b) demonstrated that BIQI performs well in terms of correlation with human perception and it is competitive with the classical full-reference method PSNR. BIQI is used for no-reference image quality assessment based on natural scene statistics (NSS). The score typically has a value between 0 and 100 with 0 representing the better quality and 100 the worst. The BIQI method classifies the image into particular distortions category by estimating the presence of distortions in the image. The amount of probability of each distortion in the image is gauged and denoted as $p_i, \{i = 1,...,5\}$. Then the quality of the image along each of these distortions is evaluated. Let $q_i, \{i = 1,...,5\}$ represent the quality scores from each of the quality assessment algorithms (corresponding to the five distortions) described in (Moorthy and Bovik, 2010b). The quality of the image is then expressed as a probability-weighted summation

$$BIQI = \sum_{i=1}^{5} p_i \cdot q_i \tag{3-20}$$

Further details on the BIQI method can be found in (Moorthy and Bovik, 2010b).

The Root Mean Square (RMS) contrast calculate the standard deviation of the error signal (Yao et al., 2005), for all the test images.

$$RMS\ \text{Contrast}=\sqrt{\frac{1}{MN}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\left(I_{ij}-I\right)^{2}} \tag{3-21}$$

Where N x M is the size of the image, I_{ij} is the reference image, I is the test image.

The mutual information metrics measures the degree of statistical dependence of two images. The mutual information of two discrete random variables X and Y can be defined as

$$I(X;Y)=\sum_{y\in Y}\sum_{x\in X}p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{3-22}$$

Where $p(x,y)$ is the probability distribution function of x and y, and $p(x)$ and $p(y)$ are the marginal probability distribution function of x and y respectively.

For an image the entropy is calculated from the image intensity histogram in which the probabilities are the histogram entries. For image registration, joint entropy is considered. Joint entropy measures the amount of information in the two images. The joint entropy $H(I, J)$ can be calculated using the joint histogram of two images. A large measure implies better quality because an optimal transformation can be gained by maximizing mutual information of the two images (Roshni et al, 2008).

For images from video frames, the quality of the reconstructed high-resolution images is evaluated by BIQI. Generally, a good algorithm is reflected by the assessment measure metrics. Nevertheless, the best performance measure remains human inspection of the reconstructed high-resolution images.

3.12 Summary

This chapter has presented an overview of super-resolution methods particularly the non-uniform interpolation approach. The steps in the non-uniform interpolation approach have been discussed in detail. The image registration method was presented which focused on feature-based image registration. SIFT, belief propagation and RANSAC algorithms for detecting and matching feature points in the image been discussed in details. Overviews of image fusion and interpolation techniques were presented discussed. The quality metrics used for assessing the performance of the methods were explained.

CHAPTER 4

4. IMAGE REGISTRATION FOR SUPER-RESOLUTION USING SCALE INVARIANT FEATURE TRANSFORM, BELIEF PROPAGATION AND RANDOM SAMPLING CONSENSUS

4.1 Introduction

Image registration is the process of aligning one image to another image of the same scene. The two images are taken from different viewpoints and/or at different times or by different sensors. Image registration is an important step in multi-frame super-resolution; indeed, accurate image registration is crucial for the effectiveness of super-resolution. In this chapter an image registration approach for super-resolution based on a combination of Scale Invariant Feature Transform (SIFT), Belief Propagation (BP) and Random Sampling Consensus (RANSAC) is proposed. The SIFT algorithm is used to detect and extract the local features in images, BP is used to match the features while RANSAC is adopted to filter out the mismatched points and then estimate the transformation matrix. The proposed method is compared with traditional SIFT to verify its accuracy and stability. Finally, the result of using the proposed approach in the super-resolution application is given and compared to the traditional approach where SIFT is used for image registration step.

This chapter demonstrate the effectiveness of SIFT with BP algorithm for image registration in super-resolution imaging for the test images, and further improve the result by applying RANSAC for eliminating the remaining mismatched points and

estimating the transformation matrix. The rest of the chapter is organised as follows. Section 4.2 describes the implementation of the proposed method. Experimental results are given in Section 4.3. Summary of the chapter is given in Section 4.4.

4.2 The proposed method

Image registration is used in super-resolution to align low resolution image frames prior to subsequent processing. A subpixel-registered image sequence of the same scene potentially contains more information than any single view alone. Image registration enables the determination of subpixel shifts hence enables the uncovering of useful information from multiple frames. In this section, an image registration that combines SIFT, Belief propagation and RANSAC techniques is described. To simplify exposition, the proposed methods will be explained assuming two low-resolution images only. Generalisation to more than two images is straightforward for example when three images are being used the third image will be registered to the results of registering the first and second image.

The proposed image registration method for super resolution is shown in Figure 4-1. It is assume that the Test image needs to be registered with the Reference image. First, the original SIFT algorithm (Lowe, 2004) is used to extract the local features from both images. The extracted features are then matched using the BP algorithm as in (Cheng et al., 2009). Next, mismatched points that remain after the BP matching are eliminated using RANSAC. Finally, the transformation matrix is estimated once all the correct matching points are established, and the image is resampled using the optimal transform model. After image registration, the relative pixel positions of all low-resolution images in the sequence in reference to the first image are identified. Then this information can be projected onto a high-resolution grid. For this task, the algorithm of (Kim and Kwon, 2008) is used, though other method can also be applied. In the next section, each of the steps will be describe in detail.

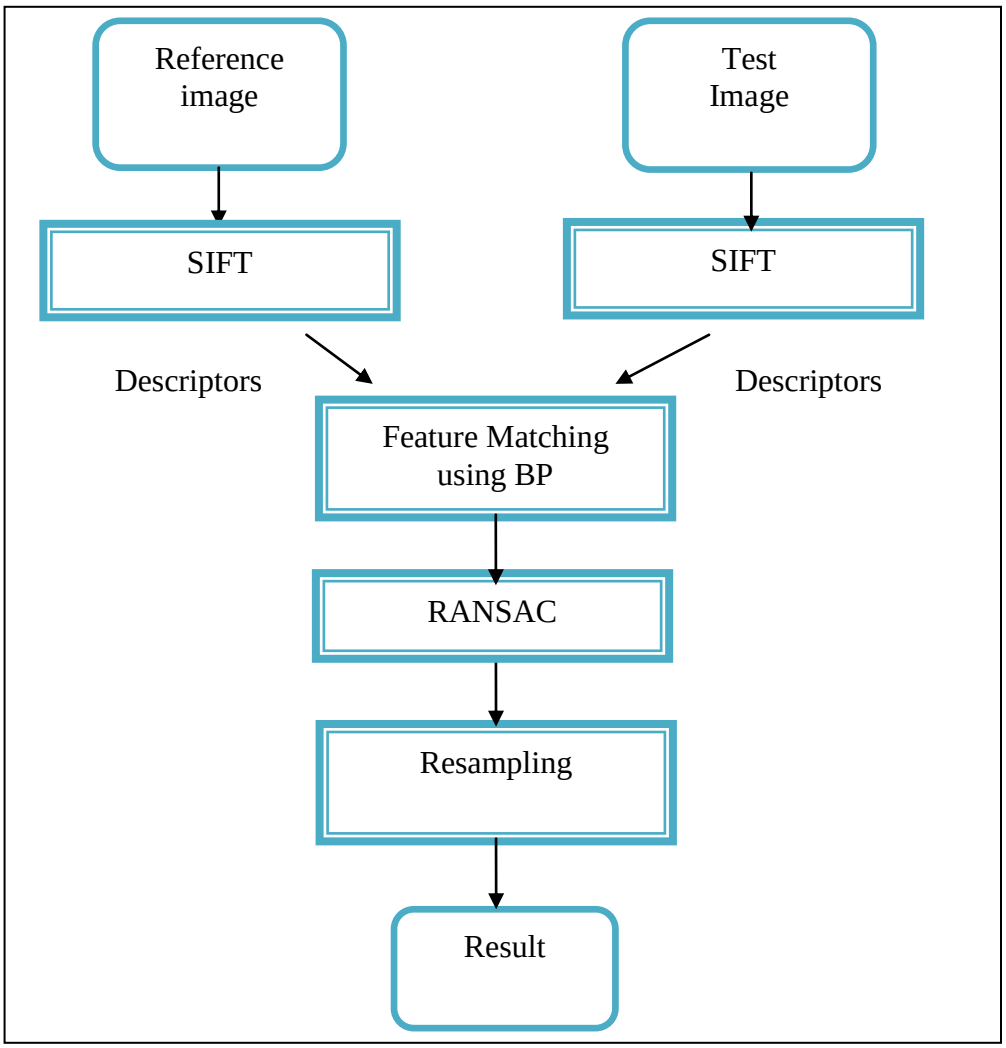


Figure 4-1: The block diagram of the image registration method.

4.2.1 SIFT feature extraction

In this step, the SIFT algorithm by (Lowe, 2004) is used for the detection of feature points. Many methods are proposed for this task (see (Zitova and Flusser, 2003)) for a survey). The SIFT algorithm has been demonstrated to be tolerant to scale and illumination changes as well as rotation. The SIFT algorithm is one of the most popular feature-based image registration methods often used in panoramic imaging, medical imaging, robotics, data mining and surveillance.

The approach of SIFT feature detection taken in our implementation is similar to the one taken by (Lowe, 2004). There are four main steps in extracting the local features:

(i) keypoints detection, (ii) keypoints localization, (iii) orientation assignment, and (iv) keypoints descriptor generation. More details on the SIFT algorithm are given in Chapter 2.

One example of the keypoints detection and matching obtained by the SIFT algorithm is shown in Figure 4-2. Keypoints are shown by circles and matches with lines.

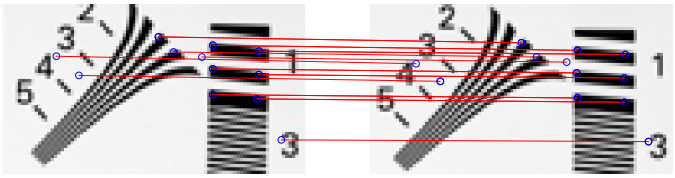


Figure 4-2: Keypoints matching of the original low resolution images using original SIFT (25 matches)

4.2.2 Descriptor matching using belief propagation

The second step is to match the feature points. The matching points are initially established using the Euclidean distance method and then the belief propagation is applied to eliminate false matches. In order to achieve a highly accurate registered image, belief propagation is employed to refine the keypoint matching and to correct for minor matching errors.

For image matching, descriptor vectors of all keypoints are stored in a database. In traditional SIFT (Lowe, 2004), matches between keypoints are found based on Euclidean distance.

In (Cheng et al., 2009), belief propagation (BP) is used in the matching process where the keypoint matching is formulated as a global optimisation problem. Detailed steps on how BP is used in the SIFT matching process can be found in (Cheng et al., 2009) and described in Chapter 2.

Figure 4-3 demonstrates that the proposed approach eliminates four incorrect matches by applying BP after the SIFT algorithm. Thus, using BP after SIFT can reduce the number of wrongly matched keypoints that can potentially improve the image registration result.

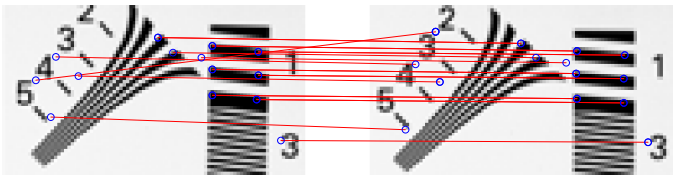


Figure 4-3: Keypoints matching using the SIFT-BP (21 matches).

4.2.3 RANSAC and transformation matrix estimation

The third step is to employ RANSAC to further refine the keypoint matching and estimating the transformation matrix. RANSAC is a robust estimator originally proposed by (Fischler and Bolles, 1981) where it was used to derive a robust model from a set of data. In (Yuan et al., 2008), RANSAC is used to filter out the incorrectly mapped points that come from errors in the SIFT model.

The correct matching features are classified into inliers and outliers using RANSAC. Inliers are the data that adhere to the model while the outliers are the data that do not. The RANSAC algorithm starts by randomly selecting sets of corresponding points. The mapping transform is found for each possible set of four keypoints at the reference image and their respective matches at the target image. Then transformation matrix is estimated using those points as follows:

$$\begin{pmatrix} x' \\ y' \\ 1' \end{pmatrix} = A \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{4-1}$$

where $x', y' \longleftrightarrow (x, y)$ are pixel point correspondences, and A is a 3x3 transformation matrix.

Using the transformation matrix, the symmetric transfer error $d(x, A^{-1}x')^2 + d(x', Ax)^2$ is calculated for every matching point, and the numbers of inliers that are less than the threshold value are counted. Here $d(x, y)$ is the Euclidean distance between points *x* and *y*. Then the same procedure is applied to the rest of the keypoints in the reference image, and spatial coordinates of transformed keypoints are compared to the coordinates of the respective keypoints in the target

image. This allows the number of keypoint pairs that fit the model within a certain tolerance to be established. The model that supports the maximum number of keypoint pairs (consensus set) within a transform model is considered as optimal. Then the model transforms the target image to the reference image, so that corresponding points in both images are spatially close to each other.

Figure 4-4 demonstrates that the approach eliminates two remaining incorrect matches by applying RANSAC after the BP algorithm. Thus, using RANSAC after BP can reduce the number of wrongly matched keypoints that can potentially improve the image registration result, and consequently the super resolution performance.

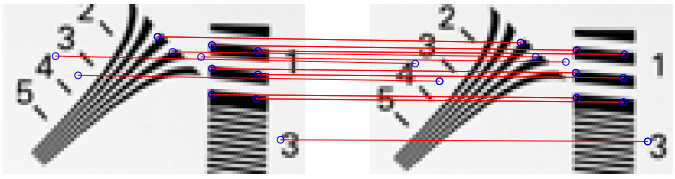


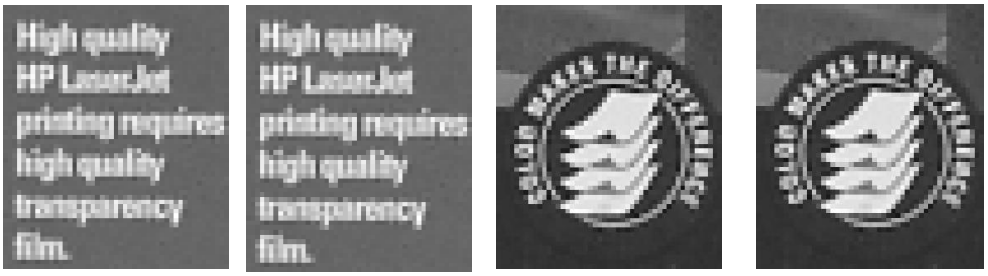
Figure 4-4: Keypoints matching using the SIFT-BP-RANSAC (19 matches).

4.3 Results

This section presents the image registration results illustrating the needs of using SIFT-BP-RANSAC for accurate registration in super-resolution and the super-resolution results obtained with the proposed registration method. The intention is to illustrate the effect of incorporating RANSAC into the SIFT-BP image registration framework. The performance was tested on simulated and real-world low resolution images.

A set of images shown in Figure 4-5 are used to test the proposed method, and in Figure 4-6 the proposed method are further tested using low-resolution video frames from (Milanfar) i.e., 'text' and 'disk'. These images were chosen since they represent a wide cross-section of image statistics. Image 'letter', 'chart1' and 'chart2' are commonly used test images for super-resolution algorithm. Indeed, images 'letters' and 'chart' contain a lot of high frequencies – letters and numbers; 'Radcliffe' is a real-world image taken from (Philbin and Zisserman), with many edges and smooth areas in between, while 'signboard' and 'girl' are real-world images that contain a combination of the above features.

Figure 4-5: Six low resolution images used in the experiments, referred to as (from left to right and top to bottom): 'letters', 'chart1', 'chart2', 'Radcliffe', 'signboard' and 'girl'.



(a) (b)

Figure 4-6 : Low resolution frames used in the experiment: (a) ‘text’ and (b) ‘disk’.

4.3.1 Image registration results

First, the feature matching results are presented. Figure 4-7 to Figure 4-12 shows the keypoint matches obtained by SIFT, SIFT-BP, SIFT-RANSAC, and SIFT-BP-RANSAC. From the figures, it is observed that there are mismatch points in the original SIFT and SIFT-RANSAC but most of the mismatched points are removed after BP. In the employed method, RANSAC is used after BP feature matching to remove the remaining outliers. Note that, although BP applied after SIFT removes many outliers, often there will still be a small number remaining. However, RANSAC applied directly after SIFT, on the other hand, results in many undetected outliers. Thus, to provide high accuracy, SIFT-BP-RANSAC is used for registration.

(a)

(b)

(c)

(d)

Figure 4-7: Keypoints matching using (a) SIFT (93 matches) (b) SIFT-RANSAC (87 matches) (c) SIFT-BP (86 matches) and (d) the proposed method (86 matches).

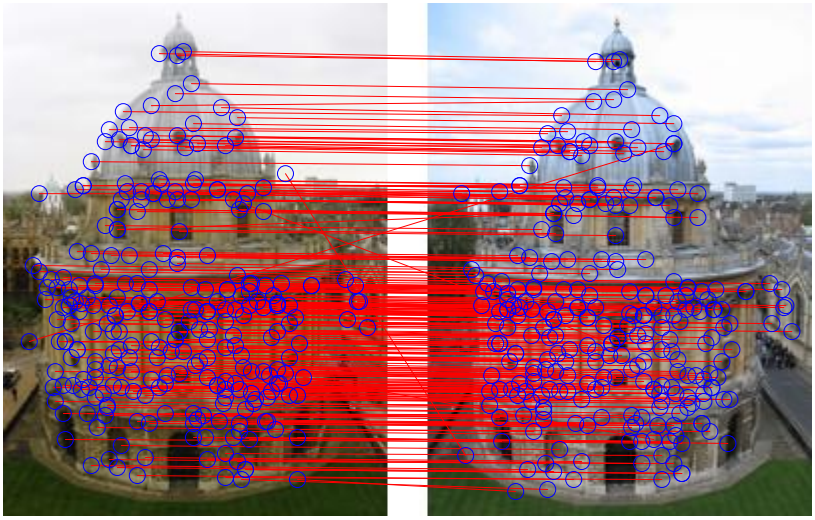
(a)

(b)

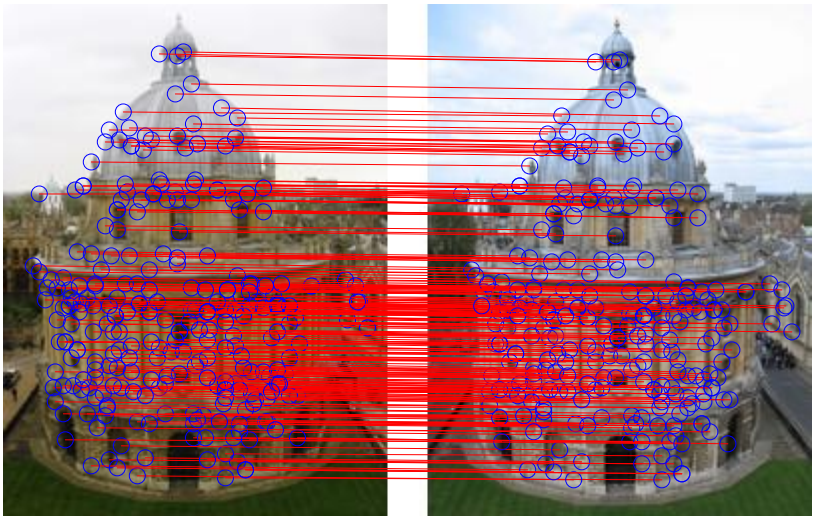
(c)

(d)

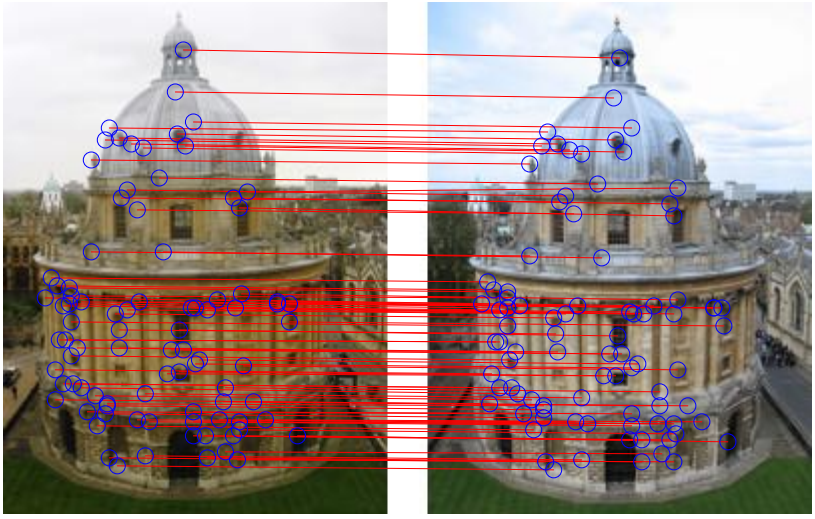
Figure 4-8: Keypoints matching using (a) SIFT (375 matches) (b) SIFT-RANSAC (373 matches) (c) SIFT-BP (288 matches) and (d) the proposed method (288 matches).



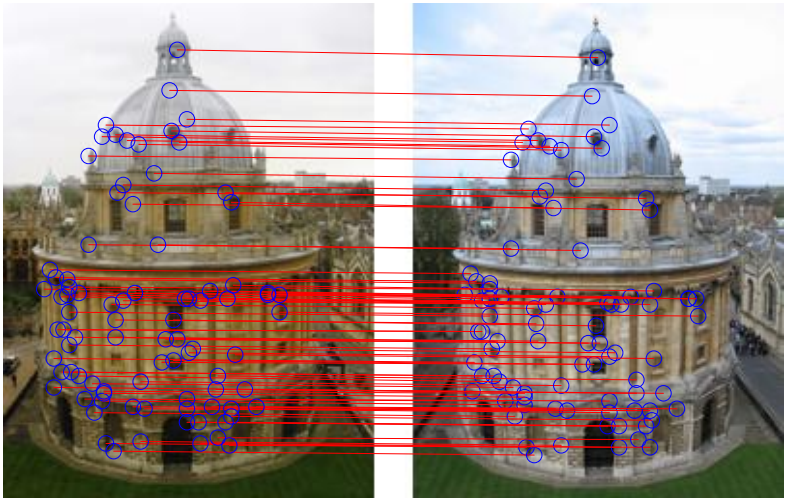
(a)



(b)

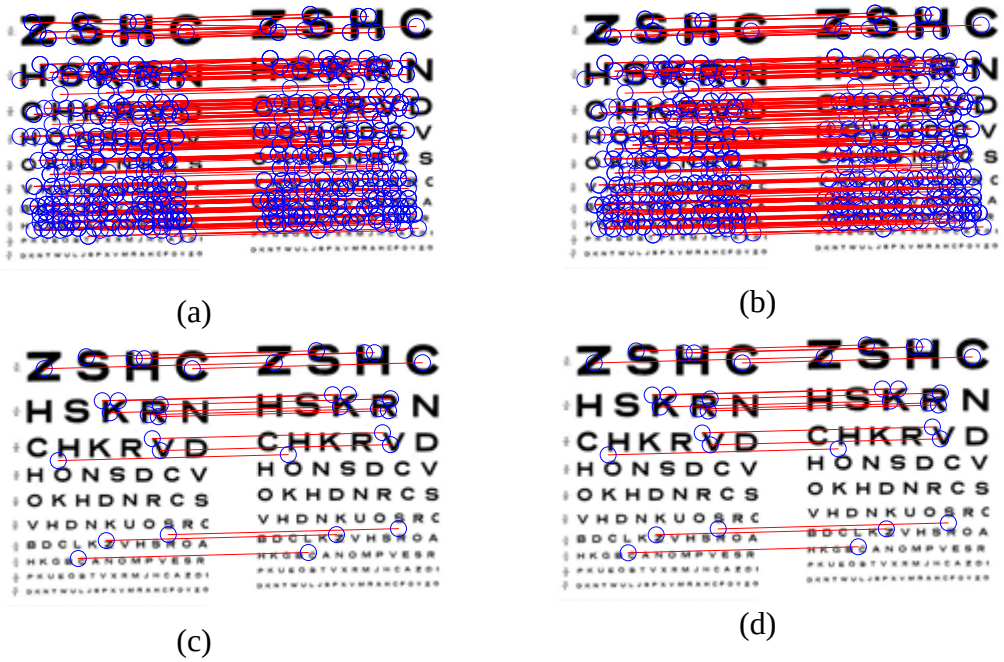


(c)



(d)

Figure 4-9: Keypoints matching using (a) SIFT (277 matches) (b) SIFT-RANSAC (258 matches) (c) SIFT-BP (89 matches) and (d) the proposed method (87 matches).



(a) (b)

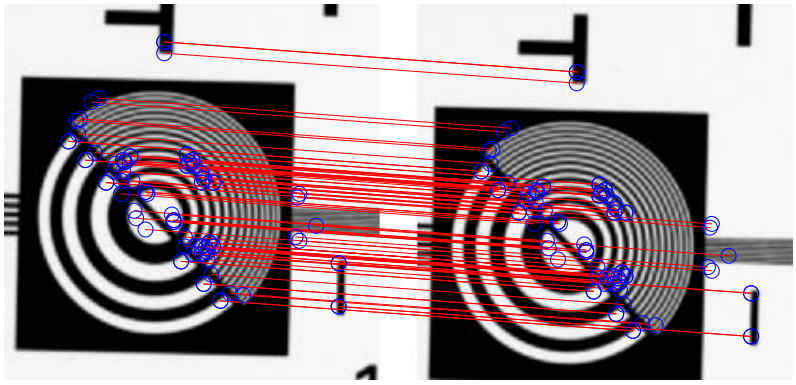
(c) (d)

Figure 4-10: Keypoints matching using (a) SIFT (353 matches) (b) SIFT-RANSAC (349 matches) (c) SIFT-BP (17 matches) and (d) the proposed method (17 matches).

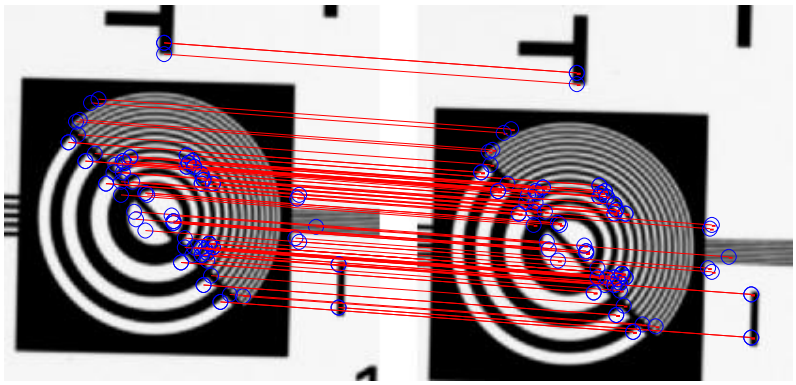
(a) (b)

(c) (d)

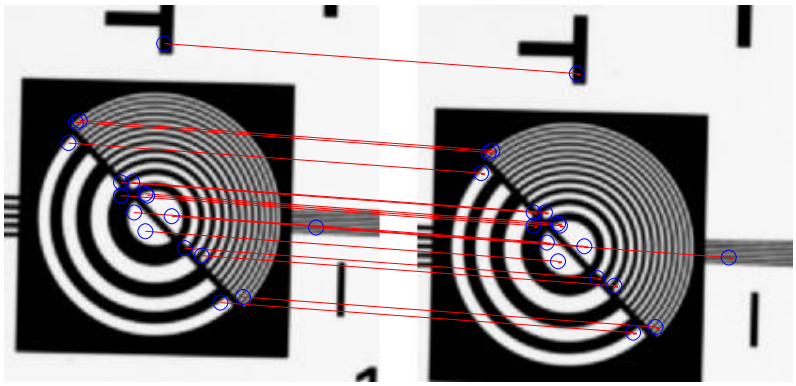
Figure 4-11: Keypoints matching using (a) SIFT (7 matches) (b) SIFT-RANSAC (6 matches) (c) SIFT-BP (7 matches) and (d) the proposed method (6 matches).



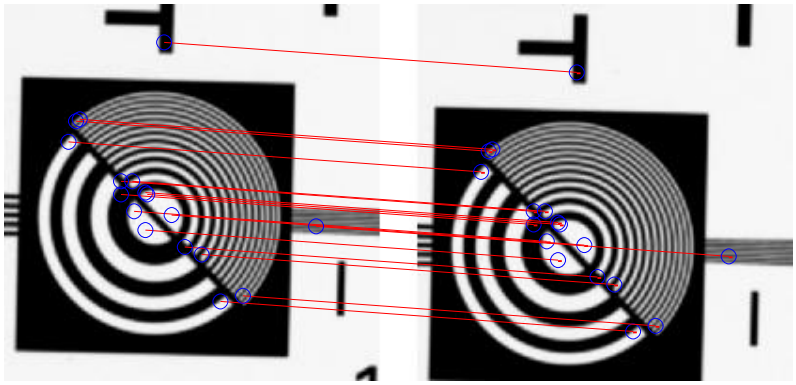
(a)



(b)



(c)



(d)

Figure 4-12: Keypoints matching using (a) SIFT (93 matches) (b) SIFT-RANSAC (93 matches) (c) SIFT-BP (17 matches) and (d) the proposed method (17 matches).

It can be seen in some cases, the number of matches for BP and BP-RANSAC are the same. Note that, although they have the same number of matches but transformation matrixes estimated from both methods are different. This can be seen from the registration results in Figure 4-13 to Figure 4-15 where the proposed method provides better results.

For RANSAC algorithm, the threshold parameters can be set for deciding the outliers in the datasets. In this experiment, the threshold value is tested until the minimum numbers of outliers are reached.

After the incorrect matches have been removed the inliers are used to solve the transformation matrix. Figure 4-13 to Figure 4-15 shows the registration results. The circles in Figure 4-13 to Figure 4-15 highlight the most obvious errors in the registered images after resampling. For example, in Figure 4-13(a) and Figure 4-13(c) the text has been distorted and the same can be said for the arrow in Figure 4-13(b). In Figure 4-13(d) however, the proposed method offers improved performance as no such artefacts can be seen in the registered image.

(a) (b)

(c) (d)

Figure 4-13: Registered image using (a) SIFT (b) SIFT-RANSAC (c) SIFT- BP (d) the proposed method

Figure 4-14 compares image registration results obtained with SIFT, SIFT-BP (Cheng et al., 2009), and SIFT-RANSAC (Yuan et al., 2008). SIFT-BP and the proposed SIFT-BP-RANSAC method gave the best results. It can be seen from Figure 4-14(c) that the registered image based on SIFT-BP has artefacts due to incorrect registration at the first rectangle close to the number 1 on the right of the image. This problem was removed by eliminating two more bad matches with RANSAC.

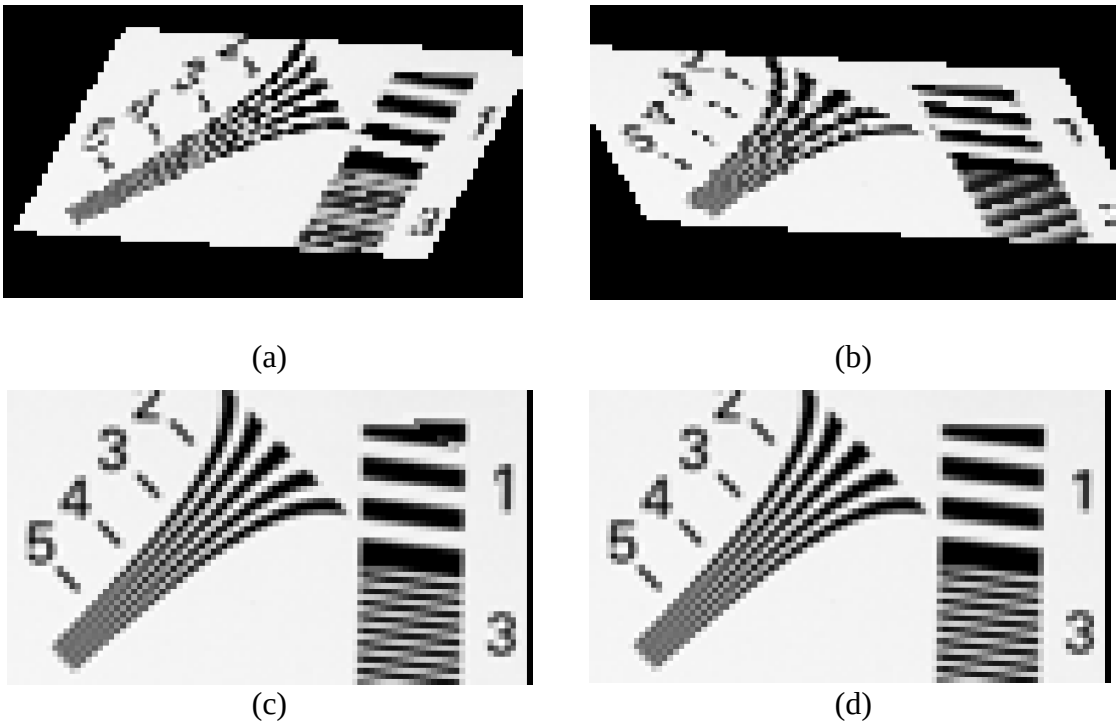
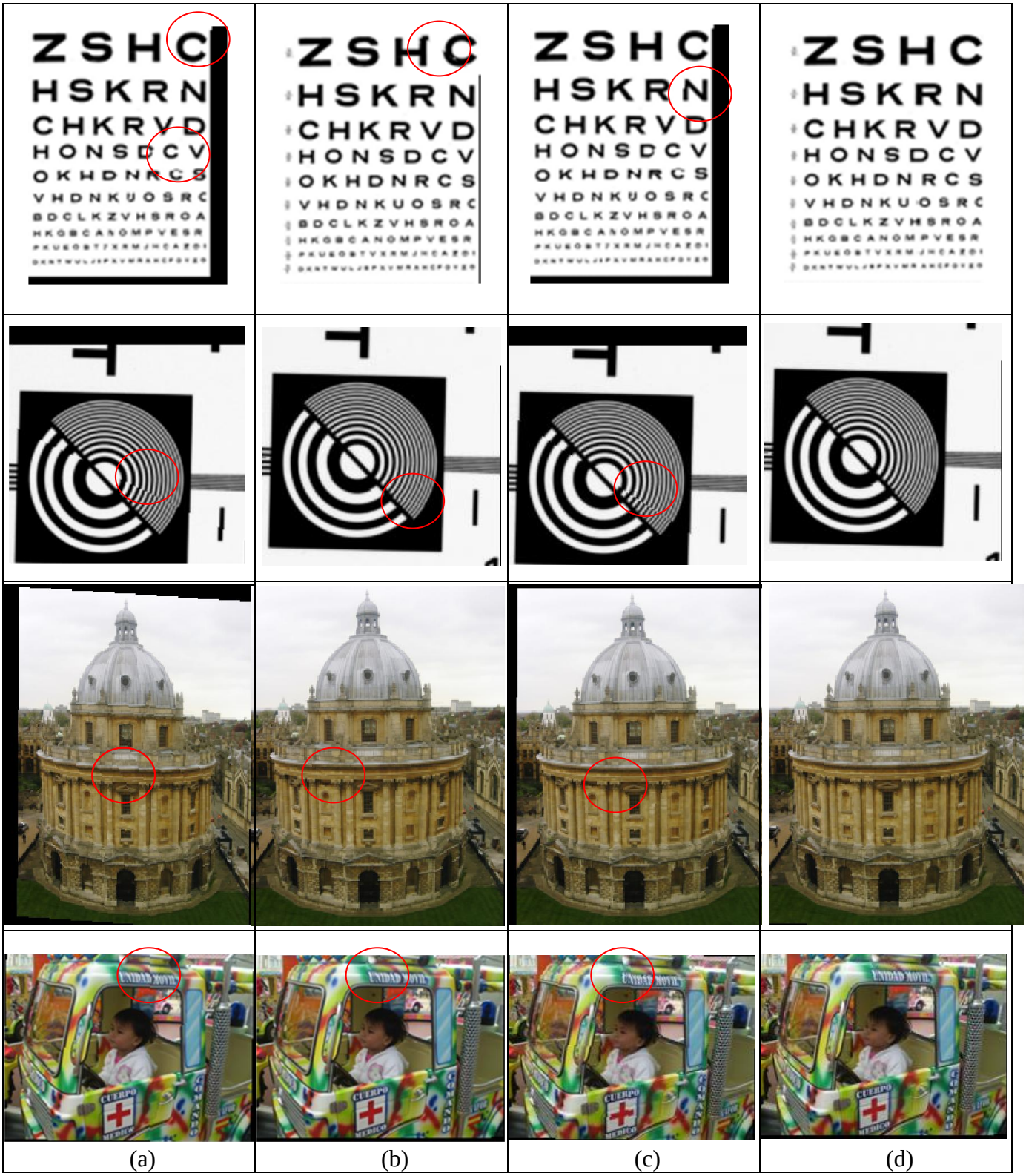


Figure 4-14: The registered image using (a) SIFT (b) SIFT- RANSAC (c) SIFT-BP (d) the proposed method

As demonstrated, applying RANSAC after BP improves matching performance and consequently provides higher image registration accuracy. This is illustrated in Figure 4-15. The circles in Figure 4-15 highlight the most obvious errors in the registered images. For example, for image "letters" in Figure 4-15(a) (top) and Figure 4-15(c) (top) the text has been distorted in Figure 4-15(b) (top). In Figure 4-15(d) however, SIFT-BP-RANSAC shows improved performance as no such artifacts can be seen in the registered image.



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

Figure 4-15 : Registered image using (a) SIFT (b) SIFT-RANSAC (c) SIFT- BP (d) the proposed method

To objectively assess the quality of the registered image, the mutual information metric is used. It measures the degree of statistical dependence of two images. A large measure implies better quality. As shown in Table 4-1, the registered image obtained by the proposed method has a larger measure.

Table 4-1: Mutual Information

Test Image	SIFT	SIFT-RANSAC	SIFT-BP	SIFT-BP-RANSAC
Radcliffe	1.4821	2.7567	1.7577	5.9988
Chart	1.0276	1.8960	1.0282	1.9098
Signboard	1.0157	2.0748	0.9451	5.7628
Girl	1.3941	1.8960	1.8769	3.2200
Letters	0.3937	1.3686	0.4053	2.6579
Chart2	1.3977	1.6158	1.4340	1.6980
Text	1.7462	2.4950	1.7909	2.9345
Disk	3.2721	3.0734	4.5783	4.6620

4.3.2 Super-resolution results

After image registration, the proposed image registration method is tested for super-resolution. In this work, the algorithm of (Kim and Kwon, 2008) is used i.e., to perform interpolation and restoration of the registered image. Note, however that any other interpolation and restoration method can be used instead of.

Figure 4-16 shows the resulting super-resolution images obtained when SIFT-BP, and SIFT-BP-RANSAC were used for image registration. The improved quality of images after super resolution with SIFT-BP and SIFT-BP-RANSAC image registration compared to the original low-resolution image shown in Figure 4-16(a) is obvious. SIFT-BP still suffers from the same artefacts as after registration.

(a)

(b) (c)

Figure 4-16: Result comparing (a) the low-resolution image with super-resolved image of (b) SIFT-BP (c) the proposed method

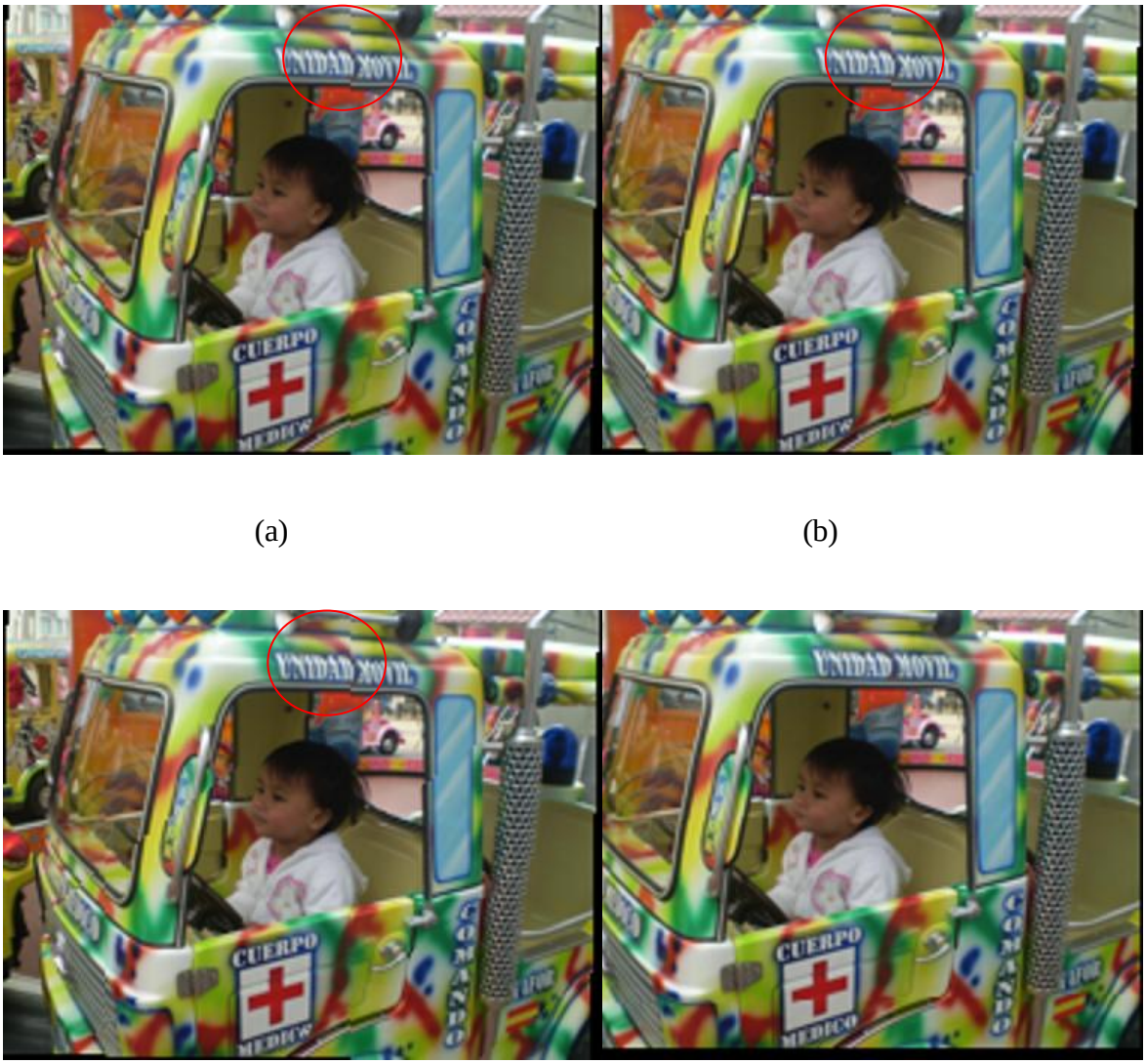
As demonstrated, applying RANSAC after BP improves image registration performance. As a direct result of the registration improvement, the performance of the super resolution algorithm is significantly enhanced. This is illustrated in Figure 4-17 and Figure 4-18 from which better super resolution performance as a result of more accurate registration. The red circles again highlight the most obvious artefacts in the resulting images. The super-resolution image obtained using SIFT-BP-RANSAC for image registration has better visual quality. Figure 4-17 and Figure 4-18 also show the importance of the image registration step, since super resolution applied to poorly registered images with SIFT and SIFT-RANSAC (shown in Figure 4-17 to Figure 4-18(a) and (b), respectively) leads to poor high resolution results as shown in Figure 4-17 to Figure 4-18(a) and (b), respectively.

(a) (b)

(c) (d)

Figure 4-17: Super resolution results with (a) SIFT (b) SIFT-RANSAC (c) SIFT-BP (d) the proposed method

Figure 4-17 shows the super-resolution results for image ‘signboard’. The circles in Figure 4-17 highlight the most obvious errors in the images. In Figure 4-17(a) and Figure 4-17(c) it is observed that the text has been distorted and the same can be said for the arrow in Figure 4-17(b). In Figure 4-17(d) however, the proposed method offers improved performance as no such artefacts can be seen in the super-resolved image. The same improvement can be seen for image ‘Girl’ in Figure 4-18(d), where the proposed method registration method able to removes the artefacts thus improved the super-resolution results.



(a) (b)

(c) (d)

Figure 4-18: Super-resolution results with (a) SIFT (b) SIFT-RANSAC (c) SIFT-BP (d) the proposed method

Further results on the improved quality of images after super-resolution with SIFT-BP-RANSAC image registration are shown in Figure 4-19 to Figure 4-21. In these figures, the left hand side shows the low-resolution images and on the right hand side are the reconstructed high-resolution images. From the figure, it is observed that the proposed method manages to enhance the super-resolution images when accurate registration is achieved. From the figure, it is observed that the super-resolution results for video frame ‘disk’ and ‘text’ are poor. This is because the resolution of these video frames is

low and the super-resolution method used in this experiment is designed for still image. Note that for the same dataset used, better visual results are presented in (Farsiu et al., 2004); however, in (Farsiu et al., 2004) twenty low-resolution video frames have been used for generating the high-resolution image compared to only two used in this experiment. The reason for testing these video frames in this experiment is to demonstrate the effectiveness of the proposed method in registering the video sequences.

Figure 4-19: (left) One of the low resolution image (right) High resolution image

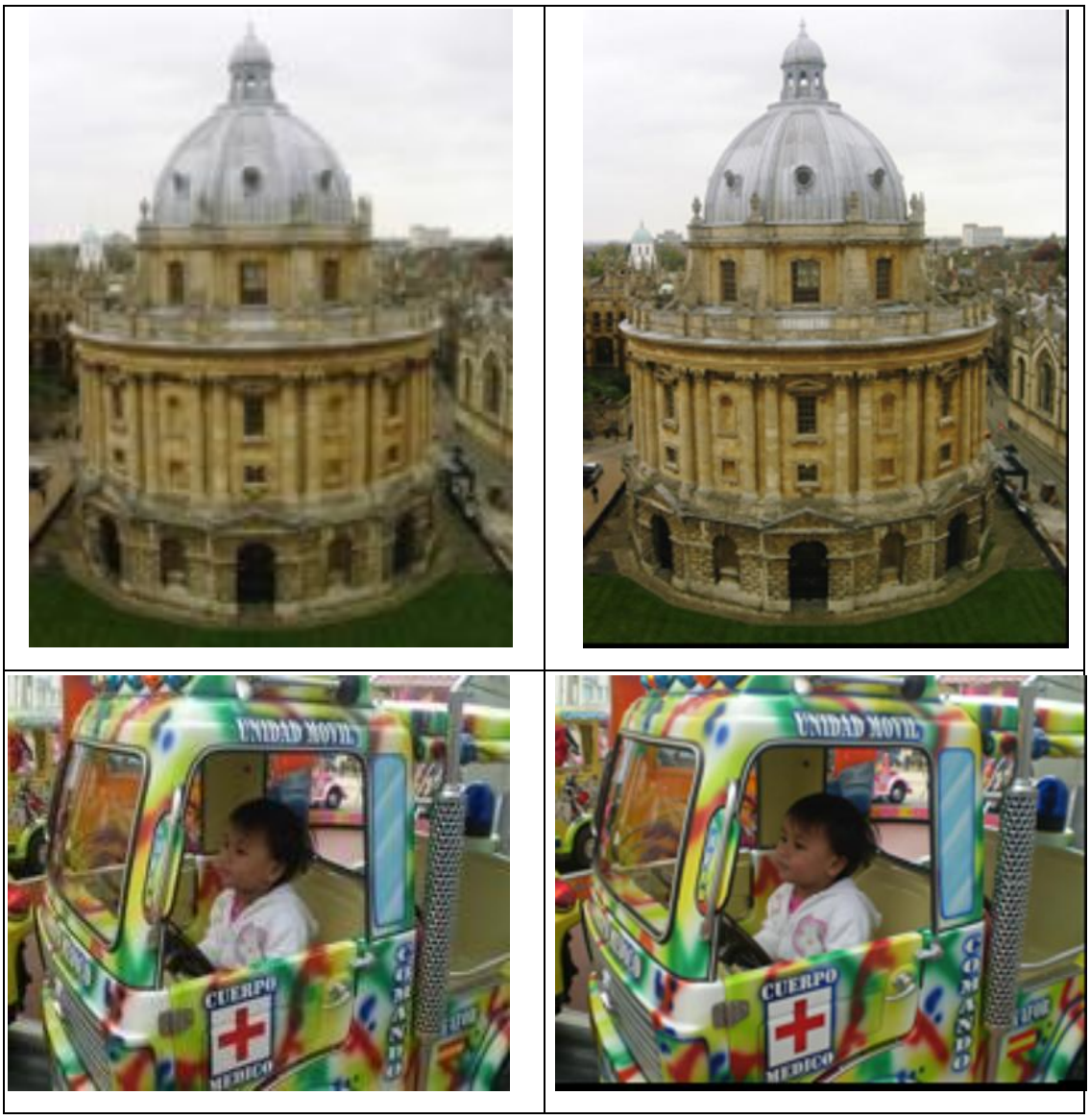


Figure 4-20: (left) One of the low resolution image (right) High resolution image

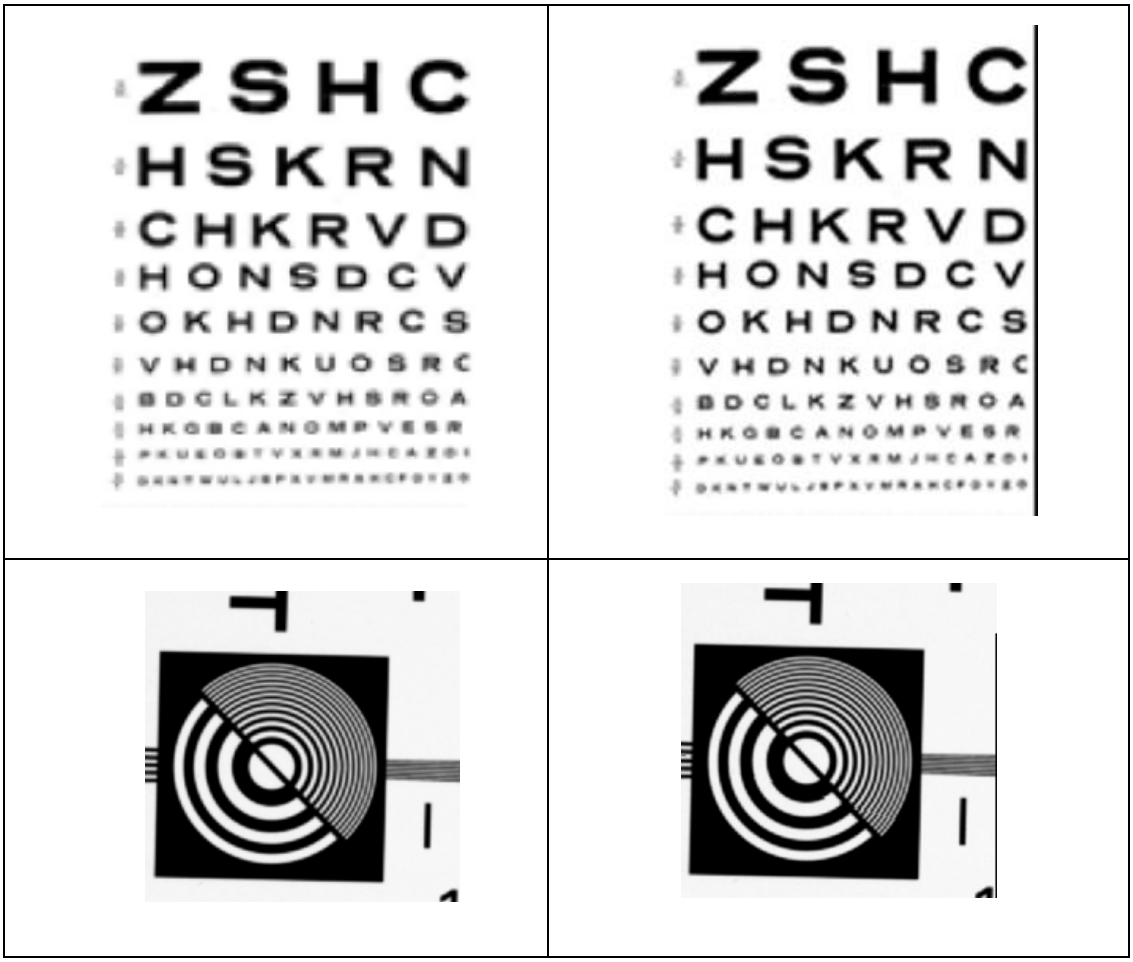


Figure 4-21: (left) One of the low resolution image (right) High resolution image

4.4 Summary

This chapter suggests using SIFT-BP-RANSAC based image registration for image super-resolution. The technique was applied on simulated and real-world images and the results are encouraging especially when compared to the traditional SIFT method. The advantage of the proposed method lies in its ability to overcome the outliers introduced in the SIFT-BP method and hence improves the estimation. The resulting super-resolution images suggest an improvement of visual quality compared to the case when SIFT, SIFT-BP or SIFT-RANSAC alone are used for image registration.

CHAPTER 5

5. SVD-BASED FUSION FOR SUPER-RESOLUTION IMAGE RECONSTRUCTION

5.1 Introduction

Singular value decomposition (SVD) has been successfully used in image processing applications such as image compression, feature extraction and object detection. This chapter presents a novel SVD image fusion approach for super-resolution. SVD image fusion is used to enhance the super-resolution results. The objective of using SVD is to integrate the important features from low resolution images into the method. The proposed method converts the registered and reference image into the SVD domain and then the images' singular values are fused based on the fusion rule before performing the interpolation. Simulation results of applying SVD-fusion prior to interpolation show significant performance improvement when compared to standard interpolation techniques and existing learning-based super-resolution approaches.

The remainder of the chapter is organised as follows. A brief introduction of SVD and its properties is presented in Section 5.2. Section 5.3 describes the implementation of the proposed SVD-based fusion technique. This method fuses the singular values of the registered and reference image in an order to maintain good contrast. Experimental results are provided in Section 5.4. In this section, results obtained from testing and comparing with bicubic interpolation and the single-image super-resolution approach of (Kim and Kwon, 2008) are presented and discussed.

The results suggest that SVD-based fusion improved super-resolution results for the test images. Finally, summary are presented in Section 5.5.

5.2 SVD

In this section, introduction and background information on SVD is presented. SVD involves the decomposition of a matrix into three individual components which has been widely used in mathematics (Horn and Johnson, 1987) and numerical computation (Golub and Loan, 1996).

SVD of an *m* x *n* matrix *X* is given by:

$$X = U_X \Sigma_X V_X^T, \tag{5-1}$$

where the columns of the *m* x *n* matrix U_X are called the *left singular vectors*, the rows of the *n* x *n* matrix V_X^T contain the elements of the *right singular vectors*, and the diagonal elements of the *n* x *n* diagonal matrix $\Sigma_X = \text{diag}(\sigma_1,...,\sigma_n)$ are called the *singular values*. Furthermore, $\sigma_k > 0$ for $1 \le k \le r$, and $\sigma_k = 0$ for $(r+1) \le k \le n$. By convention, the ordering of the singular vectors is determined by high-to-low sorting of singular values, with the highest singular value being in the upper left corner of the Σ_X matrix, that is, $\sigma_1 \ge \sigma_2 \ge ... \ge 0$. The left singular vectors are the orthogonal eigenvectors of X^TX and the right singular vectors are the orthogonal eigenvectors of XX^T. Thus the singular values Σ of *X* are the positive square roots of X^TX.

5.2.1 SVD properties

SVD has an important property that makes it useful in many applications. SVD can be used to provide the best low-rank linear approximation of the original matrix. It is possible to reduce dimensions by selecting greatest *k* singular values. The value *k* may change according to the size and the structure of the data.

The reduced matrix S_k is constructed by retaining the first k singular values. The matrices U and V are also reduced to produce matrices U_k and V_k, respectively. The matrix U_k is produced by removing (r-k) columns from the matrix U and matrix V_k is produced by removing (r-k) rows from the matrix V. Multiplying these three reduced matrices, the matrix X_k is obtained. The reconstructed matrix X_k is a matrix that is the closest approximation to the original matrix X.

Some researchers (Ranade et al., 2007) claim that the low-rank approximation of the original matrix is better than the original matrix itself because the small singular values mainly represent noise, and thus the rank-k matrix approximation represents a filtered signal with less noise.

The singular value matrix Σ_X also represents the intensity information of a given image, where the highest singular values have a great amount of image information (Zhu et al., 2008). The main motivation of this work comes from the fact that changing the highest singular values affects the illumination of the image significantly as reported in (Demirel et al., 2011).

5.3 The proposed method

In this section, the proposed SVD-based fusion for super-resolution is described for generating a high-resolution image from low-resolution images. First, an overview of the method is presented and the novelty of the technique is described. To simplify exposition, all proposed methods will be explained assuming two low-resolution images only.

The problem of generating a single high-resolution image from two low-resolution images is addressed. The proposed method consists of three steps: (i) image registration, (ii) SVD-based fusion and (iii) image interpolation. Note that the first and third steps are traditionally used in super-resolution and they are usually followed by image restoration. In the proposed method, image restoration is carried

out during the fusion step prior to interpolation. The proposed method for two low resolution images I_1 and I_2 is shown in Figure 5-1.

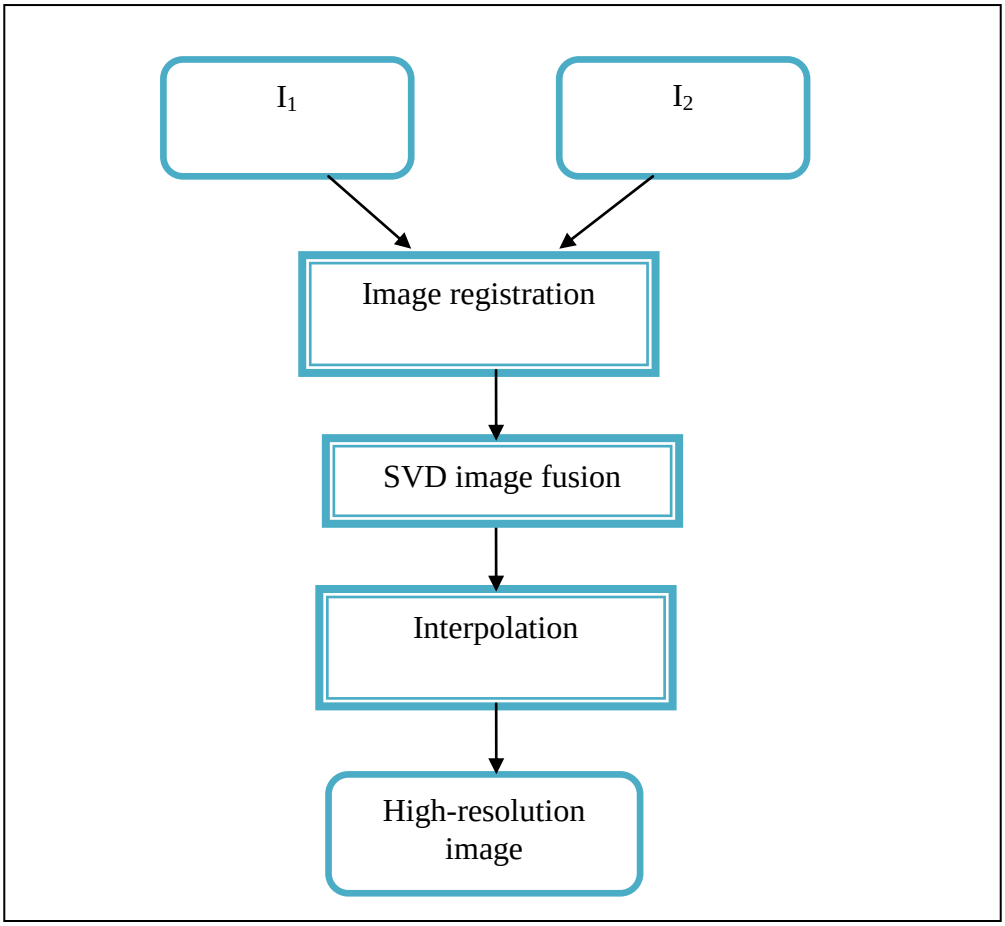


Figure 5-1: The block diagram of the proposed method with two low resolution images I_1 and I_2.

5.3.1 Image registration

Image registration enables identification of subpixel shifts and hence it extracts useful information from multiple images. Image registration starts with the original SIFT algorithm (Lowe, 2004) that is used to extract the local features from both images. The extracted features are then matched using the Belief Propagation (BP) algorithm as in (Cheng et al., 2009). Next, mismatched points that remain after the BP matching are eliminated using RANSAC (Fischler and Bolles, 1981). Finally, the transformation matrix is estimated after all the correct matching points have been established and the image is resampled using the optimal transform model. The

registered image I_r will be used in the fusion step. Details of the image registration method used in this work were described in Chapter 4 of the thesis.

5.3.2 SVD-based image fusion

Conventionally, after image registration, interpolation is performed on the registered image followed by image restoration. Instead, in this work, an image fusion is performed to integrate the information from the low-resolution images prior to the interpolation step. The objective is to maintain the important information from the low-resolution images, which would not occur if the interpolation were carried out on the registered image I_r directly. This thesis proposes the use of SVD for the image fusion step.

The proposed fusion method is shown in Figure 5-2. The SVD is applied to both I_1 and I_r images separately to highlight the features. Then pixel level image fusion is applied to obtain the SVD-based fusion results. The image fusion rule used is the maximum rule.

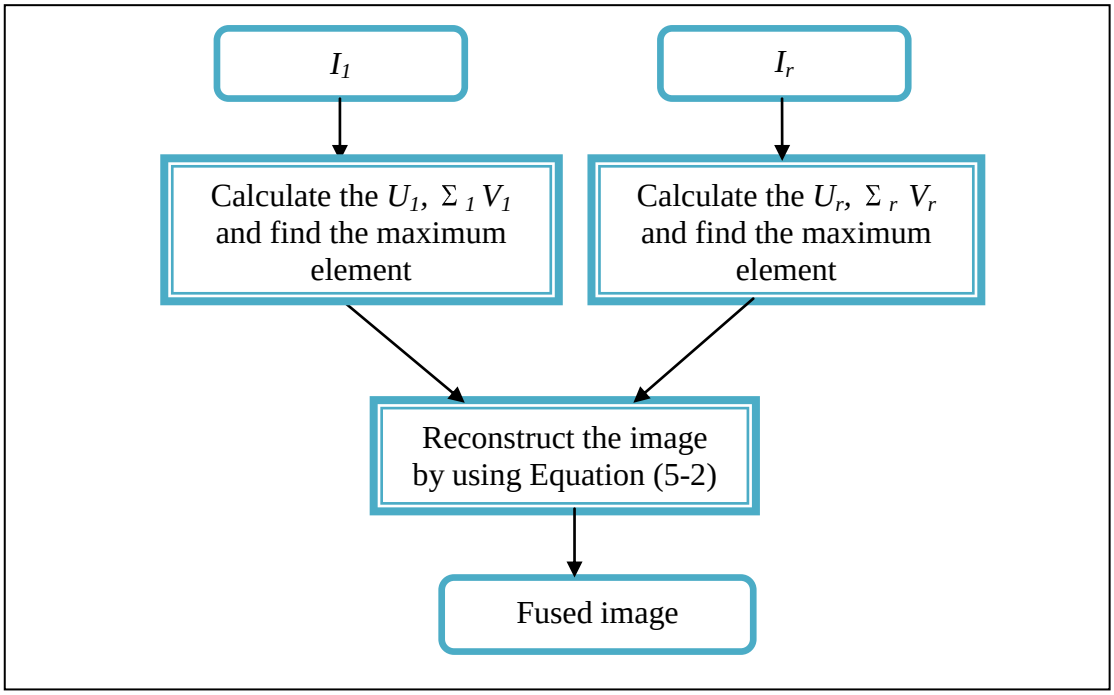


Figure 5-2: The block diagram of the proposed SVD image fusion method with two low resolution images I_1 and I_r.

The SVD of the reference image I_1 and the registered image I_r is calculated as $I_1 = U_1 \sum_1 V_1^T$ and $I_r = U_r \sum_r V_r^T$, respectively. For colour images, the decomposition is carried out for each colour matrix separately. Let the maximum singular values of matrices $\sum_1$ and $\sum_r$ be $\sigma_{1\max}$ and $\sigma_{r\max}$, respectively. Then if $\sigma_{1\max} > \sigma_{r\max}$, matrix $\sum_1$ is chosen to be used in the reconstruction of the fused image I_f; otherwise, matrix $\sum_r$ is chosen. That is,

$$I_f = U_r \sum_{\max} V_r^T \tag{5-2}$$

Where $\sum_{\max} = \sum_1$ if $\sigma_{1\max} > \sigma_{r\max}$ or $\sum_{\max} = \sum_r$, otherwise. Matrix I_f will be further used in the interpolation step. Note that, the approach can be extended to the case when more than two low resolution images are available for registration.

When $\sigma_{r\max} > \sigma_{1\max}$ the reference image is not used at all. However, from the results it is observed that in most cases $\sigma_{r\max} < \sigma_{1\max}$, which means that the fused image is formed as a combination of the reference image and the registered image. This way, higher singular values of the reference image are retained since they contain a significant amount of image information preserving useful features of the reference low resolution image. Note that the reference and test images have similar intensity.

The purpose of the SVD-based fusion step is to maintain the contrast level of the low-resolution image which is degraded in the registration process. This can be seen from Table 5-1, where for most of the test images the low-resolution reference image has better contrast than the registered image. The preservation of good contrast in the reference image enhances the reconstructed high-resolution results. Table 5-1 shows Root Mean Square (RMS) contrast, that calculate the standard deviation of the error signal (Yao et al., 2005), for all the test images. The result shows that out of eight tested images, only in one case was the contrast of the registered image higher than that of the reference image.

Table 5-1: The RMS contrast for the test images. Smaller value of the RMS contrast indicates higher contrast.

Test Image	RMS Contrast	
	Low resolution	Registered
Radcliffe	0.2528	0.2528
Chart1	0.0842	0.0871
Signboard	0.0906	0.0907
Girl	0.1315	0.1391
Letters	0.1389	0.1431
Chart2	0.3253	0.3263
Disk	0.0476	0.0491
Text	0.0368	0.0521

5.3.3 Interpolation

After the SVD-based image fusion, an interpolation step is performed. Generally, any interpolation technique can be used after SVD-based fusion. In this work, the standard bicubic interpolation and the single frame super-resolution approach proposed in (Kim and Kwon, 2008) are used.

The latter interpolation algorithm is summarised as follows:

1. First the fused image I_f is interpolated using bicubic interpolation into the desired scale.
2. Then a set of candidate images based on patch-wise regression is generated by utilising the kernel ridge regression, and a sparse basis is found by combining kernel matching pursuit and gradient descent.
3. An output image is obtained by combining all candidates based on estimated confidences for each pixel.
4. Post-processing of the regression image is performed to preserve the discontinuity at major edges, because the kernel ridge regression tends to smooth major edges.

The details can be found in (Kim and Kwon, 2008).

5.4 Results

This section presents the experimental results obtained with the proposed SVD-based fusion method with bicubic interpolation and the single-image super-resolution approach of (Kim and Kwon, 2008). The images used in the experiments are shown in Figure 5-3. The proposed method is further tested using low-resolution video frames The images used in the experiments are shown in Figure 5-3. The proposed method is further tested using low-resolution video frames from (Milanfar) i.e., 'text' and 'disk'. These images are chosen since they represent a wide cross-section of image statistics. Indeed, images 'letters' and 'chart' contain a lot of high frequencies – letters and numbers; 'Radcliffe' is a real-world image taken from (Philbin and Zisserman), with many edges and smooth areas in between, while 'signboard' and 'girl' are real-world images that contain a combination of the above features.

The super-resolution results compare the proposed method with bicubic interpolation, a multi-image super-resolution method that uses the same image registration method but without SVD-based fusion, a multi-image super-resolution method that uses the registration method of (Vandewalle et al., 2006) and robust super-resolution method of (Zomet et al., 2001) (implemented using super-resolution software of (Vandewalle et al., 2006b)) and the state-of-the-art single-image super-resolution approach of (Kim and Kwon, 2008).

Figure 5-3: Six low resolution images used in the experiments, referred to as (from left to right and top to bottom): ‘letters’, ‘chart1’, ‘chart2’, ‘Radcliffe’, ‘signboard’ and ‘girl’.

Experiment 1: In this set of experiments, the low-resolution image, I_2, was obtained from reference low-resolution image I_1. The low-resolution image I_1 is randomly translated and rotated to get low-resolution image I_2. The images were first registered, and then the SVD-based fusion was applied as explained in Section 5.3.2; finally, the bicubic interpolation or the method of (Kim and Kwon, 2008) was performed to obtain a high resolution image.

To objectively assess the quality of the reconstructed high-resolution image, a sharpness index measure (Lee et al., 2009) and blind image quality indices (BIQI) (Moorthy and Bovik, 2010b)(implemented using BIQI software release in (Moorthy and Bovik, 2010a)) as described in Chapter 3 Section 3.11 were used.

As shown in Table 5-2 and Table 5-3, the high-resolution image obtained by the SVD-based fusion with both interpolation methods has a larger sharpness index and best quality score.

Visual results are shown in Figure 5-4, Figure 5-5 and Figure 5-6 for 'letters', 'signboard', 'chart1' and 'girl' image, respectively. The proposed method is compared to the multi-image super-resolution method that does not use SVD-based fusion but uses the same registration and interpolation methods, bicubic interpolation and single-image super-resolution method of (Kim and Kwon 2008).

Table 5-2: The sharpness index (S) for the test images.

Test Image	Sharpness index (S)			
	Bicubic interpolation only	Method of (Kim and Kwon, 2008)	The proposed method with bicubic interpolation	The proposed method with the method of (Kim and Kwon, 2008)
Radcliffe	1,094,200	1,784,800	1,136,100	1,961,800
Chart1	450,700	814,210	492,400	2,998,825
Signboard	289,260	453,890	316,040	1,692,800
Girl	917,820	1,347,000	1,096,600	2,052,400
Letters	3,960	1,714,300	534,621	1,856,500
Chart2	5,162	1,859,800	15628	2,025,400

Table 5-3: The BIQI for the test images.

Test Image	BIQI score				
	Bicubic interpolation only	Robust SR (Zomet et al., 2001)	Method of (Kim and Kwon, 2008)	The proposed method with bicubic interpolation	The proposed method with the method of (Kim and Kwon, 2008)
Radcliffe	24.1505	32.6206	6.4690	6.4702	4.0347
Chart1	39.6507	6.9690	7.4113	2.9324	0.0603
Signboard	30.7851	33.7595	5.9350	13.2553	4.0351
Girl	34.5601	36.5428	6.4705	6.4498	4.0347
Letters	24.7662	49.3989	8.7549	6.4824	1.4875
Chart2	10.5394	8.6379	9.0975	6.3791	0.1112

As a direct result of the registration improvement and SVD-based fusion, the performance of the proposed super-resolution algorithm is significantly improved. This is illustrated in Figure 5-4, Figure 5-5 and Figure 5-6 from which it can be seen superior super-resolution performance. As seen in Figure 5-4(c) to Figure 5-6(c) and

Figure 5-4(f) – Figure 5-6(f) the super-resolution image obtained using the proposed method has the highest visual quality compared to the reconstructed image when no fusion had been used prior to interpolation, but the same registration is applied. This is in accordance with Table 5-2 and Table 5-3, where the high-resolution image obtained by the proposed method has the largest sharpness index and highest quality score. For Figure 5-4, it is observed that the letters in the image produced by the proposed method is sharper especially letters towards the end of the image. Note that due to the fact that the two objective measures used are highly nonlinear, and the tested image is not a natural images (as BIQI is proven to be efficient measure for natural images), for image 'letter', BIQI show significant performance advantage of bicubic interpolation method over (Kim and Kwon, 2008) compared to sharpness index measure. The proposed method also manages to improve the blur effect and enhance the image especially the edges for image 'signboard' and 'chart' as shown in Figure 5-5(f). The same improvement can be seen in Figure 5-6(f) as well, where the proposed method produces sharper and enhanced image.

(a) (b) (c)

(d) (e) (f)

Figure 5-4: Super-resolution results obtained for ‘letters’ image with: (a) bicubic interpolation, (b) SIFT-BP-RANSAC registration method with bicubic interpolation without fusion, (c) SIFT-BP-RANSAC registration method with bicubic interpolation, (d) method of (Kim and Kwon, 2008) (e) SIFT-BP-RANSAC registration method with the method of (Kim and Kwon, 2008) *without fusion*, (f) the proposed super-resolution method with the method of (Kim and Kwon, 2008)

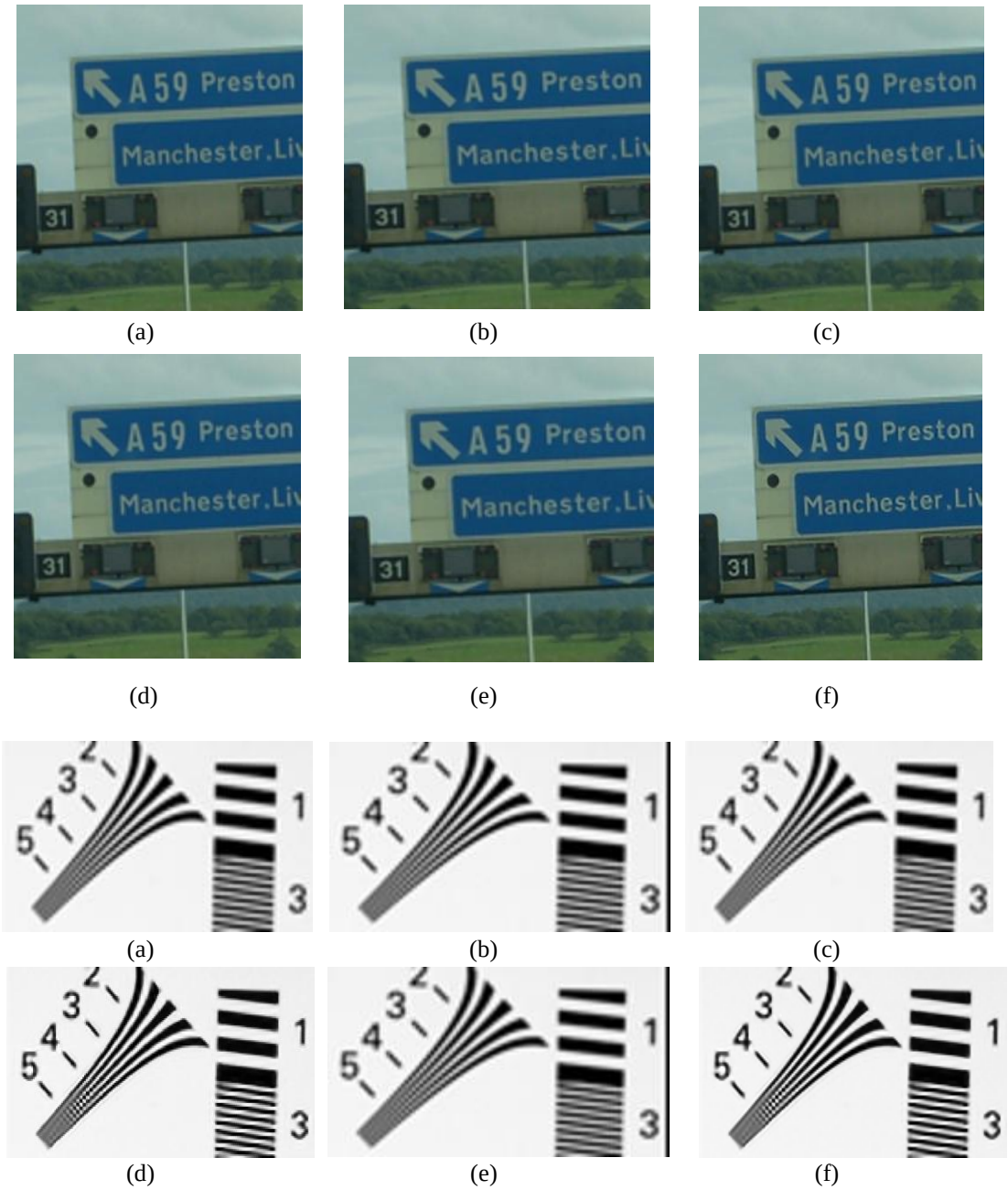


Figure 5-5: Super-resolution results obtained for 'signboard' and 'chart1' image with: : (a) bicubic interpolation, (b) SIFT-BP-RANSAC registration method with bicubic interpolation without fusion, (c) SIFT-BP-RANSAC registration method with bicubic interpolation, (d) method of (Kim and Kwon, 2008) (e) SIFT-BP-RANSAC registration method with the method of (Kim and Kwon, 2008) *without fusion*, (f) the proposed super-resolution method with the method of (Kim and Kwon, 2008)

(a)

(b)

(c)

(d)

(e)

(f)

Figure 5-6: Super-resolution results obtained for 'girl' image with: : (a) bicubic interpolation, (b) SIFT-BP-RANSAC registration method with bicubic interpolation without fusion, (c) SIFT-BP-RANSAC registration method with bicubic interpolation, (d) method of (Kim and Kwon, 2008) (e) SIFT-BP-RANSAC registration method with the method of (Kim and Kwon, 2008) *without fusion*, (f) the proposed super-resolution method with the method of (Kim and Kwon, 2008)

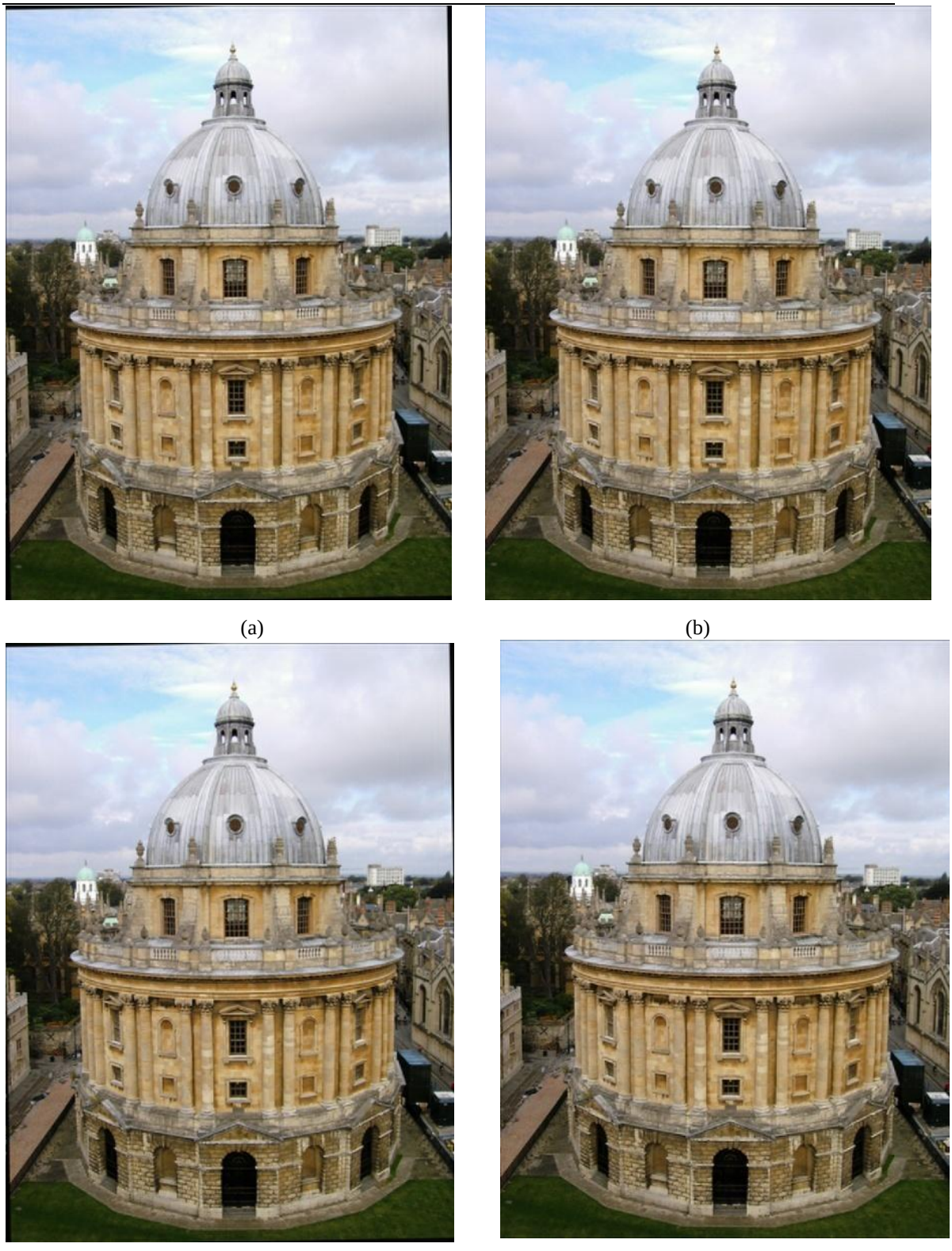
Experiment 2: In this set of experiments, the low-resolution images are created by downsampling a high-resolution image. For this experiment, image 'Radcliffe' taken from (Philbin and Zisserman) and 'Chart' are used.

Peak signal-to-noise ratio (PSNR) of the reconstructed image is used to objectively assess the quality of the reconstructed high-resolution image. The results are shown in Table 5-4, from where it is clear the gain of SVD-based fusion is roughly 8dB for 'Radcliffe' image. The proposed method suggests an improvement over method of (Kim and Kwon, 2008) and Bicubic interpolation. The PSNR value of 74.30dB was obtained for 'Radcliffe' image by the proposed method with the method of (Kim and Kwon, 2008) and 74.29dB for the proposed method with bicubic interpolation. The performance of the proposed method over the method of (Kim and Kwon, 2008) and bicubic without fusion for image 'chart' is comparable with some gain of 0.24dB.

Table 5-4: The PSNR for the test images.

Test Image	PSNR (in dB)				
	Bicubic interpolation only	Robust SR (Zomet et al., 2001)	Method of **(**Kim and Kwon, 2008)	The proposed method with bicubic interpolation	The proposed method with the method of (Kim and Kwon, 2008)
Radcliffe	66.39	30.08	66.41	74.29	74.30
Chart	65.27	31.69	65.28	65.52	65.52

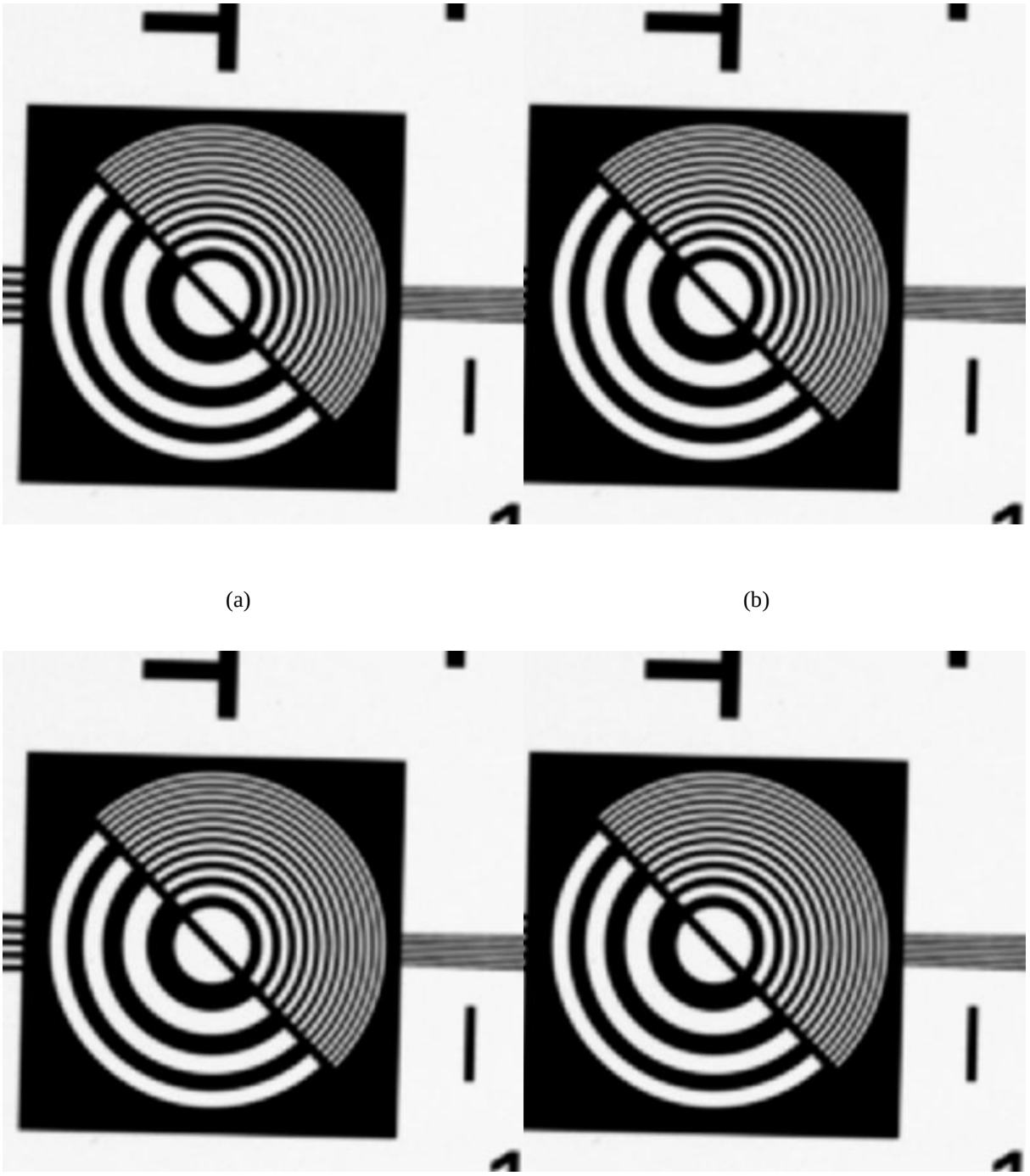
Visual results are shown in Figure 5-7(a) to (d) and Figure 5-8(a) to (d). From the figures, it is observed that the proposed fusion method shows better performance than the interpolation method and method of (Kim and Kwon, 2008) without fusion. This is because the fused image preserves all the useful information from the low-resolution. From the figures, one can see that the proposed method produces sharper images.



(a) (b)

(c) (d)

Figure 5-7: Super-resolution results obtained for ‘Radcliffe’ image with: (a) bicubic interpolation, (b) the SIFT-BP-RANSAC registration method with bicubic interpolation, (c) method of (Kim and Kwon, 2008), (d) the proposed super-resolution method with the method of (Kim and Kwon, 2008).

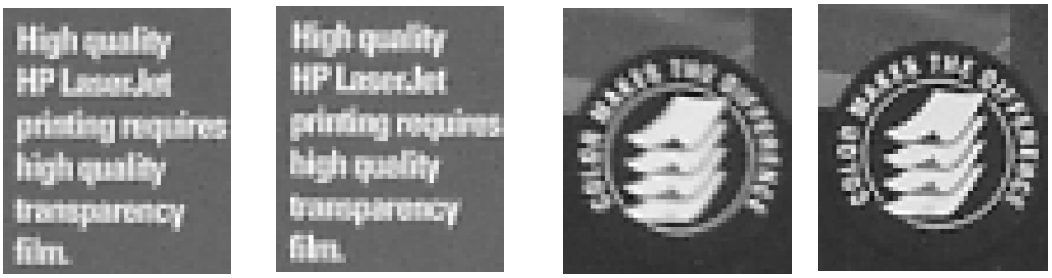


(a) (b)

(c) (d)

Figure 5-8: Super-resolution results obtained for 'Radcliffe' image with: (a) bicubic interpolation, (b) the SIFT-BP-RANSAC registration method with bicubic interpolation, (c) method of (Kim and Kwon, 2008**)**, (d) the proposed super-resolution method with the method of (Kim and Kwon, 2008).

Experiment 3: The proposed method using the first two frames from the low-resolution video sequences, 'text' and 'disk' taken from (Milanfar) and shown in Figure 5-9.



(a) (b)

Figure 5-9: The low-resolution video sequences used in the experiment: (a) 'text' and (b) 'disk'.

To objectively assess the quality of the reconstructed high-resolution image, sharpness index measure (Lee et al., 2009) and BIQI (Moorthy and Bovik, 2010b) were used. As shown in Table 5-5 and Table 5-6, the high-resolution image obtained by the SVD-based fusion with both interpolation methods has a larger sharpness index and best quality score.

Table 5-5: The BIQI for the test images.

Test Image	BIQI score				
	Bicubic interpolation only	Robust SR (Zomet et al., 2001)	Method of (Kim and Kwon, 2008)	The proposed method with bicubic interpolation	The proposed method with the method of (Kim and Kwon, 2008)
Text	28.2128	28.8263	0.9539	26.5933	0.4486
Disk	11.4249	16.8026	0.0772	7.7036	0.0633

Table 5-5 shows the BIQI score for the test images. From the table, it is observed that the proposed method provides an improvement over method of (Kim and Kwon, 2008) and Bicubic interpolation. For image 'Text', the performance of the proposed method is comparable with the original method with 1.6 gains in BIQI quality score. For image 'Disk' the proposed method gives an improvement of 3.7 quality score. Table 5-6 shows the sharpness index measure. It is observed that the proposed fusion method produces larger sharpness index than the original method without fusion.

Table 5-6: The sharpness index (S) for the test images

Test Image	Sharpness index (S)				
	Bicubic interpolation only	Robust SR (Zomet et al., 2001)	Method of (Kim and Kwon, 2008)	The proposed method with bicubic interpolation	The proposed method with the method of (Kim and Kwon, 2008)
Text	92,217	98,599	297,240	177,480	299,660
Disk	79,721	76,003	256,960	157,410	259,870

Visual results are shown in Figure 5-10. It can be seen from the figures that the proposed method provides the best visual experience. Note that due to nonlinearity of the used objective measures, BIQI and sharpness index show significant performance advantage of method of (Kim and Kwon, 2008) over bicubic interpolation, while visually, the difference is small.

Note that for the same dataset used, better visual results are presented in (Farsiu et al., 2004), however, in (Farsiu et al., 2004) twenty low-resolution video frames have been used for generating the high-resolution image compared to only two used in this experiment.

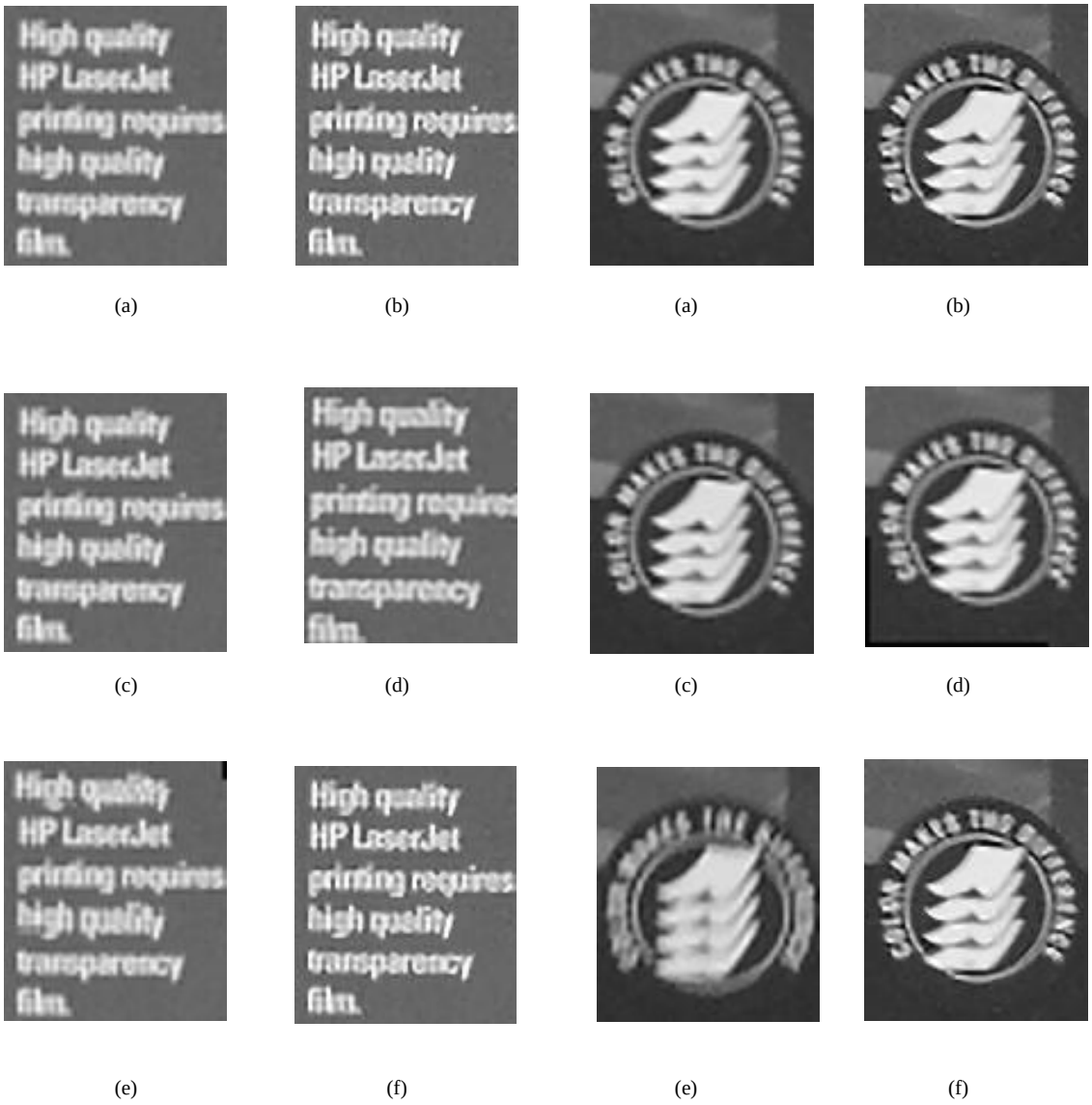


Figure 5-10: Super-resolution results with: (a) bicubic interpolation, (b) method of (Kim and Kwon, 2008**)**, (c) the SIFT-BP-RANSAC registration method with bicubic interpolation, (d) the SIFT-BP-RANSAC registration method with the method of (Kim and Kwon, 2008) *without fusion* (e) Method of (Zomet et al., 2001), (f) the proposed super-resolution method with the method of (Kim and Kwon, 2008)

Processing Time: Table 5-7 shows the processing time of the proposed method compared to bicubic interpolation and the method of (Kim and Kwon, 2008). All algorithms were run on a PC with the Pentium processor (2.50GHz) and 2GB memory. The table shows that the registration and fusion steps require roughly four seconds.

Table 5-7: Processing time comparison (Elapsed Time) [sec].

Test Image	Times(seconds)			
	Bicubic interpolation	Method of (Kim and Kwon, 2008)	The proposed method with bicubic interpolation	The proposed method with the method of (Kim and Kwon, 2008)
Radcliffe	0.09	35.77	4.05	39.73
Chart	0.09	13.39	2.57	15.87
Signboard	0.07	9.96	2.65	12.54
Girl	0.35	43.90	3.18	46.73
Letters	0.06	37.18	1.79	42.21
Chart2	0.09	64.65	1.39	74.02
Text	0.01	0.04	0.21	0.76
Disk	0.01	0.05	0.23	0.77

5.5 Summary

This chapter presents a novel SVD-based image fusion method for super-resolution image reconstruction. The proposed image registration and fusion steps seem effective in maintaining the important features of the test images and greatly improve the super-resolution results. The proposed SVD-based fusion method suggests retaining the important information in the registered low-resolution images and hence improving the super-resolution results. The proposed super-resolution scheme shows performance improvements compared to the bicubic interpolation and methods of (Zomet et al., 2001) and (Kim and Kwon, 2008), using both simulated and real-world images and video frames.

CHAPTER 6

6. SINGLE IMAGE SUPER-RESOLUTION USING SPARSE REPRESENTATION

6.1 Introduction

Single-image super-resolution with dictionary learning is a recent effective method to increase resolution of a captured image by exploiting the fact that the image is sparse in some transform domain. Several recent papers focus on single-image super-resolution see for example, (Yang et al., 2008, Glasner et al., 2009., Adler et al., 2010, Zeyde et al., 2010, Kim and Kwon, 2008). Single image super-resolution methods can be categorised into (1) the interpolation-based method and (2) the learning-based method. The learning based methods are superior, but require a large amount of training data to generate efficient learning models. Some of the learning-based methods need prior estimation of the high-resolution image.

Recently, sparse representation has been proposed as an efficient learning-based single-image super-resolution technique (Zeyde et al., 2010, Yang et al., 2008). Sparse representation is applied in many fields in signal processing including applications such as image denoising, image restoration and classification (Elad and Aharon, 2006, Huang and Aviyente, 2007, Mairal et al., 2008).

Various attempts have been made to regularise the ill-posed inverse problem of image super-resolution using sparse representations of low-resolution image patches (Yang et al., 2010, Zeyde et al., 2010). An image patch can be represented as a sparse linear combination of elements from an over-complete dictionary (Wang et al., 2011, Kim and Kwon, 2008). The idea is to seek a sparse representation for each patch of the low-resolution input, followed by exploiting this representation to

generate a high-resolution output. By jointly training two dictionaries for the low-resolution and high-resolution image patches, the sparse representation of a low-resolution image patch can be applied with the high-resolution image patch dictionary to generate a high-resolution image. Yang et al proposed a sparse representation based super-resolution method (Yang et al., 2008), based on the assumption that the high-resolution and low-resolution patches have the same sparse representation coefficients with respect to a high-resolution dictionary and a corresponding low-resolution dictionary. Recently, this w, this work has been extended and improved by (Zeyde et al., 2010).

This paper presents a novel image interpolation method that is used to improve single image super-resolution using sparse representation in (Zeyde et al., 2010). The paper improves a single-image-super-resolution result of (Zeyde et al., 2010) both visually and objectively. The remainder of the chapter is organised as follows. An overview of super-resolution using sparse representation is presented in Section 6.2. In Section 6.3, the improved single image super-resolution using sparse representation is discussed. The proposed image interpolation method is given in Section 6.3.1. Section 6.4 discusses the obtained results. Finally, summary are presented in Section 6.6.

6.2 Super-resolution via sparse representation (Zeyde et al., 2010)

Sparse theory (Candés et al., 2006) states that many natural signals can be expressed as a linear combination of elementary signals or *atoms*, where most of the representation coefficients are zero. Often, the atoms are chosen from a so called over-complete dictionary. An over-complete dictionary is a collection of atoms such that the number of atoms exceeds the dimension of the signal space. Sparse representation is widely used in noise reduction, feature extraction, pattern recognition and image super-resolution (Elad, 2010).

The single-image super-resolution problem focuses on recovery of a high-resolution image from a low-resolution one, i.e., given a low-resolution image as $z_l \in R^{N_l}$,

represented as a vector of length N_l pixels, recover higher-resolution image $y_h \in R^{N_h}$, represented as a vector, of the same scene, where $N_l < N_h$. The observation model can be represented as

$$z_l = SHy_h + v \quad \textbf{(6-1)}$$

where $S: R^{N_h} \to R^{N_l}$ is a decimation operator, $H: R^{N_h} \to R^{N_h}$ as a blurring operator and *v* is an additive Gaussian noise.

The goal of super-resolution using sparse representation is to process $z_l \in R^{N_l}$ to produce $\hat{y}_h \in R^{N_h}$ such that $\hat{y}_h \approx y_h$. Due to the Gaussianity of *v*, the maximum-likelihood estimation is obtained by the minimisation $\left\| SH\hat{y} - z_l \right\|_2$. Since *SH* is rectangular with many columns than rows, it cannot be inverted stably. Zeyde et al used Sparse-Land local model (Elad and Aharon, 2006) to overcome this problem. This model assumes that each patch from the images considered can be represented well as a linear combination of a few atoms from a dictionary using a sparse coefficient vector (Zeyde et al., 2010).

In super-resolution using sparse-representation, the low-resolution image is scale-up by an interpolation operator $Q: R^{N_l} \to R^{N_h}$ to avoid complexities due to different resolutions of the low-resolution image and the high-resolution image. The scale-up image is denoted by

$$y_l = Qz_l = Q(SHy_h + v) = QSHy_h + Qv = L^{all} y_h + \hat{v} \quad \textbf{(6-2)}$$

The algorithm of (Zeyde et al., 2010) operates on patches extracted from the low-resolution image y_l aiming to estimate the corresponding patch for the high-resolution image.

Let

$$p_k^l = R_k y_l \in R^n \quad \textbf{(6-3)}$$

be a low-resolution image patch of size $\sqrt{n}\times\sqrt{n}$ extracted by operator $R_k : R^{Nh} \to R^n$ from the image z_l in location *k*. *k* is the location in which the patches are extracted around the true pixels in the low-resolution image y_l . The sparse coefficients can be estimated from these low-resolution patches, and then using this information the dictionary can be estimated.

According to sparse theory (Candés et al., 2006), each patch from the image can be well represented sparsely by $\alpha_k \in R^m$ over the dictionary $A_h \in R^{n\times m}$. Each patch is considered to be generated by multiplying the dictionary by a sparse (mostly zero) vector of coefficients.

$$p_k^l = A_h \alpha_k \quad \textbf{(6-4)}$$

where $P_k^l \in R^n$ is the low-resolution patches, $A_h \in R^{n\times m}$ is an over-complete dictionary and $\alpha_k \in R^m$ is a vector with very few ($\ll n$) non-zero entries.

For super-resolution based on sparse representation, given a low-resolution image feature patch p_k^l, based on over-complete dictionary A_h, the sparse representation α^k is estimated. The corresponding high-resolution dictionary A_h will be multiplied with the sparse representation α^k to generate the high-resolution patch p_h^k.

$$p_h^k = A_h \alpha^k \quad \textbf{(6-5)}$$

The relation between the low-resolution and the high-resolution patches is

$$P_l^k = LP_h^k + \hat{v}_k \quad \textbf{(6-6)}$$

Where *L* is a local operator being a portion of L^{all} and $\hat{v}_k$ is the additive noise in the patch.

Associate (1-4) into (1-6) gives

$$LA_h\alpha_k = LP_k^h = p_k^l - \hat{v}_k \quad \textbf{(6-7)}$$

Implying that

$$\left\| P_k^l - LA_h\alpha \right\|_2 \le \varepsilon \quad \text{(6-8)}$$

Where ε is related to the noise power σ of $\hat{v}$.

From (6-8), the low-resolution patch can be represented by the same sparse vector α^k over the effective dictionary $A_l = LA_h$ with the controlled error ε_l. Based on this for a given low-resolution patch p_k^l, the sparse representation α^k can be estimated then the high-resolution patch p_h^k are generated by multiplying the sparse representation α^k by the dictionary A_h (Zeyde et al., 2010) The process of super-resolution based on over-complete sparse representation involves the construction of the dictionary as well as solving the sparse decomposition problem.

6.3 The proposed method

In this section, the proposed image interpolation for single-image super-resolution using sparse representation is described for generating a high-resolution image from a low-resolution image. The proposed method consists of two steps: (i) image interpolation for prior estimate of the high-resolution image (ii) single image super-resolution using sparse representation (Zeyde et al., 2010). Figure 6-1 shows the block diagram of the proposed method. First the low-resolution image z_l is scaled up to the original size using the proposed interpolation method. This step provides image with stronger edges. The proposed interpolation method improves the estimation of the high-resolution image prior to the feature extraction phase in the single image super-resolution using sparse representation (Zeyde et al., 2010). Finally the enhanced image is used to reconstruct high-resolution image using method of (Zeyde et al., 2010). Next, an overview of the overall single image super-resolution using sparse representation is described in details.

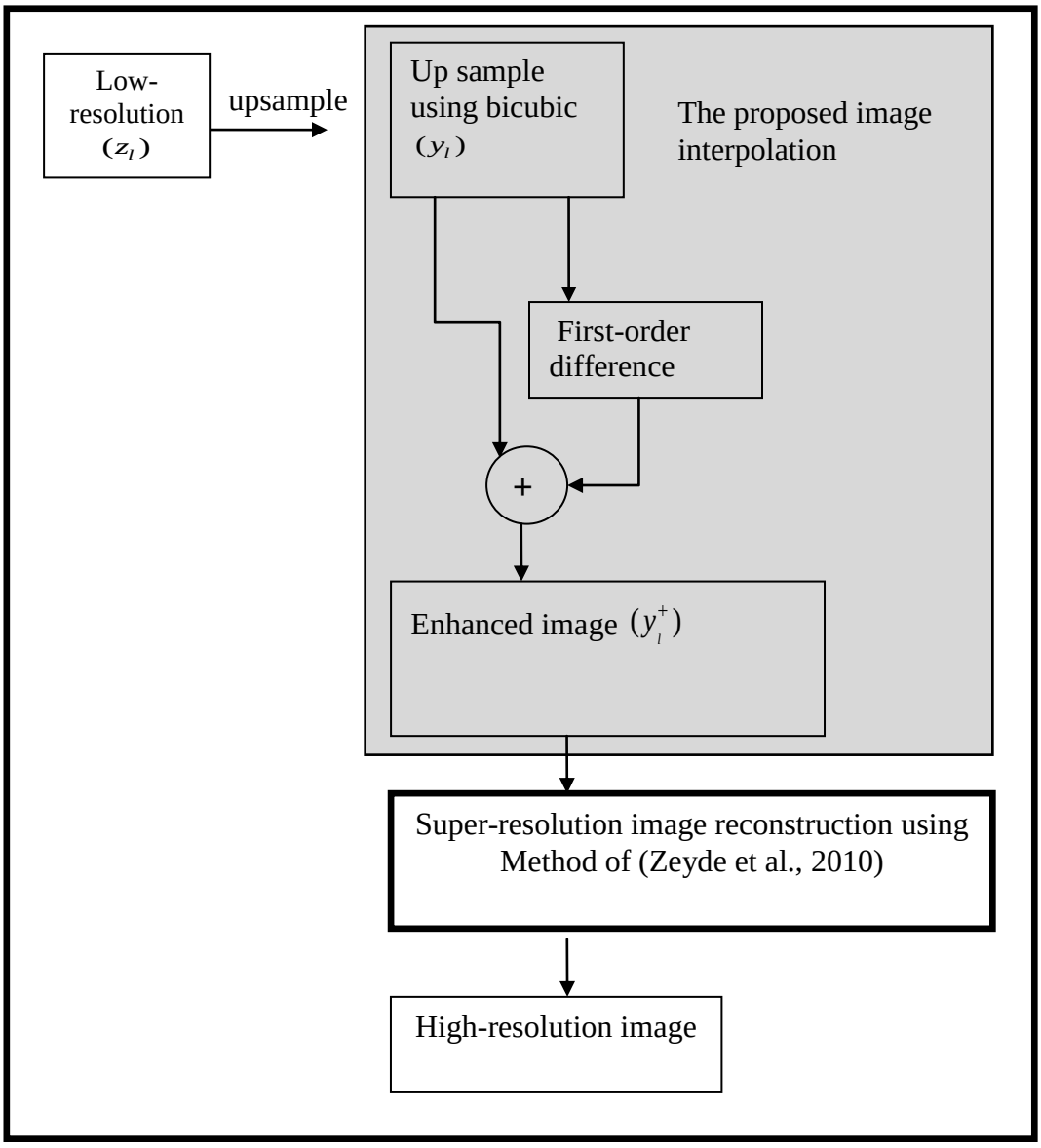


Figure 6-1: The block diagram of the improved single image super-resolution using sparse representation

6.3.1 The proposed image interpolation method

In this section, the proposed image interpolation method is described for generating an interpolated image from a low-resolution image. The low-resolution patches are gathered from the interpolated image to avoid the complexity cause by different resolution as well as to avoid coordinate ambiguity. The proposed method consists of three steps: (i) bicubic interpolation on the low-resolution image (ii) calculation of

the first order difference of the interpolated image and (iii) the resulting image is then added to the interpolated image.

First, the low-resolution image z_l is first interpolated using bicubic interpolation.

$$y_l = Qz_l \tag{6-9}$$

where matrix Q denotes interpolation operator that produces the interpolated image y_l. Then first-order difference of the interpolated image is calculated using

$$y_e = [y_l(2:m,:) - y_l(1:m-1,:)] \tag{6-10}$$

which returns a matrix of row differences. The resulting image y_e is then added to the interpolated image y_l .

$$y_l^+ = y_l + y_e \tag{6-11}$$

The proposed method enhances the interpolated image so that the edge information is more evident in the image. The enhanced image will provides better features for generation of image patches. The edges are the locations where there is discontinuity in image pixel intensity. The locations where the discontinuities happen can be found by applying the first-order difference equation on the interpolated image.

The images used to test the proposed method are listed in Table 6-1. The original high-resolution image is downsample to create a low-resolution image. Note that, the test images used in this experiment did not contain noise. This low-resolution image is then interpolated using the proposed method. Adding the first-order difference to the interpolated image y_l will enhance the image. Figure 6-2 shows how this linear combination technique enhances the edges for image 'Child'. To objectively assess the quality of the proposed interpolation method sharpness index measure (Lee et al., 2009) and PSNR are used. Table 6-1 shows the sharpness index measure and PSNR comparing the bicubic interpolation and the proposed interpolation method. One can see from the table that the images obtained using the proposed method has larger sharpness index and PSNR than bicubic interpolation. It will be shown later that this

improvement will provide better features for the image patches and dictionary learning phase in single-image super-resolution using sparse representation in (Zeyde et al., 2010) and will improve the super-resolution results.

(a) (b)

Figure 6-2: Results of interpolated the low-resolution image using (a) Bicubic interpolation (b) the proposed method.

Table 6-1: The sharpness index (S) and PSNR for the test images.

Test Image (Size)		Sharpness index (S)		PSNR (dB)	
		Bicubic Interpolation	The proposed method	Bicubic Interpolation	The proposed Method
Lena	(170x170)	3.5389e+005	3.9712e+005	27.84	28.85
Text	(120x120)	4.0476e+005	4.7814e+005	18.01	18.26
Peppers	(85x85)	1.4965e+005	1.6510e+005	23.05	23.58
Barbara	(170x170)	3.8864e+005	4.3444e+005	26.74	26.95
Building	(266x266)	1.6412e+006	1.7757e+006	23.68	24.27
Child	(170x170)	3.6104e+005	3.9958e+005	27.62	28.47

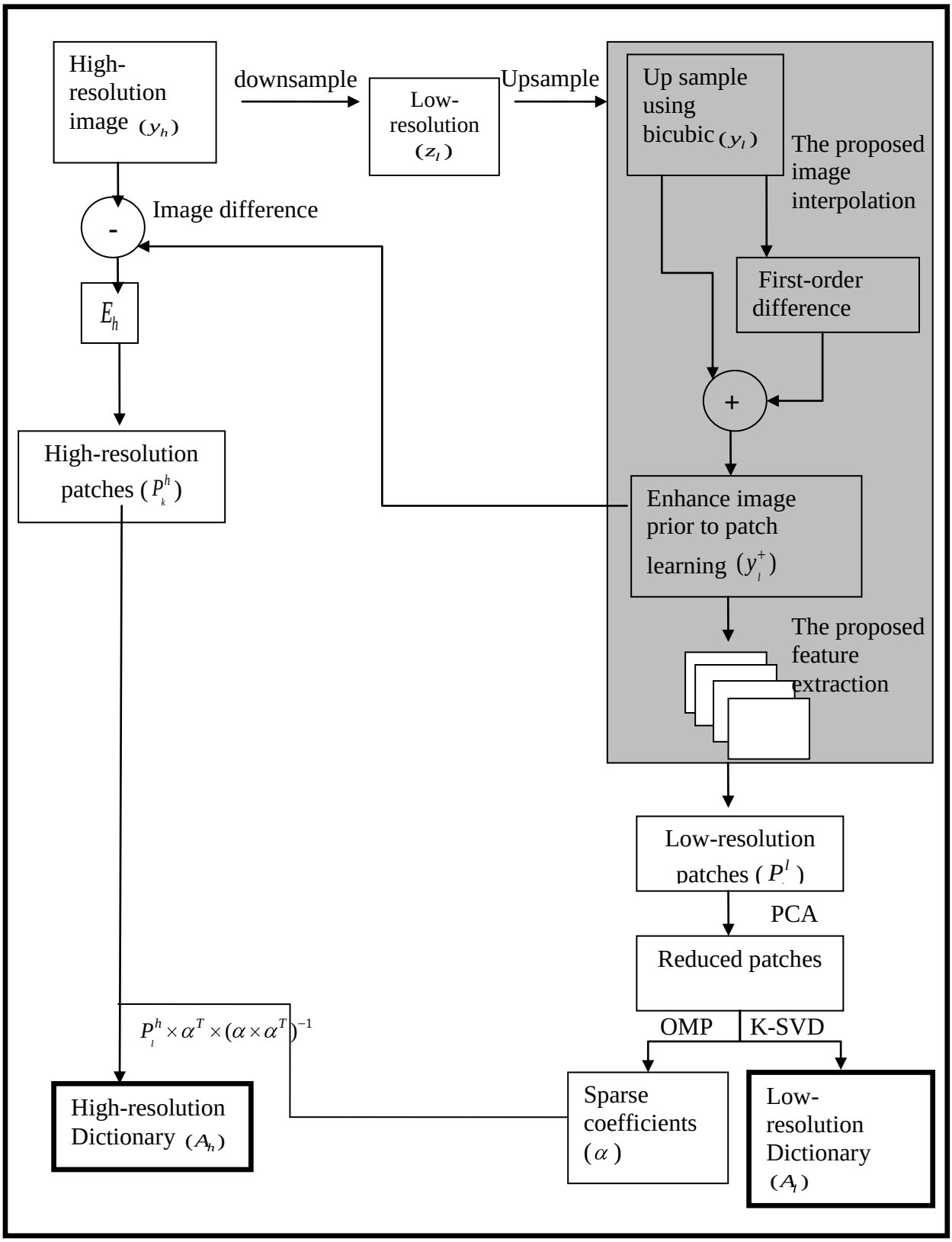


Figure 6-3: The block diagram of the training phase.

6.4 Super-resolution using sparse representation

As in (Zeyde et al., 2010), there are two ways of obtaining a high-resolution image with the proposed method: (i) using the ground truth image to acquire the over-complete dictionaries of the image patches and (ii) bootstrapping, where only the low-resolution image is used to obtain the dictionary and image patches. The following subsection describes the single image super-resolution using the ground truth image followed by the bootstrapping method.

6.5 Super-resolution using ground truth image

As in (Zeyde et al., 2010), the proposed method uses two phases in obtaining a high resolution image from a low-resolution image: (i) training phase and (ii) reconstruction phase. The method extracts two sparse dictionaries from low and high-resolution training image patches. Then these dictionaries are used to super-resolve the low-resolution test image. Each of the steps in the proposed method will be described next.

6.5.1 Training phase

The training phase consists of three main steps: (i) training set construction, (ii) feature extraction and (iii) dictionary training. Figure 6-3 shows the block diagram of the training phase. The proposed improvement is shown as the grey box in Figure 6-3.

Training set construction: The training phase starts by generating a low-resolution image from the ground truth image. The ground truth image y_h is blurred and downsampled. This will produce a degraded low-resolution image z_l. This process is essentially the same as in (Zeyde et al., 2010). Then the image z_l is scaled up to the original size using bicubic interpolation to avoid the complexity cause by different resolution as well as to avoid coordinate ambiguity. The upscale image y_l has the same resolution as y_h but lacks of textures including blur edges. In order to enhance the upscale image, sharpening the upscale image prior to the feature extraction phase is considered. The upscale image is enhanced using the proposed image interpolation

as described in Section 6.3.1. The enhanced image will provide better features for extraction of the image patches and the dictionary learning phase. This process will significantly improve the super-resolution results.

This step produces two pairs of training images y_h, y_l^+ i.e high-resolution and enhanced low-resolution images. The next step is to extract pairs of matching patches from these two training images.

Features extraction: The features are extracted from the two training images i.e. the high-resolution image and the enhanced low-resolution image. Different features for the low-resolution image patch have been used in the literature. The high-pass filter has been used to extract the edge information in (Freeman et al., 2000). While in (Sun et al., 2003) a set of Gaussian derivative filters are used to extract the contours in the low-resolution patches. (Chang et al., 2004) used the first- and second-order gradients of the patches as the representation. (Zeyde et al., 2010) used the first- and second-order gradients of the patches as the representation. In this work, the first order difference equation is proposed as a filter to extract the features for the low-resolution patch due to their simplicity and effectiveness.

Prior to patches extraction, the pre-processing for high-resolution image is done to remove the low-frequencies from the high-resolution image by computing the difference images $E_h = y_h - y_l^+$. Computing the difference of these images will remove the low-frequencies and retain the coefficients which correspond to the high frequency information. The reason is to extract the critical details in the image (Zeyde et al., 2010). The low-resolution image is filtered using the first-order difference equation to extract the local features that correspond to their high-frequency information. The first-order difference equation returns a matrix of row differences and then the transpose matrix is calculated. The step is repeated to produce four filtered images.

The low-resolution image patches are extracted from the filtered images while the high-resolution image patches are extracted from the difference image E_h. The whole image is divided into same-sized patches with several pixels overlap.

The high resolution patches P_h are then extracted from the difference image E_h using 9×9 block size which is 81 vectors at one time. These vectors are arranged in 81×1 matrix. The total size for the high-resolution patches is calculated as:

$$P_k^h\,(size)=[n^2,\ \ N/s-2*dd\ ^2] \tag{6-12}$$

where *n* is the block size of the image patch, *dd* is the displacement between current and previous block, *s* is a scale-down factor and *N* is the number of columns in the image. This high-resolution patch will be used to create the high-resolution dictionary in the dictionary learning phase. The process to gather the high-resolution image patches is similar to (Zeyde et al., 2010) except that the difference image is obtained from the enhanced image and the original high-resolution image.

For low-resolution image, patches are selected from the features extraction in the four filtered image. The corresponding low-resolution patches are extracted from the same locations in the four filtered image using the same block size i.e. (9×9 pixels) which is 4×81 vectors. These vectors are arranged in 324×1 matrix. There are four filtered image so these feature vectors are concatenated into one vector as the final representation of the low-resolution patches. The total size for the low-resolution patches is

$$P_k^l\,(size)=[4*n^2,\ \ N/s-2*dd\ ^2] \tag{6-13}$$

These low-resolution patches will be used to create the low-resolution dictionary.

Dimensionality reduction: The dimensionality reduction of the input low-resolution patches is used to reduce dimension of the low-resolution patches and to reduce computational complexity in the subsequent training and super-resolution algorithms. The Principal component analysis (PCA) algorithm is applied on these patches, finding the components with the strong energy. The eigenvalues of the low-resolution patches are sorted into the descending order and only the coefficients larger than 0.999 will be selected to be put in the reduced feature vector. This is to preserve 99.9% of the energy. The projection that transforms the patch $\tilde{p}_k^l \in R^{nR}$ (vector of length *nR* pixels) to its reduced feature vector, $p_k^l \in R^{nl}$ (vector of length *nl* pixels) is denoted by $B \in R^{nl\times nR}$, $p_k^l \in B\tilde{p}_k^l$. The reduced patches are used in the

dictionary learning phase. Table 6-2 shows the number of coefficients larger than 0.999 in the low-resolution patches for the method of (Zeyde et al., 2010) before and after PCA and the proposed method. It can be seen that the proposed filtering method effectively reduces the dimensions of the low-resolution patches by selecting only the components with strong energy. Reducing the feature vector (dimensions) will lead to less computation.

Table 6-2: Dimensionality reduction comparison

Image	Method of (Zeyde et al., 2010) before PCA	Method of (Zeyde et al., 2010) after PCA	The proposed method
Peppers	81	29	23
Lena	81	29	24
Barbara	81	29	24
Text	81	28	25
Building	81	28	25
Child	81	41	37

Dictionary Learning: The objective of the dictionary learning methods is to find a dictionary that provides the sparsest representation for each example in the data set. The state-of-the-art algorithms to find the dictionary are the method of optimal directions (MOD) (Engan et al., 1999) and K-SVD (Aharon et al., 2006). Both methods are iterative methods in which they alternate between the sparse-coding and the dictionary update steps. More details on the dictionary learning using K-SVD and OMP can be found in (Aharon et al., 2006), (Zeyde et al., 2010).

Similar to (Zeyde et al., 2010), the K-SVD dictionary training algorithm (Aharon et al., 2006) is used for the dictionary learning for both the high-resolution and the low-resolution images. The K-SVD is applied to the low-resolution patches for low-resolution dictionary learning.

$$A_l, \{\alpha^k\} = \arg\min_{A_l, \{\alpha^k\}} \sum_k \left\| p_k^l - A_l \alpha^k \right\|^2 \quad \textbf{s.t.} \ \left\| \alpha^k \right\|_0 \le L \quad \forall k \tag{6-14}$$

L in (6-14) is the number of atoms for the sparse coefficients α^k. Then perform the sparse coding using the Orthogonal Matching Pursuit (OMP) algorithm (Engan et al., 1999) to obtain the sparse coefficients α. OMP is used as in (Zeyde et al., 2010) because of its simplicity and fast execution. Then the sparse coefficients are used to estimate the high-resolution dictionaries. The dictionary is defined to be the one that minimises the mean approximation error

$$A_h = \arg\min_{D_h} \sum_k \left\| p_k^h - A_h \alpha^k \right\|_2^2 = \arg\min_{A_h} \left\| P_h - A_h Q \right\|_2^2 \quad \textbf{(6-15)}$$

where the matrix A_h is constructed with the high-resolution training patches p_h^k as its columns, and Q contains α^k as its columns as in (Zeyde et al., 2010, Yang et al., 2008).

The resulting high-resolution dictionaries are used to update the high-resolution image patches. The high-resolution patch p_h^k can be recovered by approximating $p_l^h \approx A_h \alpha^k$. To recover p_h^k, the sparse representation vector that corresponds to the low-resolution patch is multiplied by the high-resolution dictionary.

6.5.2 Reconstruction Phase

This section describes the reconstruction phase for single image super-resolution using sparse representation. Figure 6-6 shows the block diagram of the reconstruction phase. The reconstruction phase consists of three main steps: (i) training set construction, (ii) feature extraction and (iii) high-resolution image construction. In the reconstruction phase, the test low-resolution image z_l is to be magnified. The generation of the low-resolution image is similar to the construction of low-resolution images step in the training phase. Then the proposed image interpolation described in Section 6.3.1 is performed prior to patch learning in order to enhance the upscale image. After that, the low-resolution image patch as explained in Section 6.5.1 is gathered. These two steps produce the low-resolution patches of the test image. Then the low-resolution image patches and the low-resolution

dictionary from the training phase are used to create the sparse coefficients. The OMP algorithm (Engan et al., 1999) is used for sparse coding of the low-resolution test image. The rest of the reconstruction phase is the same as in (Zeyde et al., 2010).

Then the resulting α^k is used to estimate the high-resolution patches. The approximated high-resolution patches p_k^h are obtained as

$$p_k^h \approx A_h \alpha^k \quad \textbf{(6-16)}$$

The sparse representation vectors are multiplied by high resolution dictionary A_h and the approximated high resolution patches p_h. The high resolution estimate y_h^* is obtained from P_k^h by solving the following minimisation with respect to y_h.

$$y_h^* = \arg\min_{y_h^*} \sum_k \left\| R_k(y_h^* - y_l) - p_k^h \right\|_2^2 \quad \textbf{(6-17)}$$

where R_k is the operator which extracts a patch of size *n x n* from the high-resolution image in location *k*. The output of this phase is the final super-resolved image y_h^*.

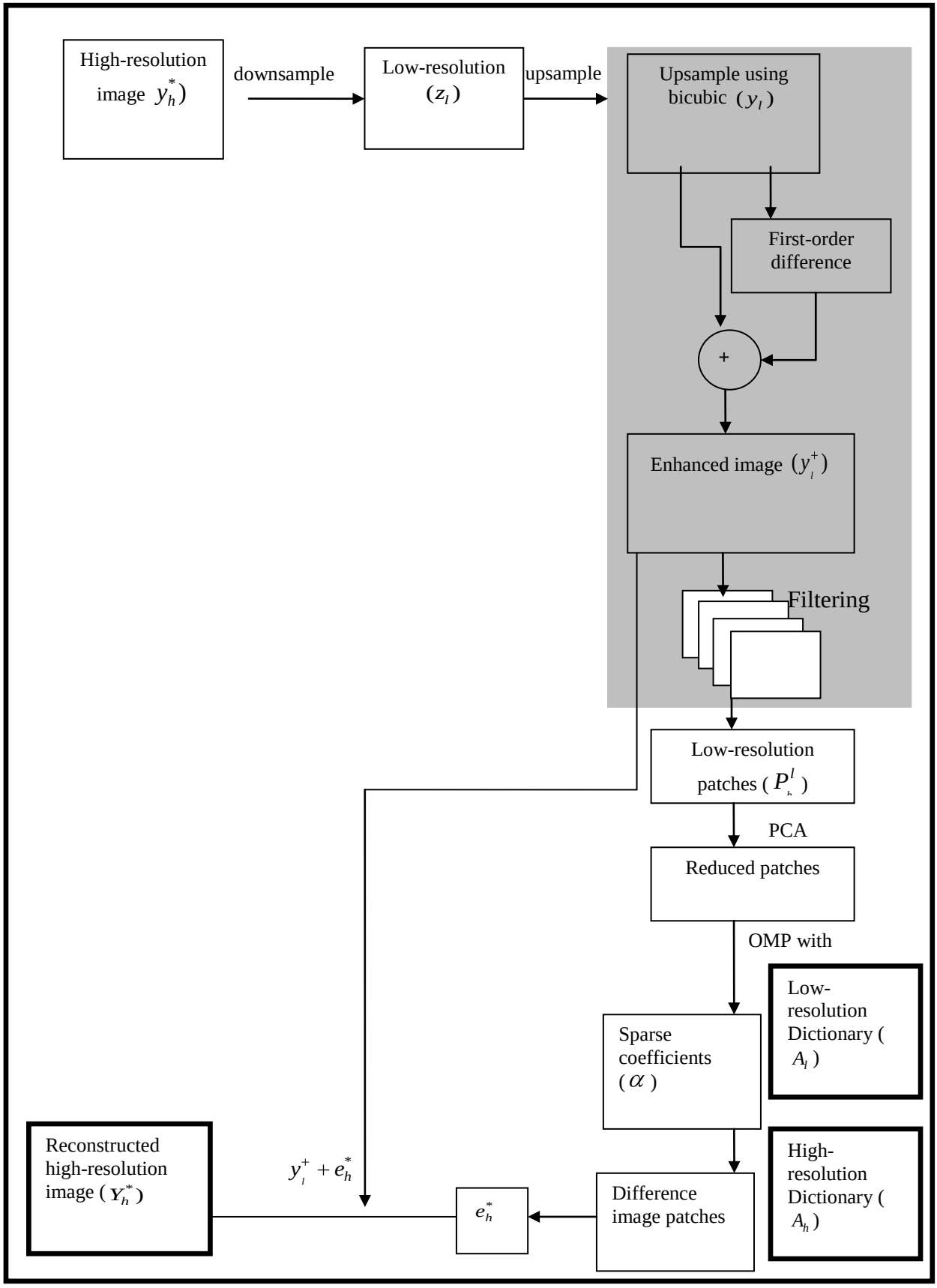


Figure 6-4: Block diagram of the reconstruction phase

6.6 Bootstrapping Method

The bootstrapping method learns the dictionaries from the low-resolution patches without the need for the ground truth image. The method is applicable to real scenarios in which there is no access to the ideal image. In this method the low-resolution test image is used to train and "bootstrap" the test image as proposed by (Glasner et al., 2009.). The low-resolution image is downsampled by 0.5 to create the low-low-resolution image. This method extracts two sparse dictionaries from these low-resolution images. Then these dictionaries are used to super-resolve the low-resolution test image. The improved bootstrapping method by the inclusion of the proposed interpolation method prior to patch learning and filtering the low-resolution image for feature extraction are shown in the grey box in Figure 6.7.

Similar to the super-resolution using ground truth image explained in Section 3.2.1, the proposed image interpolation described in Section 6.3.1 is performed prior to patch learning (shown in shaded box in Figure 6.7) in order to enhance the upscale image. Then the gathering of low-resolution image patch as explained in Section 6.5.1 is implemented here. The same reconstruction phase as explained in Section 6.5.2 is used to reconstruct the high-resolution image for the improved bootstrapping method.

6.6.1 The training phase

The training phase consists of three main steps: (i) training set construction, (ii) feature extraction and (iii) dictionary training. Figure 6-7 shows the block diagram of the training phase for the bootstrapping method. The same feature extraction and dictionary training steps as described in Section 6.5.1 are adapted in the bootstrapping method.

Training set construction: This step is the same as in (Zeyde et al., 2010). In order to train the dictionaries of the low-resolution test image and the downsampled version of the low-resolution image, the low-resolution image is downsampled by 0.5 and then upsampled back to its original size to serve as an ideal image. The trained dictionaries are used to enable the reconstruction phase, which scales up the low-

resolution image z_l to the high-resolution image y_h. The high-resolution image y_h is only used to generate a low-resolution image for testing purposes, and it is not needed for training the dictionaries.

This step produces a low-resolution image z_{ll} which is then scaled up back to the original size using bicubic interpolation. The upscale image y_l has the same resolution as y_h. Then, the proposed image interpolation described in Section 6.3.1 is applied to the upscale image y_l. This improvement provides better features for the dictionary learning phase and is significantly improve the learned dictionaries.

The remaining steps in the training phase are similar to the super-resolution method using the ground truth image as described in Section 6.5.

6.6.2 Reconstruction Phase

The same reconstruction phase as explained in Section 6.5.2 is used to reconstruct the high-resolution image for the improved bootstrapping method.

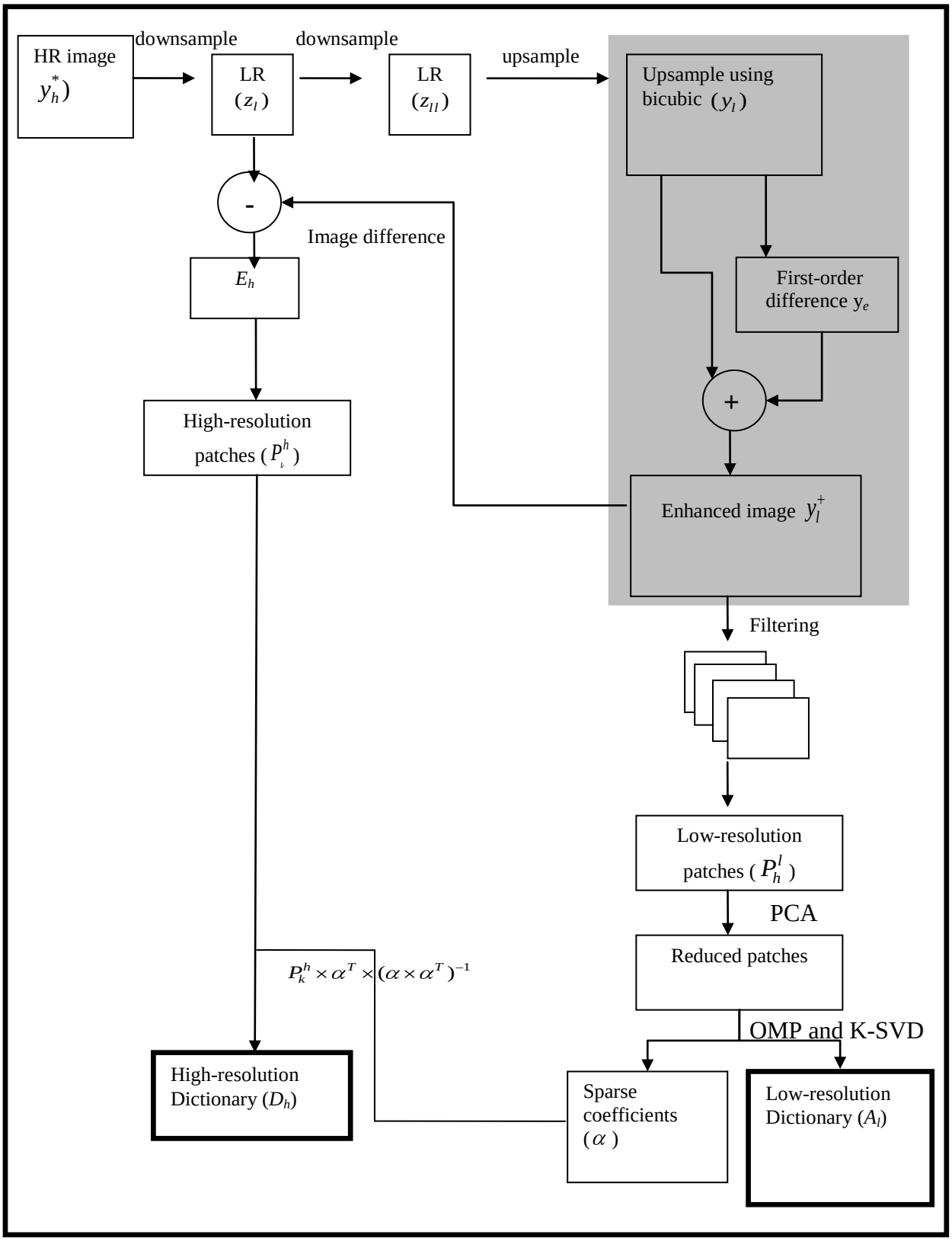


Figure 6-5: Block diagram of the training phase for bootstrapping method.

6.7 Experimental results

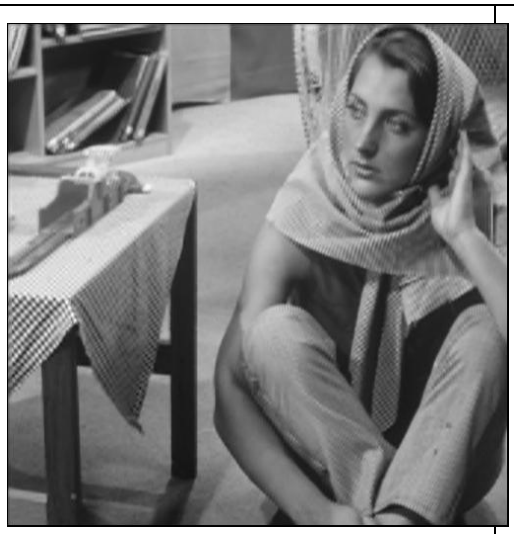
This section presents super-resolution results obtained by the proposed method. First, the result of super-resolution in which the high-resolution image is used for training is presented and then bootstrapping. The images used in the experiments are shown in Figure 6-8. The high-resolution image and the corresponding low-resolution image are used to extract the patches. The block size for patch extraction is 9x9 pixels and the overlapping between block is 3 pixels. These parameters are used to compare the results of the proposed method with (Zeyde et al., 2010). Then these patches are used for dictionary learning where the outputs are low-resolution and high-resolution dictionaries and sparse coefficients. The K-SVD algorithm (Aharon et al., 2006) is used for dictionary learning. The number of iterations for K-SVD is 40. In this work, the proposed method is compared with the method in (Zeyde et al., 2010). The results are evaluated both visually and quantitatively in terms of PSNR.

Figure 6-6: Six images used in the experiments, referred to as (from left to right and top to bottom): ‘text’, ’lena’, ‘peppers’, ‘barbara’, ‘child’ and ‘building’.

6.7.1 Super-resolution results using high-resolution dictionary

In this experiment, the high-resolution image and the low-resolution image are used to gather the image patches and to create the dictionary. The low-resolution image is generated by applying the blurring filter both horizontally and vertically to the high-resolution image and downsampling it by 0.25. Figure 6-7 shows the results for image 'barbara'. Figure 6-7 (a) shows the results of method (Zeyde et al., 2010) and (b) the result of the proposed method. To show the detail after the super-resolution, the magnified version of the selected region of (a) and (b) are shown in (c) and (d). From the results one can see that the textures in the image (d) have been recovered quite well compared to the method of (Zeyde et al., 2010) in (c). The clear improvement is shown by the circle.

Figure 6-8 shows the results for image 'text'. This figure compares the result against the ground truth. To show the detail after super-resolution, the extraction of the selected region after super-resolution is shown in the right-hand side of the figure. The result of the high-resolution image obtained by the proposed method shown in (c) is sharper than the result of (b), which has many blurred effects around the text. Notice that the proposed method significantly improves the image resolution with acceptable blur effect. The proposed method is compared with the method of (Zeyde et al., 2010) quantitatively in terms of PSNR as shown in the figure. The comparison in terms of PSNR also indicates that the proposed method shows improvement of 3dB.



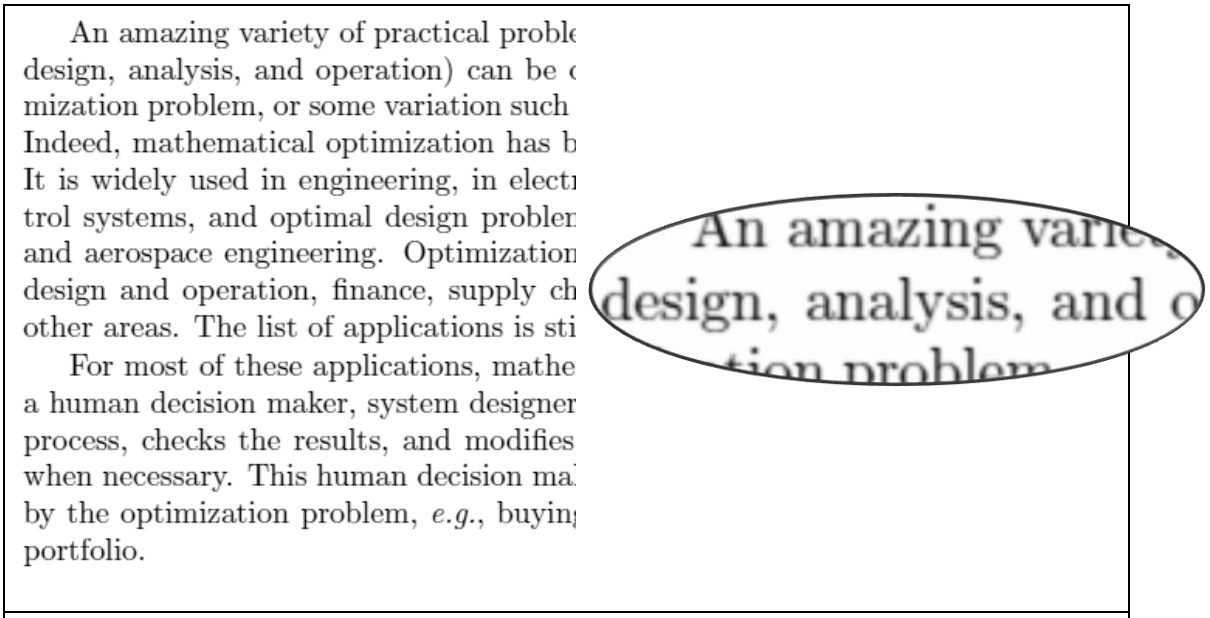
(a)Method of (Zeyde et al., 2010)(PSNR=23.96 dB)

(b) Proposed method (PSNR=25.81 dB)

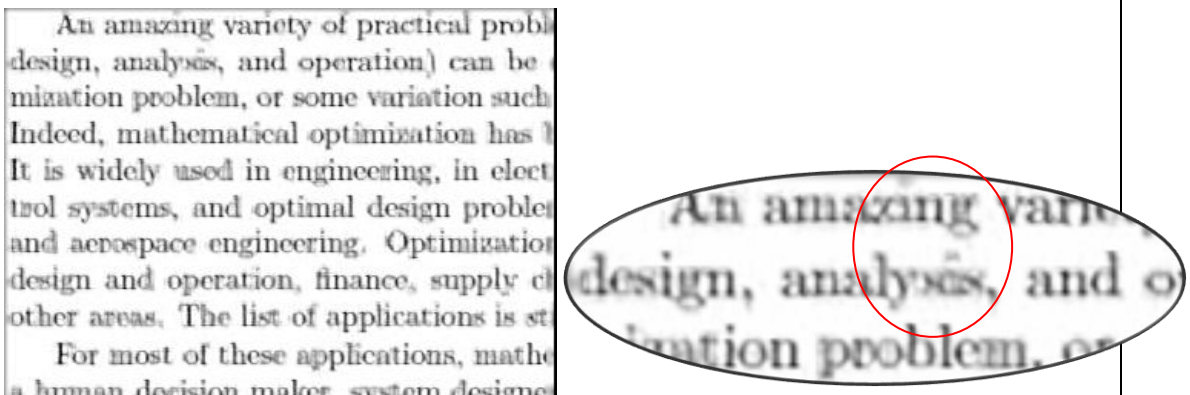
(c) Magnified version of (a)

(d) Magnified version of (b)

Figure 6-7: Super-resolution results for image 'barbara' and its PSNR (a) Method of (Zeyde et al., 2010) (b) The proposed method (c) Magnified version of (a) (d) Magnified version of (b).



(a)Original Image



(b) Method of (Zeyde et al., 2010) (PSNR=15.84 dB)

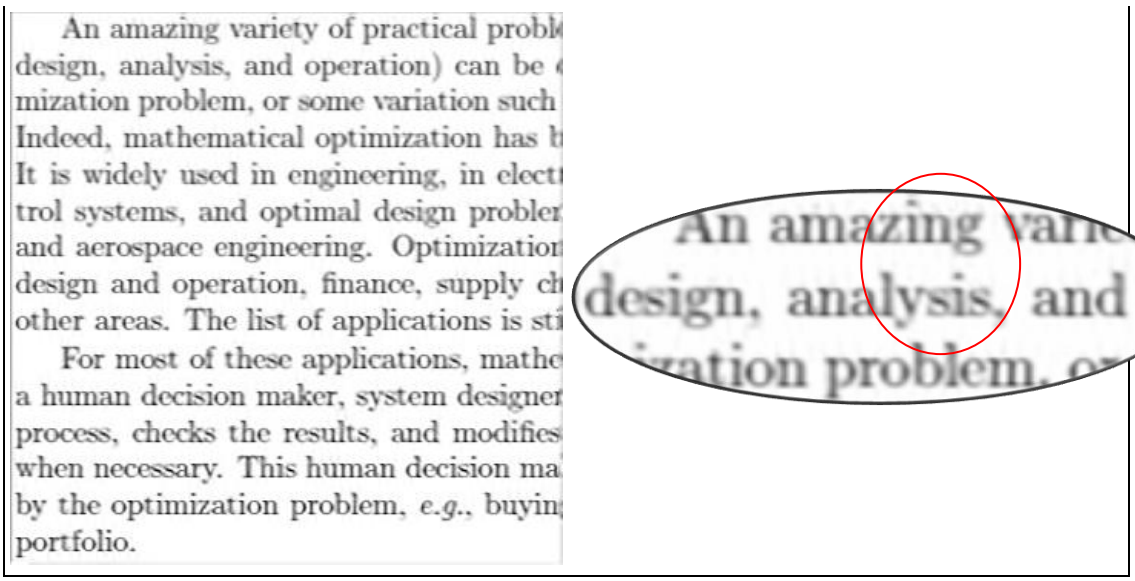


Figure 6-8: Super-resolution results for image ‘text’ and its PSNR (a) Original image (b) Method of (Zeyde et al., 2010) (c) Proposed method (PSNR=19.68 dB)

The proposed method is further tested on other images, i.e., ‘peppers’, ‘lena’ and ‘child’. The results are shown in Figure 6-9 and 6-10. Notice that in Figure 6-9 the output of the proposed method (right) have high-resolution details and have higher PSNR compared to the method of (Zeyde et al., 2010). Figure 6-9(b) shows the result of applying the proposed method to image ‘child’. The visual improvement can be seen in the image for the proposed method with PSNR improvement of 7dB.

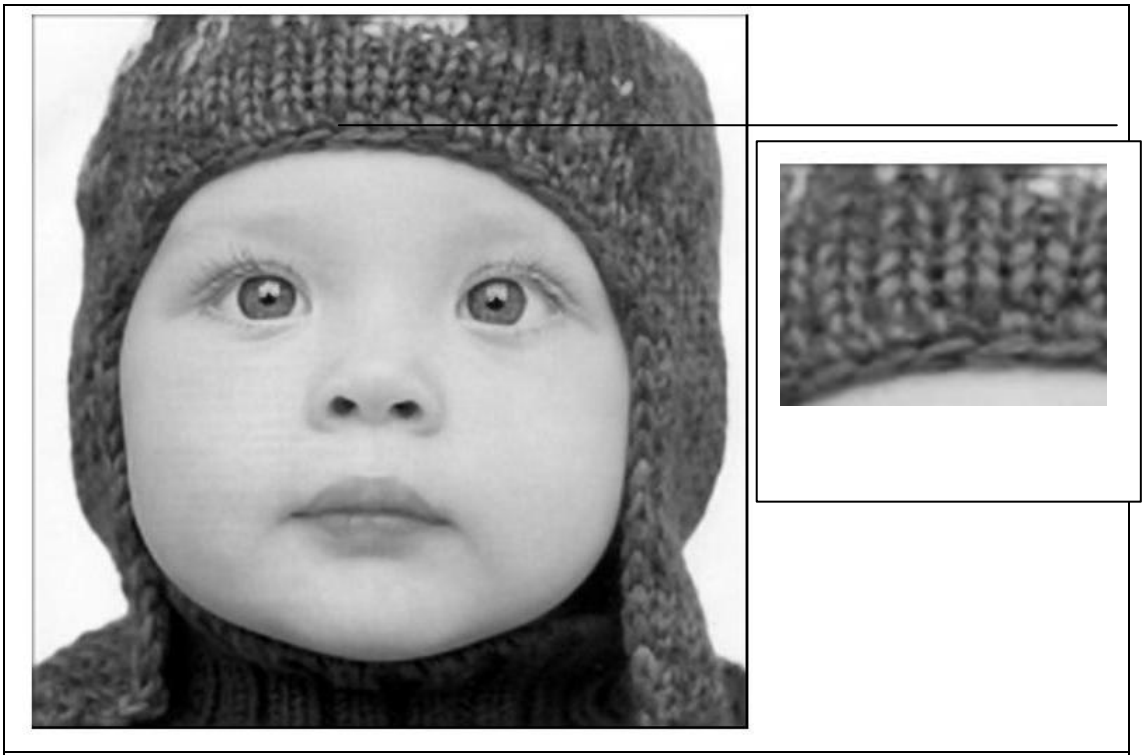
(a) PSNR=24.65dB

(b) PSNR=27.65

(a) PSNR=28.62

(b) PSNR=33.32

Figure 6-9: Super-resolution results and its PSNR: (top) ‘peppers’ (bottom) ‘lena’ (a) Method of (Zeyde et al., 2010) (b) The proposed method



(a) (PSNR=23.66)

(b) (PSNR=30.09)

Figure 6-10: Super-resolution results for ‘child’ and its PSNR: (a) Method of (Zeyde et al., 2010)(b) The proposed method

The PSNR value for the test images are shown in Table 6-3. The improved method has higher PSNR value for all of the test image compare to Method of (Zeyde et al., 2010).

Table 6-3: The PSNR (dB) for the test images.

Test Image (Size)		PSNR (dB)	
		Method of (Zeyde et al., 2010)	The proposed method
Lena	(170x170)	28.62	33.32
Text	(120x120)	15.84	19.68
Peppers	(85x85)	24.65	27.65
Barbara	(170x170)	23.96	25.81
Building	(266x266)	21.98	25.27
Child	(170x170)	23.66	30.09

6.7.2 Super-resolution results using bootstrapping

In this section, the result for the bootstrapping approach for single image super-resolution is presented. For this experiment, the patches and dictionaries are trained using only the low-resolution image. The low-resolution image is scaled down by 0.5 and upscale again for the training phase. Figures 6-11 and 6-12 show the results of applying the improved bootstrapping method to image ‘barbara’ and ‘text’, respectively. One can see that the results are visually comparable for both image as well as the PSNR. Note that, the result in Figure 6-7 and Figure 6.11 are not comparable because the size of the low-resolution images for each experiment is not the same. Furthermore the first method uses the ground truth high-resolution image for patches and dictionary learning while the bootstrapping method only uses the low-resolution image.

(a) PSNR=24.53 dB

(b) PSNR=24.65 dB

(c)

(d)

Figure 6-11: Super-resolution results for image ‘barbara’ and its PSNR (a) Method of (Zeyde et al., 2010) (b) The proposed method (c) Magnified version of (a) (d) Magnified version of (b)

(a)

(b) PSNR=16.97dB

(c) PSNR=17.1361

Figure 6-12: Super-resolution results for image ‘text’ and its PSNR (a) Original image (b) Method of (Zeyde et al., 2010)(c) The proposed method

Figure 6-13 and Figure 6-14 shows the results for ‘peppers’, ‘lena’, ‘child’ and ‘building’. Notice that in these figures the output of the proposed method (right) has better image quality and have higher PSNR compare to the method of (Zeyde et al., 2010).

(a) PSNR=24.65)

(b) PSNR=25.75

(a) PSNR=28.62

(b) PSNR=30.11

Figure 6-13: Super-resolution results and its PSNR: (top) ‘peppers’ (bottom) ‘lena’ (a) Method of (Zeyde et al., 2010) (b) The proposed method

(a) PSNR=32.59

(b) PSNR=32.96

(a) PSNR=27.81

(b) PSNR=26.98

Figure 6-14: Super-resolution results for image ‘building’ and its PSNR (a) Method of (Zeyde et al., 2010) (b) The proposed method

The PSNR value for the test images are shown in Table 6-4. The improved method has higher PSNR value for most of the test image except for image ‘building’.

Table 6-4: The PSNR (dB) for the test images.

Test Image (Size)		PSNR (dB)	
		Method of (Zeyde et al., 2010)	The proposed method
Lena	(170x170)	28.62	30.11
Text	(120x120)	16.97	17.13
Peppers	(85x85)	24.65	25.75
Barbara	(170x170)	24.53	24.65
Building	(266x266)	27.51	26.98
Child	(170x170)	32.59	32.96

Processing time: Table 6-5 shows the processing time of the improved bootstrapping method compared to the bootstrapping method of (Zeyde et al., 2010). All algorithms were run on a PC with the Pentium processor (2.50GHz) and 2GB memory. One can see from the table that the improved method effectively reduces the dimensions and requires less processing time by 10.36s in average over all six images.

Table 6-5: Complexity comparison (Elapsed Time) [sec].

Test Image (Size)		Times(seconds)	
		Method of (Zeyde et al., 2010)	The proposed method
Barbara	(256x256)	125.44	118.63
Text	(180x180)	93.96	92.37
Lena	(256x256)	119.97	116.25
Building	(400x400)	158.76	143.74
Peppers	(128x128)	97.90	88.41
Child	(256x256)	117.57	116.23
	Average	117.96	107.60

6.8 Summary

This chapter presents a single image super-resolution using sparse-representation and a novel image interpolation based on linear combination of the original patches with their first-order difference. The proposed image interpolation before patch learning step effectively enhances the prior estimation of the high-resolution image and method of (Zeyde et al., 2010). The proposed method enhanced the important features as well as the image patches and improved the super-resolution results. The proposed method to extract the low-resolution patches also improves the processing time. The proposed super-resolution scheme suggests performance improvements compared to methods of (Zeyde et al., 2010) for the test images.

CHAPTER 7

7. CONCLUSION

This chapter presents the summary of the contributions and future work. The remainder of the chapter is organised as follows. A summary of contribution is discussed in Section 7.1. The recommendation and future work is presented in Section 7.2.

7.1 Summary of contributions

The thesis has developed the techniques for combining information contained in multiple images of the same scene into a single high-resolution image. Different parts of the super-resolution problem separately been addressed in this thesis and SVD-based fusion before performing interpolation or single-image super-resolution is introduced. The advantages are that the solution to each phase can be tailored more easily and to avoid complex models. There are three main aspects of this thesis: multiple image registration, image fusion and single-image super-resolution.

For image registration, an image registration approach for super-resolution based on a combination of Scale Invariant Feature Transform (SIFT), Belief Propagation (BP) and Random Sampling Consensus (RANSAC) is proposed. The SIFT algorithm is used to detect and extract the local features in the low-resolution images. BP is used to match the features while RANSAC is adopted to filter out the mismatched points and then estimate the transformation matrix. The technique has been applied to synthetic and real-world images. The proposed method seems effective for the removal of the outliers introduced in the SIFT-BP method. The method produces better results for the test images in terms of visual quality compared to the case when SIFT, SIFT-BP or SIFT-RANSAC alone are used for image registration.

For image fusion, a novel SVD-based image fusion approach for super-resolution has been proposed. SVD-based image fusion enhances the super-resolution results by

integrating the significant features from low-resolution images. The proposed method converts the registered and reference image into the SVD domain and then the images' singular values are fused based on the fusion rule before performing the interpolation. In the proposed method, the maximum fusion rule is used because the highest singular values convey a great amount of image information representing the intensity information of a given image. Significant performance improvement can be seen from the results when compared to standard interpolation techniques and existing learning-based super-resolution approaches.

For single-image super-resolution, a novel image interpolation based on a linear combination of the original patches with their first-order derivates has been proposed to improve the method of single image super-resolution using sparse representation. The proposed method has enhanced the prior estimation of the high-resolution image while maintaining important image features. The improved method effectively enhances the prior estimation and reduces the computational time and hence improves the super-resolution results compared to state-of-the-art. The proposed method shows clear performance improvements compared to the methods of (Zeyde et al., 2010), using the test images.

7.2 Recommendation and future work

Although the proposed method was shown to performed well there is always opportunity for improvement. The following areas are of interest for potential further investigation:

- In this thesis, the different parts of the super-resolution algorithm are addressed separately and SVD-based fusion is introduced before performing interpolation or single-image super-resolution. The disadvantage of this approach is in terms of processing time, since each stage needs the input from the previous stage to process. In the future, the work can be extended by simultaneously perform fusion and super-resolution using sparse representations. This will make the algorithm more elegant and will reduce the processing time.

- The proposed method successfully reconstructs a high-resolution image from low-resolution images. The problem could be extended to reconstruct high-resolution video sequences rather than single images of the scene.
- Currently the algorithm for single image super-resolution presented in Chapter 6 only operates on grayscale images. In future work, the plan is to adapt the algorithm to operate with colour images. Basically, this can be achieved by applying the algorithm to the different colour planes. The extension of this should consider the correlation among the different colour channels as a prior knowledge.

Author Publications

Conference

1. Nasir, H., Stankovic, V., Marshall, S., Image registration for super resolution, *Proc. Eusipco-2010 18th European Signal Processing Conference*, Aalborg, Denmark, August 2010.
2. Nasir, H., Stankovic, V., Marshall, S., SVD-based fusion for super resolution image reconstruction, IEEE International Conference on Signal and Image Processing Applications pp. 393 - 398 Nov 2011

Journal

1. Nasir, H., Stankovic, V., Marshall, S., SVD-based fusion for super-resolution image reconstruction Elsevier Journal of Signal Processing: Image Communication vol 27 pp. 180-191 Feb 2012

Journal in preparation

2. Nasir, H., Stankovic, V., Marshall, S., Single Image super-resolution using sparse representation (to be submitted to IEEE Transaction on Image Processing)

REFERENCES

Acharya, T. & Tsai, P.-S. 2007. Computational foundations of image interpolation algorithms. *Ubiquity,* 2007**,** 1-17.

Adler, A., Hel-Or, Y. & Elad, M. 2010. A shrinkage learning approach for single image super-resolution with overcomplete representations. *Proceedings of the 11th European conference on Computer vision: Part II.* Heraklion, Crete, Greece: Springer-Verlag.

Aharon, M., Elad, M. & Bruckstein, A. 2006. K -SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *Signal Processing, IEEE Transactions on,* 54**,** 4311-4322.

Allebach, J. & Wong, P. W. 1996. Edge-directed interpolation. *In:* Image Processing, 1996. Proceedings., International Conference on, 16-19 Sep 1996 1996. 707-710 vol.3.

Amintoosi, M., Fathy, M. & Mozayani, N. 2009a. Image Registration for Super-Resolution using SIFT Key-points. *In proceeding of: 17th Iranian Conference on Electrical Engineering.*

Amintoosi, M., Fathy, M. & Mozayani, N. 2009b. Precise image registration with structural similarity error measurement applied to superresolution. *EURASIP J. Adv. Signal Process,* 2009**,** 1-7.

Baboulaz, L. & Dragotti, P. L. 2007. Local Feature Extraction for Image Super-Resolution. *In:* Image Processing, 2007. ICIP 2007. IEEE International Conference on, Sept. 16 2007-Oct. 19 2007 2007. V - 401-V - 404.

Baker, S. & Kanade, T. 1999. *Super-Resolution Optical Flow* [Online]. Available: citeulike-article-id:8982484 http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.5655[Accessed].

Besag, J. 1986. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society,* B-48**,** 259-302.

Bhushan, D. B., Sowmya, V. & Soman, K. P. 2010. Super Resolution Blind Reconstruction of Low Resolution Images Using Framelets Based Fusion. *In:* Recent Trends in Information, Telecommunication and Computing (ITC), 2010 International Conference on, 12-13 March 2010 2010. 100-104.

Borman, S. & Stevenson, R. L. 1998a. Spatial resolution enhancement of low-resolution image sequences. A comprehensive review with directions for future research. University of Notre Dame, Tech. Rep.,.

Borman, S. & Stevenson, R. L. 1998. Super-resolution from image sequences-a review. *In:* Midwest Symposium on Circuits and Systems, 9-12 Aug 1998 1998b. 374-378.

Bose, N. K. & Chappalli, M. B. 2004. A second-generation wavelet framework for super-resolution with noise filtering. *International Journal of Imaging Systems and Technology,* 14**,** 84-89.

Bose, N. K., Kim, H. C. & Valenzuela, H. M. 1993. Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes. *In:* Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on, 27-30 April 1993 1993. 269-272 vol.5.

Brown Jr, J. 1981. Multi-channel sampling of low-pass signals. *Circuits and Systems, IEEE Transactions on,* 28**,** 101-106.

Brown, M. & Lowe, D. G. 2007. Automatic Panoramic Image Stitching using Invariant Features. *Int. J. Comput. Vision,* 74**,** 59-73.

Candés, E., Romberg, J. & Tao, T. 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theory,* 52**,** 489–509.

Candocia, F. M. & Principe, J. C. 1999. Super-resolution of images based on local correlations. *Neural Networks, IEEE Transactions on,* 10**,** 372-380.

Cernekova, Z., Kotropoulos, C., Nikolaidis, N. & Pitas, I. 2005. Video shot segmentation using fusion of SVD and mutual information features. *In:* Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on, 23-26 May 2005 2005. 3849-3852 Vol. 4.

Cha, Y. & Kim, S. 2007. PDE-Based Interpolation Methods for Image Super Resolution. *In:* Future Generation Communication and Networking (FGCN 2007), 6-8 Dec. 2007 2007. 214-219.

Chang, H., Yeung, D.-Y. & Xiong, Y. 2004. Super-resolution through neighbor embedding. *In:* Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 27 June-2 July 2004 2004. I-275-I-282 Vol.1.

Chang, P. C. & Wu, T. L. 2007. Region Weighted Satellite Super-resolution Technology. *Proc. of National Symposium on Telecommunications (NST).* Taipei, Taiwan.

Chappalli, M. B. & Bose, N. K. 2005. Simultaneous noise filtering and super-resolution with second-generation wavelets. *Signal Processing Letters, IEEE,* 12**,** 772-775.

Cheng, S., Stankovic, V. & Stankovic, L. 2009. Improved sift-based image registration using belief propagation. *In:* IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.

Datsenko, D. & Elad, M. 2007. Example-based single document image super-resolution: a global MAP approach with outlier rejection. *Multidimensional Syst. Signal Process.,* 18**,** 103-121.

Demirel, H. & Anbarjafari, G. 2011. Image Resolution Enhancement by Using Discrete and Stationary Wavelet Decomposition. *Image Processing, IEEE Transactions on,* 20**,** 1458-1460.

Demirel, H., Ozcinar, C. & Anbarjafari, G. 2010 Satellite Image Contrast Enhancement Using Discrete Wavelet Transform and Singular Value Decomposition. *Geoscience and Remote Sensing Letters, IEEE,* 7**,** 333-337.

El-Khamy, S. E., Hadhoud, M. M., Dessouky, M. I., Salam, B. M. & El-Samie, F. E. A. 2005a. A new super-resolution image reconstruction algorithm based on wavelet fusion. *In:* Radio Science Conference, 2005. NRSC 2005. Proceedings of the Twenty-Second National, 15-17 March 2005. 195-204.

El-Khamy, S. E., Hadhoud, M. M., Dessouky, M. I., Salam, B. M. & El-Samie, F. E. A. 2005b. Regularized super-resolution reconstruction of images using wavelet fusion. *Optical Engineering,* 44**,** 097001-10.

Elad, M. 2010. Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer.

Elad, M. & Aharon, M. 2006. Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries. *Image Processing, IEEE Transactions on,* 15**,** 3736-3745.

Elad, M. & Feuer, A. 1994. On restoration and super-resolution for continuous image sequence-adaptive filtering approach. *Int. Rep. 942,*.

Engan, K., Aase, S. O. & Hakon Husoy, J. 1999. Method of optimal directions for frame design. *In:* Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on,. 2443-2446 vol.5.

Fan, C., Zhu, J., Gong, J. & Kuang, C. 2006. POCS Super-Resolution Sequence Image Reconstruction Based on Improvement Approach of Keren Registration Method. *In:* Intelligent Systems Design and Applications, 2006. ISDA '06. Sixth International Conference on, 16-18 Oct. 2006. 333-337.

Farsiu, S., Robinson, M. D., Elad, M. & Milanfar, P. 2004. Fast and robust multiframe super resolution. *Image Processing, IEEE Transactions on,* 13**,** 1327-1344.

Fischler, M. A. & Bolles, R. C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM,* 24**,** 381-395.

Freeman, W. T., Jones, T. R. & Pasztor, E. C. 2002. Example-based super-resolution. *Computer Graphics and Applications, IEEE,* 22**,** 56-65.

Freeman, W. T., Pasztor, E. C. & Carmichael, O. T. 2000. Learning Low-Level Vision. *Int. J. Comput. Vision,* 40**,** 25-47.

Frey, B. & Mackay, D. 1998. A revolution: Belief propagation in graphs with cycles. *In:* In Neural Information Processing Systems, 1998. 479-485.

Gehani, A., Reif, J., Craiger, P. & Shenoi, S. 2007. Super-Resolution Video Analysis for Forensic Investigations Advances in Digital Forensics III. Springer Boston.

Glasner, D., Bagon, S. & Irani, M. 2009. Superresolution from a single image. *ICCV.*

Golub, G. H. & Loan, C. F. V. 1996. *Matrix Computation*, John Hopkins University Press.

Gunturk, B. K., Altunbasak, Y. & Mersereau, R. M. 2004. Super-resolution reconstruction of compressed video using transform-domain statistics. *Image Processing, IEEE Transactions on,* 13**,** 33-43.

Hao, S., Lin, L., Weiping, Z. & Limin, L. 2009. Location and Super-Resolution Enhancement of License Plates Based on Video Sequences. *In:* Information Science and Engineering (ICISE), 2009 1st International Conference on, 26-28 Dec. 2009. 1319-1322.

Hardie, R. 2007. A Fast Image Super-Resolution Algorithm Using an Adaptive Wiener Filter. *IEEE Transactions on Image Processing,* 16**,** 2953-2964.

He, Y., Yap, K.-H., Chen, L. & Chau, L.-P. 2007. A Nonlinear Least Square Technique for Simultaneous Image Registration and Super-Resolution. *Image Processing, IEEE Transactions on,* 16**,** 2830-2841.

Hill, P., Canagarajah, N. & Bull, D. 2002. Image fusion using complex wavelets. *13th British Machine Vis. Conf. BMVC-2002.*

Horn, R. A. & Johnson, C. R. 1987. *Matrix Analysis*, Cambridge University Press.

Huang, K. & Aviyente, S. 2007. Sparse Representation for Signal Classification. *NIPS*, 19.

Huang, T. S. & Tsay, R. Y. 1984. Multiple frame image restoration and registration. *Advances in Computer Vision and Image Processing,* 1**,** 317-339.

Irani, M. & Peleg, S. 1991. Improving resolution by image registration. *CVGIP Graphical Models and Image Processing***,** 231–239.

Ji, H. & Fermüller, C. 2006. Wavelet-based super-resolution reconstruction: theory and algorithm. *Proceedings of the 9th European conference on Computer Vision - Volume Part IV*. Graz, Austria: Springer-Verlag.

Jia, H. & Zhang, X. 2009. Facial feature extraction and recognition based on Curvelet transform and SVD. *In:* Apperceiving Computing and Intelligence Analysis, 2009. ICACIA 2009. International Conference on, 23-25 Oct. 2009. 104-107.

Jing, G., Shi, Y. & Lu, B. 2010. Single-Image Super-Resolution Based on Decomposition and Sparse Representation. *2010 International Conference on Multimedia Communications.*

Keren, D., Peleg, S. & Brada, R. 1998. Image Sequence Enhancement Using Sub-pixel Displacements. *In:* Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1988. 742-746.

Kim, K. I. & Kwon, Y. 2008. Example-Based Learning for Single-Image Super-Resolution. *Proceedings of the 30th DAGM symposium on Pattern Recognition.* Munich, Germany: Springer-Verlag.

Kim, S. P., Bose, N. K. & Valenzuela, H. M. 1990. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *Acoustics, Speech and Signal Processing, IEEE Transactions on,* 38**,** 1013-1027.

Kindermann, R. 1980. *Markov Random Fields and Their Applications (Contemporary Mathematics ; V. 1),* Amer Mathematical Society.

Kschischang, F., Frey, B. & Loeliger, H.-A. 2001. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory,* 47**,** 498-519.

Lanir, J., Maltz, M., Yatskaer, I. & Rotman, S. R. 2006. Comparing Multispectral Image Fusion Methods for a Target Detection Task. *In:* Information Fusion, 2006 9th International Conference on, 10-13 July 2006. 1-5.

Lee, W.-L., Yang, C.-C., Wu, H.-T. & Chen, M.-J. 2009. Wavelet-based Interpolation Scheme for Resolution Enhancement of Medical Images. *Journal of Signal Processing Systems,* 55**,** 251-265.

Lertrattanapanich, S. & Bose, N. K. 2002. High resolution image formation from low resolution frames using Delaunay triangulation. *Image Processing, IEEE Transactions on,* 11**,** 1427-1441.

Lewis, J. J., Callaghan, R. J., Nikolov, S. G., Bull, D. R. & Canagarajah, N. 2007. Pixel- and region-based image fusion with complex wavelets. *Information Fusion,* 8**,** 119-130.

Li, F., Fraser, D. & Jia, X. 2007. Wavelet Domain Deblurring and Denoising for Image Resolution Improvement. *In:* Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society on, 3-5 Dec. 2007. 373-379.

Li, F., Jia, X. & Fraser, D. 2008. Universal HMT based super resolution for remote sensing images. *In:* Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, 12-15 Oct. 2008. 333-336.

Li, H., Manjunath, B. S. & Mitra, S. K. 1994. Multi-sensor image fusion using the wavelet transform. *In:* Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference, 13-16 Nov 1994. 51-55 vol.1.

Li, X. & Orchard, M. T. 2001. New edge-directed interpolation. *Image Processing, IEEE Transactions on,* 10**,** 1521-1527.

Lin, F. C., Fookes, C. B., Chandran, V. & Sridharan, S. 2005. Investigation into optical flow super-resolution for surveillance applications. *APRS Workshop on Digital Image Computing: Pattern Recognition and Imaging for Medical Applications.* Brisbane.

Liyakathunisa. 2009. Super Resolution Reconstruction of Compressed Low Resolution Images Using Wavelet Lifting Schemes. *In:* KUMAR, C. N. R. & ANANTHASHAYANA, V. K., eds., 2009. 629-633.

Liyakathunisa & Ananthashayana, V. K. 2009. Super resolution blind reconstruction of low resolution images using wavelets based fusion. *Proceedings of World Academy of Science, Engineering and Technology,* 40**,** 177-181.

Loeliger, H. A. 2004. An introduction to factor graphs. *Signal Processing Magazine, IEEE,* 21**,** 28-41.

Lowe, D. G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision,* 60**,** 91-110.

Lucas, B. D. & Kanade, T. 1981. An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2.* Vancouver, BC, Canada: Morgan Kaufmann Publishers Inc.

Madhusudhan, T. & Pais, A. R. 2007. Generation of Super-Resolution Video from Low Resolution Video Sequences: A Novel Approach. *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007) - Volume 03.* IEEE Computer Society.

Mairal, J., Elad, M. & Sapiro, G. 2008. Sparse Representation for Color Image Restoration. *Image Processing, IEEE Transactions on,* 17**,** 53-69.

Martins, A. L. D., Homem, M. R. P. & Mascarenhas, N. D. A. 2007. Super-Resolution Image Reconstruction using the ICM Algorithm. *In:* Image Processing, 2007. ICIP 2007. IEEE International Conference on, Sept. 16 2007-Oct. 19 2007. IV - 205-IV - 208.

Milanfar, P. *MDSP Super Resolution and Demosaicing Datasets* [Online]. Available: http://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html [Accessed].

Miravet, C. A. & Rodriguez, F. B. A. 2005. Accurate and robust image superresolution by neural processing of local image representations *superresolution, neural networks, image sequence processing* 3696**,** 499-506.

Mooij, J. & H. Kappen 2007. Sufficient conditions for convergence of the sum-product algorithm. *Information Theory, IEEE Transactions on,* 53**,** 4422–4437.

Moorthy, A. K. & Bovik, A. C. 2010 *BIQI Software Release,* [Online]. Available: http://live.ece.utexas.edu/research/quality/biqi.zip [Accessed].

Moorthy, A. K. & Bovik, A. C. 2010b. A Two-Step Framework for Constructing Blind Image Quality Indices. *Signal Processing Letters, IEEE,* 17**,** 513-516.

Nguyen, N. & Milanfar, P. 1999. A Wavelet-Based Interpolation-Restoration Method for Superresolution. *Circuits, Systems, and Signal Processing.*

Ni, K. S. & Nguyen, T. Q. 2007. Image Superresolution Using Support Vector Regression. *Image Processing, IEEE Transactions on,* 16**,** 1596-1610.

P. Hill, Canagarajah, N. & Bull, D. 2002. Image fusion using complex wavelets. *13th British Machine Vis. Conf. BMVC-2002.*

Pang, Y., Yu, N., Zhang, R., Rong, J. & Liu, Z. 2004. Fusion of SVD and LDA for face recognition. *In:* Image Processing, 2004. ICIP '04. 2004 International Conference on, 24-27 Oct. 2004. 1417-1420 Vol.2.

Papoulis, A. 1977. Generalized sampling theorem. *IEEE Trans. Circuits Syst.,* 24**,** 652–654.

Park, S., Park, M. & Kang, M. 2003. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine,* 20**,** 21-36.

Patil, V. H., Bormane, D. S. & Pawar, V. S. 2008. Super Resolution Using Neural Network. *In:* Modeling & Simulation, 2008. AICMS 08. Second Asia International Conference on, 13-15 May 2008. 492-496.

Patti, A. J., Sezan, M. I. & Tekalp, A. M. 1997. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *Image Processing, IEEE Transactions on,* 6**,** 1064-1076.

Philbin, J. & Zisserman, A. *Oxford Database* [Online]. Available: http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/index.html [Accessed].

R.S.Blum & Z.Liu 2005. *Multi-sensor image fusion and its applications,* CRC press.

Ranade, A., Mahabalarao, S. S. & Kale, S. 2007. A variation on SVD based image compression. *Image and Vision Computing,* 25**,** 771-777.

Ravishankar, S., Reddy.C, N. & M.V.Joshi 2011. Single Image Super Resolution Using Sparse Image and GLCM Statistics as Priors. *Directory of Open Access Journals (DOAJ).*

Repperger, D. W., Pinkus, A. R., Farris, K. A., Roberts, R. G. & Sorkin, R. D. 2009. Investigation of image fusion procedures using optimal registration and SVD algorithms. *In:* Aerospace & Electronics Conference (NAECON), Proceedings of the IEEE 2009 National, 21-23 July 2009. 231-235.

Robinson, D., Farsiu, S. & Milanfar, P. 2009. Optimal Registration Of Aliased Images Using Variable Projection With Applications To Super-Resolution. *Comput. J.,* 52**,** 31-42.

Robinson, M. D., Chiu, S. J., Toth, C. A., Izatt, J., Lo, J. Y. & Farsiu, S. 2010. Novel applications of super-resolution in medical imaging. *In:* MILANFAR, P. (ed.) *Super-Resolution Imaging.* CRC Press.

Rockinger, O. & Fechner, T. 1998. Pixel-level image fusion: the case of image sequences. *SPIE Proceedings***,** 378–388.

S. Mallat. Super-resolution bandlet upconversion for HDTV. Available: http://www.cmap.polytechnique.fr/~mallat/papiers/whitepaper.pdf.

Schultz, R. R. & Stevenson, R. L. 1996. Extraction of high-resolution frames from video sequences. *Image Processing, IEEE Transactions on,* 5**,** 996-1011.

Seong, Y.-M. & Park, H. 2008. Superresolution technique for planar objects based on an isoplane transformation. *Opt. Eng,* 47.

Shannon, C. E. 1949. Communication in the presence of noise. *Proc. Institute of Radio Engineers,* vol. 37**,**, 10–21.

Shen, L. & Sun, Q. 2004. Biorthogonal wavelet system for high-resolution image reconstruction. *Signal Processing, IEEE Transactions on,* 52**,** 1997-2011.

Srinivas, C. & Srinath, M. D. 1990. A stochastic model-based approach for simultaneous restoration of multiple miss-registered images. *SPIE,* 1360**,** 1416-1427.

Srivastava, A., Lee, A. B., Simoncelli, E. P. & Zhu, S.-C. 2003. On Advances in Statistical Modeling of Natural Images. *Journal of Mathematical Imaging and Vision*.

Stark, H. & Oskoui, P. 1989. High-resolution image recovery from image-plane arrays, using convex projections. *J. Opt. Soc. Am. A,* 6**,** 1715.

Sun, J., Xu, Z. & Shum, H.-Y. 2011. Gradient Profile Prior and Its Applications in Image Super-Resolution and Enhancement. *Image Processing, IEEE Transactions on,* 20**,** 1529-1542.

Sun, J., Zheng, N.-N., Tao, H. & Shum, H.-Y. 2003. Image hallucination with primal sketch priors. *In:* Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, 18-20 June 2003. II-729-36 vol.2.

Susu, Y., Weisi, L., Eeping, O. & Zhongkang, L. 2005. Contrast signal-to-noise ratio for image quality assessment. *In:* Image Processing, 2005. ICIP 2005. IEEE International Conference on, 11-14 Sept. 2005. I-397-400.

Takeda, H., Farsiu, S. & Milanfar, P. 2007. Kernel Regression for Image Processing and Reconstruction. *Image Processing, IEEE Transactions on,* 16**,** 349-366.

Tam, W.-S., Kok, C.-W. & Siu, W.-C. 2010. Modified edge-directed interpolation for images. *J. Electron. Imaging 19*.

Tang, C., Dong, Y. & Su, X. 2008. Automatic Registration Based on Improved SIFT for Medical Microscopic Sequence Images. *Proceedings of the 2008 Second International Symposium on Intelligent Information Technology Application - Volume 01*. IEEE Computer Society.

Tang, J. 2004. A contrast based image fusion technique in the DCT domain. *Digital Signal Processing,* 14**,** 218-226.

Tekalp, A. M., Ozkan, M. K. & Sezan, M. I. 1992. High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. *In:* Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, 23-26 Mar 1992. 169-172 vol.3.

Tian, J. & Ma, K.-K. 2011. A survey on super-resolution imaging. *Signal, Image and Video Processing,* 5**,** 329-342.

Ur, H. & Gross, D. 1992. Improved resolution from subpixel shifted pictures. *CVGIP: Graph. Models Image Process.,* 54**,** 181-186.

Vandewalle, P., Sbaiz, L., Susstrunk, S. & Vetterli, M. 2006. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP J. Appl. Signal Process.,* 2006**,** 233-233.

Vandewalle, P., Süsstrunk, S. & Vetterli, M. Super-resolution Software.

Villanueva, L. G., Callico, X, M., G., Tobajas, F., Lo, Pez, S., De Armas, V., Pez, J. F. & Sarmiento, R. 2010. Medical Diagnosis Improvement Through Image Quality Enhancement Based on Super-Resolution. *In:* Digital System

Design: Architectures, Methods and Tools (DSD), 2010 13th Euromicro Conference on, 1-3 Sept. 2010. 259-262.

Vrigkas, M., Nikou, C. & Kondi, L. P. 2011. On the improvement of image registration for high accuracy super-resolution. *In:* Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, 22-27 May 2011. 981-984.

W. H. Press, Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 2002. *Numerical Recipes in C++: The Art of Scientific Computing, 2nd ed*, Cambridge University Press, New York.

Wang, J., Liang, J., Hu, H., Li, Y. & Feng, B. 2007. Performance Evaluation of Infrared and Visible Image Fusion Algorithms for Face Recognition. *Proceedings of the International Conference on Intelligent Systems and Knowledge Engineering.*

Wang, J., Zhu, S. & Gong, Y. 2011. Resolution enhancement based on learning the sparse association of image patches. *Pattern Recognition Letters,* 31**,** 1-10.

Wittman, T. 2005. Mathematical Techniques for Image Interpolation *Oral Exam Paper.* Dept. of mathematics, Univ. of Minnesota, 2005.

Yang, J., J.Wright, S.Huang, T. & Ma, Y. 2008. Image super-resolution as sparse representation of raw image patches. *IEEE Conference on Computer Vision and Pattern Recognition (2008).*

Yang, J., Wright, J., Huang, T. S. & Ma, Y. 2010. Image Super-Resolution Via Sparse Representation. *Image Processing, IEEE Transactions on,* 19**,** 2861-2873.

Yao, S., Lin, W., Ong, E. & Lu, Z. 2005. Contrast signal-to-noise ratio for image quality assessment. Image Processing. *ICIP 2005. IEEE International Conference on.*

Yedidia, J. S., Freeman, W. T. & Weiss, Y. 2003. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium.* Morgan Kaufmann Publishers Inc.

Yin, H., Li, S. & Fang, L. 2012. Simultaneous image fusion and super-resolution using sparse representation. *Information Fusion*.

Yu, W., Yao, H., Liu, X., Ji, R., Sun, X. & Xu, P. 2011. Contextual dictionaries for image super resolution. *ACM International Conference Proceeding Series.*

Yuan, Z., Yan, P. & Li, S. 2008. Super resolution based on scale invariant feature transform. *In:* Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on, 7-9 July 2008. 1550-1554.

Zeyde, R., Michael, E. & Matan, P. 2010. On single image scale-up using sparse-representations. *Proceedings of the 7th international conference on Curves and Surfaces.* Avignon, France: Springer-Verlag.

Zhang, H., Zhang, Y. & Huang, T. S. 2011. Efficient sparse representation based image super resolution via dual dictionary learning. *In:* Multimedia and Expo (ICME), 2011 IEEE International Conference on, 11-15 July 2011. 1-6.

Zhang, Z. & Blum, R. S. 1999. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of the IEEE,* 87**,** 1315-1326.

Zhu, W., Zhou, Y., Chen, J., Sun, B. & Hou, G. 2008. Objective Evaluation of Remote Sensing Image Fusion based on the Singular Value Decomposition. *In:* Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, 7-11 July 2008. III - 447-III - 450.

Zhu, Z., Luo, J. & Shen, Z. 2010. Automatic remote sensing image registration based on SIFT descriptor and image classification. *In:* Geoinformatics, 2010 18th International Conference on, 18-20 June 2010. 1-5.

Zitova, B. & Flusser, J. 2003. Image registration methods: a survey. *Image and Vision Computing,* 21**,** 977-1000.

Zomet, A., Rav-Acha, A. & Peleg, S. 2001. Robust super-resolution. *In:* Computer Vision and Pattern Recognition, 2001. CVPR 2001.