SUPER-RESOLUTION TECHNIQUES FOR HIGH-RESOLUTION IMAGE RECONSTRUCTION

TIAN YUSHUANG

2014

NANYANG TECHNOLOGICAL UNIVERSITY

SUPER-RESOLUTION TECHNIQUES FOR HIGH-RESOLUTION IMAGE RECONSTRUCTION

TIAN YUSHUANG

SCHOOL OF ELECTRICAL & ELECTRONIC ENGINEERING

2014

SUPER-RESOLUTION TECHNIQUES FOR HIGH-RESOLUTION IMAGE RECONSTRUCTION

TIAN YUSHUANG

TIAN YUSHUANG

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2014

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

_____ _____

DATE TIAN YUSHUANG

To my family, for all your love.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor, Associate Professor Kim-Hui Yap, for his patient guidance and invaluable support to help me undertake the research in this thesis. He has always provided helpful instructions and sincere encouragement throughout my four-year doctoral training. His technical knowledge, insightful scientific vision and rigorous scholarship have benefited me greatly.

I would like to thank Associate Professor Kap-Luk Chan, my co-supervisor, for his kind guidance and valuable support. My sincere thank also goes to Dr. Yu He, my senior and friend, who has given me supportive help in my research work.

Further, I would like to thank my parents for teaching me the value of hard work and supporting me throughout my life. Their endless love is always the greatest inspiration for me. Special thanks to my girlfriend, Miss Nini Liu, a caring and supportive partner always.

Finally, I would like to thank Nanyang Technological University for awarding me the research scholarship. I would like to acknowledge the supports and assistance of my colleagues and the staff in the Media Technology Laboratory (MTL). I learned a lot from our discussions and will cherish the friendship forever.

I dedicate this thesis to all of you.

Table of Contents

Abstract

This thesis investigates a number of techniques and algorithms for super resolution (SR) image reconstruction. In digital imaging applications, high resolution (HR) images with more details are often desired for human interpretation or machine learning of the image contents. In many commercial digital cameras, however, the quality of the captured images may be limited by many real-life constraints, including aliasing effects, lens or motion blurring, color filter and shot noise. Compared with upgrading hardware, performing software post-processing on the captured degraded low resolution (LR) images is an economical and effective alternative solution to obtain a HR image. SR is such an image processing technique which can reconstruct a HR image from one or a sequence of LR counterparts.

In the first part, several algorithms are proposed to handle multi-frame HR image reconstruction from a sequence of LR images which contain an overlapping content of the same scene. The basic principle of multi-frame SR is to combine the unique information exploited from each of the LR images to obtain a HR image. Therefore, the two main issues for successful multi-frame SR reconstruction are exploring the motion among the LR images followed by fusing the LR images based on the estimated motion parameters. Outliers occurred in the LR images may decrease the accuracy of image registration and subsequently introduce artifacts to the reconstructed HR image. In Chapter 3, we propose a L_1-norm approach to handle SR reconstruction from multiple LR images that may contain outliers. The adopted L_1-norm technique is robust to the registration error and outliers existing in the captured LR images. Further, image registration and SR are combined into a single process and they are performed

simultaneously. An iterative scheme ensures the improved estimates of HR image and motion parameters are obtained progressively. Another fundamental problem for multi-frame SR is the selection of a specific motion model to describe the relationship of the LR images. Most of the existing SR approaches assume the relative motion between LR images to be translational or sometimes rotational. This in-plane motion model may not be applicable in such SR applications when relative zooming motion exists among the captured LR images. In Chapter 4, we develop a joint image registration and SR framework for LR images with zooming motion. A general motion model consisting of translation, rotation and zooming is adopted to handle more real-life applications. The proposed method performs image registration and zooming SR simultaneously, instead of regarding them as two disjoint steps. Further, the proposed method develops an adaptive weighting scheme in the HR reconstruction.

The second part of this thesis focuses on single frame SR problems. Different from multi-frame SR, less information about the scene is provided for single frame SR. Therefore, strong prior information of the target image is required in order to obtain a stable solution. In Chapter 5, we propose a reconstruction-based single frame vehicle license plate SR approach using learning patches. The example patches are selected from a specific training database which consists of example images sharing similar characteristic with the target image. A Maximum *a posteriori* (MAP) framework is adopted and the example patches are incorporated into the cost function as prior information to regularize the final solution. Different from the conventional hard-binary learning based methods which use a threshold to select the patch or not, the proposed method employs a soft weighting strategy based on the character prior information of license plate. The underlying idea is that patches from the categories with higher matching scores will be given higher weightings, and verse vice.

In the third part, we deal with color SR reconstruction problems. It is a challenging task to reconstruct a HR color image from a sequence of LR mosaiced color images. Conventional approaches usually can be considered as a two-stage process of color demosaicing followed by color image SR. However, these approaches tend to produce suboptimal solutions as the color artifacts arising from the color demosaicing process may be propagated to the subsequent HR reconstruction. In Chapter 6, we propose a joint framework to combine color demosaicing and color SR into a single process. To suppress the color artifacts, a color correlation term is adopted to ensure consistency amongst all the color channels. An edge-directed total variation (TV) regularization technique is developed to preserve the high-frequency details of the desired HR image, especially in the regions consisting of edges. The edge information is explored based on the reconstructed HR image in the previous iteration. An iterative scheme ensures the improved HR color image is obtained.

Key Abbreviations

1-D	One-dimensional
2-D	Two-dimensional
AM	Alternating minimization
AWGN	Additive white Gaussian noise
BTV	Bilateral total variation
CCD	Charged-coupled device
CFA	Color filter array
CFT	Continues Fourier transform
CG	Conjugate gradient
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
EM	Expectation maximization
FGW	First-generation wavelet
FP	Fixed point
GCV	Generalized cross-validation
GMRF	Gaussian Markov random field
HR	High-resolution
IBP	Iterative backprojection
IDFT	Inverse discrete Fourier transform
K-NN	*K*-nearest-neighbor
LIPSOL	Linear interior point solver
LLE	Locally linear embedding
LMMSE	Linear minimum mean square-error estimation
LP	Linear programming
LPC	Locality preserving constraints
LR	Low-resolution
MAP	Maximum *a posteriori*

MASD	Maximum average square difference
ML	Maximum likelihood
MMSE	Minimum mean squared error
MRF	Multichannel recursive filtering
NLS	Nonlinear least squares
NMES	Normalized mean squared error
NN	Neural network
NPE	Nonlinear-parametric-estimation
PCA	Principle component analysis
PCG	Preconditioned conjugate gradient
PDE	Partial differential equation
PDR	Parametric double regularization
POCS	Projection-onto-convex-sets
PSF	Point-spread function
PSNR	Peak signal-to-noise ratio
PVE	Partial volume effect
RGB	Red, Green and Blue
ROI	Region of interest
RS	Remote sensing
SAR	Spatial auto-regression
SC	Single channel
SGW	Second-generation wavelet
SIMO	Single-input multiple-output
SISO	Single-input single-output
SNR	Signal-to-noise ratio
SR	Super-resolution
SVD	Singular value decomposition
SVM	Support vector machine
TV	Total variation
VP	Variable projection

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Nowadays, digital imaging devices (*e.g.* digital camera) are becoming popular for image acquisition, including both consumer electronics and industrial applications. Different from their analog counterparts, the digital camera utilizes the charge-coupled device (CCD) instead of chemical films to capture the information of scene that is passed through the optical system. The basic measure of a digital image is image resolution, including pixel resolution and spatial resolution. Pixel resolution refers to the pixel count in digital images and spatial resolution is defined as the number of *independent* pixel values per unit area. Two images with different pixel resolutions are shown in Figure 1.1. From the figure, it is noted that the right image has a higher pixel resolution than the left one, but is of worse spatial resolution. In most cases, spatial resolution plays a more important role than pixel resolution for human interpretation of the image content. Therefore, our interests mainly focus on the spatial resolution of images. For simplicity, the word "resolution" refers to spatial resolution in the following thesis.

(a) (b)

Figure 1.1 Illustrations of pixel resolution and spatial resolution. (a) A test image with higher spatial resolution but lower pixel resolution, (b) A test image with higher pixel resolution but lower spatial resolution.

Among the current digital imaging applications, high resolution (HR) images with more details are desired, but may not be available due to various practical limitations. Blurring, frequency aliasing and noise are common degradations found in digital imaging systems. For instance, consider a video traffic surveillance system which tracks moving vehicles through a digital camera as illustrated in Figure 1.2 (a). The captured video frame may be degraded by optical blurring since the imaging scenes may be out-of-focus as shown in Figure 1.2(b), or by motion blurring due to rapid vehicle motion as shown in Figure 1.2(c), respectively. During the spatial sampling of the continuous scene, the surveillance system may potentially produce a downsampled image in Figure 1.2(d) due to the limited resolution of the optics and transducer. Further, under the poor lighting conditions, the imaging sensor array (e.g. charge-coupled device (CCD)) will be subject to various noises including thermal and shot noise. An example of the

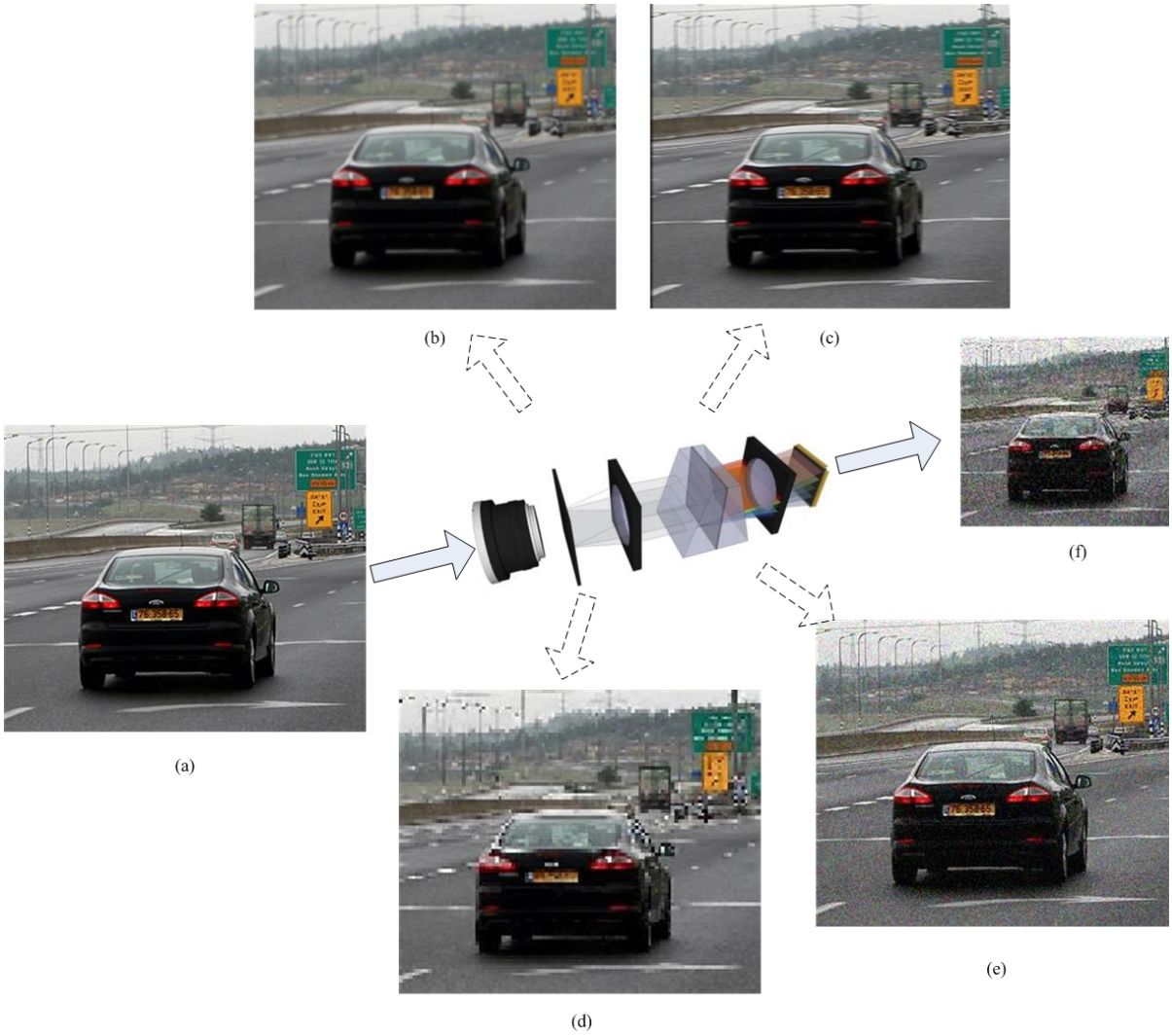


Figure 1.2 Effects of various imaging degradations. (a) Original image, (b) Out-of-focus blurred image, (c) Motions blurred image, (d) Downsampled image, (e) Noisy Image, (f) Final captured LR image.

noisy image is shown in Figure 1.2(e). The quality of the final captured images through the imaging devices could be degraded severely as shown in Figure 1.2(f).

There are mainly two approaches to alleviate the aforementioned degradations: (i) hardware-based methods and (ii) software-based methods. The most intuitive solution is to increase the spatial resolution by employing a CCD chip with more pixels. However, the denser

CCD array (with smaller pixels) can only improve the image quality to a certain degree due to the practical constraints. The presence of shot noise, which is unavoidable in any digital imaging system, is associated with a lower limit on the pixel size of CCD [92],[132]. The limit of pixel size arises from the fact that while reducing the sensor pixel size increases the resolution of the acquired image, the amount of light available for each pixel decreases accordingly, thus leading to noise degradation in the image quality. Apart from the practical physical constraints due to the noise, production cost is another concern in using high precision optical and imaging sensors in many commercial applications. Therefore, hardware improvement is not always a feasible option. This motivates the study of employing software approaches as an effective and economical alternative to overcome the hardware limitations.

One example of the software-based HR image reconstruction techniques is image super-resolution (SR), which has been developed rapidly over the last two decades. The goal of image SR is to reconstruct an image with higher resolution than the captured image frames. From the viewpoint of *information* theory, the HR image contains more information, in terms of the scene details, than any LR image. Hence, the main issue of SR reconstruction is to explore the sub-pixel information from the blurred, overlapped and noisy LR frames. Based on the LR frame number, SR can be generally classified into multi-frame SR and single-frame SR. For multi-frame SR, each LR image must have a unique "look" of the scene and the non-redundant information in each of the LR images is exploited and combined to reconstruct a HR image. In contrast to multi-frame SR, single-frame SR reconstructs a HR image from a single LR image and the non-redundant information directly from the LR image is insufficient. Therefore, single frame SR adopts an indirect way to obtain the additional information learned from training database that consists of images similar to the target HR image.

SR post-processing can overcome the limitation of hardware to obtain a HR image of considerable good quality. It can be employed in various applications, ranging from high-end applications of medical imaging and remote sensing to low-end applications of video security surveillance, and so on.

In medical imaging applications, high resolution images are required because higher resolution could provide more accurate information for doctor to understand the anatomy [37]. Therefore, the abnormalities of human bodies would be detected as early as possible, which could increase the survival rate of therapy for some illness, such as cancer. One example is the magnetic resonance imaging (MRI) used for head imaging as shown in Figure 1.3. In medical image processing, the partial volume effect (PVE) is a direct problem due to the limited resolution of the MRI machine. The reason of PVE is that two adjacent tissues' interface may occur within a single voxel. Because of PVE, the exact location of boundaries may be shifted, which leads to a major obstacle for MRI brain segmentation in medical image processing. However, due to economic cost or physical limit, image quality cannot be improved by replacing the imaging hardware in some practical situations. Therefore, SR can be considered as a feasible option to extend the existing medical imaging resolution constraints.

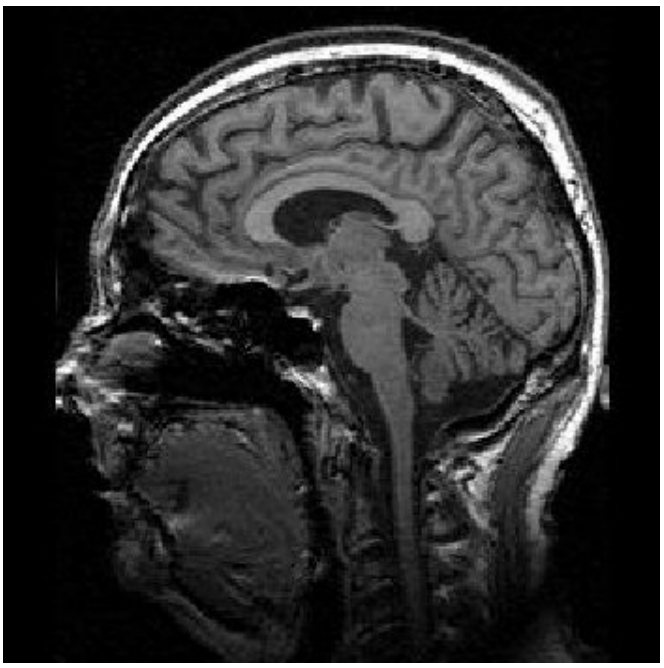


Figure 1.3 A MRI image of a head.

Satellite remote sensing (RS) is another important high-end application which produces various imagery for scientific analysis (e.g. intraurban) [82]. One example is shown in Figure 1.4. Spatial resolution enhancement for satellite images is usually required in the field of RS, especially for the task of recognizing objects of size near the spatial resolution limitation of the optical system. As discussed above, the direct hardware approach is to update the whole system using longer focal ratios and increasing the pixel density of CCD chips. But this may not be applicable as larger and more advanced orbital platforms are required. Moreover, the cost of a manned space mission is far more expensive than that of a common repair on earth. Therefore, using software approach, such as SR, is an economical alternative solution to obtain higher spatial resolution images at affordable costs.

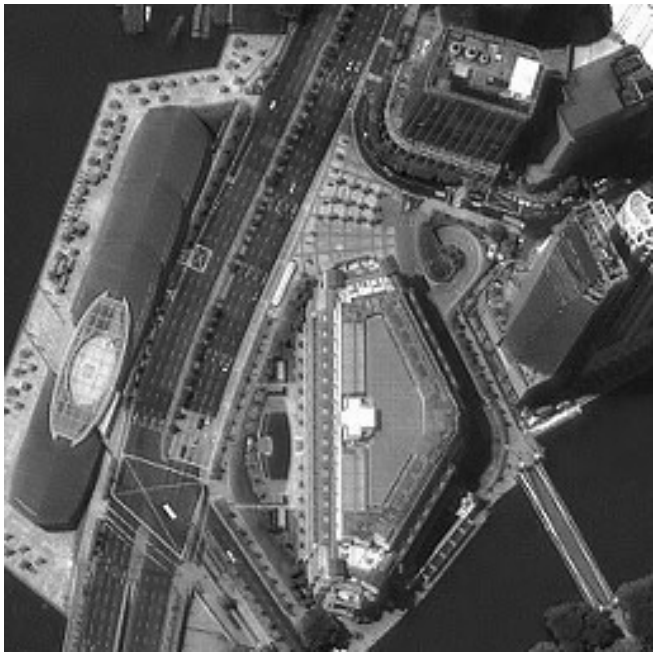


Figure 1.4 A remote sensing satellite image of buildings.

The most common civilian application of SR is the video security surveillance. Surveillance activities often require zooming into a region of interest (ROI) in an image such as a suspect's face or the vehicle license plate number. Because of hardware limitations of acquisition equipments, however, the zoomed ROI image would suffer from the pixel artifacts and contain insufficient details. Taking the picture in Figure 1.5 as an example, a suspect in a theft is recorded by the security surveillance system. But we cannot identify the details of his face due to the low spatial resolution of the video camera. Another example is the security surveillance system applied in a car park of shopping mall, as shown in Figure 1.6. The digital camera is a 16 megapixel camera. However, we cannot identify the people and vehicles when zooming in the ROI. This is because the number of available pixels in ROI is limited although the whole scene has a large pixel resolution. Therefore, employing the software post-processing is a good alternative when compared with hardware promotion.

Figure 1.5 A suspect in a theft captured by the surveillance system.

Figure 1.6 A zooming-in ROI in an image captured by the car park surveillance system.

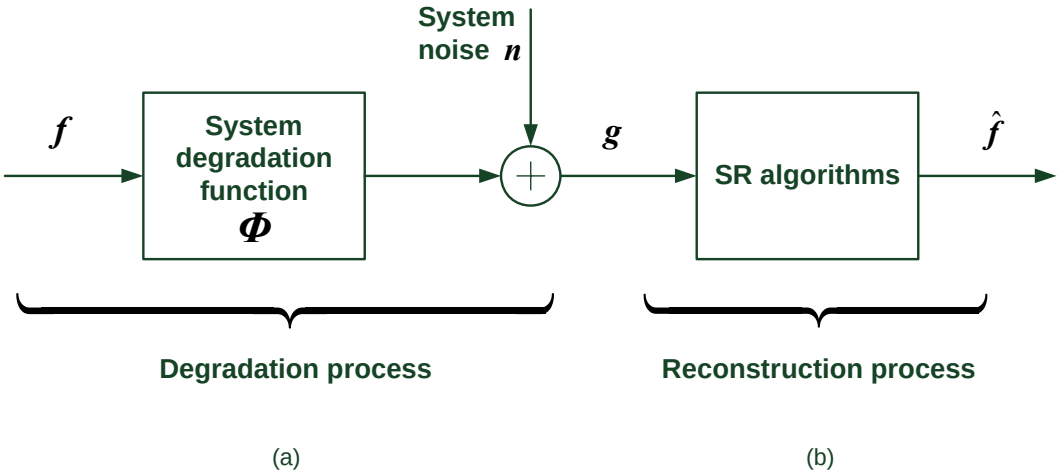


Figure 1.7 The block diagram of degradation and reconstruction: (a) forward process; (b) inverse process.

1.2 Objective

The main goal of this thesis is to develop efficient super-resolution (SR) algorithms and techniques to reconstruct a high resolution (HR) image from the captured degraded low resolution (LR) images. It covers the interesting topics in SR areas, including multi-frame SR, single-frame SR, and color SR. As illustrated in Figure 1.7(a), the degradation process can be modeled in a mathematical formation as

$$\boldsymbol{g} = \boldsymbol{\Phi f} + \boldsymbol{n} \tag{1.1}$$

where $\boldsymbol{g}$ denotes the captured image, $\boldsymbol{f}$ represents the original scene, and $\boldsymbol{n}$ is the additive noise. $\boldsymbol{\Phi}$ is defined as the degradation operator describing the property of the digital imaging system. It consists of down-sampling, blurring and sometimes shifting and color filtering.

The purpose of HR image reconstruction is to estimate the original HR image $\hat{\boldsymbol{f}}$ from the existing captured LR image $\boldsymbol{g}$, as shown in Figure 1.7(b). In real-life applications, however, a sufficient number of LR images may not be available due to some capturing constraints. In such cases, the sparse matrix $\boldsymbol{\Phi}$ would have high condition number and the inverse process of estimating HR image $\hat{\boldsymbol{f}}$ would be an ill-posed problem. More than one HR images $\hat{\boldsymbol{f}}$ could be mapped to the initial LR image $\boldsymbol{g}$ and a unique solution is not guaranteed [35]. Therefore, regularization techniques are often utilized to constraint the reconstruction cost function.

1.2.1 Multi-frame SR

Multi-frame SR refers to an inverse process to reconstruct a HR image, given a sequence of LR images. These LR images can be captured by multiple imaging devices at the same time or a single imaging device at different times. In this thesis, we mainly focus on the latter case as the majority of real-life SR applications can be classified into this category. There are two necessary prerequisites for multi-frame SR. One is that each of the LR images contains an overlapping content of the same scene. The other is that there are relative sub-pixel motions between the LR images. To illustrate the principle of multi-frame SR, a simplistic example which considers relative motion and downsampling only is shown in Figure 1.8. From the figure, it can be observed that the HR image pixels f_1, f_2, f_3, and f_4 are covered fully or partially by the LR image pixels g_1, g_2, g_3, and g_4. Hence, given LR pixels, these HR pixels can be calculated by solving the following equations

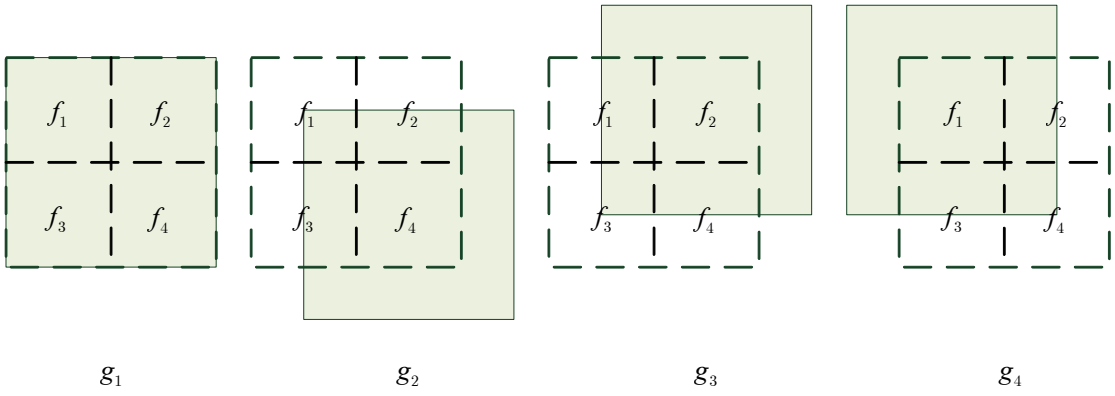


Figure 1.8 An example of multi-frame SR.

$$\begin{aligned} g_1 &= \Phi_{1,1} f_1 + \Phi_{1,2} f_2 + \Phi_{1,3} f_3 + \Phi_{1,4} f_4 \\ g_2 &= \Phi_{2,1} f_1 + \Phi_{2,2} f_2 + \Phi_{2,3} f_3 + \Phi_{2,4} f_4 \\ g_3 &= \Phi_{3,1} f_1 + \Phi_{3,2} f_2 + \Phi_{3,3} f_3 + \Phi_{3,4} f_4 \\ g_4 &= \Phi_{4,1} f_1 + \Phi_{4,2} f_2 + \Phi_{4,3} f_3 + \Phi_{4,4} f_4 \end{aligned} \tag{1.2}$$

where $\Phi_{i,j}$ is the weight indicator proportional to the contribution of HR pixel f_j to LR pixel g_i.

Generally, the degradation process consists of blurring and system noise along with motion and downsampling. The degradation model can be stated in a matrix-vector form as

$$\boldsymbol{g}_i = \boldsymbol{D}_i \boldsymbol{H}_i \boldsymbol{S}_i \boldsymbol{f} + \boldsymbol{n}_i , \qquad 1 \le i \le N \tag{1.3}$$

where $\boldsymbol{D}_i$ is the downsampling operator, $\boldsymbol{H}_i$ is the matrix representing blurring, and $\boldsymbol{S}_i$ denotes the shifting operator. $\boldsymbol{n}_i$ represents the system noise. Based on the imaging model in (1.3), therefore, the main issue in the multi-frame SR problem is to explore the relative motion between the LR images through registration. The problem can be divided into two sub-problems of how to describe the relationship of the LR images using an appropriate motion model and how to estimate the motion parameter based on the specific motion model.

To tackle the first problem, we need to adopt a general observation model which can describe the motion between the LR images. As the result of flexible relationship of LR images, it is difficult to use a specific motion model to describe the relative motion. In the region of interest (ROI), however, the motion can be considered as a simple global motion and a specific motion model is adequate to model the motion. Conventional methods adopt a global translational motion model. However, this model may not be accurate in SR applications involving motion of translation, rotation and zooming. Therefore, it is necessary to incorporate a general motion model into the SR reconstruction, which takes into account both flexibility and simplicity.

Registration plays an important role in image SR approaches. The aim of image registration is to estimate the motion parameters using various techniques, such as FFT (fast Fourier transform). Conventional methods usually ignore the registration error and assume the motion parameters estimated by current registration approaches as error-free. This assumption is not practical as the registration techniques are conducted on the LR images. The aliasing effects may reduce the registration accuracy and lead to suboptimal solution in the following HR reconstruction. Therefore, a unified framework that combines the two stages into a single process and performs them simultaneously is desirable.

1.2.2 Single-frame SR

Single-frame SR is a process to reconstruct the HR image from a single observation. With a small modification of ignoring the motion matrix $\boldsymbol{S}$, the degradation model can be modeled as

$$\boldsymbol{g} = \boldsymbol{DHf} + \boldsymbol{n} \tag{1.4}$$

where $\boldsymbol{D}$ is the downsampling operator, $\boldsymbol{H}$ is the matrix representing blurring, and $\boldsymbol{n}$ represents the system noise. It is noted that the available information from the LR observations is limited for single frame SR when compared with multi-frame SR. Therefore, strong prior information of the desired HR image should be imposed to guarantee a stable solution. Inspired by pattern recognition, a feasible way is to learn the characteristic of the desired HR image from a database containing similar HR images. Hence, single frame SR can be regarded as learning based SR for special domain image reconstruction, including face images and text images.

HR vehicle license plate (VLP) image reconstruction is an interesting SR application. It can be used in video surveillance systems of large parking areas, for instance airport or shopping mall car parks where only one LR image can be captured at times. In this case, single frame SR is a better choice when compared with multi-frame SR. A direct approach is to estimate the HR image by simple interpolation, such as bicubic interpolation. Other works center on imposing specific constraints on the HR image, such as edge-preserving prior. However, the insufficiency of initial information increases the ill-posedness of the reconstruction problem. License plate image is of some specific characteristics when compared with general images. For example, license plate consists of alphanumerics with rigid format. The HR image reconstruction process can benefit by learning from the database containing similar images.

1.2.3 Color SR

Color image SR is defined as a process to reconstruct a HR color image from the captured LR color images. In most of the real life applications, the captured images contain color information which can be divided into three components of red (R), green (G) and blue (B). However, in

order to reduce the production cost, most of the digital imaging devices employ a CFA (color filter array) before the CCD chip. The aliasing effects due to the discrete signal sampling could lead to suboptimal demosaicing and introduce color artifacts in the subsequent HR reconstruction. Hence, it is more practical to develop SR algorithms directly on the mosaiced LR color images to reconstruct a HR color image. The degradation of color image can be described as

$$\boldsymbol{g}_i = \boldsymbol{D}_C \boldsymbol{D}_i \boldsymbol{H}_i \boldsymbol{S}_i \boldsymbol{f} + \boldsymbol{n}_i , \qquad 1 \le i \le N \tag{1.5}$$

where $\boldsymbol{D}_i$ is the downsampling operator, $\boldsymbol{H}_i$ is the matrix representing blurring, and $\boldsymbol{S}_i$ denotes the shifting operator. $\boldsymbol{D}_C$ denotes the CFA operator. $\boldsymbol{n}_i$ represents the system noise.

Most of the conventional SR methods are designed for gray-scale images or full color images which can be considered as the combination of three color channels. It is noted that there is only one kind of color information at each pixel location of the raw CFA image. Applying the conventional SR approaches on the raw CFA images directly may produce color artifacts. Therefore, an existing process of HR reconstruction from the raw CFA images can be divided into two stages of color demosaicing and conventional gray-scale SR. As the result of suboptimal demosaicing performed on mosaiced LR images, this two-stage process may produce unsatisfactory results during the HR reconstruction. Therefore, this motivates the combination of demosaicing and SR into a single process.

1.3 Contributions

The main contribution of this thesis is to develop a set of algorithms and techniques to reconstruct a HR image from the LR observations. These methods include four aspects, namely, L_1 norm nonlinear SR from LR images with potential outliers, multi-frame zooming SR, single frame SR for license plate image reconstruction, and color image SR.

(1) L_1-norm nonlinear SR from LR images

We propose a nonlinear L_1-norm SR method to jointly perform image registration and HR image reconstruction from LR image with outliers. An L_1-norm cost function is proposed to preserve the sharp-feature of edges and alleviate the effect of outliers. We then formulate the minimization of the cost function into a linear programming (LP) problem. It is noted that the proposed SR approach can integrate image registration and HR reconstruction into a single-step of estimation, hence alleviating the shortcoming of the two-stage process of registration followed by HR reconstruction. The desired HR image along with the motion parameters are estimated progressively over the iterations. Experiments show that the proposed algorithm can effectively handle the initial registration errors as well as outliers in the LR images in HR image reconstruction.

(2) Multi-frame zooming SR

We propose an iterative method for joint image registration and SR reconstruction by fusing the captured LR observations with relative zooming motion. The proposed algorithm combines image registration and zooming SR reconstruction into a single process of estimation. An iterative technique serves to ensure that progressively better estimates of motion parameters and HR image can be obtained. This is more promising when compared with other two-stage SR methods as the estimation of motion parameters could benefit from the information of the

reconstructed HR image. Different from conventional methods that focus mainly on translation and possibly rotation, the proposed method takes into account the relative zooming between the acquired LR images. The adopted motion model enables the proposed method to handle more real-life applications involving motion consisting of translation, rotation and zooming. Further, the proposed method also develops an adaptive weighting scheme in the HR reconstruction. Experimental results demonstrate the effectiveness of the proposed method in handling joint registration and zooming SR.

(3) Single-frame SR for license plate image reconstruction

We propose a learning based single-frame SR framework to address the problem of vehicle license plate image reconstruction. The proposed method introduces a new iterative technique that integrates car plate character prior information and HR reconstruction into license plate SR. Different from the conventional methods utilizing a hard-decision strategy for training patches, the proposed method adopts a soft learning prior which firstly estimates the importance of different learning samples, and then integrates this information into the subsequent HR license plate image reconstruction. Furthermore, a scheme that utilizes both global and local characteristics is introduced for performing license plate HR reconstruction. Experimental results show that the proposed method can handle license plate SR effectively.

(4) Color image SR

We propose a new framework to perform joint color image SR and demosaicing from mosaiced LR color image directly. The main feature of the proposed method is that it combines the color SR and demosaicing into a single process. To alleviate the effects of color errors in HR reconstruction, a new data fidelity term and a color correction term are incorporated into the proposed cost function. Through solving the minimization problem, color image SR and

demosaicing can be performed simultaneously. Moreover, the proposed method introduces an adaptive edge-directed bilateral total variation (BTV) regularization term to preserve the high frequency information in the reconstructed HR image. An iterative scheme enables the reconstructed HR image to be improved progressively. Experimental results demonstrate the effectiveness of the proposed method in handling color SR.

1.4 Organization of the Thesis

The remainder of the thesis is organized as follows:

In Chapter 2, a comprehensive literature review about the recent SR algorithms and techniques are given. These SR approaches cover the subjects of multi-frame SR, single-frame SR and HR color image reconstruction.

In Chapter 3, we firstly present the conventional SR algorithms which are sensitive to initial registration error and outliers in the LR image. To handle this issue, a nonlinear L_1-norm approach for joint image registration and SR is proposed. Experimental results demonstrate that the proposed SR algorithm can achieve good performance of HR reconstruction in the presence of outliers in the captured LR images.

In Chapter 4, we firstly show that the conventional super-resolution (SR) methods typically formulate the SR problem as a two-stage process, namely registration of LR images and subsequent HR reconstruction. As the registration algorithms in these two-stage processes experience various degrees of errors, this could degrade the quality of the subsequent HR reconstruction. In view of this, we develop a new approach of jointly performing registration and SR reconstruction. Both simulated and real-life experimental results demonstrate the effectiveness of the proposed method.

In Chapter 5, we firstly present that the hard binary weighting strategy used by current learning-based method is inconsistent with practical applications. Then, a new single frame SR framework using soft learning prior is provided, which can be applied in license plate image reconstruction. The method integrates image SR with car plate character prior information to perform license plate SR. Experimental results show that the proposed method is effective in handling license plate SR.

In Chapter 6, we firstly show that most of the conventional SR methods are typically designed for gray-scale images. These single-input single-output (SISO) model-based approaches tend to suffer from color artifacts in applications when the LR images are captured by a single CCD chip equipped with a CFA (Color Filter Array). In view of this, we develop a new framework of joint demosaicing and SR to reconstruct a HR color image from a sequence of LR mosaiced color images directly. Experimental results show that the proposed method can handle color SR effectively.

In Chapter 7, the primary contributions toward efficient HR image reconstruction are summarized. Several possible avenues of future research are discussed.

Chapter 2

Literature Review

In this chapter, a comprehensive review of techniques and methods for image super-resolution (SR) is provided. These approaches are broadly categorized under multi-frame SR methods, single-frame SR methods and high-resolution (HR) color image reconstruction methods.

2.1 Multi-frame SR Approaches

In many survey articles [92], [132], [26], there are mainly two categories of approaches to reconstruct a HR image from multi-frame LR images. One class of methods is those that are performed in some transformed domains (e.g. the frequency domain and wavelet domain). The other class of methods directly processes the LR images in the spatial domain.

2.1.1 Frequency Domain Approaches

The first SR algorithm was proposed by Tsai and Huang [134] in 1984, which is based on the shifting and aliasing properties of the Fourier transform. In their work, the HR image is reconstructed by using a sequence of LR images with exactly known global translational displacement. However, their method does not take into account the effects of sensor blurring and observation noise during the image acquisition process. Later, an extension of Tsai and

Huang' initial work for blurred and noisy LR images was proposed by Kim *et al.* [62]. They developed a weighted recursive least-squares algorithm for solving the linear equations in the frequency domain. They further refined their work in [64], [63] to address the ill-posed inverse problem by introducing the Tikhonov regularization technique. In [105], Rhee and Kang developed a discrete cosine transform (DCT) approach. Different from conventional frequency domain methods, their work adopts the regularization technique in spatial domain to obtain a stable solution. A block-based parallel structure is adopted to improve the computational efficiency and reduce the memory requirement. Generally, the frequency domain approaches discussed above assume the original HR image to be band-limited, which will constrain their applications on a generalized under-sampled model. An improved frequency domain-based method using a cyclostationary spectral signal model was developed by Prendergast and Nguyen in [99]. In their work, the minimum mean squared error (MMSE) reconstruction technique [100] is employed to find the solution of the SR problem. The major advantage of the aforementioned frequency-domain approaches is their low computational complexity and simplicity of implementation. However, the assumption of global translational observation model and linear space invariant blur is an obstacle which constrains them from real-life applications. Further, it is noted that *a prior* knowledge is difficult to be applied in the frequency domain.

Wavelet is a useful technique for modern signal processing due to its capability of multi-scale representation, including image SR reconstruction. Similar to Fourier transform, the wavelet technique deals with signals in the wavelet domain. Inspired by the success of wavelet for one dimensional signal processing, Nguyen and Milanfar extended the wavelet technique for two dimensional image signals in their works [86], [87]. They utilized the regularity and structure of the interlaced data and proposed an efficient implementation of wavelet interpolation

for SR application. These approaches can be regarded as the first-generation wavelet (FGWs) techniques. However, as the result of infinite-domain definition, it is difficult for FGWs to handle boundaries of finite image signal.

As the second-generation wavelets (SGWs) techniques possess the capability to deal with bounded domains and arbitrary boundary conditions of a digital image which has a finite number of pixels, they are more suitable for image SR application when compared with FGWs discussed above. Instead of employing dilations and translation of mother function to represent the data sampling, SGWs utilize a spatial-domain lifting scheme which consists of three basic operations, namely, prediction, splitting and updating. In [10], SGWs were firstly developed to solve image SR problems by Bose *et al*. In their work, HR image reconstruction from noisy LR images is restricted in semi-regular sampling lattices which are defined as the tensor product of 1D irregular sampling lattices. However, the limitation of their work is that the motion of LR images is only translational. In order to overcome the constraint of translational motion model, 2D prediction-and-update operators were employed to deal with arbitrary sampling lattices in [9]. Later in [17], Chappalli *et al*. analysed their previous work in [9] by investigating the relationship between the threshold level and the reconstructed image quality. The optimal threshold is selected by taking into account the impact of blur and noise.

2.1.2 Nonuniform Interpolation Approaches

An intuitive spatial-domain based method is the non-uniform interpolation [2], [68], which considers the LR image $\boldsymbol{g}$ as non-uniform spaced sample points of HR image $\boldsymbol{f}$:

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} \tag{2.1}$$

where matrix $\boldsymbol{A}$ is linear operator representing the nonuniform sampling process and its elements are determined by the impulse response function of the imaging system. In [2], the HR image reconstruction consists of two components: (i) missing points interpolation followed by (ii) interpolated image deblurring. The major advantage of non-uniform interpolation methods is their low computational load. However, they assume the imaging system as an idea linear model and do not take into account the blur effects. In reality, the image sampling matrix $\boldsymbol{A}$ which is a function of the characteristics of the imaging system may include effects such as spatial and temporal blurring caused by the lens and sensor, leading to nonideal samples in the observed LR images. Furthermore, they require the strict assumption that the imaging system is known in advance and the size and type of blurs are identical for all LR images.

2.1.3 Iterative Back-projection (IBP) Approaches

Given the relationship between the estimated HR image and the observed LR images as prior knowledge, it becomes possible to estimate the output of an unknown imaging system. Based on this idea, the back-projection of the residual error between the estimated image and the observed image can be used to form an updated estimation of the HR image. When the updated HR estimation has smaller error compared with the earlier estimation, the HR estimate is better than the original one. This process will continue iteratively until the estimate reaches minima. The iterative SR reconstruction can be formulated as follows:

$$\boldsymbol{f}^{i+1} = \boldsymbol{f}^{i} + \boldsymbol{H}\left(\boldsymbol{g} - \boldsymbol{W}\boldsymbol{f}^{i}\right) \tag{2.2}$$

where $\boldsymbol{f}^i$ denotes the HR image in the ith iteration, $\boldsymbol{W}$ represents the degradation on the HR image and $\boldsymbol{H}$ is a back-projection matrix which projects the predicated error $\left(\boldsymbol{g}-\boldsymbol{W}\boldsymbol{f}^i\right)$ onto the estimate of HR image.

S. Peleg *et.al*. [96] proposed an IBP algorithm applied to SR reconstruction. In their work, the estimated HR image is first back-projected to simulate the LR image. Then, the projected LR images are compared with the observed data to calculate the predicted error. Finally, the HR estimation would be improved based on the predicted error. Later, taking into account the point spread function of the imaging system, Irani and Peleg [47] extended their earlier work [96] by back-projecting the error between the simulated LR images and the LR observations. The advantage of the IBP method is that a more general motion model could be adopted. However, the method is unable to ensure a stable unique solution due to the inherent ill-posed nature of the inverse problem. Further, the back-projection matrix $\boldsymbol{H}$ is difficult to construct.

2.1.4 Projection onto Convex Sets (POCS) Approaches

Another prominent spatial-domain approach for SR reconstruction is developed based on the projection-onto-convex-sets (POCS) technique. Stark and Oskoui [122] first proposed a POCS-based method for SR reconstruction problem. In their work, the SR solution space is restricted to the intersection of different constraints, including data fidelity with respect to observations, finite energy of images and statistical characteristics of specific images (texts or faces) etc. Defining the constraint set as $\mathbb{R}_i$ for each constraint ($\boldsymbol{P}_i$ is the projection of estimation $\boldsymbol{f}$ onto set $\mathbb{R}_i$), the estimated HR image $\boldsymbol{f}$ is restricted to $\cap_{i=1}^{m}\mathbb{R}_i (1 \le i \le m)$ and it can be obtained recursively as follows:

$$\boldsymbol{f}^{i+1} = \boldsymbol{P}_m \boldsymbol{P}_{m-1} \cdots \boldsymbol{P}_1 \boldsymbol{f}^i . \tag{2.3}$$

To extend their work, Patti and Altunbasak [93] enhanced the previous POCS-based method by using higher-order interpolation. Their approach involves edge adaptive constraints to produce sharp edges and reduce edge ringing. The advantage of POCS-based method is its ease with which *a prior* constraint can be incorporated. However, it is noted that the final solution is nonunique and dependent on the initial guess. Further, the algorithm converges slowly. To combine the benefits of the stochastic estimators, and the advantages of the POCS-based methods, Elad and Feuer [23] proposed a hybrid algorithm to ensure a single optimal solution for the HR image reconstruction.

2.1.5 Maximum Likelihood (ML) Approaches

Instead of treating the unknown HR image and the LR observations as deterministic, stochastic approaches assume that the desired HR image, the captured LR images, and degradation (blur and noise) are realizations of stochastic processes which are related via stochastic equations. Based on the Maximum Likelihood (ML) statistical estimation theory, the estimation of the desired HR image $\boldsymbol{f}$ can be performed by maximizing the conditional probability density function of measurements $\boldsymbol{g}$ as follows:

$$\begin{aligned} \boldsymbol{f} &= \arg\max_{f \in \Re} \left\{ P(\boldsymbol{g} | \boldsymbol{f}) \right\} \\ &= \arg\max_{f \in \Re} \left\{ \ln\left(P\left(\boldsymbol{g} | \boldsymbol{f} \right) \right) \right\} \end{aligned} \tag{2.4}$$

where $\boldsymbol{f}$ is the desired HR image and $P(\boldsymbol{g}|\boldsymbol{f})$ represents the conditional probability density function.

To enhance the resolution of multi-frame infrared image data, a stochastic SR approach based on ML technique was investigated in [12]. The authors assume that the HR image is of Poisson distribution and use the expectation maximization (EM) algorithm to estimate translational motion and a HR image simultaneously. The advantage of ML method is the flexible spatial-domain observation model when compared with interpolation methods. However, due to the inherent ill-posed nature of the SR inverse problem, there may be infinitely many possible solutions for (2.4).

2.1.6 Maximum a Posteriori (MAP) Approaches

An effective way for solving the ill-posed SR problem is to use *a prior* knowledge, which can constrain the HR image space. It is noted that Bayesian estimation is particularly well suited to solve the ill-posed inverse problems as they include *a prior* constraint in the form of prior probability density functions [97]. To estimate the target HR image $\boldsymbol{f}$, the Maximum *a posteriori* (MAP) estimator can be formulated as

$$\begin{aligned} \boldsymbol{f} &= \arg\max_{f} \left\{ P(\boldsymbol{f}|\boldsymbol{g}) \right\} \\ &= \arg\max_{f} \left\{ P(\boldsymbol{g}|\boldsymbol{f}) P(\boldsymbol{f}) \right\} \\ &= \arg\max_{f} \left\{ \ln\left(P(\boldsymbol{g}|\boldsymbol{f}) \right) + \ln\left(P(\boldsymbol{f}) \right) \right\}. \end{aligned} \tag{2.5}$$

The distinct feature of the MAP methods is that they will define the prior probability $P(\boldsymbol{f})$ to reflect the characteristics of the desired HR image.

A general model for the prior distribution $P(\boldsymbol{f})$ can be defined as a Markov random field (MRF) which is usually characterized by Gibbs distribution [33]

$$P(\boldsymbol{f}) = \frac{1}{\alpha}\exp\{-\lambda G(\boldsymbol{f})\} \tag{2.6}$$

where α is the normalizing constant, λ represents "temperature" parameter and $G(\boldsymbol{f})$ denotes the energy function of the image $\boldsymbol{f}$. Because of the analytical property of Gaussian distribution, Gaussian Markov random field (GMRF) is often utilized for image prior [39], [41]

$$P(\boldsymbol{f}) \propto \exp\left(-\frac{1}{2}\boldsymbol{f}^T\boldsymbol{C}\boldsymbol{f}\right) \tag{2.7}$$

where $\boldsymbol{C}$ represents a symmetric positive matrix. However, one disadvantage of GMRF prior for SR reconstruction is that GMRF tends to produce over-smoothness in the reconstructed HR result with blurred edges.

In [130], Tian and Ma exploited a Markov Chain Monte Carlo (MCMC) technique to generate samples from the posterior probability distribution $P(\boldsymbol{f}|\boldsymbol{g})$ to compute the MAP estimation. Later, in order to preserve the edge information in the reconstructed HR image, an edge adaptive MCMC SR approach was further proposed in [129]. Recently, Tae Oh and Jay Kuo developed a locally-adaptive stochastic technique for SR reconstruction [127]. In their work, a piecewise auto-regressive (PAR) process is utilized to model the 2D textual image and its parameters are calculated based on a non-local (NL) training procedure.

To produce a stable and good quality result, R. Molina *et al*. [83] also formulated the distribution of HR image $\boldsymbol{f}$ by a simultaneously spatial autoregressive (SAR) model

$$P(\boldsymbol{f}|\alpha) \propto \exp\left(-\frac{1}{2}\alpha\boldsymbol{f}^T\boldsymbol{C}\boldsymbol{f}\right) \tag{2.8}$$

where α is a parameter measuring the smoothness of the HR image $\boldsymbol{f}$. The expectation maximization (EM) technique is employed to compute α iteratively. However, this model requires the additional calculation of the unknown parameter α when compared with GMRF method. Furthermore, it is computationally demanding for general SR problems as it requires computation of the trace of the inverse of matrix $\boldsymbol{C}$.

In [51], the stationary image models of both MRF and SAR were discussed. The paper presents a technique for SR imaging of the same scene from several observations at different camera zoom settings. Nevertheless, either the MRF or SAR prior imposes smoothness uniformly across the whole image, leading to blurred edges in the reconstructed HR result. To address this problem, Wood *et al*. [139] proposed a non-stationary image prior which provides the flexibility to model spatially varying correlation.

For sharper edges in the reconstructed HR image, Schultz and Stevenson [112], [113] adopted a discontinuity-preserving model using the Huber-Markov Gibbs prior

$$P(\boldsymbol{f}) \propto \exp -\frac{1}{\lambda}\sum_{c\in C} V_c(\boldsymbol{f}) \tag{2.9}$$

where λ represents the "temperature" parameter of the density, $V_c(\bullet)$ denotes some function of a local group of points c called cliques and C is the set of all cliques through the image $\boldsymbol{f}$.

The advantage of MAP approaches is that it is easy to incorporate prior information to the HR image reconstruction process. Further, under the assumption of Gaussian noise and convex prior model, a global optimization can be reached. The disadvantage is its high computational cost.

2.1.7 Deterministic Regularization Approaches

Similar to MAP methods, the deterministic regularization approaches aim to estimate the HR image by minimizing the cost function

$$\boldsymbol{f} = \arg\min_{f} \left\{ \left\| \boldsymbol{g} - \boldsymbol{W}\boldsymbol{f} \right\|_P^Q + R(\boldsymbol{f}) \right\} \tag{2.10}$$

where $\left\| \bullet \right\|_P$ denotes the L_P (P=1 or 2) norm, $\left\| \boldsymbol{g} - \boldsymbol{W}\boldsymbol{f} \right\|_P^Q$ represents the data fidelity to the observed data and $R(\boldsymbol{f})$ imposes constraints on the estimated HR image. Smooth constraint such as Tikhonov regularization is often adopted in formulation $\left\| \boldsymbol{\Gamma}\boldsymbol{f} \right\|_2^2$, where $\boldsymbol{\Gamma}$ denotes the high pass filter. It is noted that, when P=2 in the data fidelity term, function (2.10) utilizing Tikhonov regularization is a special case of the more general MAP formulation under the assumption of both Gaussian noise and Gaussian image prior. Recently, a very fast SR reconstruction algorithm has been developed in [24] for LR images with pure translational motion and common space invariant blur. Farsiu *et al*. [28] further proposed a fast and robust SR algorithm by using L_1 norm, for both the regularization term and the data fidelity term. It is shown that this method is robust to errors from the estimation of motion or blurs, and results in images with sharp edges.

2.2 Single-frame HR Image Reconstruction

Most HR reconstruction techniques and algorithms are proposed for multi-frame image SR. However, at times, where there is only one observed LR image available, single-frame SR/interpolation is required to reconstruct a HR image. In this section, we will review the related

work, which can be broadly classified into three groups: interpolation-based approaches, reconstruction-based approaches and learning-based approaches.

2.2.1 Interpolation-based Approaches

A classical interpolation-based method is bi-cubic interpolation technique which models the image intensity in a local spatial image neighborhood using piecewise polynomial. However, this approach tends to produce over-smoothness, especially in the edge regions containing high-frequency features. In order to overcome this problem, several edge-based methods have been proposed, including adaptive bi-cubic interpolation technique which utilizes edge directions to improve interpolation [110]. In [45], Hong *et al*. developed an adaptive interpolation method based on the local feature of different regions of the desired HR image. In their work, they employ discrete cosine transform (DCT) technique to extract the edge information of each local region. The interpolation strategy is adaptively determined based on the local edge information and different for each local area. Recently, an edge-directed interpolation method was proposed by Li *et al*. in [70]. The underlying principle is that the local covariance coefficients used for HR reconstruction can be estimated from the single LR image based on the geometric duality of HR and LR images. The edge direction based interpolation is adaptively tuned based on the estimated covariance. Although the methods discussed above have low computational load, they need to estimate heuristic threshold values and filter weights. Thus, the results are sensitive to changes in values of these parameters.

2.2.2 Reconstruction-based Approaches

Under the assumption that the down-sampled version of HR estimate is close to the input single LR image, many reconstruction-based single-frame SR methods have be developed. Similar to

their multi-frame counterparts, the reconstruction-based approaches formulate the single-frame SR problem into the optimization of a cost function

$$\min_{\boldsymbol{f}} \| \boldsymbol{W}\boldsymbol{f} - \boldsymbol{g} \|_2^2 + R(\boldsymbol{f}) \tag{2.11}$$

where $\| \boldsymbol{W}\boldsymbol{f} - \boldsymbol{g} \|_2^2$ represents the data fidelity term and $R(\boldsymbol{f})$ is denoted as the regularization term. The regularization term serves to ensure a stable solution to the inverse problem by incorporating the characteristic of the desired HR image as prior information [49]. In [126], Suresh *et al*. applied an adaptive edge-based regularization term to preserve the sharp features in the reconstructed HR result. Similarly, Chen *et al*. employed *Tikhonov* regularization technique to constrain the desired HR image during the SR reconstruction in [18]. In their work, singular value decomposition (SVD) technique and Kronecker product technique are used to reduce the computational cost. Generally, most of the regularization terms adopted in the methods above concentrate on smoothness priors. Further, the regularization term becomes increasingly less effective as the magnification factor increases during the HR image reconstruction [5].

2.2.3 Learning-based SR Approaches

Learning-based SR approaches use training samples from a general image database to perform single-frame SR [128], [149]. The most popular approaches are based on the idea of *pattern recognition* techniques.

As shown in Figure 2.1, an example of the training databases consists of a set of HR image patches and a corresponding set of LR versions obtained by applying degradation matrix on each HR images. The basic idea of learning-based SR methods is as follows. Given a small patch of

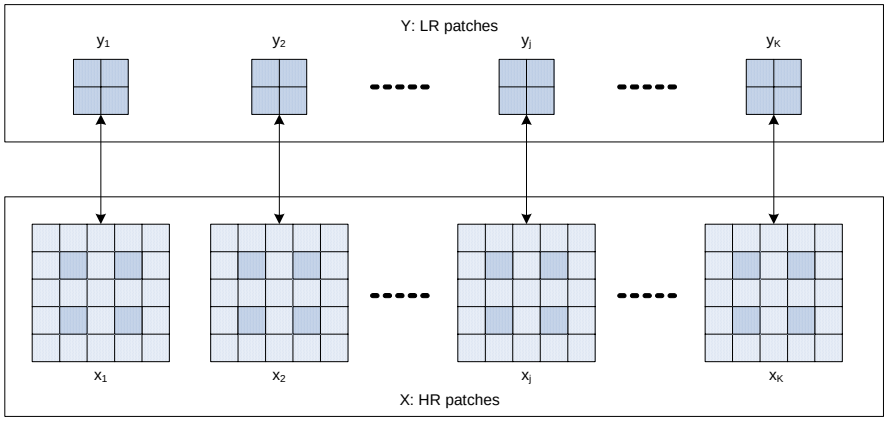


Figure 2.1 Training database

LR image, first compare it with LR examples in the database to find all the similar ones. Then, their corresponding HR patches could be employed for the HR reconstruction as they provide the HR details that represent the characteristics of the target HR image.

Freeman *et al*. [31], [30] first proposed a learning-based framework for image SR, which uses a Markov network model, as shown in Figure 2.2, to reconstruct the desired HR image. In his work, it is assumed that the probability of any selected HR patch for each node $\boldsymbol{x}$ is proportional to the product of all sets of transition function ψ which represent the possible states of each pair of neighboring hidden nodes, and observation function ϕ which represents each observation to the underlying hidden states

$$P(\boldsymbol{x}|\boldsymbol{y}) \propto \prod_{ij} \psi_{ij} \prod_{i} \phi_i \,. \tag{2.12}$$

This approach is a direct way to utilize the selected example patches to reconstruct the HR image. In order to find the appropriate LR example patches $\boldsymbol{y}$, large quantity of training sets with the similar characteristic of the target HR image is required. To alleviate the computational

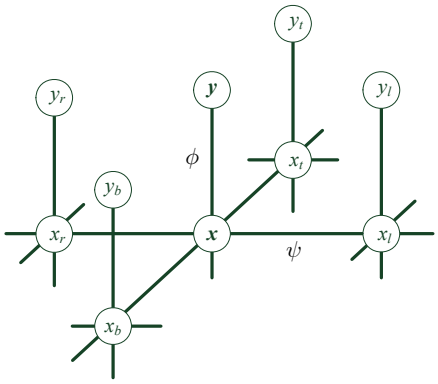


Figure 2.2 Markov network model for HR image: x represents the underlying HR image patch; y represents the observed LR image patch.

burden of recognition process, Sun *et al*. [125] employed the low-dimension primal sketches as the recognition elements to explore the proximate training candidates, and then the Iterative Back-projection technique is applied to constraint the consistency of the training candidates in the HR reconstruction.

The locally linear embedding (LLE) technique [109] developed recently provides a power tool for simplifying the nonlinear process (2.12) into a local linear expression for manifold. In [15], Chang *et al*. proposed a manifold learning-base method which uses the neighbor embedding of the HR patches to reconstruct each patch of the HR image. Their work is under the assumption that the feature spaces for LR and HR images have the similar local linear geometric structure. Following this work line, Chang *et al*. [16] utilized a three-layer Markov network to model the nonlinear manifold correlation between the LR patches and their corresponding HR patches. The desired HR image can be obtained by solving a direct energy minimization problem. Similarly, Yang *et al*. extended the sparse representation technique into the neighbor embedding and calculate the weighting for neighboring patches by L_1 norm minimization [144], [145]. These methods are developed under the assumption that HR patch pairs share the identical local

geometry with the corresponding LR counterparts. However, a recent work in [8] found that this assumption does not always hold true. To avoid the confusion caused by non-local reconstruction, they introduced a Locality Preserving Constraints (LPC) technique to guarantee the consistency and the locality on the coupled patches. The advantage of the learning-based methods discussed above is that the information learning from training examples can reconstruct the HR details of the target image well with any magnification ratio, while the conventional reconstruction based methods have an upper boundary.

It is noted that the learning-based approaches are more effective in handling domain-specific SR problems, such as HR text or face image reconstruction. In [60], Kim proposed a learning based SR approach to increase the resolution of the noiseless text image. A weighted *k*-nearest neighbor search technique is adopted to estimate the most matched HR patches for HR reconstruction. Following this idea, Dalley *et al*. [20] developed a full-Bayesian approach with an explicit noise model. To increase training efficiency, they employ a simple table lookup strategy instead of the nearest-neighbor search. An adaptive interpolation algorithm using spatially-variable filter technique was proposed in [19]. The mean square error (MSE) filter is modeled as a function of edge patterns, whose parameters can be determined nonlinearly from the rendered text. The method combines the efficiency of linear interpolation with the power of example training. In [91], Park *et al*. proposed a prior model based on text images in the Bayesian SR framework. The local characteristics of the text can be preserved by the adopted Markov Random Field (MRF) technique. Recently, a text SR approach based on the neighbor embedding technique was developed by Smith in [117].

Face SR is another popular technique due to the broad application of security surveillance system. The concept of face hallucinating was first proposed by Baker and Kanade in their work

[4]. Later, they developed a face hallucinating approach which reconstructs the high frequency details of the face by learning from a pyramid-structure database. The use of Principle Component Analysis (PCA) techniques for face SR purposes has been investigated in [138]. Under the assumption that the principal components remain unchanged across the scale, the desired HR image is reconstructed using the interpolated eigenimages of LR observation. To reconstruct the detail information of facial images captured at a distance, Park and Lee [90] proposed a top-down learning resolution enhancement method involving face representation using a linear combination of shape and texture prototypes, followed by an error back projection procedure. It is robust to sensor noise, incompleteness of input images and occlusion by other objects. Generally, to encode the reconstructed HR image using a linear image structure is a simple and direct idea. However, it ignores part of the high-order statistics or details which are particularly important for the reconstruction process. Without encoding a fixed relationship between HR and LR image patches, Kim *et al*. [61] developed an unsupervised learning SR approach using a model obtained from kernel PCA. Based on this idea, Chakrabarti *et al*. [13] employed the kernel PCA to derive prior knowledge for the face feature. The prior information for frontal face is extracted in a computationally efficient manner and incorporated into a MAP framework to improve the quality of the SR face. Extending Freeman's work [31], Stephenson and Chen [123] developed an adaptive MRF technique for learning-based face SR. In their work, both the observation and transition functions are region-dependent. In [75], Liu *et al*. developed a dual associative learning method to explore dependencies of the LR face data and HR face data in the hidden parameter space instead of the observation space. Recently, Yang *et al*. [144], [145] have proposed a novel approach to single-frame image SR, based on sparse signal representation. They seek a sparse representation for each patch of the LR image, and then use the coefficients

of this representation to produce the target HR image. Although the aforementioned learning-based algorithms are able to provide good visual results, they usually cannot handle the real-time applications, since its performance depends on the training database that is usually huge, leading to high computational cost.

2.3 HR Color Image Reconstruction

Most of the conventional SR methods focus on gray-scale HR image reconstruction. In real-life applications, however, many commercial cameras employ a single CCD chip with a color filter, such as CFA (color filter array), to capture the R, G, and B color information of the original scene. In the case with CFA, each pixel of the captured color LR image represents only one color information. In order to reconstruct a HR color image for the initial scene, various approaches which are performed on the raw mosaiced color LR images have been proposed. These methods can be classified into two steps: color image demosaicing and color image SR.

2.3.1 Color Image Demosaicing

A variety of demosaicing techniques have been proposed to estimate the full resolution color image based on the mosaiced color images. The basic demosaicing approach is to apply interpolation [48], [111], [7] to each channel of the color image. These techniques include bilinear interpolation, cubic interpolation [148] and edge based interpolation [143]. The single-channel algorithms can produce satisfactory results in smooth regions of a color image. However, they tend to introduce color artifacts, especially in regions containing high-frequency features. This is because there are correlations between different color channels of most natural images.

To alleviate the effects of color artifacts on the demosaiced color image, both the intra-channel and inter-channel dependences should be utilized during the demosaicing process. These methods can be classified into two categories, namely, sequential demosaicing methods and simultaneous demosaicing methods.

The motivation behind performing demosaicing in a sequential process is that the green (G) channel is less aliased when compared with the other two channels. Hence, an interpolated full resolution G channel can help to restore the red (R) and blur (B) channels. In [84], Muresan and Parks proposed an improved edge-based demosaicing method using optimal recovery technique. In their work, the interpolation of G channel is modeled as an example-based quadratic optimization problem. In [54], a spatially adaptive demosaicing algorithm was developed by Kakarala and Baharav. The direction of interpolation at each pixel is estimated by computing the Jacobian matrix of the color map and neighborhood voting. A two-step process consisting of missing color information interpolation and post-processing of color correlation was proposed by Lu and Tan in [77]. The missing value of one color plane is obtained by a weighted sum of estimates from four different interpolation directions. In the post-processing step, an adaptive median filter is selectively applied to image regions containing sharp edges. In [141], Wu and Zhang proposed a new primary-consistent soft-decision (PCSD) approach for color demosaicing. In order to keep the consistency of different interpolated color channels, multiple estimates of the missing color pixel value were calculated under different assumptions on edge directions. Then an optimal statistical technique is adopted to select the best one based on the testing results of different interpolation assumptions. In their later extended work [150], the missing green samples are adaptively estimated using an optimal fusion of both horizontal and vertical directions by the linear minimum mean square-error estimation (LMMSE) technique.

An important assumption for color image demosaicing is that the color channel difference (R-G and B-G) is constant within a local region for real-life images. Although this assumption holds in the flat regions within the boundary of an image, it has been extensively applied in the interpolation of R and B color channels. In [79], Lukac and Plataniotis developed a color demosaicing method using an edge-sensing mechanism to preserve the sharp edges and high-frequency details of the captured image. Their work is based on a vector-spectral model which can describe the spectral characteristics of RGB channels for the true visual scene. In [81], a similar edge-directed approach discussed above was applied to estimate the full resolution color-difference signals (R-G and B-G) based on the prior information from the interpolated full resolution G channel. Then, G channel is added back to recover the R and B color channels. Recently, Yuk *et al.* proposed a color demosaicing approach which adaptively selects interpolation direction. The principle behind their work is that the directions of color-difference spaces are similar and the potential direction is selected as the one which has higher similarity score. Generally, the color artifacts will be reduced as the demosaicing methods which are adaptively conducted considering the fined edge direction information. Correspondingly, the computational cost will increase due to the additional estimation of edge direction.

The methods discussed above can be considered as sequential process. One disadvantage is that the color artifacts arising from the interpolation of G channel may be propagated to the subsequent R and B channel interpolations. Therefore, applying a color model estimating the RGB channel simultaneously could be an alternative approach. In [38], Gupta and Chen proposed a vector demosaicing algorithm that interpolates missing colors based on the color vector that minimizes the sum of distances to the surrounding pixels. Basically, the selected color vector is a vector median of the surrounding pixels, which can preserve sharp edges. Neural

network (NN) technique is also applied to describe the color model based on learning the relationship between missing color pixels and their neighboring pixels. In [76], for instance, a multiple NN approach was proposed by Long and Huang. Treating the low-gradient flat regions and high-gradient steep regions differently, they adopt an adaptive scheme that conduct more complex neural networks on steep regions in larger sizes of neighborhoods and verse vice. Generally speaking, the above described demosaicing techniques can be regarded as image interpolation approaches and the number of pixels contained in each color channel produced by these approaches is equal to the CCD resolution. A high resolution is often desirable for display and post-processing. Therefore, color image SR techniques are developed to tackle this problem.

2.3.2 Color Image SR

A number of works have been developed to address the problem of color image SR [71], [88], [114], [133], [36], [25]. Based on the Elastic net theory [157], Li and Dai proposed a regularization approach to reconstruct an enhanced resolution image [71]. In their work, the data fidelity term is measured by a linear combination of L_2 norm and L_1 norm. In [88], Ni. *et al*. proposed a learning based super-resolution algorithm for color images by using a stochastic classification-regression framework. In [114], Shah and Zakhor proposed a multi-frame algorithm to enhance the spatial resolution of frames in color video sequences. Their technique specifically utilizes the color component of the video signal to compensate the inaccuracies of the motion vector. Similar to [114], Tom and Katsaggelos proposed an iterative algorithm for enhancing the resolution of color image sequences. In order to improve the HR reconstruction, all three color channels are employed to estimate the motion and to compensate for it [133]. In [36], an approach consisting of SR reconstruction and registration performed directly on raw

"color mosaic" images was proposed. This method is based on a generalized formulation of SR that simultaneously performs both resolution enhancement and color demosaicing.

Chapter 3

Nonlinear L_1-norm Joint Image Registration and Super-resolution

3.1 Introduction

Multi-frame super-resolution (SR) is an inverse process to reconstruct a high-resolution (HR) image of the original scene by fusing a sequence of low-resolution (LR) observations. Typically, subpixel motion exists amongst these LR images and each LR image bears the unique partial information of the original scene [134]. The underlying principle is that if an enough number of such LR images are available, a HR image for the original scene can be reconstructed by performing SR algorithms on these LR observations. In image SR, at times, the region of interest (ROI) may contain a few outliers, namely, a very small number of the LR images which have appearance different from the majority. An example of outlier is shown in Figure 3.1, corresponding to the fourth image with a pencil. In situation when the number of the LR observations is limited, these outlier frames cannot be ignored in the HR reconstruction. However, outliers existing in the LR images may degrade the registration accuracy and

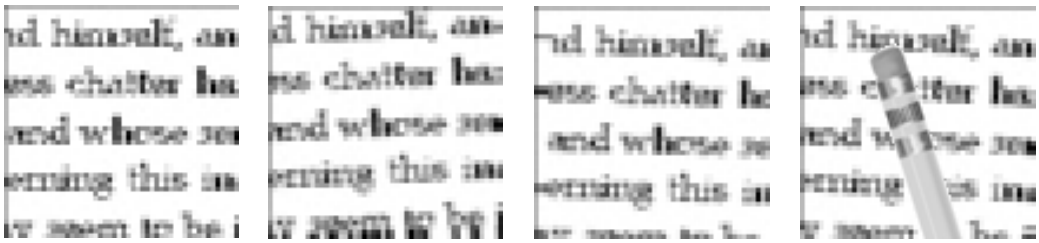


Figure 3.1 Four LR "text" images with one outlier frame.

subsequent HR reconstruction. Therefore, this motivates the study of SR algorithms performed on LR images with outliers in this chapter.

The development of SR algorithms has been an active research area in image processing in recent years [92], [26], [50], [147]. Early SR algorithms have been typically performed in the frequency domain [134], [62], [105]. These methods are centered on the aliasing and shifting properties of the Fourier transform. They are computationally attractive. However, the constraint of the global translational motion model often limits their applications. Therefore, recent research [120], [101], [44], [131], [154], [146], [115], [73] tends to concentrate on spatial domain techniques as a more generic degradation model can be applied in the spatial domain. Many SR methods adopt a two-stage approach of (i) performing registration of the captured LR images followed by (ii) solving the inverse problem to reconstruct the HR image. In this chapter, we refer to these methods as two-stage disjoint methods. However, these two-stage approaches ignore the registration error and assume that the estimates of motion parameters obtained by current registration approaches are error-free. Generally speaking, accurate motion estimation is difficult to achieve, which could lead to suboptimal results in the following reconstructed HR image [3]. Thus, several approaches are developed to alleviate the effects of poor registration.

Various robust SR algorithms have been developed to alleviate the effects of outliers in the HR reconstruction. It is well-known that median-based estimator techniques have the capability

to discard the local inconsistent outliers, such as highlights, while keep the rest of consistent contents during the HR reconstruction process. Based on this idea, Zomet *et al*. proposed a median filter in [156], which can be performed iteratively on the LR images to obtain the final HR image. Considering the inconsistent pixels in the region of interest (ROI) areas as noise, Farsiu *et al*. [27] demonstrated that the Laplacian distribution is more suitable to describe the outlier noise when compared with a Gaussian distribution. In their later work [28], therefore, an L_1-norm regularization method was proposed to reconstruct HR image while preserving the sharp edges. The method is able to suppress the artifacts due to the registration error or outliers up to a certain degree. However, a fundamental issue remains, that is the method does not improve upon the initial registration during the image reconstruction, hence making the method suboptimal.

Another class of SR methods aims to model/alleviate the registration inaccuracy in the SR problem. Lee *et al.* propose a channel adaptive regularization method in [67] by modeling the registration errors using Gaussian noise. Using a translation-only motion model, D. Robinson *et al*. [106] proposed an alternate variable-projections (VP) based framework for image SR reconstruction. In [39], an iterative method using alternating minimization (AM) algorithm was proposed to calculate the motion parameters and the HR image alternately. Expectation-maximization (EM) technique was utilized to solve the SR reconstruction problem based on the formulation of Bayesian and Maximum *a posteriori* (MAP) [140]. To enhance the registration precision, Wang and Xue [137] proposed an algorithm that performs registration in HR domain employing regularized mean curvature upsampled images. Recently, a nonlinear least squares technique has also been proposed in [43] to calculate the motion parameters and the HR image simultaneously. It is noted that most of the SR methods discussed above are sensitive to the outliers in the LR images. The reconstruction of the HR image often suffers from errors due to

inaccurate registration. A reliable SR method, therefore, should be able to take these errors into consideration.

In view of this, we propose a nonlinear L_1-norm SR method to jointly perform image registration and HR image reconstruction. Unlike previous works which focus on translation alone, this work will adopt a motion model which consists of translation and rotation. An L_1-norm cost function is developed to preserve the sharp feature of edges and to alleviate the effect of outliers. We then formulate the minimization of the cost function into a linear programming (LP) problem. It is noted that the proposed method could integrate image registration and HR reconstruction into a single-step of estimation, hence alleviating the shortcoming of the projection-onto-convex-sets (POCS) based approaches, for instance the AM method. The desired HR image along with the unknown motion parameters are estimated progressively over the iterations.

The rest of this chapter is organized as follows. The problem formulation is presented in Section 3.2. An iterative nonlinear L_1-norm HR reconstruction algorithm for joint image registration and super-resolution is developed in Section 3.3. Experimental results on both simulated and real-life images are given and discussed in Section 3.4. Finally, the summary of this chapter is provided in Section 3.5.

3.2 Problem Formulation

Firstly, let us consider the LR image acquisition process. Without loss of generality, let us assume $f(x,y)$ and $g_k(u,v)$ represent the underlying HR image and the *k*th observed LR image, respectively. The LR image $g_k(u,v)$ of sized $M_g \times M_g$ is obtained by rotating the HR $M_f \times M_f$ size image $f(x,y)$ by θ_k, shifting it by a translational vector $\boldsymbol{s}_k = [s_{xk}, s_{yk}]$, then blurring the

rotated-and-shifted result by a point-spread-function (PSF) h_k, followed by down-sampling it by a factor ρ. Therefore, the LR image capturing process can be modeled as

$$g_k(u,v) = (f(x\cos\theta_k - y\sin\theta_k + s_{xk}, x\sin\theta_k + y\cos\theta_k + s_{yk}) \otimes h_k)\downarrow_\rho + n_k(u,v) \tag{3.1}$$

where $\otimes$ denotes the operator of 2-D (2-Dimensional) convolution, $\downarrow_\rho$ represents the operator of down-sampling by a decimation factor ρ and n_k denotes the additive noise. Rewriting (3.1) in the form of matrix, the image capturing process for the SR problem can be commonly modeled as [28], [43]

$$\boldsymbol{g}_k = \boldsymbol{W}_k \boldsymbol{f} + \boldsymbol{n}_k, \qquad (1 \le k \le N) \tag{3.2}$$

where $\boldsymbol{f}$, $\boldsymbol{g}_k$ and $\boldsymbol{n}_k$ denote the lexicographically-ordered original HR image, the kth observed LR image and system noise, respectively. $\boldsymbol{W}_k$ represents the linear mapping that includes motion, blur and down-sampling. Generally, we assume that the first LR image is the reference image. Then, matrix $\boldsymbol{W}_k$ can be expressed as [43]

$$\boldsymbol{W}_k = \begin{cases} \boldsymbol{D}\boldsymbol{H}_1 & (k=1); \\ \boldsymbol{D}\boldsymbol{H}_k\boldsymbol{S}_k(\boldsymbol{\alpha}_k), \ \ \boldsymbol{\alpha}_k = [\cos\theta_k, \sin\theta_k, s_{xk}, s_{yk}]^T & (2 \le k \le N) \end{cases} \tag{3.3}$$

where $\boldsymbol{D}$ is the identical down-sampling operator for each of LR images. $\boldsymbol{H}_k$ represents the blurring operator caused by the combined effect of the camera lens and sensors. $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)$ $(2 \le k \le N)$ is the geometric motion (rotation and translation) operator for the kth LR image. θ_k, s_{xk} and s_{yk} denote the rotational angle, translational shift in the x-direction and y-direction, respectively. In this chapter, the SR reconstruction aims to produce a HR image of the

original scene by the fusion of multiple captured LR images whose motion parameters $\boldsymbol{\alpha}_k$ are unknown.

3.3 Joint Estimation using Nonlinear Method

3.3.1 L_1-norm Cost Function

SR is considered as an ill-posed problem, since a small amount of noise in measurement will result in large perturbation in the final solution. A least squares (L_2-norm) approach is commonly used in SR. However, it has been shown in [28] that the least squares estimation is considered as a mean estimator, while the L_1-norm approach is considered as a median estimator. The results obtained by using the L_1-norm approach are less sensitive to the outliers in the LR images. Therefore, we adopt the L_1-norm technique in this work. However, as opposed to the method in [28] that performs one-off registration on the LR images, the proposed SR algorithm performs the registration of LR images iteratively utilizing the progressively improved HR estimate.

The following cost function is minimized to estimate the HR image $\boldsymbol{f}$ and the motion vector $\boldsymbol{\alpha}_k$:

$$E(\boldsymbol{\alpha}_k, \boldsymbol{f}) = \| r_1(\boldsymbol{f}) \|_1 + \sum_{k=2}^{N} \| r_k(\boldsymbol{\alpha}_k, \boldsymbol{f}) \|_1 + \lambda \| \boldsymbol{L}\boldsymbol{f} \|_1 \tag{3.4}$$

where $\|\cdot\|_1$ denotes the L_1-norm, $r_1(\boldsymbol{f}) = \boldsymbol{g}_1 - \boldsymbol{W}_1\boldsymbol{f}$ and $r_k(\boldsymbol{\alpha}_k, \boldsymbol{f}) = \boldsymbol{g}_k - \boldsymbol{W}_k\boldsymbol{f}$ $(2 \le k \le N)$ are defined as the residual vectors. In this cost function, the first two terms together are the data fidelity. It has been shown in [28] that minimization using L_1-norm in the data fidelity term is less sensitive to the model and data errors. The last term represents a regularization functional

that introduces stability into the HR estimate. Matrix $\boldsymbol{L}$ can be expressed as $[\boldsymbol{L}_x^T \ \boldsymbol{L}_y^T]^T$, where $\boldsymbol{L}_x$ and $\boldsymbol{L}_y$ are the operators that represent the first derivative in the horizontal direction and vertical direction, respectively. λ represents the regularization parameter which servers to control the relative contributions of the data fidelity term and the regularization term. The parameter λ is determined empirically for the experiments in Section 3.4.

3.3.2 Joint Estimation of Motion Parameters and HR Image

The proposed SR algorithm aims at estimating the unknown motion parameters $\boldsymbol{\alpha}_k$ and the desired HR image $\boldsymbol{f}$ simultaneously. Because of the inclusion of rotational motion, however, the minimization problem in (3.4) is nonlinear with respect to vector $[\boldsymbol{\alpha}_k^T, \boldsymbol{f}^T]^T$. In this work, we derive an approximate linear expression for $r_k(\boldsymbol{\alpha}_k, \boldsymbol{f})$. Let us define $\triangle \boldsymbol{f}$ as a small change in the HR image $\boldsymbol{f}$, and $\triangle \boldsymbol{\alpha}_k$ as a small change in the motion vector $\boldsymbol{\alpha}_k$. The residual vector $\boldsymbol{r}_k(\boldsymbol{\alpha}_k, \boldsymbol{f})$ can be linearized with respect to $\triangle \boldsymbol{\alpha}_k$ and $\triangle \boldsymbol{f}$ as follows:

$$
\begin{aligned}
&\boldsymbol{r}_k(\boldsymbol{\alpha}_k + \triangle \boldsymbol{\alpha}_k, \boldsymbol{f} + \triangle \boldsymbol{f}) \\
&= \boldsymbol{g}_k - \boldsymbol{D}\boldsymbol{H}_k \boldsymbol{S}_k(\boldsymbol{\alpha}_k + \triangle \boldsymbol{\alpha}_k)(\boldsymbol{f} + \triangle \boldsymbol{f}) \\
&= \boldsymbol{g}_k - \boldsymbol{D}\boldsymbol{H}_k \boldsymbol{S}_k(\boldsymbol{\alpha}_k + \triangle \boldsymbol{\alpha}_k)\boldsymbol{f} - \boldsymbol{D}\boldsymbol{H}_k \boldsymbol{S}_k(\boldsymbol{\alpha}_k + \triangle \boldsymbol{\alpha}_k)\triangle \boldsymbol{f} \\
&= \boldsymbol{g}_k - \boldsymbol{D}\boldsymbol{H}_k \boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f} - \boldsymbol{D}\boldsymbol{H}_k \boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f})\triangle \boldsymbol{a}_k \\
&\quad - \boldsymbol{D}\boldsymbol{H}_k \boldsymbol{S}_k(\boldsymbol{\alpha}_k)\triangle \boldsymbol{f} - \boldsymbol{D}\boldsymbol{H}_k \boldsymbol{J}_k(\boldsymbol{\alpha}_k, \triangle \boldsymbol{f})\triangle \boldsymbol{\alpha}_k \\
&\simeq \boldsymbol{r}_k(\boldsymbol{\alpha}_k, \boldsymbol{f}) - \boldsymbol{D}\boldsymbol{H}_k \boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f})\triangle \boldsymbol{\alpha}_k - \boldsymbol{D}\boldsymbol{H}_k \boldsymbol{S}_k(\boldsymbol{\alpha}_k)\triangle \boldsymbol{f}
\end{aligned}
\tag{3.5}
$$

where $\boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f})$ represents the Jacobian of the *k*th shifted HR image $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}$ with respect to $\boldsymbol{\alpha}_k$. It should be noted that due to the adopted non-translational motion model, the motion operator $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)$ is no longer Toeplitz. Thus, the *k*th shifted HR image $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}$ is now nonlinear

with respect to motion parametric vector $\boldsymbol{\alpha}_k$. That is to say, we cannot find an equivalent matrix $\boldsymbol{X}$ which satisfies $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f} = \boldsymbol{X}\boldsymbol{\alpha}_k$. To simplify the explanation, the derivation of the Jacobian matrix $\boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f})$ will be given in the Appendix at the end of this chapter. We ignore the second-order term $\boldsymbol{D}\boldsymbol{H}_k\boldsymbol{J}_k(\boldsymbol{\alpha}_k, \triangle\boldsymbol{f})\triangle\boldsymbol{\alpha}_k$ in (3.5) since its value is generally negligible for small $\triangle\boldsymbol{f}$ and $\triangle\boldsymbol{\alpha}_k$.

The problem for directly estimating the motion parameter $\boldsymbol{\alpha}_k$ and the HR image $\boldsymbol{f}$ by minimizing (3.4) could be formulated into the minimization problem for the increment of motion parameter $\triangle\boldsymbol{\alpha}_k$ and the increment of HR image $\triangle\boldsymbol{f}$ as follows:

$$\min_{\triangle\boldsymbol{\alpha}_k, \triangle\boldsymbol{f}} \| \boldsymbol{P}\underline{\boldsymbol{f}} - \boldsymbol{b} \|_1$$

$$\text{where } \boldsymbol{P} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{W}_1 \\ \boldsymbol{J}_2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{W}_2 \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J}_N & \boldsymbol{W}_N \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \lambda\boldsymbol{L} \end{bmatrix}, \underline{\boldsymbol{f}} = \begin{bmatrix} \triangle\boldsymbol{\alpha}_2 \\ \vdots \\ \triangle\boldsymbol{\alpha}_N \\ \triangle\boldsymbol{f} \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} \boldsymbol{r}_1(\boldsymbol{f}) \\ \boldsymbol{r}_2(\boldsymbol{\alpha}_2, \boldsymbol{f}) \\ \vdots \\ \boldsymbol{r}_N(\boldsymbol{\alpha}_N, \boldsymbol{f}) \\ -\lambda\boldsymbol{L}\boldsymbol{f} \end{bmatrix}. \tag{3.6}$$

The proposed iterative framework for jointly performing image registration and HR reconstruction is shown in Figure 3.2. During the initialization, $\boldsymbol{\alpha}_k$ can be computed by employing an existing image registration method in [136], and the HR image $\boldsymbol{f}$ can be obtained by resizing the reference LR image using bicubic interpolation. The minimization problem in (3.6) can be considered as a linear programming (LP) problem. The details of the optimization will be explained in the following Section 3.3.3. The estimates of HR image and motion parameters are updated progressively until convergence is reached.

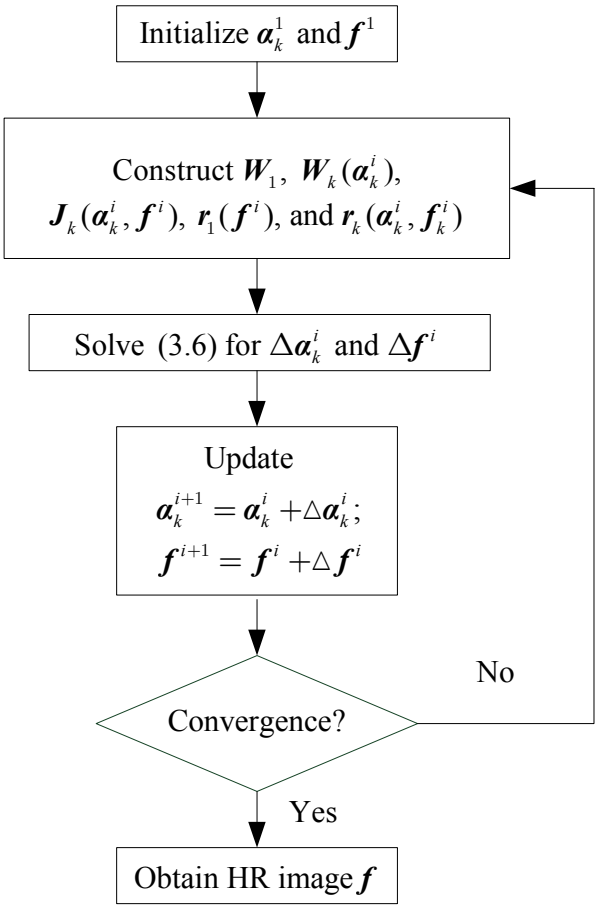


Figure 3.2 Schematic diagram of the proposed method

3.3.3 Optimization of the L_1-norm Function

It is noted that the optimization problem in (3.6) cannot be solved easily using methods such as the steepest descent method in [28]. Therefore, extending the idea in [32], we formulate the problem as a LP problem as follows. Let $\boldsymbol{u} = \boldsymbol{P}\underline{\boldsymbol{f}} - \boldsymbol{b}$, $\boldsymbol{u}^+ = \max(\boldsymbol{u}, \boldsymbol{0})$, and $\boldsymbol{u}^- = \max(-\boldsymbol{u}, \boldsymbol{0})$, where the operator max($\boldsymbol{m}$,$\boldsymbol{n}$) returns the larger elements between the vectors $\boldsymbol{m}$ and $\boldsymbol{n}$. Then $\boldsymbol{u}$ can be expressed as $\boldsymbol{u}^+ - \boldsymbol{u}^-$. The minimization problem can now be rewritten as

$$\min_{\Delta\boldsymbol{\alpha},\Delta\boldsymbol{f},\boldsymbol{u}^+,\boldsymbol{u}^-} \boldsymbol{1}^T\boldsymbol{u}^+ + \boldsymbol{1}^T\boldsymbol{u}^- \quad \text{subject to} \quad \begin{cases} \boldsymbol{P}\underline{\boldsymbol{f}} - \boldsymbol{b} = \boldsymbol{u}^+ - \boldsymbol{u}^-, \\ \boldsymbol{u}^+, \boldsymbol{u}^- \geq 0. \end{cases} \tag{3.7}$$

where $\boldsymbol{1}$ denotes the vector of all ones with appropriate size. The equation above can be written as a LP problem in a standard form as follows:

$$\min_{\boldsymbol{x}} \boldsymbol{q}^T\boldsymbol{x} \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \quad \text{where } \boldsymbol{q} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{1} \\ \boldsymbol{1} \end{bmatrix},\ \boldsymbol{x} = \begin{bmatrix} \underline{\boldsymbol{f}} \\ \boldsymbol{u}^+ \\ \boldsymbol{u}^- \end{bmatrix} \text{ and } \boldsymbol{A} = \begin{bmatrix} \boldsymbol{P} & -\boldsymbol{I} & \boldsymbol{I} \end{bmatrix}. \tag{3.8}$$

Let us denote $\boldsymbol{0}$ as the vector of all zeros with appropriate size. Then we can solve the LP problem using linear interior point solver (LIPSOL) [151], which is based on a primal-dual interior point method. After solving (3.8), we can use the solution $\underline{\boldsymbol{f}}$ to update $\triangle\boldsymbol{\alpha}_k$ and $\triangle\boldsymbol{f}$ as shown in Figure 3.2.

3.4 Experimental Results

In this section, we will illustrate the effectiveness of our proposed method using both simulated and real-life experiments. We compared the results reconstructed by our proposed method with the results by other three SR methods of two-stage SR approach in [28], AM SR method similar to [39] and our recently proposed L_2-norm method in [43].

To evaluate the performance of the methods, the objective metric of peak signal-to-noise ratio (PSNR) is employed for the estimated HR image $\hat{\boldsymbol{f}}$, which is defined as follows:

$$PSNR(\hat{\boldsymbol{f}}) \equiv 10\log_{10}\frac{255^2 M_f N_f}{\left\|\hat{\boldsymbol{f}}-\boldsymbol{f}\right\|^2} \tag{3.9}$$

where '^' denotes the current estimate of the HR image.

3.4.1 SR on LR Images with Various Additive Noise

In the simulated experiments, the HR reconstruction of a "Circuit board" image from multiple LR images at different noise levels and motion parameters were carried out. The simulation number was set to 15 at each noise level. The 90×90 "Circuit board" image in Figure 3.5 (a) was chosen for testing. Based on the imaging model in (3.1), 5 different LR images were generated with rotational angles randomly selected from a uniform distribution over [-10, 10] degrees and translations randomly selected from a uniform distribution over [-5, 5] pixels. Subsequently, the rotated-and-shifted HR images were blurred by a uniform kernel before down-sampled by a decimation factor of 2. The downsampled LR images were further degraded by additive white Gaussian noise (AWGN) to produce different noise levels of signal-to-noise ratios (SNRs) of 45, 35 and 15 dB. Note that a SNR of 15 dB for image SR can be considered as a very noisy environment. Finally, we added salt and pepper noise to 1% of the pixels in only one of the LR images shown in Figure 3.5 (b). The salt-and-pepper noise in this LR image can be considered as the outlier in the LR images. For all four methods, we calculated the initial motion parameters utilizing the existing registration algorithm in [136]. The regularization parameter λ is determined empirically based on 8 images, as shown in Figure 3.3. The proposed method is applied to these images with regularization parameter ranging from 10^{-6} to 0.1 and the average PSNR is given in Figure 3.4. It can be observed that the average PSNR increases from $\lambda=10^{-1}$ to $\lambda=10^{-2}$ and then decreases from $\lambda=10^{-2}$ to $\lambda=10^{-6}$ gradually. The best regularization parameter is

obtained at $\lambda=10^{-2}$. The regularization parameter obtained is used in the subsequent experiments, including the real-life experiment.

Figure 3.3 Image set: (a) “Circuit board”, (b) “Bridge”, (c) “Window”, (d) “Barbara”, (e) “RS-house”, (f) “Lighthouse”, (g) “Lena”, (h) “House”.

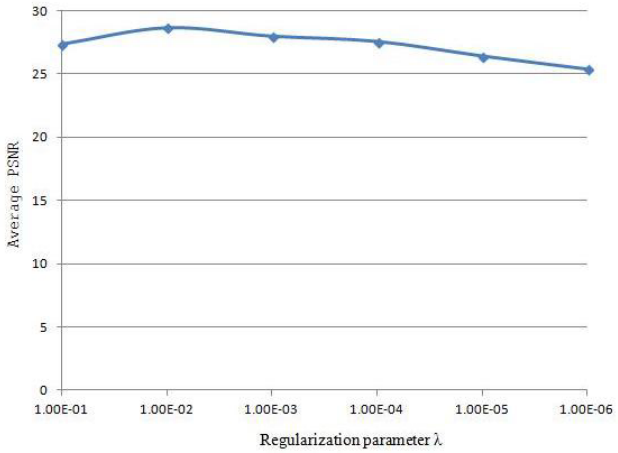


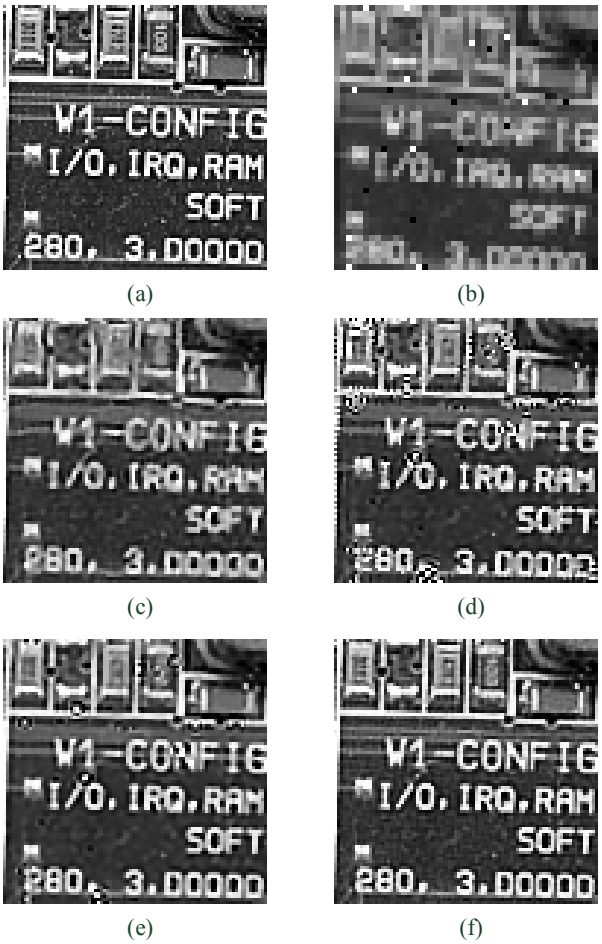
Figure 3.4 The graph of average PSNR against regularization parameter.

In our proposed method, we considered that the proposed algorithm had converged when the estimated HR image satisfies the convergence criterion $\| \boldsymbol{f}^i - \boldsymbol{f}^{i-1} \|^2 / \| \boldsymbol{f}^{i-1} \|^2 < 10^{-6}$. The average peak signal-to-noise ratios (PSNRs) of the reconstructed HR images for different SNR noise levels are given in Table 3.1. The PSNR obtained are based on average of 15 different runs. The reconstructed HR images of the experiment at 35 dB noise level using the four methods are given in Figure 3.5 (c)-(f).

It can be observed that the two-stage disjoint SR method shows unsatisfactory performance, as it does not improve upon the initial registration during the HR image reconstruction. Further, the results obtained using the AM and the L_2-norm methods, as shown in Figure 3.5 (d)-(e), both suffer from the outlier LR image that contains the salt-and-pepper noise. On the other hand, the proposed method can produce superior HR reconstruction when compared with the other three methods. This is because the proposed nonlinear SR method adopts a cost function consisting of L_1-norm data fidelity term and L_1-norm regularization term. The former is used to preserve the details of edges and textures, while the latter provide robustness with respect to outliers. The proposed method improves estimated motion parameters iteratively, which further improves the quality of the desired high-resolution (HR) image progressively. Therefore, the proposed method outperforms the two-stage disjoint L_1-norm SR method [28] and the L_2-norm based SR methods [39], [43], in particular when the LR images contain a few outliers. The visual inspection is consistent with the objective performance metric given in Table 3.1 which shows that the proposed method offers the best result. However, the proposed method tends to produce noisy HR results when the LR images contain large additive noise. This is because the L_1-norm

Table 3.1 Average PSNR of the reconstructed HR image against the SNR of the LR images.

SNR	Two-stage SR method [28] (dB)	AM SR method [39] (dB)	L_2 norm method [43] (dB)	Proposed SR method (dB)
45 dB	16.37	17.67	19.35	27.54
35 dB	16.31	17.60	19.07	24.85
15 dB	15.58	17.11	18.06	18.36



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

Figure 3.5 SR for LR images degraded by AWGN and salt & peppers noise. (a) Original HR "Board" image, (b) The scaled-up version of LR image degraded by salt & peppers noise, (c) HR result by the two-stage SR method [28], (d) HR result by the AM method [39], (e) HR result by the L_2 norm method [43], (f) HR result by the proposed method.

approaches are more sensitive to noise when compared with L_2-norm approaches.

To further demonstrate the performance of the proposed method, we employed the remote sensing "Aerial" image in Figure 3.6(a) as the test image. The experimental setting was similar to that in the previous experiment. Firstly, 5 LR images were generated with different rotational angles of (0, 9.42, 7.68, -6.59, -2.75) degrees and translations of ((0, 0), (3.76, 2.37), (-3.64, -4.89), (3.94, -3.01), (-2.01, 1.61)) pixels. Next, the rotated-and-shifted HR images were blurred by a 3×3 uniform kernel before down-sampled by a decimation factor of 2. The downsampled LR images were further degraded by AWGN to produce SNR of 35dB. Finally, we added salt and pepper noise to 1% of the pixels in only one of the LR images shown in Figure 3.6(b). The proposed method along with the three other SR methods of two-stage SR approach [28], AM method [39] and L_2-norm method [43] were performed on the LR images. The reconstructed HR images using all the four methods are shown in Figure 3.6(c)-(f). It can be observed that the proposed method can produce better HR result which contains less artifacts, when compared with the other methods.

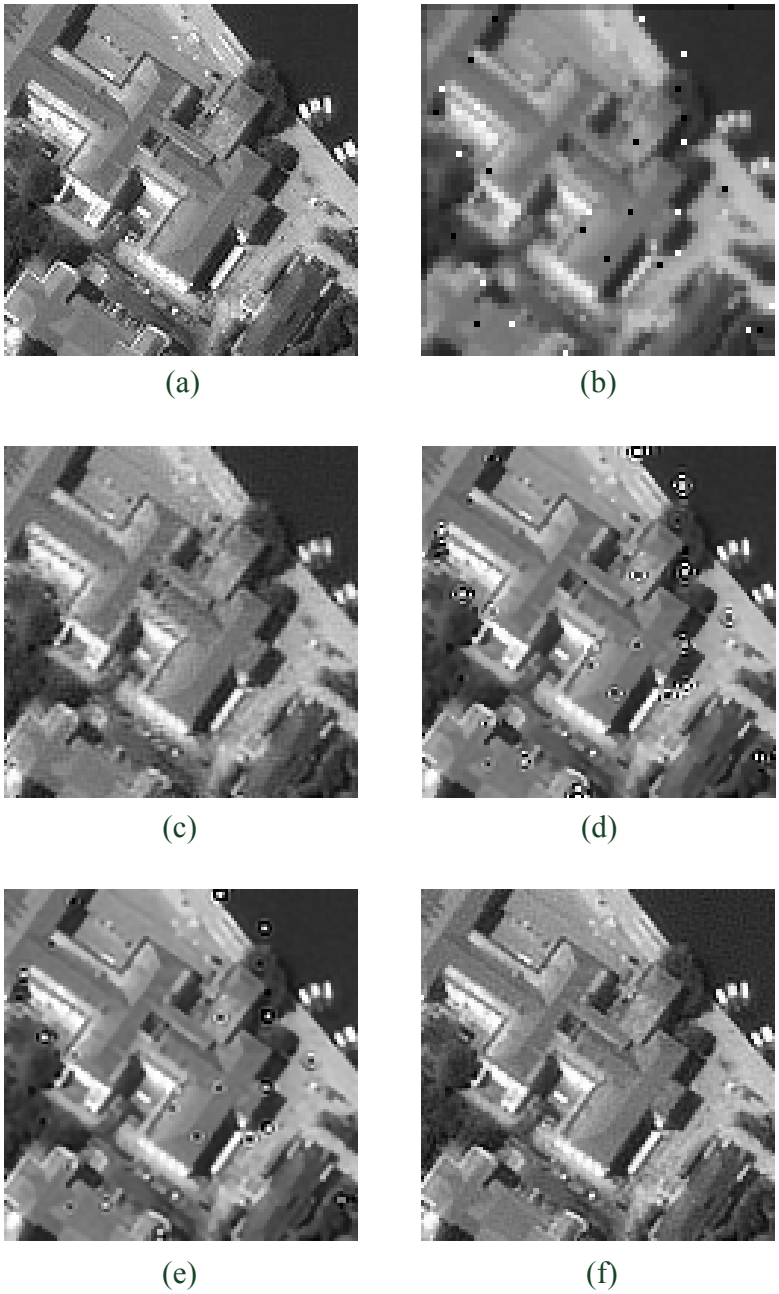


Figure 3.6 SR for LR images degraded by AWGN and salt & peppers noise. (a) Original HR "Aerial" image, (b) The scaled-up version of LR image degraded by salt & peppers noise, (c) HR result by the two-stage SR method [28], (d) HR result by the AM method [39], (e) HR result by the L_2 norm method [43], (f) HR result by the proposed method.

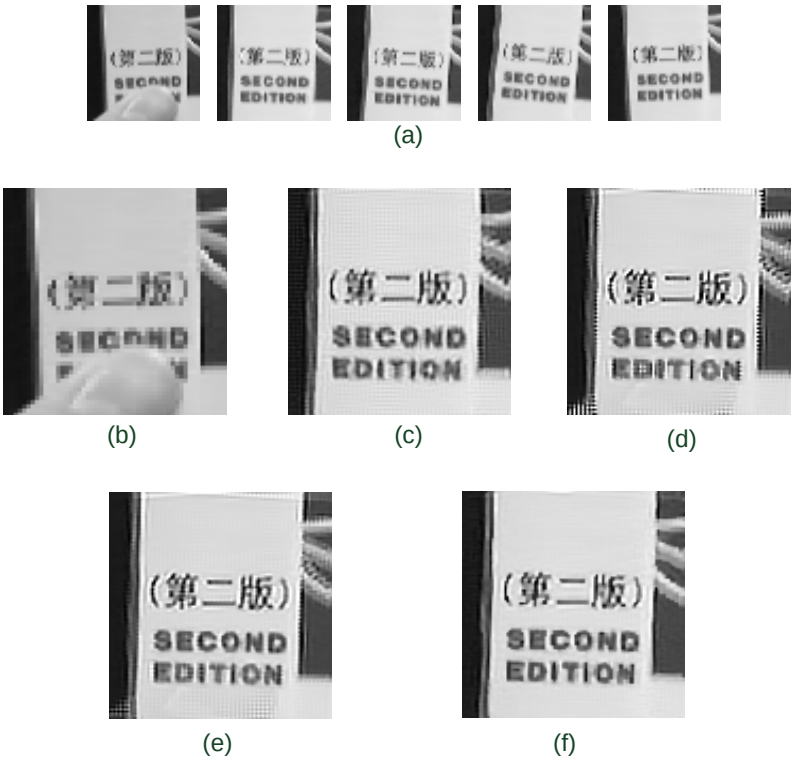


Figure 3.7 SR for real-life "book" images. (a) 5 captured LR images, (b) The scaled-up version of LR image partially obscured by a finger, (c) HR result by the two-stage SR method [28], (d) HR result by the AM method [39], (e) HR result by the L_2 norm method [43], (f) HR result by the proposed method.

3.4.2 Experiments on Real-life Images

We also conducted a real-life experiment. The five "book" images were captured by a web-camcorder with relative translational and rotational motion, which are shown in Figure 3.7(a). It is noted that one of the LR images is partially obscured by a finger. This is considered as an outlier in the LR images. We set the size of the reconstructed HR image to be twice of the LR image. The results obtained using all methods are shown in Figure 3.7(c)-(f). It can be seen that other comparative methods can produce comparable HR results in regions without outlier (namely the finger region). However, the proposed method is superior to other methods in the

region that contains the outlier, i.e. the "EDITION" text region. It is noted that the HR results by other comparative methods suffer from the artifacts due to the outlier. Comparison reveals that the approach is superior in handling real-life image SR, especially in the presence of outlier.

Finally, we will employ computational time of the experiment in Figure 3.5 to show the proposed method's complexity. The L_2 norm method in [43] is used as the benchmark for comparison. The simulation environments of all the experiments above are given as follows: CPU Core Q9400-2.67GHz, 3.25 GB RAM, Windows XP OS (operation system) and MATLAB R2010b. The total computational time is about 96.1s for our proposed algorithm while 23.5s for the L_2-norm algorithm. Even though the proposed method is computationally more expensive, the improvement in the quality of the reconstructed HR image, nevertheless, justifies this additional cost.

3.5 Summary

This chapter presents an effective nonlinear algorithm for joint image registration as well as image SR reconstruction. The method involves development of an effective L_1 norm cost function, formulation of the SR problem into a joint optimization problem, and solving it using a LP framework. The algorithm has the capability of performing image registration and HR image reconstruction simultaneously along with improving them progressively. Experimental results show that the proposed algorithm can achieve good HR reconstruction with the existence of outliers in the LR images.

3.6 Appendix

DERIVATION OF THE JACOBIAN MATRIX $\boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f})$

In this appendix, the derivation of the Jacobian matrix $\boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f})$ will be presented. $\boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f})$ can be expressed as follows:

$$\boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f}) = \frac{\partial(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f})}{\partial \boldsymbol{\alpha}_k}. \tag{3.10}$$

Similar to the discussion in our previous work [43], the motion operator $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)$ is no longer Toeplitz. Thus, it is difficult to obtain $\frac{\partial(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f})}{\partial \boldsymbol{\alpha}_k}$ directly due to the fact that the rotated-and-shifted HR image $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}$ is nonlinear with respect to motion parametric vector $\boldsymbol{\alpha}_k$. To tackle this problem, bilinear interpolation is used to derive the underlying relationship between $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}$ and $\boldsymbol{\alpha}_k$.

Firstly, we will explain the formation of the rotated-and-shifted HR images $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}$ from the desired HR image $\boldsymbol{f}$. Let us assume (x_1, y_1) and (x_k, y_k) represent the coordinates of the reference HR grid and the *k*th rotated-and-shifted HR grid, respectively. They are denoted by solid circles in Figure 3.8. The change of coordinate $(\triangle x_k, \triangle y_k)$ is defined as

$$\begin{aligned} \triangle x_k &= x_k - x_1 = x_1 \cos\theta_k - y_1 \sin\theta_k + s_{xk} - x_1; \\ \triangle y_k &= y_k - y_1 = x_1 \sin\theta_k + y_1 \cos\theta_k + s_{yk} - y_1. \end{aligned} \tag{3.11}$$

It is noted that there are four kinds of relative positions between the coordinates (x_1, y_1) and (x_k, y_k), as shown in Figure 3.8. From the figure, we can see the pixel (x_k, y_k) in the *k*th

rotated-and-shifted HR grid as well as its four surrounding pixels, including bottom-right f_{br}, bottom-left f_{bl}, top-right f_{tr} and top-left f_{tl}, in the reference HR grid. Rewriting (3.11) in vector-matrix form, we have

$$\begin{bmatrix}\triangle \boldsymbol{x}_k \\ \triangle \boldsymbol{y}_k\end{bmatrix} = \begin{bmatrix}\boldsymbol{x}_k \\ \boldsymbol{y}_k\end{bmatrix} - \begin{bmatrix}\boldsymbol{x}_1 \\ \boldsymbol{y}_1\end{bmatrix} = \boldsymbol{C}\boldsymbol{\alpha}_k - \begin{bmatrix}\boldsymbol{x}_1 \\ \boldsymbol{y}_1\end{bmatrix}$$
$$\text{where } \boldsymbol{C} = \begin{bmatrix}\boldsymbol{x}_1 & -\boldsymbol{y}_1 & \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{y}_1 & \boldsymbol{x}_1 & \boldsymbol{0} & \boldsymbol{1}\end{bmatrix} \textit{ and } \boldsymbol{\alpha}_k = [\cos\theta_k \ \sin\theta_k \ s_{xk} \ s_{yk}]^T. \tag{3.12}$$

$\triangle \boldsymbol{x}_k$, $\triangle \boldsymbol{y}_k$, $\boldsymbol{x}_k$, $\boldsymbol{y}_k$, $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$ are the vectors which represents the discrete and lexicographically ordered displacements of $\triangle x_k$, $\triangle y_k$, x_k, y_k, x_1 and y_1, respectively. ***1*** and ***0*** represent vectors consisting of all ones and zeros, respectively. From (3.12), it is noted that $\left[\triangle \boldsymbol{x}_k^T, \triangle \boldsymbol{y}_k^T\right]^T$ is linear with respect to $\boldsymbol{\alpha}_k$. $\partial\left(\left[\triangle \boldsymbol{x}_k^T, \triangle \boldsymbol{y}_k^T\right]^T\right) \Big/ \partial \boldsymbol{\alpha}_k$ can be obtained as

$$\frac{\partial\left(\left[\triangle \boldsymbol{x}_k^T, \triangle \boldsymbol{y}_k^T\right]^T\right)}{\partial \boldsymbol{\alpha}_k} = \frac{\partial\left(\boldsymbol{C}\boldsymbol{\alpha}_k\right)}{\partial \boldsymbol{\alpha}_k} = \boldsymbol{C}. \tag{3.13}$$

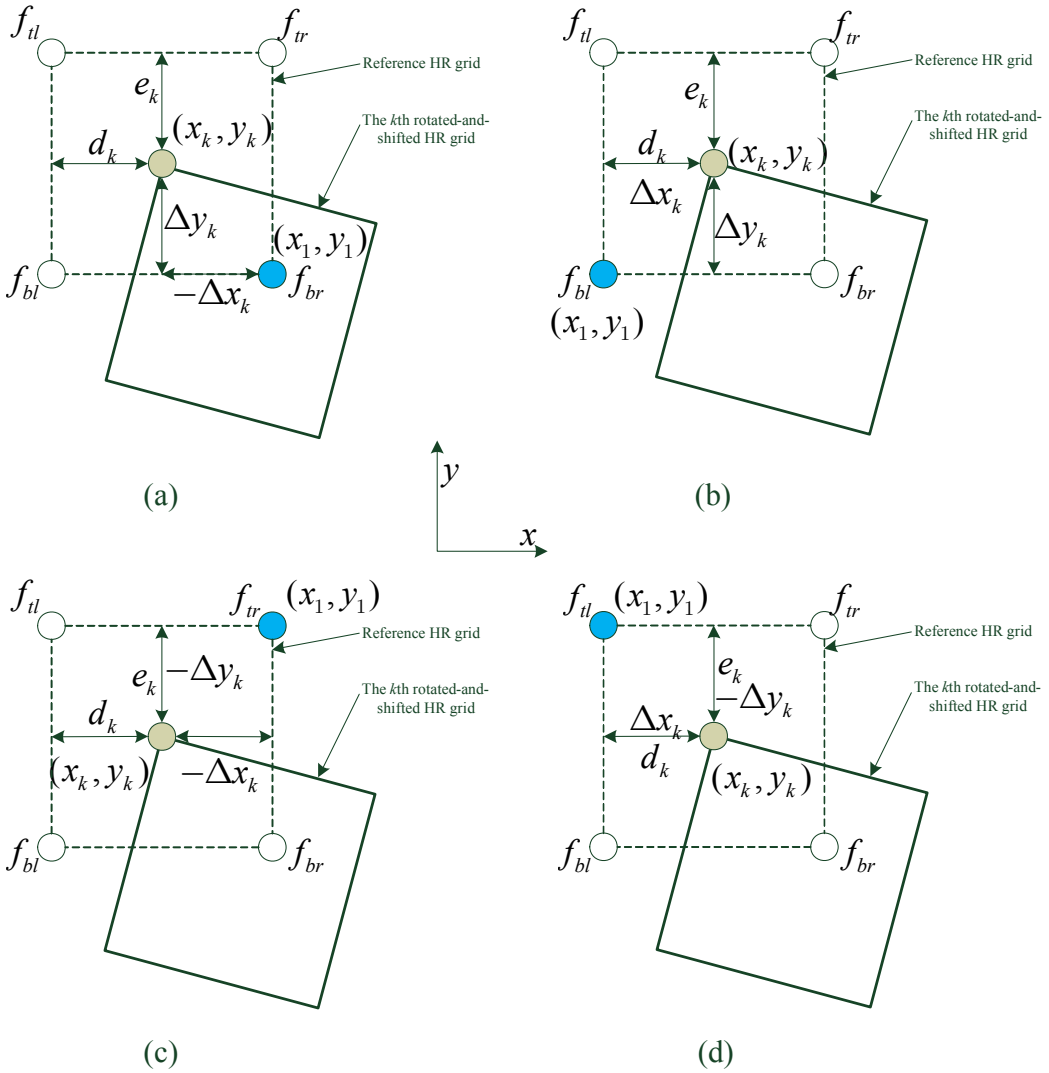


Figure 3.8 Four possible relative positions between (x_k, y_k) and (x_1, y_1)

Subsequently, we need to calculate $\partial\left(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)\Big/\partial\left(\left[\triangle\boldsymbol{x}_k^T,\triangle\boldsymbol{y}_k^T\right]^T\right)$. In the kth HR grid, the value of pixel (x_k, y_k) can be determined by four neighboring pixel values f_{br}, f_{bl}, f_{tr} and f_{tl}, which is shown in Figure 3.8. The four points are denoted as the neighboring reference HR pixels which surround the rotated-and-shifted pixel (x_k, y_k) at the locations of bottom-right, bottom-left, top-right and top-left, respectively. We denote $[d_k, e_k]^T$ as the distance vector

between the pixel (x_k, y_k) and the corresponding top-left reference HR pixel. Considering the four kinds of coordinate positions in Figure 3.8(a)-(d), d_k and e_k can be seen to satisfy the following relationships

$$(a)-(d):\ \begin{cases} d_k = \triangle x_k - \text{floor}(\triangle x_k); \\ e_k = \triangle y_k - \text{floor}(\triangle y_k) \end{cases} \tag{3.14}$$

where floor(•) represents the rounding functional which outputs the nearest integer not larger than the input number. Utilizing bilinear interpolation, the rotated-and-shifted HR image $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}$ can be calculated by

$$\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f} = \boldsymbol{d}_k \odot (\boldsymbol{1}-\boldsymbol{e}_k) \odot \boldsymbol{f}_{bl} + \boldsymbol{d}_k \odot \boldsymbol{e}_k \odot \boldsymbol{f}_{br} + (\boldsymbol{1}-\boldsymbol{d}_k) \odot (\boldsymbol{1}-\boldsymbol{e}_k) \odot \boldsymbol{f}_{tl} + (\boldsymbol{1}-\boldsymbol{d}_k) \odot \boldsymbol{e}_k \odot \boldsymbol{f}_{tr} \tag{3.15}$$

where $\odot$ denotes an operator representing entry-by-entry multiplication. Vectors $\boldsymbol{f}_{tl}$, $\boldsymbol{f}_{tr}$, $\boldsymbol{f}_{bl}$, $\boldsymbol{f}_{br}$, $\boldsymbol{d}_k$ and $\boldsymbol{e}_k$ represent the lexicographically-ordered displacements of $f_{tl}, f_{tr}, f_{bl}, f_{br}, d_k$ and e_k, respectively. We will use $\boldsymbol{E}_k$ to denote $\partial\left(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}\right) \Big/ \partial\left(\left[\triangle \boldsymbol{x}_k^T, \triangle \boldsymbol{y}_k^T\right]^T\right)$ for simplicity. Combining (3.14) and (3.15), $\boldsymbol{E}_k$ can then be expressed as

$$\begin{aligned} \boldsymbol{E}_k &= \frac{\partial\left(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial\left(\left[\boldsymbol{d}_k^T, \boldsymbol{e}_k^T\right]^T\right)} \frac{\partial\left(\left[\boldsymbol{d}_k^T, \boldsymbol{e}_k^T\right]^T\right)}{\partial\left(\left[\triangle \boldsymbol{x}_k^T, \triangle \boldsymbol{y}_k^T\right]^T\right)} \\ &= \left[diag\left\{(\boldsymbol{1}-\boldsymbol{e}_k) \odot (\boldsymbol{f}_{bl}-\boldsymbol{f}_{tl}) + \boldsymbol{e}_k \odot (\boldsymbol{f}_{br}-\boldsymbol{f}_{tr})\right\}, diag\left\{(\boldsymbol{1}-\boldsymbol{d}_k) \odot (\boldsymbol{f}_{tr}-\boldsymbol{f}_{tl}) + \boldsymbol{d}_k \odot (\boldsymbol{f}_{br}-\boldsymbol{f}_{bl})\right\} \right] \end{aligned} \tag{3.16}$$

as $\partial\left(\left[\boldsymbol{d}_k^T, \boldsymbol{e}_k^T\right]^T\right) \Big/ \partial\left(\left[\triangle \boldsymbol{x}_k^T, \triangle \boldsymbol{y}_k^T\right]^T\right)$ is equal to identity matrix based on (3.14). Similarly, we can obtain the same equation (3.16) for all the four cases in Figure 3.8(a)-(d). Therefore, substituting (3.13) and (3.16) into(3.10), $\boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f})$ can then be obtained as

$$\boldsymbol{J}_k(\boldsymbol{\alpha}_k, \boldsymbol{f}) = \frac{\partial(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f})}{\partial \boldsymbol{\alpha}_k} = \frac{\partial\left(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial\left(\left[\triangle \boldsymbol{x}_k^T, \triangle \boldsymbol{y}_k^T\right]^T\right)} \frac{\partial\left(\left[\triangle \boldsymbol{x}_k^T, \triangle \boldsymbol{y}_k^T\right]^T\right)}{\partial \boldsymbol{\alpha}_k} = \boldsymbol{E}_k \boldsymbol{C} \,. \tag{3.17}$$

Chapter 4

Joint Image Registration and Super-resolution from Low-resolution Images with Zooming Motion

4.1 Introduction

This chapter focuses on the multi-frame SR reconstruction problem involving a motion model consisting of relative zooming. Conventional super resolution (SR) methods [62], [135], [107] typically adopt motion model consisting of only translation and possibly rotation. This may constrain their applications when relative zooming exists among the acquired LR images. Further, many existing super-resolution (SR) methods [120], [101], [44], [131], [154], [146], [115], [73] typically formulate the SR problem as a two-stage process, namely 1) image registration based on a specific motion model to calculate the motion parameters, followed by 2) HR reconstruction that incorporates the estimated motion parameters into an inverse estimation. Generally, accurate registration is a challenging task for image SR reconstruction as the motion parameters are calculated from the captured LR images. The aliasing effect among the LR images may reduce the accuracy of registration. Therefore, various SR algorithms have been proposed to alleviate the effect of registration error on the final estimated HR image.

A number of SR methods [28], [156], [67], [40] employ various techniques to improve the HR estimation with the existing registration error. In [28], Farsiu *et al*. proposed a L_1-norm SR algorithm which is able to suppress the artifacts due to the registration error. He *et al*. [40] proposed an approach that considers inaccurate estimates of motion parameters and model the registration error in the HR reconstruction. Other works try to tackle the problem by improving the estimation of motion parameters [39], [106], [140]. In [39], Hardie *et al*. proposed a maximum *a posteriori* (MAP) based framework to calculate the HR image and motion parameters iteratively. In their work, alternating minimization (AM) technique is adopted to solve the cost function with respect to the HR image and motion parameters. In [106], Robinson *et al*. proposed a variable projection (VP) based SR method and further developed an efficient Fourier domain implementation of the motion estimation for global translational motion model. Typically, these SR approaches consider registration and HR reconstruction as two disjoint processes. Moreover, they are designed under the assumption that all LR images have the identical spatial resolution. This may reduce the potential SR improvement for such cases when there is relative zooming motion between LR images, as illustrated in Figure 4.1. It is difficult to estimate the accurate registration parameters of the captured LR images by using a two-dimensional (2-D) in-plane motion model. Hence, this motivates the study of techniques that can be used to deal with zooming motion between the captured images, including optical flow techniques [115] and parametric motion models [51], [69], [52], [85].

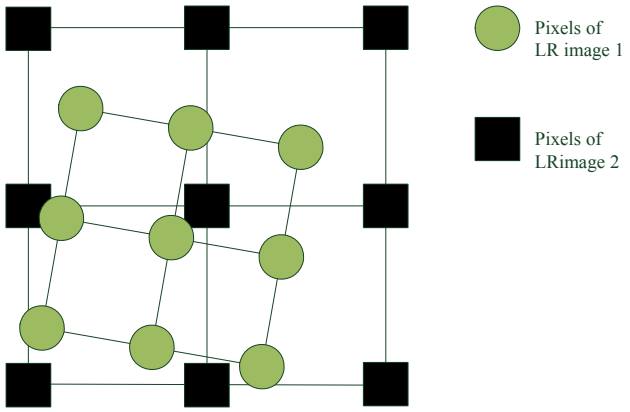


Figure 4.1 Illustration of the relative zooming motion between two LR images

Amongst the existing zooming SR literatures, Li proposed a SR approach for synthetic zooming [69]. In his work, all the LR images are related to each other by employing a line-geometry model and the zooming factors are estimated based on this model. Joshi *et al*. [51] presented a zooming SR algorithm using both the Markov random field (MRF) as well as a simultaneous autoregressive (SAR) model under the assumption that the SR field is homogenous. Similar to [52], the HR image is estimated from multiple LR images of the same scene with continuous zooming. Recently, a zooming-based SR approach with total variation (TV) prior has been proposed in [85]. Generally, these methods can be considered as the two-stage disjoint SR methods as discussed previously.

In view of this, we propose an iterative method for joint registration and SR reconstruction based on the captured LR observations with relative zooming motion. The proposed algorithm integrates zooming SR reconstruction along with image registration into a single process of estimation. An iterative technique serves to ensure that progressively better estimates of motion parameters and HR image can be obtained. This is more promising when compared with other two-stage SR methods as the estimation of motion parameters could benefit from the information

of the reconstructed HR image. Different from conventional methods, such as [43], which focus mainly on translation and possibly rotation, the proposed method takes into account the relative zooming between the acquired LR images. The adopted motion model enables the proposed method to handle more practical applications involving motion consisting of translation, rotation and zooming. Further, the proposed method also develops an adaptive weighting scheme and incorporates it into the HR reconstruction. As the captured LR images provide different degree of information to the HR reconstruction, the adopted weighting strategy takes this into consideration.

The rest of this chapter is organized as follows. The problem formulation of image SR is presented in Section 4.2. The proposed cost function is introduced in Section 4.3. An iterative algorithm using the regularized nonlinear least squares (NLS) technique is developed in Section 4.4. Experimental results on simulated and real-life images are presented and discussed in Section 4.5. A brief conclusion is given in Section 4.6.

4.2 Problem Formulation

First of all, let us consider the generation of LR images under the motion model consisting of zooming, translation and rotation. Similar to the definitions in the previous chapter, $f(x, y)$ and $g_k(u, v)$ are assumed to represent the underlying HR image and the kth observed LR image ($1 \leq k \leq N$), respectively. During the image acquisition process, the LR image $g_k(u, v)$ of size $M_g \times N_g$ is obtained by zooming the original HR image $f(x, y)$ of size $M_f \times N_f$ by a zooming factor l_k, rotating it by an angle θ_k and shifting it by a translational vector $[s_{xk}, s_{yk}]$, then followed by blurring the zoomed-and-rotated-and-shifted HR image $f_k(x, y)$ by a point-spread

function (PSF) $h_k \otimes h_c$ and down-sampling the result by a factor ρ. Finally, the LR image is degraded by additive white Gaussian noise (AWGN) $n(u,v)$. The process can be expressed as [43], [42]

$$\begin{aligned} g_k(u,v) &= \left(f(l_k x\cos\theta_k - l_k y\sin\theta_k + s_{xk}, l_k x\sin\theta_k + l_k y\cos\theta_k + s_{yk}) \otimes h_k \otimes h_c \right)\downarrow_\rho + n(u,v) \\ &= \left(f_k(x,y) \otimes h_k \otimes h_c \right)\downarrow_\rho + n(u,v) \end{aligned} \tag{4.1}$$

where $\otimes$ represents the 2D convolution operator. $\downarrow_\rho$ denotes the down-sampling operator with a decimation factor ρ, where ρ is employed to represent the resolution ratio between the desired HR image and the acquired LR images. In this work, we denote the identical decimation factor for both vertical and horizontal directions. h_k represents the blur operator related to camera lens while h_c represents the blur operator due to a square-region integration of light intensity.

As in our work, we assume the lens condition (h_k) to be known and identical for all LR images. Then using a matrix-vector form, the imaging process in (4.1) can be modeled as

$$\boldsymbol{g} = \boldsymbol{DHS}(\boldsymbol{\alpha})\boldsymbol{f} + \boldsymbol{n} \tag{4.2}$$

where $\boldsymbol{f}$ represents the lexicographically ordered original image $f(x,y)$. The vectors $\boldsymbol{g} = [\boldsymbol{g}_1^T, ..., \boldsymbol{g}_N^T]^T$ and $\boldsymbol{n} = [\boldsymbol{n}_1^T, ..., \boldsymbol{n}_N^T]^T$ represent the discrete, concatenated and lexicographically ordered displacements of $g_k(u,v)$ and $n_k(u,v)$, respectively. The matrix $\boldsymbol{D}$ denotes the identical down-sampling operator for all LR images. The matrix $\boldsymbol{H} = [\boldsymbol{H}_1^T, ..., \boldsymbol{H}_N^T]^T$ and $\boldsymbol{H}_k (1 \le k \le N)$ denotes the corresponding matrix representing the blurring operator $h_k \otimes h_c$. The matrix $\boldsymbol{S}(\boldsymbol{\alpha})$ is formed by nonlinear, differentiable functions with respect to the unknown

motion parameter $\boldsymbol{a}$, where $\boldsymbol{a} = [\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \cdots, \boldsymbol{a}_N^T]^T$. Hence, $\boldsymbol{S}(\boldsymbol{a}) = [\boldsymbol{S}(\boldsymbol{a}_1)^T, \boldsymbol{S}(\boldsymbol{a}_2)^T, \cdots, \boldsymbol{S}(\boldsymbol{a}_N)^T]^T$, where matrix $\boldsymbol{S}(\boldsymbol{\alpha}_k)\,(1 \le k \le N)$ denotes the operator of relative motion for the kth LR image. The vector $\boldsymbol{\alpha}_k$ represents the unknown motion parameter of the kth LR image and $\boldsymbol{\alpha}_k = [l_k \cos\theta_k,\ l_k \sin\theta_k,\ s_{xk},\ s_{yk}]^T$. In this work, we consider the initial estimates of motion parameters $\boldsymbol{\alpha}_k$ to contain certain degrees of error. This assumption is more realistic, since accurate estimate of motion parameters for SR reconstruction is hard to calculate, particularly in the early stage. In this chapter, the purpose of SR reconstruction is to fuse N LR observations, whose motion parameters $\boldsymbol{\alpha}$ are unknown, to produce the HR image $\boldsymbol{f}$ of the original scene. Because the first LR image is denoted as the reference, $4\times(N-1)$ motion parameters along with the desired HR image are required to be estimated.

4.3 Proposed Cost Function

The main ideas of the proposed joint image registration and zooming SR include integration of registration and SR into an iterative process, fusion of multiple LR images with different resolutions, and incorporation of useful *a prior* information for the desired HR image. Joint image registration and zooming SR is an ill-posed inverse problem. Thus, regularization for the HR image is considered in order to achieve a stable solution.

Given the imaging model in (4.1), we propose the following cost function to estimate the desired HR image $\boldsymbol{f}$ along with the unknown motion parametric vector $\boldsymbol{\alpha}$:

$$E(\boldsymbol{\alpha}, \boldsymbol{f}) = \left\| \boldsymbol{V}(\boldsymbol{g} - \boldsymbol{DHS}(\boldsymbol{\alpha})\boldsymbol{f}) \right\|^2 + \lambda \boldsymbol{T}(\boldsymbol{f}) \tag{4.3}$$

where $\boldsymbol{g}=[\boldsymbol{g}_1^T,\cdots,\boldsymbol{g}_N^T]^T$, $\boldsymbol{H}=[\boldsymbol{H}_1^T,\cdots,\boldsymbol{H}_N^T]^T$ and $\boldsymbol{S}(\boldsymbol{\alpha})=\left[\boldsymbol{S}(\boldsymbol{\alpha}_1)^T,\boldsymbol{S}(\boldsymbol{\alpha}_2)^T,\cdots,\boldsymbol{S}(\boldsymbol{\alpha}_N)^T\right]^T$. The matrix $\boldsymbol{S}(\boldsymbol{\alpha}_k)$ represents the motion operator consisting of zooming, rotation and translation. The first term in (4.3) is the data fidelity term. $\boldsymbol{V}=diag\{[\boldsymbol{V}_1,\boldsymbol{V}_2,\cdots,\boldsymbol{V}_N]\}$ represents the channel weights and $\boldsymbol{V}_k=V_k\boldsymbol{I}$ ($1\le k\le N$) is the indicator of reliability for the kth channel (the kth LR image). The second term in (4.3) is regularization functional which introduces stability into the estimates of HR image $\boldsymbol{f}$. λ is regularization parameter that serves to control the relative contribution between the data fidelity term and the regularization term. For HR image $\boldsymbol{f}$, total variation (TV) technique is adopted to constraint its estimate

$$\boldsymbol{T}(\boldsymbol{f})=\int_{\Omega}|\nabla\boldsymbol{f}|dxdy\,. \tag{4.4}$$

It is noted that the nonlinear partial equations (PDEs) is required to solve during the minimization of (4.3), due to the incorporation of TV norm $\boldsymbol{T}(\boldsymbol{f})$. To alleviate this problem, we develop a L_2 norm approximation for $\boldsymbol{T}(\boldsymbol{f})$,

$$\boldsymbol{T}(\boldsymbol{f})=\int_{\Omega}\frac{\|\nabla\boldsymbol{f}\|^2}{|\nabla\boldsymbol{f}|}dxdy=\sum_{m=-1}^{1}\sum_{n=0}^{1}\sum_{x,y}\frac{(f(x+m,y+n)-f(x,y))^2}{\mu_{m,n}} \tag{4.5}$$

where $\mu_{m,n}$ is defined as $\mu_{m,n}=\sqrt{m^2+n^2}\sqrt{(f(x+m,y+n)-f(x,y))^2+\gamma}$. Here, $\gamma>0$ is the regularization constant [120]. The fixed-point technique is adopted to calculate the coefficient $\mu_{m,n}$ employing the HR image $\boldsymbol{f}$ estimated in the previous iteration. Therefore, the right-hand side of (4.5) can be expressed in a matrix-vector form as

$$\begin{aligned}&\sum_{m=-1}^{1}\sum_{n=0}^{1}\sum_{x,y}\frac{(f(x+m,y+n)-f(x,y))^2}{\mu_{m,n}}\\&=\sum_{m=-1}^{1}\sum_{n=0}^{1}\boldsymbol{f}^T\boldsymbol{U}_{m,n}^T\boldsymbol{W}_{m,n}\boldsymbol{U}_{m,n}\boldsymbol{f}\\&=\boldsymbol{f}^T\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{f}=\|\boldsymbol{L}\boldsymbol{f}\|^2\end{aligned} \tag{4.6}$$

where matrix $\boldsymbol{W}_{m,n}=diag\{\boldsymbol{c}\}$ and vector $\boldsymbol{c}$ represents the lexicographically ordered $1/\mu_{m,n}$. $\boldsymbol{U}_{m,n}$ represents the first-order derivative operator on the desirable HR image. Therefore, the TV norm $\boldsymbol{T}(\boldsymbol{f})$ can be constructed as $\|\boldsymbol{L}\boldsymbol{f}\|^2$ using half-quadratic scheme. During the minimization process in Section 4.4.3, we only need to construct $\boldsymbol{L}^T\boldsymbol{L}$, which can be calculated by $\sum_{m=-1}^{1}\sum_{n=0}^{1}\boldsymbol{U}_{m,n}^T\boldsymbol{W}_{m,n}\boldsymbol{U}_{m,n}$.

In the proposed method, registration as well as zooming SR are incorporated into a single optimization procedure considering the estimation of zooming, rotation and translation parameters. This could enhance the SR performance in applications where relative zooming exists between the captured LR images.

4.4 Iterative Joint Registration and Zooming Super-resolution

4.4.1 Development of the Proposed Joint Registration and SR Framework using Nonlinear Least Square Technique

We will develop a new framework for simultaneous image registration and zooming SR in this section. Typically, image registration in existing zooming SR approaches is performed on the captured LR images. However, the subsequent HR reconstruction in these approaches relies on the initially estimated motion parameters heavily. To alleviate this dependency, we propose an

iterative method to jointly perform image registration and image SR progressively. The minimization problem (4.3) can be rewritten as

$$\min_{\boldsymbol{\alpha},\boldsymbol{f}} \left\| \begin{matrix} \boldsymbol{V}\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f}) \\ \sqrt{\lambda}\boldsymbol{L}\boldsymbol{f} \end{matrix} \right\|^2 \tag{4.7}$$

where $\|\bullet\|$ denotes the L_2-norm and $\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f}) = \boldsymbol{g} - \boldsymbol{DHS}(\boldsymbol{\alpha})\boldsymbol{f}$ represents the fidelity residual vector.

As $\boldsymbol{\alpha}$ consists of rotation and zooming parameters, the residual vector $\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})$ is nonlinear with respect to motion parametric vector $\boldsymbol{\alpha}$. Therefore, to solve the minimization problem in (4.7), we extend the nonlinear parametric estimation (NPE) technique to derive a linear approximation for residual vector $\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})$. Let $\Delta\boldsymbol{\alpha}$ and $\Delta\boldsymbol{f}$ denote small changes in the unknown motion vector $\boldsymbol{\alpha}$ and the desired HR image $\boldsymbol{f}$, respectively. The residual vector $\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})$ can be linearized with respect to $\Delta\boldsymbol{\alpha}$ and $\Delta\boldsymbol{f}$ as follows:

$$\begin{aligned} \boldsymbol{r}(\boldsymbol{\alpha}+\Delta\boldsymbol{\alpha},\boldsymbol{f}+\Delta\boldsymbol{f}) &= \boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f}) + \nabla_{\alpha}\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})\Delta\boldsymbol{\alpha} + \nabla_{f}\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})\Delta\boldsymbol{f} + O(\Delta\boldsymbol{\alpha},\Delta\boldsymbol{f}) \\ &\simeq \boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f}) + \nabla_{\alpha}\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})\Delta\boldsymbol{\alpha} + \nabla_{f}\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})\Delta\boldsymbol{f} \end{aligned} \tag{4.8}$$

where $\nabla_{\alpha}\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})$ and $\nabla_{f}\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})$ represent the gradient matrix of residual vector $\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})$ with respect to vectors $\boldsymbol{\alpha}$ and $\boldsymbol{f}$, respectively. $O(\Delta\boldsymbol{\alpha},\Delta\boldsymbol{f})$ is the higher-order term that is ignored. The approximation in (4.8) is used to transform a nonlinear problem into a linear problem, with the assumption that the estimates of $\Delta\boldsymbol{\alpha}$ and $\Delta\boldsymbol{f}$ are sufficiently small.

The gradient, $\nabla_{\alpha}\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})$ and $\nabla_{f}\boldsymbol{r}(\boldsymbol{\alpha},\boldsymbol{f})$ can be expressed as

$$\begin{aligned}\nabla_{\boldsymbol{\alpha}} \boldsymbol{r}(\boldsymbol{\alpha}, \boldsymbol{f}) &= \frac{\partial\left(\boldsymbol{g} - \boldsymbol{DHS}(\boldsymbol{\alpha})\boldsymbol{f}\right)}{\partial \boldsymbol{\alpha}^T} \\ &= -\boldsymbol{DH}\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\right)}{\partial \boldsymbol{\alpha}^T},\end{aligned} \tag{4.9}$$

$$\begin{aligned}\nabla_{\boldsymbol{f}} \boldsymbol{r}(\boldsymbol{\alpha}, \boldsymbol{f}) &= \frac{\partial\left(\boldsymbol{g} - \boldsymbol{DHS}(\boldsymbol{\alpha})\boldsymbol{f}\right)}{\partial \boldsymbol{f}^T} \\ &= -\boldsymbol{DH}\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\right)}{\partial \boldsymbol{f}^T} \\ &= -\boldsymbol{DHS}(\boldsymbol{\alpha}).\end{aligned} \tag{4.10}$$

Combining with (4.9) and (4.10), (4.8) can be expressed as

$$\boldsymbol{r}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha}, \boldsymbol{f} + \Delta\boldsymbol{f}) = \boldsymbol{r}(\boldsymbol{\alpha}, \boldsymbol{f}) - \boldsymbol{DHG}\Delta\boldsymbol{\alpha} - \boldsymbol{DHS}(\boldsymbol{\alpha})\Delta\boldsymbol{f} \tag{4.11}$$

where $\boldsymbol{G} = \partial\left(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\right)/\partial\boldsymbol{\alpha}^T$ and it denotes the derivative of the HR image $\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}$ with respect to $\boldsymbol{\alpha}$.

Let $\boldsymbol{\alpha} = \boldsymbol{\alpha}^i + \Delta\boldsymbol{\alpha}^i$ and $\boldsymbol{f} = \boldsymbol{f}^i + \Delta\boldsymbol{f}^i$, where $\Delta\boldsymbol{\alpha}^i$ is the increment for the current motion vector estimate $\boldsymbol{\alpha}^i$ and $\Delta\boldsymbol{f}^i$ is the increment for the current HR image estimate $\boldsymbol{f}^i$. The minimization problem (4.7) can be written as (4.12a). Next, based on the first-order approximation in (4.8), (4.12a) can then be simplified as (4.12b)

$$\min_{\Delta\boldsymbol{\alpha}^i, \Delta\boldsymbol{f}^i} \left\| \begin{matrix} \boldsymbol{V}^i \boldsymbol{r}(\boldsymbol{\alpha}^i + \Delta\boldsymbol{\alpha}^i, \boldsymbol{f}^i + \Delta\boldsymbol{f}^i) \\ \sqrt{\lambda}\boldsymbol{L}(\boldsymbol{f}^i + \Delta\boldsymbol{f}^i) \end{matrix} \right\|^2 \tag{4.12a}$$

$$\simeq \min_{\Delta\boldsymbol{\alpha}^i, \Delta\boldsymbol{f}^i} \left\| \begin{pmatrix} \boldsymbol{V}^i\boldsymbol{DHG}^i & \boldsymbol{V}^i\boldsymbol{DHS}(\boldsymbol{\alpha}^i) \\ 0 & \sqrt{\lambda}\boldsymbol{L} \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{\alpha}^i \\ \Delta\boldsymbol{f}^i \end{pmatrix} + \begin{pmatrix} -\boldsymbol{V}^i\boldsymbol{r}(\boldsymbol{\alpha}^i, \boldsymbol{f}^i) \\ \sqrt{\lambda}\boldsymbol{L}\boldsymbol{f}^i \end{pmatrix} \right\|^2. \tag{4.12b}$$

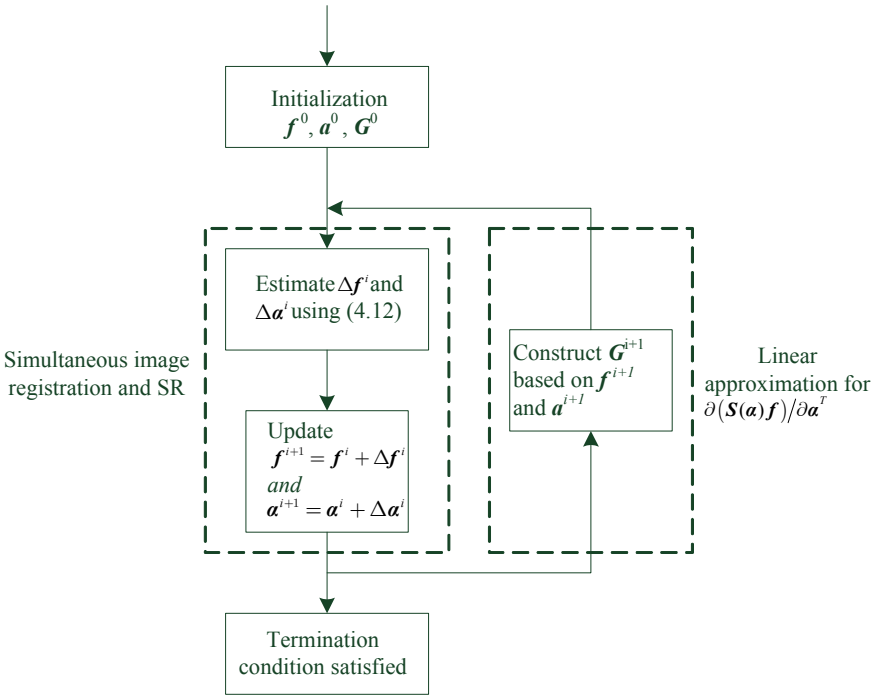


Figure 4.2 Overview of the proposed method.

Based on the above derivation, therefore, we have transformed the original SR problem for directly estimating HR image $\boldsymbol{f}^i$ and motion parameter $\boldsymbol{\alpha}^i$ by minimizing (4.7) into the minimization problem for calculating the HR image increment $\Delta\boldsymbol{f}^i$ and the motion parameter increment $\Delta\boldsymbol{\alpha}^i$ in (4.12). This is a new technique to address the zooming SR problem.

In order to calculate the increment $\Delta\boldsymbol{f}^i$ and $\Delta\boldsymbol{\alpha}^i$, we need to focus on the derivation of an explicit expression for $\partial\left(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\right)/\partial\boldsymbol{\alpha}^T$. Because of the inclusion of zooming and rotation, $\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}$ is nonlinear with respect to $\boldsymbol{\alpha}$. Therefore, a linear approximation for $\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}^i)\boldsymbol{f}^i\right)/\partial\boldsymbol{\alpha}^T$ is developed to handle this problem. Under the assumption that the initial estimates of the motion parameters are reasonably close to their true values, an iterative technique that performs

reconstruction of the HR image and estimation of motion parameters is developed. The overview of the proposed algorithm is given in Figure 4.2. We will discuss the derivation of linear approximation for $\partial(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f})/\partial\boldsymbol{\alpha}^T$ in the following section.

4.4.2 Derivation of Linear Approximation for $\partial(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f})/\partial\boldsymbol{\alpha}^T$

To describe the process clearly, we will first introduce the formation of the *k*th zoomed, rotated and shifted HR images $\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}$. The relative position between the *k*th zoomed, rotated and shifted HR image $\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}$ and the reference HR image $\boldsymbol{f}$ is shown in Figure 4.3. We denote $f_k(i,\ j)$ to be the pixel of $\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}$ and $(x(i,\ j),\ y(i,\ j))$ to be its corresponding coordinates in the reference HR image $\boldsymbol{f}$. Thus, $(x(i,\ j),\ y(i,\ j))$ can be defined as

$$\begin{aligned} x(i,j) &= l_k i\cos\theta_k - l_k j\sin\theta_k + s_{xk} \\ y(i,j) &= l_k i\sin\theta_k + l_k j\cos\theta_k + s_{yk} \end{aligned}. \tag{4.13}$$

Using bilinear interpolation, the pixel $f_k(i,\ j)$ can be expressed as a linear combination of surrounding pixels of the reference HR image $\boldsymbol{f}$,

$$\begin{aligned} f_k(i,j) = {} & (1-d_k(i,j))(1-e_k(i,j))f_{tl} + (1-d_k(i,j))e_k(i,j)f_{tr} \\ & + d_k(i,j)(1-e_k(i,j))f_{bl} + d_k(i,j)e_k(i,j)f_{br} \end{aligned} \tag{4.14}$$

where $d_k(i,j) = x(i,j) - \text{floor}(x(i,j))$ and $e_k(i,j) = y(i,j) - \text{floor}(y(i,j))$. floor(•) denotes the rounding functional which outputs the nearest integer not larger than the input number.

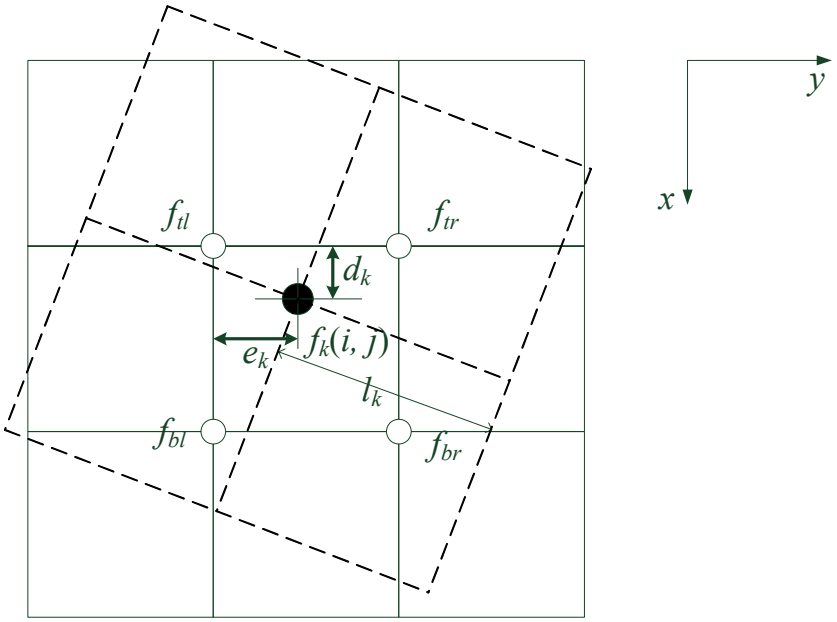


Figure 4.3 The relative position between the *k*th HR grid and the reference HR grid; $f_k(\,i_k\,,j_k\,)$ denotes the HR pixels and $f_{tl}\,,f_{tr}\,,f_{bl}$ and f_{br} are its surrounding pixels in the reference HR grid.

The derivation of linear approximation for $\partial\big(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\big)/\partial\boldsymbol{\alpha}^T$ will be given based on the formation of $\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}$ discussed above. We assume that HR images $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}$ are independent to each other. Therefore, $\partial\big(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\big)/\partial\boldsymbol{\alpha}^T$ can be expressed as

$$
\begin{aligned}
\boldsymbol{G} &= \frac{\partial\big(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\big)}{\partial\boldsymbol{\alpha}^T} \\
&= \frac{\partial\Big(\big(\boldsymbol{S}(\boldsymbol{\alpha}_1)\boldsymbol{f}\big)^T,\ \cdots,\ \big(\boldsymbol{S}(\boldsymbol{\alpha}_N)\boldsymbol{f}\big)^T\Big)^T}{\partial\big(\boldsymbol{\alpha}_1{}^T,\ \cdots,\ \boldsymbol{\alpha}_N{}^T\big)} \\
&= \begin{bmatrix} \dfrac{\partial\big(\boldsymbol{S}(\boldsymbol{\alpha}_1)\boldsymbol{f}\big)}{\partial\boldsymbol{\alpha}_1{}^T} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \dfrac{\partial\big(\boldsymbol{S}(\boldsymbol{\alpha}_N)\boldsymbol{f}\big)}{\partial\boldsymbol{\alpha}_N{}^T} \end{bmatrix}
\end{aligned}
\tag{4.15}
$$

where $\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)/\partial\boldsymbol{\alpha}_k^{\,T}$ of size $\rho^2 M_g N_g \times 4$ is the derivative of the kth HR image $\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}$ with respect to $\boldsymbol{\alpha}_k$. Based on the chain rule, $\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)/\partial\boldsymbol{\alpha}_k^{\,T}$ can be expressed as

$$\frac{\partial \boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}}{\partial \boldsymbol{\alpha}_k^{\,T}} = \frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial\left(\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T, \quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]\right)} \frac{\partial\left(\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T, \quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]^T\right)}{\partial \boldsymbol{\alpha}_k^{\,T}} \tag{4.16}$$

where $\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})$ and $\boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})$ are $\rho^2 M_g N_g$ length vectors representing the lexicographically ordered displacement of $x(i,j)$ and $y(i,j)$, respectively.

Firstly, we will introduce the derivation of $\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)/\partial\left(\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T, \quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]\right)$. For simplicity, we use $\boldsymbol{E}_k$ to represent it in the following derivation. As the direct derivation of $\boldsymbol{E}_k$ is difficult to obtain, we will utilize the chain rule again and divide it into two parts

$$\boldsymbol{E}_k = \underbrace{\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial\left(\left[\boldsymbol{d}_k^{\,T}, \quad \boldsymbol{e}_k^{\,T}\right]\right)}}_{a} \underbrace{\frac{\partial\left(\left[\boldsymbol{d}_k^{\,T}, \quad \boldsymbol{e}_k^{\,T}\right]^T\right)}{\partial\left(\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T, \quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]\right)}}_{b}. \tag{4.17}$$

Hence, we can utilize the dependency between $\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}$ and $\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T, \quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]^T$ provided in (4.13) and (4.14) to derive $\boldsymbol{E}_k$. To simplify the demonstration, the calculations of part (a) and (b) in (4.17) will be explained in the Appendix at the end of this chapter. The final derivation result following appendix is given as follows:

$$\boldsymbol{E}_k = [diag\{(1-\boldsymbol{e}_k)\odot(\boldsymbol{f}_{bl}-\boldsymbol{f}_{tl})+\boldsymbol{e}_k\odot(\boldsymbol{f}_{br}-\boldsymbol{f}_{tr})\},\ diag\{(1-\boldsymbol{d}_k)\odot(\boldsymbol{f}_{tr}-\boldsymbol{f}_{tl})+\boldsymbol{d}_k\odot(\boldsymbol{f}_{br}-\boldsymbol{f}_{bl})\}]. \tag{4.18}$$

Next, we will determine $\partial\left(\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T,\quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]^T\right)\Big/\partial\boldsymbol{\alpha}_k{}^T$. As illustrated in Figure 4.3, $f_k(i,j)$ is denoted as the pixel of the *k*th zoomed, rotated and shifted HR image $\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}$. (i,j) and $\left(x(i,\ j),\ y(i,\ j)\right)$ are denoted as the corresponding coordinates of pixel $f_k(i,j)$ in the HR image $\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}$ and reference HR image $\boldsymbol{f}$. Using the matrix-vector form, we can rewrite (4.13) as

$$\begin{bmatrix}\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})\\ \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})\end{bmatrix}=\boldsymbol{C}\boldsymbol{\alpha}_k \tag{4.19}$$

where $\boldsymbol{C}=\begin{bmatrix}\boldsymbol{i} & -\boldsymbol{j} & \boldsymbol{1} & \boldsymbol{0}\\ \boldsymbol{j} & \boldsymbol{i} & \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}$ and $\boldsymbol{\alpha}_k=[l_k\cos\theta_k,\ l_k\sin\theta_k,\ s_{xk},\ s_{yk}]^T$. $\boldsymbol{i}$ and $\boldsymbol{j}$ are the $\rho^2 M_g N_g$ length column vectors which represent the lexicographically-ordered displacements of *i* and *j*, respectively. From (4.19), it is clear that $\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T,\quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]^T$ is linear with respect to $\boldsymbol{\alpha}_k$. Thus, $\partial\left(\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T,\quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]^T\right)\Big/\partial\boldsymbol{\alpha}_k{}^T$ can be expressed as

$$\frac{\partial\left(\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T,\quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]^T\right)}{\partial\boldsymbol{\alpha}_k{}^T}=\frac{\partial\left(\boldsymbol{C}\boldsymbol{\alpha}_k\right)}{\partial\boldsymbol{\alpha}_k{}^T}=\boldsymbol{C}\,. \tag{4.20}$$

Substituting (4.20) and (4.18) into (4.16), we can obtain a simplified expression for $\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)\Big/\partial\boldsymbol{\alpha}_k{}^T$ which is $\boldsymbol{E}_k\boldsymbol{C}$.

Finally, the derivative of HR image $\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}$ with respect to $\boldsymbol{\alpha}$ can be expressed as follows:

$$\boldsymbol{G}=\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\right)}{\partial\boldsymbol{\alpha}^T}=\boldsymbol{E}\widehat{\boldsymbol{C}} \tag{4.21}$$

where $\boldsymbol{E}=\begin{bmatrix}\boldsymbol{E}_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \boldsymbol{E}_N\end{bmatrix}$ and $\widehat{\boldsymbol{C}}=\begin{bmatrix}\boldsymbol{C} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \boldsymbol{C}\end{bmatrix}$.

4.4.3 Optimization Procedure

In this section, we will calculate the solution for the minimization problem in (4.12). Substituting (4.21) into (4.12), the optimization problem in (4.12) is equivalent to solving the following equation

$$\boldsymbol{A}^i\underline{\boldsymbol{f}}^i=\boldsymbol{b}^i \tag{4.22}$$

where $\boldsymbol{A}^i=\begin{pmatrix}\widehat{\boldsymbol{C}}^T\left(\boldsymbol{E}^i\right)^T\boldsymbol{H}^T\boldsymbol{D}^T\left(\boldsymbol{V}^i\right)^T\boldsymbol{V}^i\boldsymbol{D}\boldsymbol{H}\boldsymbol{E}^i\widehat{\boldsymbol{C}} & \widehat{\boldsymbol{C}}^T\left(\boldsymbol{E}^i\right)^T\boldsymbol{H}^T\boldsymbol{D}^T\left(\boldsymbol{V}^i\right)^T\boldsymbol{V}^i\boldsymbol{D}\boldsymbol{H}\boldsymbol{S}(\boldsymbol{\alpha}^i)\\ \boldsymbol{S}(\boldsymbol{\alpha}^i)^T\boldsymbol{H}^T\boldsymbol{D}^T\left(\boldsymbol{V}^i\right)^T\boldsymbol{V}^i\boldsymbol{D}\boldsymbol{H}\boldsymbol{E}^i\widehat{\boldsymbol{C}} & \boldsymbol{S}(\boldsymbol{\alpha}^i)^T\boldsymbol{H}^T\boldsymbol{D}^T\left(\boldsymbol{V}^i\right)^T\boldsymbol{V}^i\boldsymbol{D}\boldsymbol{H}\boldsymbol{S}(\boldsymbol{\alpha}^i)+\lambda\boldsymbol{L}^T\boldsymbol{L}\end{pmatrix}$,

$\underline{\boldsymbol{f}}^i=\begin{pmatrix}\Delta\boldsymbol{\alpha}^i\\ \Delta\boldsymbol{f}^i\end{pmatrix}$ and $\boldsymbol{b}^i=\begin{pmatrix}\widehat{\boldsymbol{C}}^T\left(\boldsymbol{E}^i\right)^T\boldsymbol{H}^T\boldsymbol{D}^T\left(\boldsymbol{V}^i\right)^T\boldsymbol{V}^i\boldsymbol{r}(\boldsymbol{\alpha}^i,\boldsymbol{f}^i)\\ \boldsymbol{S}(\boldsymbol{\alpha}^i)^T\boldsymbol{H}^T\boldsymbol{D}^T\left(\boldsymbol{V}^i\right)^T\boldsymbol{V}^i\boldsymbol{r}(\boldsymbol{\alpha}^i,\boldsymbol{f}^i)-\lambda\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{f}^i\end{pmatrix}$.

The calculation of the closed-form solution for (4.22) requires inverting the matrix $\boldsymbol{A}^i$. As the computational cost of matrix inversion is high, therefore, we adopt a numerical approach using preconditioned-conjugate-gradient (PCG) optimization in [35]. We use Matlab's linear optimization routine `pcg` to calculate the final solution of (4.22). The convergence criterion of PCG is set as $R_{relative}<10^{-12}$, where $R_{relative}$ is the relative residual defined as $\left\|\boldsymbol{b}^i-\boldsymbol{A}^i\underline{\boldsymbol{f}}^i\right\|/\left\|\boldsymbol{b}^i\right\|$. Finally, the estimated $\Delta\boldsymbol{\alpha}^i$ and $\Delta\boldsymbol{f}^i$ will be utilized to update the motion parametric vector $\boldsymbol{\alpha}^{i+1}=\boldsymbol{\alpha}^i+\Delta\boldsymbol{\alpha}^i$ and the reconstructed HR image $\boldsymbol{f}^{i+1}=\boldsymbol{f}^i+\Delta\boldsymbol{f}^i$ in the next iteration.

The overall complexity of the proposed algorithm can be analyzed by considering the two steps that dominate the computation, namely, construction of gradient matrix $\partial\left(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\right)/\partial\boldsymbol{\alpha}^T$ and joint registration and SR. In this analysis, the cost of a multiplication operation is assumed to be the same as an addition operation, in line with the assumption adopted in [154]. In one iteration, the computational cost for each of the two steps above is O($M_g N_g$) and O(J ($M_f N_f$ + 4 N)), respectively, where J is the number of iterations in PCG optimization. As $M_g N_g$ and $4N < M_f N_f$, implying that O($M_g N_g$) < O(J ($M_f N_f$ + 4 N)) = O($J M_f N_f$), the overall complexity of the proposed algorithm is O($I J M_f N_f$), where I is the number of iterations for the algorithm. It is noted that the complexity cost of the traditional zooming SR methods [85] is approximate O($I J M_f N_f$). Therefore, the computational cost of the proposed method is comparable to the traditional zooming SR methods.

4.4.4 Estimation of Adaptive Channel Weights

The weight $V_k (1 \le k \le N)$ captures the reliability information for the kth LR image into the HR reconstruction. Various algorithms have been used to estimate V_k, including JMAP [155], which uses weighting arising from the distance of the current frame to the reference. In this work, we consider different degree of information rendered by the captured LR images to the final reconstruction. Hence, the residual errors of the LR images are incorporated into the weighting scheme. Therefore, we employ the following

$$V_k = \frac{\dfrac{1}{\left\|\boldsymbol{g}_k - \boldsymbol{DHS}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right\|}}{\dfrac{1}{N}\displaystyle\sum_{k=1}^{N}\dfrac{1}{\left\|\boldsymbol{g}_k - \boldsymbol{DHS}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right\|}} = \frac{V_{average}}{\left\|\boldsymbol{g}_k - \boldsymbol{DHS}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right\|} \tag{4.23}$$

where $V_{average}$ is the average residual value defined as $\frac{N}{\sum_{k=1}^{N}\frac{1}{\left\|\boldsymbol{g}_k-\boldsymbol{DHS}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right\|}}$.

Based on (4.23), it is noted that the channel weight V_k is inversely proportional to the residual value $\left\|\boldsymbol{g}_k-\boldsymbol{DHS}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right\|$. This is because the LR images with small residual value can provide more reliable information than those with large residual value. In each iteration, the channel weights V_k are calculated based on the current estimates of motion parametric vector $\boldsymbol{\alpha}$ and HR image $\boldsymbol{f}$. In the following iteration, the new weights are then incorporated into (4.22).

4.5 Experimental Results

We will demonstrate the effectiveness of the proposed method in this section. Four images in Figure 4.4 are selected as the test images. We conduct various experiments and compare the results obtained by the proposed method with four other methods, namely, the L_1-norm method [28], two zooming-based SR methods of the Joshi's [51] and the Michael's [85] methods, and the AM method [39]. Finally, real-life experiments are also conducted to illustrate the performance of the proposed method. The algorithm will continue until the following criterion is satisfied

$$\frac{\left\|\boldsymbol{f}^{i}-\boldsymbol{f}^{i-1}\right\|}{\left\|\boldsymbol{f}^{i-1}\right\|}<10^{-6}. \tag{4.24}$$

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

Figure 4.4 Test images: (a) “Bridge” image; (b) “Window” image; (c) “Hill” image; (d) “Lena” image.

To evaluate the performance of the methods, two objective metrics of normalized mean-square-errors (NMSE) and peak signal-to-noise ratio (PSNR) are employed for the estimated motion parametric vector $\hat{\boldsymbol{\alpha}}$ and the estimated HR image $\hat{\boldsymbol{f}}$, respectively. The definition of the NMSE for the estimated motion parametric vector is given as follows:

$$NMSE(\hat{\boldsymbol{\alpha}}) \equiv 100\frac{\left\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right\|^2}{\left\|\boldsymbol{\alpha}\right\|^2} \tag{4.25}$$

where ‘^’ denotes the current estimate of motion parametric vector.

4.5.1 HR Reconstruction for Multiple Images with Low-level Zooming Motion

Various experiments have been conducted to demonstrate the effectiveness of the proposed method in handling motion model that consists of rotation, translation and zooming. Based on the LR image generation model (4.1), it is noted that the LR image’s resolution is inversely proportional to its zooming factor. In this section, we will examine image SR under low-level zooming motion when the zooming ratio of the highest resolution LR image to the lowest resolution LR image is between 1 and 1.1. This describes that the captured scene has a gradual

resolution change amongst the LR images. As only small low-level zooming motion is considered in this section, we have included the Farsiu's shift-and-add method in [28] for comparison, which is a fast alternative to the main method in [28] when the motion is translational. This serves to demonstrate the importance of including the zooming motion into the problem formulation even when the amount of zooming involved is small.

The "Bridge" image in Figure 4.4(a) was chosen as the test image. Based on the imaging model in (4.1), 9 LR images were generated with zooming factors randomly selected from a uniform distribution over [0.95, 1.05], rotational angles randomly selected from a uniform distribution over [-5°, 5°] and translations randomly selected from a uniform distribution over [-3, 3] pixels. The first LR image was set as the reference and its zooming factor as 1. The shifted HR image was then blurred by a 3×3 Gaussian blur to simulate the blur operator $h_k \otimes h_c$, followed by a down-sampling operator with the decimation factor of $\rho = 2$. Finally, these LR images were degraded by additive white Gaussian noise (AWGN) to produce a signal-to-noise ratio (SNR) at 35dB.

The initial motion parameter estimation consists of rotation and zooming estimation by method [104], followed by translation estimation by method [29]. Both method [104] and [29] are frequency-domain based registration algorithms. We employ an effective approach to choose an order-of-magnitude estimate of $\lambda = 10^{-4}$, similar to the algorithm adopted in [121]. The initial estimate of HR image for our method can then be calculated by minimizing (4.3). For the Farsiu's shift-and-add method, we have conducted various experiments with regularization parameters ranging from 1×10^{-5} to 1 and find that the best parameter is $\lambda=1\times10^{-1}$. Therefore, we have used the best regularization parameter to obtain the result for the Farsiu's shift-and-add method in Figure 4.5. 4 samples of the LR images and the scaled-up version of the most zoomed

in LR image are shown in Figure 4.5(a) and (b), respectively. The result obtained using the Farsiu's shift-and-add method [28] is shown in Figure 4.5(c). It can be seen that there are some artifacts near the edges. This illustrates that relative zooming that exist among the LR images cannot be neglected even when it is small. Next, the Joshi's method [51] and our proposed algorithm are performed on the LR images and the results are given in Figure 4.5(d) and (e), respectively. Comparison shows that the proposed method can achieve better HR reconstruction. This is because the proposed method performs joint registration and zooming SR iteratively, where more accurate registration parameters are incorporated into the HR reconstruction when compared with the Joshi's method [51]. A selected region of the HR images in Figure 4.5(c)-(e) is enlarged in Figure 4.5(f)-(h) for closer examination. The PSNRs of all the methods above are given in Table 4.1. From the table, it can be observed that our proposed method outperforms the other two methods. The objective performance measure further demonstrates that the proposed method is efficient in handling motion model consists of zooming, rotation and translation.

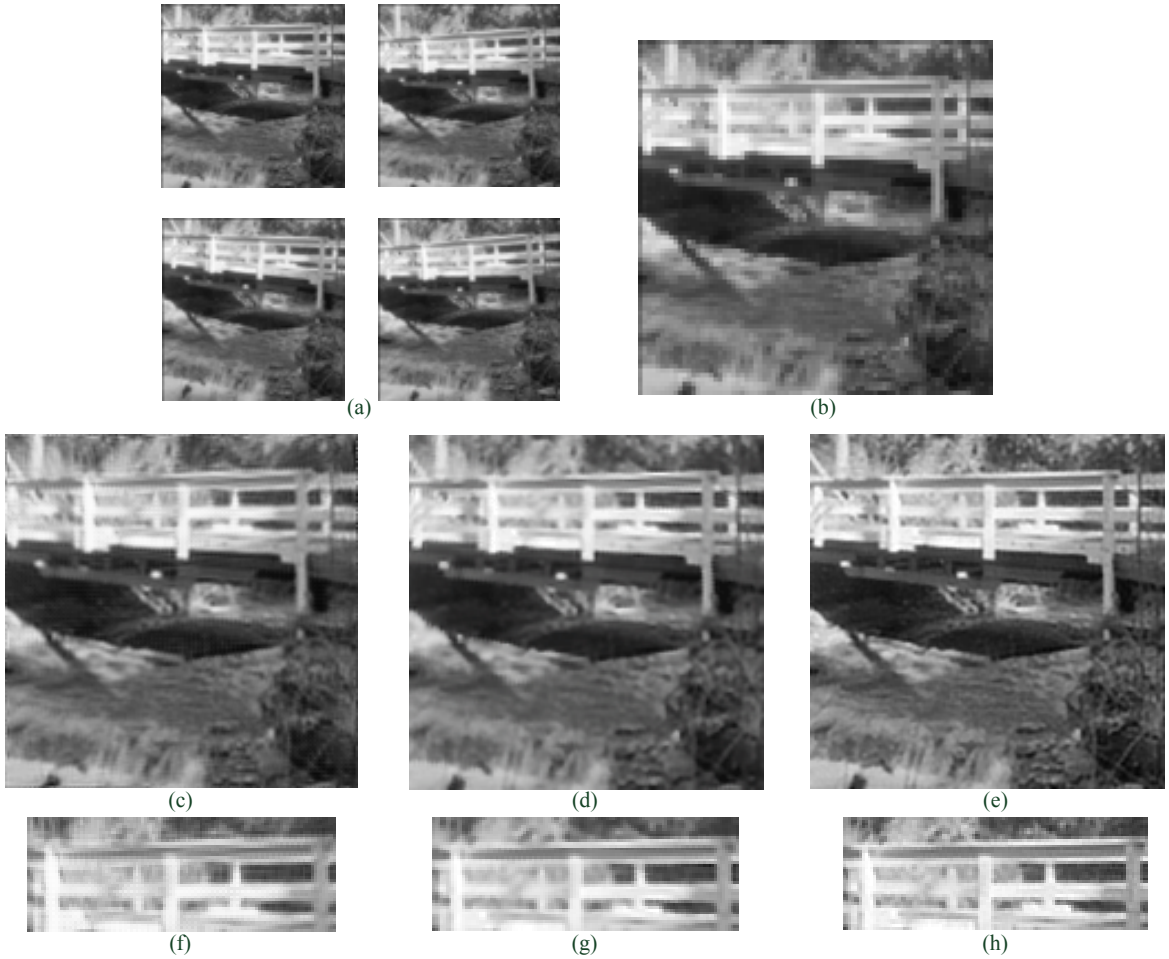


Figure 4.5 SR on the "Bridge" image. (a) 4 samples of the LR images, (b) The scaled-up version of the most zoomed-in LR image, (c) HR result by Farsiu's shift-and-add method [28], (d) HR result by Joshi's method [51], (e) HR result by our proposed method, (f)-(h) Selected enlarged region of (c)-(e), respectively.

Table 4.1 Comparison of PSNR

Farsiu' shift-and-add method [28]	Joshi's method [51]	Our proposed method
26.56	27.75	30.85

4.5.2 HR Reconstruction for Multiple Images with Moderate Zooming Motion

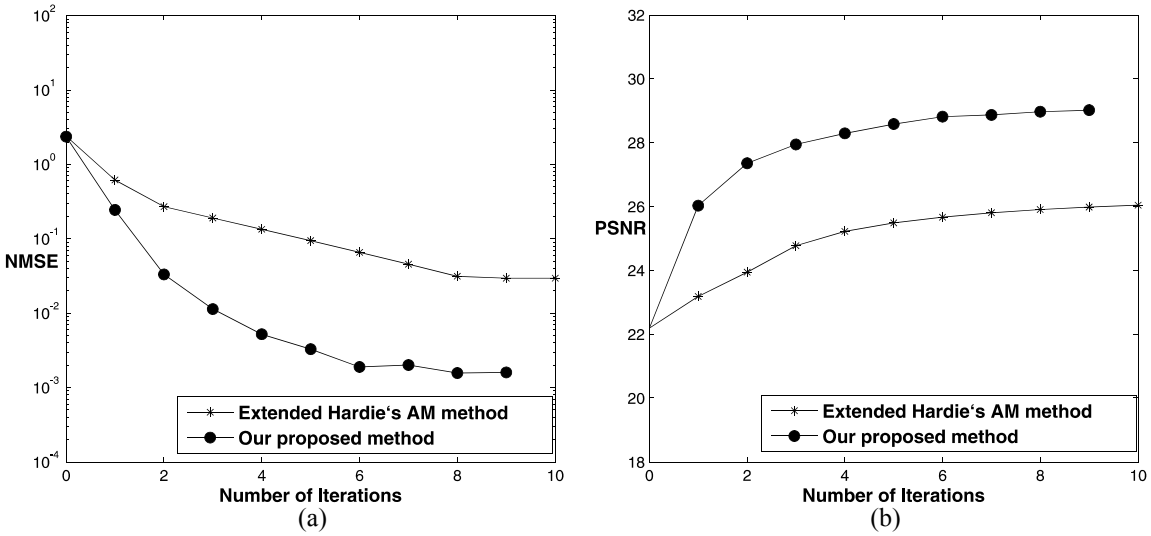
In this section, we focus on reconstruction of multiple LR images with moderate zooming motion. This describes scenarios where there is a moderate change in the resolution of the captured LR images, e.g. a vehicle is moving towards the camera. The moderate zooming motion is considered when the zooming ratio of the highest resolution LR image to the lowest resolution LR image is between 1.1 and 2. As the original Hardie's AM method [39] is designed based on the in-plane motion model, therefore, in order to have a fair comparison, we have extended the original Hardie's method by using our motion model consisting of zooming, rotation and translation. Various experiments have been conducted to verify the effectiveness of the proposed algorithm. These experiments were performed on LR images with moderate zooming motion, where the zooming factors were randomly selected from a uniform distribution over [0.75, 1.25]. The "Window" image in Figure 4.4(b) was chosen as the test image. The same experimental setup as in the former experiment was used. 4 samples of the LR images are shown in Figure 4.6(a). The scaled-up version of the most zoomed-in LR image is shown in Figure 4.6(b). The proposed algorithm is performed on the LR images and the result is given in Figure 4.6(e). It can be observed that the proposed method can restore significant amount of details. The results obtained using the Joshi's two-stage zooming SR method [51] and the extended AM method [39] at the best regularization parameter of $\lambda=1\times10^{-3}$ are shown in Figure 4.6(c) and (d), respectively. It is observed that both methods have lower visual quality (e.g. less details near the window/wall letters and some artifacts near the edges regions) when compared with the proposed method. The comparison shows that the proposed method can achieve superior HR image reconstruction, since it combines image registration and zooming SR reconstruction into a single process of

Figure 4.6 SR on the "Window" image. (a) 4 samples of the LR images, (b) The scaled-up version of the most zoomed-in LR image, (c) HR result by Joshi's method [51], (d) HR result by the extended AM method [39], (e) HR result by our proposed method, (f) HR result using known exact motion parameters.

estimation. This is different from other approaches such as the AM method [39] which projects the cost function onto the HR image and motion parameter domains one a time, and then minimize them iteratively. Subsequently, we further compared our result with the HR image reconstructed by using known exact motion parameters, as shown in Figure 4.6(f). It can be observed that the HR image reconstructed using the proposed method is similar to that reconstructed using the exact motion parameters. The PSNRs of HR results by all methods are given in Table 4.2. The comparison shows that the proposed method can achieve better PSNR, since it combines image registration and SR into a single process of estimation. The iterative

Table 4.2 PSNRs of the HR results

Joshi's method [51]	Extended AM method [39]	**Proposed method**
22.31	26.19	**28.82**



(a) (b)

Figure 4.7 Objective measurements: (a) NMSE of the estimated motion parameters; (b) PSNR of the reconstructed image.

framework enables more accurate motion parameters to be progressively incorporated into the HR reconstruction. The objective measure further confirms that the proposed method is superior to the other two comparative methods. The NSME of the intermediate estimated motion parameters and the PSNR of the intermediate reconstructed image by the extended AM method and the proposed method are given in Figure 4.7. From the figures, it is noted that the objective performance measure confirms that the proposed method is superior to the extended Hardie's AM methods.

Table 4.3 Mean and standard deviation of PSNR and NMSE for various motions

Image	Zooming				Zooming and translation				Zooming, translation and rotation			
	Joshi' method [52]		**Proposed method**		Michael' method [85]		**Proposed method**		Extended AM method [39]		**Proposed method**	
	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR
Bridge	1.51 (0.293)	27.44 (0.428)	**0.0047 (0.00113)**	**29.71 (0.414)**	1.81 (0.2882)	27.21 (0.488)	**0.0029 (0.00154)**	**30.04 (0.479)**	0.0712 (0.00388)	28.68 (0.527)	**0.0035 (0.00179)**	**29.72 (0.406)**
Window	1.79 (0.3113)	25.07 (0.443)	**0.0036 (0.00115)**	**28.76 (0.413)**	2.02 (0.295)	25.21 (0.517)	**0.0022 (0.00149)**	**28.90 (0.508)**	0.0502 (0.00367)	26.17 (0.644)	**0.0024 (0.00169)**	**28.81 (0.527)**
Hill	1.49 (0.328)	27.88 (0.470)	**0.0046 (0.00120)**	**29.67 (0.373)**	1.98 (0.293)	27.02 (0.511)	**0.0045 (0.00152)**	**29.63 (0.494)**	0.0971 (0.00389)	28.28 (0.638)	**0.0035 (0.00172)**	**29.66 (0.521)**
Lena	1.19 (0.349)	30.04 (0.503)	**0.0039 (0.00128)**	**33.38 (0.342)**	2.89 (0.266)	29.11 (0.527)	**0.0032 (0.00153)**	**33.37 (0.511)**	0.0864 (0.00217)	31.78 (0.681)	**0.0027 (0.00176)**	**33.43 (0.517)**

4.5.3 SR on LR Images with Various Motions

We conducted several Monte-Carlo simulations to demonstrate the effectiveness of our proposed method in handling SR reconstruction from LR images with different motions. The Monte-Carlo simulation was repeated 15 times for each motion. The proposed method takes into account the relative zooming, rotational and translational motion amongst the captured LR images. Existing SR methods are only designed for one or two motions above. In this context, the experiments are divided into three groups based on the motion between the LR images, namely, (i) zooming motion only, (ii) zooming and translation, and (iii) zooming, translation and rotation. The same experimental setup as in Section 4.5.2 was used to conduct the experiments. For a fair comparison, our proposed method is compared with only the respective method that can handle the motion model in each case. Table 4.3 shows the mean and standard deviations (given in brackets) of NMSE and PSNR obtained by our proposed method and other methods. Comparison shows that the proposed method is more flexible and effective in handling SR reconstruction under different motions.

4.5.4Experiments on Real-life Images

We also conducted various real-life experiments to evaluate the performance of the proposed method. Ten LR images were captured using a web camera. Four samples of the LR image are shown in Figure 4.8(a). It can be seen that there exists relative translation, rotation and zooming among the LR images. The scaled-up version of the most zoomed in LR image is shown in Figure 4.8(b). The first LR image is set as reference and a decimation factor of 2 is selected in the following experiment. The registration methods in [104] and [29] were again employed to

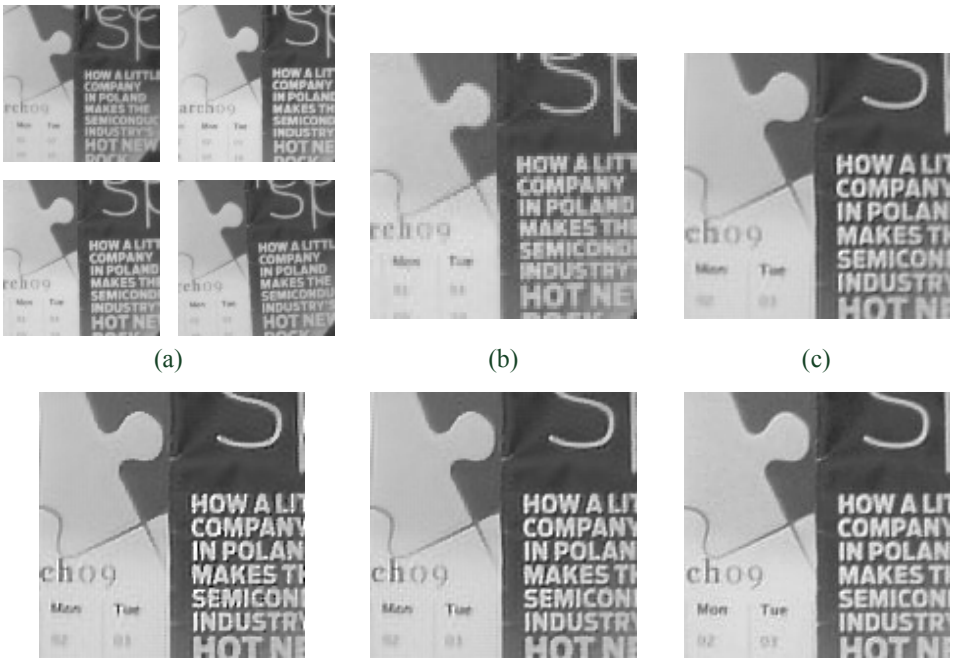


Figure 4.8 SR on real-life images. (a) 4 samples of the LR images, (b) The scaled-up version of the most zoomed-in LR image, (c) HR result by two-stage zooming SR method [51], (d) HR result by the extended AM method [39], (e) HR result by our proposed method, (f) Ground truth.

estimate the initial zooming, rotation and translation parameters, respectively. Subsequently, the two-stage zooming SR method [51], the extend AM method [39] and our proposed method were

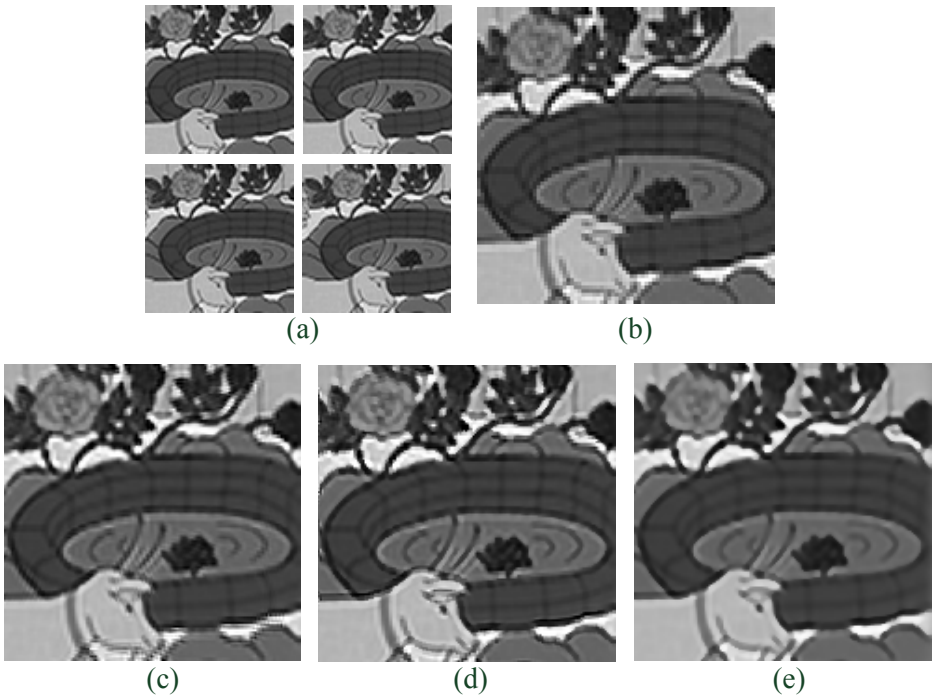


Figure 4.9 SR on video sequence. (a) 4 samples of the LR images, (b) The scaled-up version of the most zoomed-in LR image, (c) HR result by two-stage zooming SR method [51], (d) HR result by the extended AM method [39], (e) HR result by our proposed method.

performed on the captured images and the reconstructed results are given in Figure 4.8(c)-(e), respectively. From the figures, it can be seen that the proposed method has recovered considerable detail information of image when compared with the extended AM and two-stage methods. Meanwhile, it is observed that the result by our proposed method contains less artifacts than other methods. In order to provide a fair comparison, an image captured with the resolution of the HR image is used as the ground truth in Figure 4.8(f). Comparison shows that the proposed method can handle real-life SR well.

4.5.5 Experiments on Video Sequences

Finally, we conducted experiments on a popular real video sequence, namely, the *Mobile* sequence. Ten 80×80 size LR images from frame 27 to frame 36 were selected as the test materials. Four samples of the LR images are shown in Figure 4.9(a). The scaled-up version of the most zoomed in LR image is shown in Figure 4.9(b). Two-stage zooming SR method [51], the extended AM method [39] and our proposed method were performed on the LR frames and the results are given in Figure 4.9(c)-(e). From the comparison, it can be observed that the proposed method is superior to the other two methods in handling SR reconstruction from video sequences.

4.6Summary

In this chapter, we develop a new technique to perform SR reconstruction from LR images with relative zooming motion. Different from most existing two-stage zooming SR methods, the proposed algorithm combines image registration and HR reconstruction into a single process and adopt an iterative framework to improve the estimated HR image and registration parameter progressively. In addition, the proposed method utilizes a motion model consisting of zooming, rotation and translation and a linear approximation technique is developed to solve the arising problem. Both simulated and real-life experiments demonstrate that our proposed method is efficient in performing zooming SR reconstruction.

The proposed method, however, may produce unsatisfactory HR results if the initial estimates of the motion parameters are far from the true value. Nevertheless, various experimental results show that the adopted motion estimation methods [29], [104] can produce reasonable initial estimates of motion parameters to achieve satisfactory results in our algorithm.

4.7 Appendix

CALCULATION OF $\boldsymbol{E}_k$

The calculation of $\boldsymbol{E}_k$ in (4.18) will be explained in this appendix. Using the chain rule, $\boldsymbol{E}_k$ can be divided into two components as follows:

$$\boldsymbol{E}_k = \underbrace{\frac{\partial\left(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial\left(\left[\boldsymbol{d}_k{}^T, \quad \boldsymbol{e}_k{}^T\right]\right)}}_{a} \underbrace{\frac{\partial\left(\left[\boldsymbol{d}_k{}^T, \quad \boldsymbol{e}_k{}^T\right]^T\right)}{\partial\left(\left[\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T, \quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T\right]\right)}}_{b}. \tag{4.26}$$

For part (a):

It can be calculated as follows [35]:

$$\begin{aligned}
\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial\left(\left[\boldsymbol{d}_k{}^T, \quad \boldsymbol{e}_k{}^T\right]\right)} &= \left[\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial \boldsymbol{d}_k{}^T}, \quad \frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial \boldsymbol{e}_k{}^T}\right] \\
&= \left[\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial d_k(1,1)}, \quad \cdots \quad \frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial d_k(\rho M_g, \rho N_g)}, \quad \frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial e_k(1,1)}, \quad \cdots \quad \frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial e_k(\rho M_g, \rho N_g)}\right] \\
&= \begin{bmatrix}
\frac{\partial f_k(1,1)}{\partial d_k(1,1)} & 0 & \cdots & 0 & \frac{\partial f_k(1,1)}{\partial e_k(1,1)} & 0 & \cdots & 0 \\
0 & \frac{\partial f_k(2,1)}{\partial d_k(2,1)} & 0 & \vdots & 0 & \frac{\partial f_k(2,1)}{\partial e_k(2,1)} & 0 & \vdots \\
\vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 \\
0 & \cdots & 0 & \frac{\partial f_k(\rho M_g, \rho N_g)}{\partial d_k(\rho M_g, \rho N_g)} & 0 & \cdots & 0 & \frac{\partial f_k(\rho M_g, \rho N_g)}{\partial e_k(\rho M_g, \rho N_g)}
\end{bmatrix}
\end{aligned} \tag{4.27}$$

where $\dfrac{\partial f_k(i,j)}{\partial d_k(i,j)} = \dfrac{\partial\begin{pmatrix}\left(1-d_k(i,j)\right)\left(1-e_k(i,j)\right)f_{tl} + \left(1-d_k(i,j)\right)e_k(i,j)f_{tr} \\ + d_k(i,j)\left(1-e_k(i,j)\right)f_{bl} + d_k(i,j)e_k(i,j)f_{br}\end{pmatrix}}{\partial d_k(i,j)}$

$$= \left(1-e_k(i,j)\right)\left(f_{bl}-f_{tl}\right) + e_k(i,j)\left(f_{br}-f_{tr}\right)$$

and $$\frac{\partial f_k(i,j)}{\partial e_k(i,j)} = \frac{\partial\begin{pmatrix}(1-d_k(i,j))(1-e_k(i,j))f_{tl} + (1-d_k(i,j))e_k(i,j)f_{tr} \\ + d_k(i,j)(1-e_k(i,j))f_{bl} + d_k(i,j)e_k(i,j)f_{br}\end{pmatrix}}{\partial e_k(i,j)}.$$

$$= (1-d_k(i,j))(f_{tr} - f_{tl}) + d_k(i,j)(f_{br} - f_{bl})$$

Hence, (4.27) can be rewritten in a simple form as

$$\frac{\partial(\boldsymbol{S}(\boldsymbol{\alpha}_k)\boldsymbol{f})}{\partial([\boldsymbol{d}_k{}^T, \quad \boldsymbol{e}_k{}^T])}$$

$$= [diag\{(1-\boldsymbol{e}_k)\odot(\boldsymbol{f}_{bl} - \boldsymbol{f}_{tl}) + \boldsymbol{e}_k \odot (\boldsymbol{f}_{br} - \boldsymbol{f}_{tr})\},\ diag\{(1-\boldsymbol{d}_k)\odot(\boldsymbol{f}_{tr} - \boldsymbol{f}_{tl}) + \boldsymbol{d}_k \odot (\boldsymbol{f}_{br} - \boldsymbol{f}_{bl})\}]. \tag{4.28}$$

For part (b):

We follow the idea in [1] to obtain

$$\frac{\partial\left([\boldsymbol{d}_k{}^T, \quad \boldsymbol{e}_k{}^T]^T\right)}{\partial\left([\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T, \quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T]\right)} = \frac{\partial\left([\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T, \quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T]^T\right)}{\partial\left([\boldsymbol{x}(\boldsymbol{i},\boldsymbol{j})^T, \quad \boldsymbol{y}(\boldsymbol{i},\boldsymbol{j})^T]\right)} = \boldsymbol{I} \tag{4.29}$$

where $\boldsymbol{I}$ is an identity matrix of size $2\rho^2 M_g N_g \times 2\rho^2 M_g N_g$.

Combining (4.28) and (4.29), we can finally obtain the expression of $\boldsymbol{E}_k$ in (4.18).

Chapter 5

Vehicle License Plate Super-resolution Using Soft Learning Prior

5.1 Introduction

Image super-resolution (SR) is a process to reconstruct high-resolution (HR) image from single or multiple low-resolution (LR) observations. An important application of image SR is reconstruction of HR vehicle license plate (VLP) from LR observations. This could be used in surveillance systems or traffic regulation enforcement, amongst others. Image SR generally can be categorized into single-frame or multi-frame SR, where a HR image is reconstructed from a single LR image or multiple LR images, respectively. This chapter will focus on single-frame license plate SR as at times, only a single frame of the license plate image can be captured due to various imaging constraints.

Different single-frame image SR or interpolation methods have been proposed over the years. Classical interpolation methods such as bi-cubic interpolation use piecewise polynomials to model smoothness in image intensity of the local spatial image neighborhood [7]. However, these methods tend to produce over-smoothness near the edges and textured regions. Hence, other methods have been developed to handle this issue. Adaptive bi-cubic interpolation methods use local features such as edge direction to improve interpolation. Inverse gradient [46] and

warped distance [103] have been employed to determine the weights of bi-cubic interpolation. In [45], the edge information of each local region is extracted by using the discrete cosine transform (DCT). Different edge types are identified and used to determine different interpolation strategies for each area. The basic idea of [70] is to estimate the HR local covariance coefficients from the LR counterpart based on their geometric duality. The edge-directed interpolation is tuned based on the covariance. It is noted that all the methods discussed involve estimation of heuristic threshold values and filter weights. Thus, the results are sensitive to changes in values of these parameters.

Regularization methods formulate the single-frame SR problem into the optimization of a cost function consisting of a data fidelity term and a regularization term. The regularization term incorporates some prior information of the image models to achieve a stable solution to the problem [55]. In [22], a smoothness prior is used and the experimental results show that it is superior to linear minimum mean squared error (LMMSE) interpolation algorithm and maximum entropy technique. In [126], the authors apply a discontinuity adaptive regularization term to preserve the edges in the SR reconstruction. A framework of constrained least squares minimization using singular value decomposition (SVD) is proposed in [18]. However, it is noted that most the regularization terms used in the methods above are centered on smoothness priors. They do not incorporate specific prior information about the targeted types of images that are being reconstructed. To address this issue, a text-based interpolation method using L_1-norm regularization term is developed in [146]. However, the regularization term becomes increasingly ineffective as the magnification factor of the HR image reconstruction increases [5],[74].

Learning-based SR methods use training samples and models to perform single-frame SR. An advantage of these methods is that even when the magnification ratio of the HR image to the LR image is large, the information obtained from learning examples can still restore the details of the HR images effectively [30]. Through the learning process, a lookup table between the HR patches and their corresponding LR patches is built. Given a representative learning dataset, these methods can work well to produce sharper images. A maximum a posteriori (MAP) model using learning-based prior has been introduced in [21]. The authors use nearest-neighbor (NN) search to find the HR patches. The learning term then integrates the details information based on relationship between the example HR and LR patches. Existing learning-based methods use binary weighting schemes, where the weights are taken as '0' for totally irrelevant samples and '1' for fully relevant samples. The main shortcoming of these methods is that they are prone to artifacts when unsuitable HR patches are chosen for learning. The binary strategy tends to produce artifacts as the HR patches cannot be integrated partially in the cases when the patches only bear partial resemblance to the target LR patches. The HR patches either have to be incorporated fully or discarded outright. Further, these methods consider the HR reconstruction from a local perspective. They do not utilize the global information of the alpha-numerics existed in the VLP when performing reconstruction.

In view of this, we propose a new single-frame SR framework to address the issues above. The main novelties of the proposed method are threefold: (i) It introduces a new iterative technique that integrates car plate character prior information into license plate SR reconstruction. (ii) It proposes a soft learning prior that estimates the importance of different learning samples, and integrates this information into HR reconstruction of the license plate. This is as opposed to conventional hard-decision methods, which either incorporate the training sample fully, or

discard them outright. (iii) It introduces a scheme that utilizes both global and local characteristics when performing license plate HR reconstruction. This is in contrast with the existing learning methods that use local information alone. Experimental results show that the proposed method can outperform other well-known methods.

The organization of this chapter is as follows. The problem formulation is introduced in Section 5.2. The overview of the proposed framework is explained in Section 5.3. An iterative example-based SR algorithm using soft learning is developed in Section 5.4. Experimental results on both simulated and real-life images are presented and discussed in Section 5.5. A brief conclusion and future work is given in Section 5.6.

5.2 Problem Formulation

The single-frame LR image can be modeled as applying blurring and down-sampling to the original HR image, followed by addition of noise. The observation model can be expressed in matrix-vector form as

$$\boldsymbol{g} = \boldsymbol{W}\boldsymbol{f} + \boldsymbol{n} \tag{5.1}$$

where $\boldsymbol{f} = [f_1, f_2, \cdots f_{DN_1 \times DN_2}]^T$, $\boldsymbol{g} = [g_1, g_2, \cdots g_{N_1 \times N_2}]^T$, and $\boldsymbol{n} = [n_1, n_2, \cdots n_{N_1 \times N_2}]^T$ represent the HR image of sized $DN_1 \times DN_2$, the observed LR image of sized $N_1 \times N_2$ and the additive noise of sized $N_1 \times N_2$, respectively. The parameter D denotes the decimation factor, and matrix $\boldsymbol{W}$ with dimension $N_1 N_2 \times DN_1 DN_2$ represents the operator that includes the blurring and down-sampling process.

The objective of single-frame SR is to reconstruct $\boldsymbol{f}$ from the observed $\boldsymbol{g}$. The probability density functions (pdfs) of $\boldsymbol{g}$ and $\boldsymbol{f}$ are given by $p(\boldsymbol{g})$ and $p(\boldsymbol{f})$, respectively. The development

of the new soft learning cost function is explained as follows. To estimate the HR image, we will determine the argument $\boldsymbol{f}$ that maximizes the *a posteriori* probability $p(\boldsymbol{f}|\boldsymbol{g})$. Appling the Bayesian theorem, we have

$$p(\boldsymbol{f}|\boldsymbol{g}) = \frac{p(\boldsymbol{g}|\boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{g})}. \tag{5.2}$$

As the denominator $p(\boldsymbol{g})$ is a common term that does not affect the maximization of $p(\boldsymbol{f}|\boldsymbol{g})$, we will drop it in the subsequent algorithm development. Assuming $\boldsymbol{n}$ is additive white Gaussian with variance σ_n^2, then the likelihood $p(\boldsymbol{g}|\boldsymbol{f})$ can be expressed as

$$p(\boldsymbol{g}|\boldsymbol{f}) \propto \exp\left\{-\frac{\|\boldsymbol{g}-\boldsymbol{W}\boldsymbol{f}\|_2^2}{2\sigma_n^2}\right\}. \tag{5.3}$$

We represent the *prior* of the HR image $p(\boldsymbol{f})$ in an exponential form, which is the general format of Gibbs distribution:

$$p(\boldsymbol{f}) \propto \exp\{-A(\boldsymbol{f})\}. \tag{5.4}$$

The term $A(\boldsymbol{f})$ is a non-negative energy function, with smaller values for more probable signals, and vice versa.

Taking the logarithm of $p(\boldsymbol{f}|\boldsymbol{g})$, the MAP estimation can be formulated as minimizing the following cost function:

$$L(\boldsymbol{f}) = \|\boldsymbol{g}-\boldsymbol{W}\boldsymbol{f}\|_2^2 + \lambda A(\boldsymbol{f}) \tag{5.5}$$

where λ is the regularization parameter that controls the relative contribution between the data fidelity term $\|\boldsymbol{g}-\boldsymbol{W}\boldsymbol{f}\|_2^2$ and the HR image prior $A(\boldsymbol{f})$. As this chapter focuses on HR reconstruction of license plates in a domain-specific application, we propose a new soft learning cost function that incorporates the characteristics of license plate to perform learning:

$$L(\boldsymbol{f}) = \|\boldsymbol{g}-\boldsymbol{W}\boldsymbol{f}\|_2^2 + \lambda \sum_{i,j} \sum_{k\in\Omega[i,j]} w_k^{[i,j]} \left\| \boldsymbol{R}_{[i,j]}\boldsymbol{f} - \boldsymbol{x}_k^{[i,j]} \right\|_2^2 . \tag{5.6}$$

The new cost function employs a soft learning prior instead of conventional smoothness prior. The first term in the cost function represents the data fidelity term, while the second term is the learning-based functional which incorporates the learning samples. The operator $\boldsymbol{R}_{[i,j]}$ extracts an $m\times m$ block from the image $\boldsymbol{f}$, and compares it with the HR training patches $\boldsymbol{x}_k^{[i,j]}$ at location [*i,j*]. The inner summation is conducted over all HR patches $\boldsymbol{x}_k^{[i,j]}$ where $k\in\Omega[i,j]$ is the patch index and $\Omega[i,j]$ is the set of all selected HR patches at location [*i,j*]. The outer summation runs through all the pixel locations of the HR image. λ is the regularization parameter that controls the relative contribution between the data fidelity term and the learning functional. A small λ value will cause the reconstructed HR image to emphasize on the data fidelity term, while a large λ value will provide more emphasis on learning towards the training samples. $w_k^{[i,j]}$ is the soft learning weight for a selected training patch.

The main feature of the proposed method is that it uses a soft learning prior in the cost function. As opposed to existing learning approaches such as [21] where binary weighting schemes are used (i.e. the weight $w_k^{[i,j]}$ is either 0 or 1), the proposed framework uses a soft learning framework where the weight $w_k^{[i,j]}\in[0,1]$ is assigned to a continuous value between 0

and 1. The weight is estimated based on the relevance/resemblance of the training patches with respect to the current patch under reconstruction. This will ensure a more meaningful incorporation of information from learning patches, and reduce the number of outliers (dissimilar patches) from being included into the training process. Further, the proposed method incorporates character prior information to improve the performance of the system. The character prior information will be used to estimate the importance of learning patches which is reflected in $w_k^{[i,j]}$. An advantage of the proposed method is that it will take both the global and local information into consideration when performing HR reconstruction. Here, the global information refers to the high-level information such as which alphanumeric the current patch may belong to. On the other hand, the local information refers to the low-level patch structural information which is focused on local similarity of the current patch with respect to the sample learning patches. In contrast with the conventional learning-based method which relies on local patch matching alone, the method will combine the global information of character prior with the local patch matching to reach an ideal compromise in performing image SR.

5.3 Overview of the Proposed Framework

The overview of the proposed soft learning license image SR is given in Figure 5.1. It consists of two iterative steps of HR image reconstruction and VLP character prior integration. Firstly, the initial HR image $\boldsymbol{f}_0$ is estimated using bi-cubic interpolation of the LR image. The initial estimate $\boldsymbol{f}_0$ will then be passed into the soft weight estimation to determine the importance of the learning patches. It is noted that in the first iteration, a binary weight estimation scheme for the learning patches will be used to reconstruct the HR image. For the subsequent iterations, a

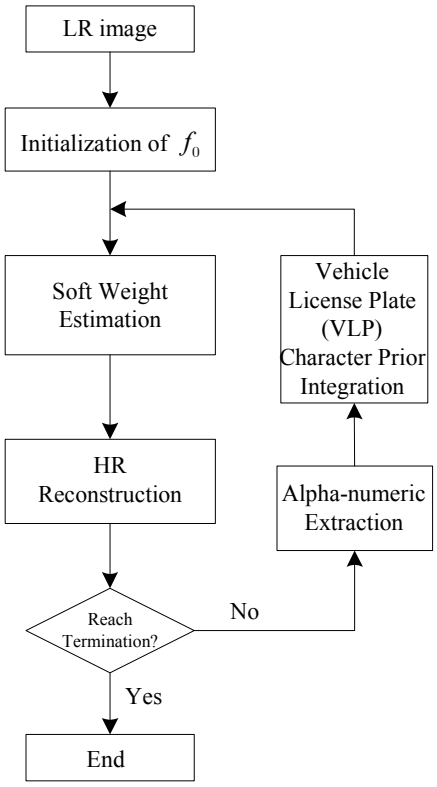


Figure 5.1 Flowchart for the proposed framework

soft weight estimation scheme using the inputs from VLP will be employed to determine the importance of the learning patches. The HR reconstruction of the license image using the soft learning will then performed by minimizing the proposed cost function using conjugate gradient optimization. The reconstructed HR image is then passed to alphanumeric extraction when the image is segmented into different characters. These characters are then passed through the VLP character prior integration which exploits the soft prior information of each segmented alphanumerics. They are used to estimate the soft weight estimation of the HR learning patches. The iterative process of HR image reconstruction and VLP prior information integration will continue so that progressively better HR image can be obtained. The process will continue until convergence or a maximum number of iterations is reached.

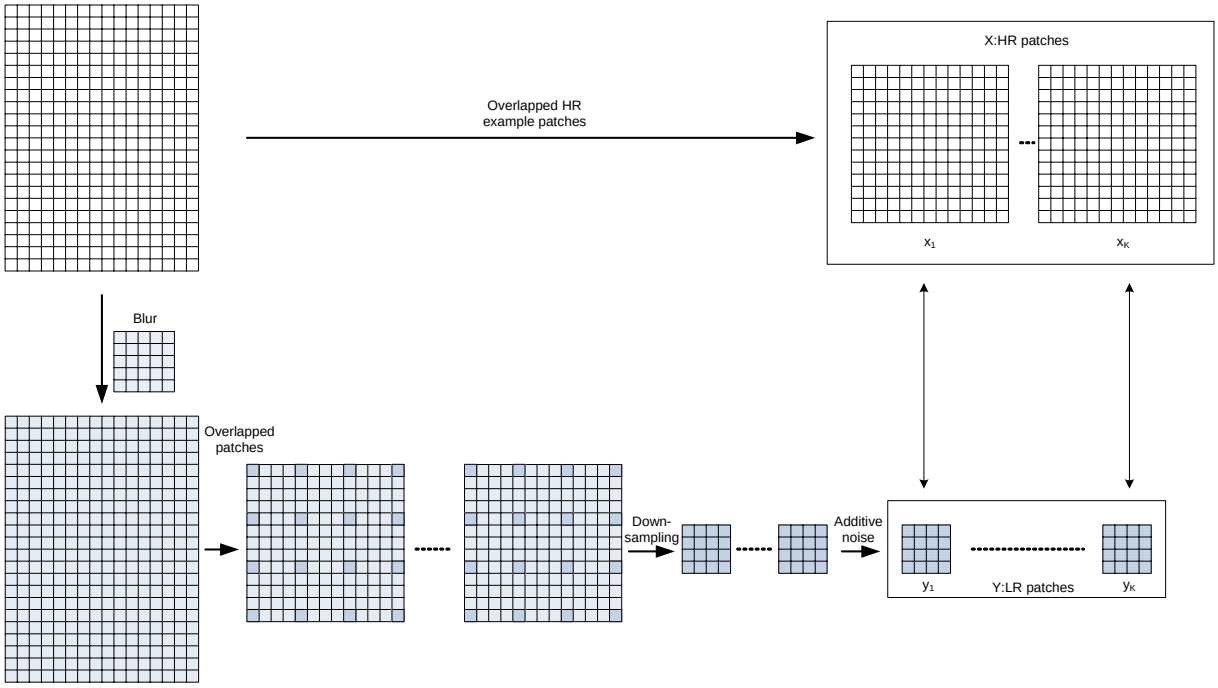


Figure 5.2 Creation of the patch learning database

5.4 Iterative HR Reconstruction using Soft Learning

5.4.1 Creation of Learning Patch Database

In many domain-specific applications such as SR of license plates and facial images, the use of training examples has been shown to be effective. This is because the target images are known to possess certain characteristics that can be utilized in HR reconstruction. For example, in the case of VLP HR reconstruction, the VLPs are consisted of 36 alphanumeric (26 letters and 10 digits) in many countries. In order to exploit this information, learning-based approaches will generate sets of training patches for the LR images and the corresponding HR images. This work will follow the standard procedure of generating the LR and HR learning patch database.

Figure 5.2 shows the creation of learning patch database. First, for each HR alphanumeric image of sized $K \times L$, a set of overlapped HR patches of sized $m \times m$ is sampled. Then, their corresponding LR patches of sized $n \times n$ are obtained as follows: (i) Apply the uniform blur to the HR images to produce a set of overlapped blurred patches of sized $m \times m$, (ii) downsample the blurred patches by a factor D, followed by an additive noise. A database containing corresponding HR and LR patch pairs is generated. The pairs of HR-LR patches in the database are partitioned into 36 alphanumeric categories.

5.4.2 VLP Character Prior Integration

The main objective of VLP character prior integration is to exploit the information that the characters on the license plate are consisted of alphanumerics. Hence the outcome of the character prior integration can be used to provide appropriate soft learning weights for different HR patches during the HR reconstruction. The VLP character prior integration consists of three steps: image preprocessing [152], character segmentation and prior information extraction. The first two steps serve to extract individual characters from the captured license image. Then, character prior information extraction is performed on the segmented characters to provide the prior information of the car plate (i.e. the likelihood of top-matched alphanumerics). Different techniques including neural network (NN) techniques [116], [89] and support vector machine (SVM) [102], [56], have been used. However, these methods need extensive training and they are usually computational intensive. As the VLP consists of alphanumerics with rigid format, computationally efficient technique such as template matching is used in this case. The cross-

correlation scores between the normalized alphanumeric image and each character prototype are calculated [108]. The normalized cross-correlation score is given as follows:

$$\rho = \frac{(\boldsymbol{f} - \boldsymbol{\mu}_f)^T (\boldsymbol{v} - \boldsymbol{\mu}_v)}{\left\| \boldsymbol{f} - \boldsymbol{\mu}_f \right\|_2 \left\| \boldsymbol{v} - \boldsymbol{\mu}_v \right\|_2} \tag{5.7}$$

where $\boldsymbol{f}$ represents the individual segmented alphanumerics in the license plate, $\boldsymbol{v}$ represents each of the 36 character templates used in VLP, $\boldsymbol{\mu}_v$ and $\boldsymbol{\mu}_f$ are the uniform vectors formed by the average gray level of $\boldsymbol{v}$ and $\boldsymbol{f}$, respectively. The character prior information extraction is centered on the computation of the normalized cross-correlation value ρ for all the templates $\boldsymbol{v}$. The value ρ is used as an input to estimate the soft weights $w_k^{[i,j]}$ in the next section.

5.4.3 Soft Learning Weight Estimation

Conventional learning-based approaches employ crisp binary decision when determining whether a HR patch should be used in learning. In other words, the patch is either considered to be fully relevant or totally irrelevant in guiding the local HR reconstruction. This approach is clearly inconsistent with real-world observations as the HR learning patches often contain only partial resemblance to the HR image under reconstruction.

In view of this, this section will introduce a soft learning weight estimation scheme that determines which HR patches will be chosen for learning and what weight $w_k^{[i,j]} \in [0,1]$ will be assigned to these learning patches. The main idea is that those HR patches originating from the top-matched character templates will be assigned greater weight during the soft learning process. During the experiments, it is observed that the characters are often classified into one of the top three categories. For instance, the character “O” is likely to be classified as “O” or “0” and the

digit “6” is likely to be classified as “6”, “5” or “S”. Therefore, the HR patches arising from these three character templates should be given higher weights to guide the solution towards these characters.

The process of soft weight estimation is explained as follows. First, the LR license image is partitioned into various overlapping LR patches. The current target LR patch is then compared with all the LR patches in the database. If the difference between the target LR patch with the LR patches in the database is less than a predefined threshold, the corresponding HR patches in the database will be used in the learning. In current binary weighting scheme such as [21], all the selected HR patches will be assigned equal weight. In other words, the binary scheme considers all the shortlisted HR patches to be fully and equally relevant while all the other unselected patches to be totally irrelevant. This approach, clearly, is non-ideal as it tends to include outliers as part of the learning patches. Further, the approach tends to focus on the local matching and ignore the global information, namely which character template is the current patch most likely to belong to.

In this work, the normalized cross-correlation scores ρ are calculated as a measure of likelihood that the current patch will belong to different character templates. The larger the score ρ, the more likely the current patch will belong to a certain character template. Hence, they are used to determine the soft weights of the learning patches. The top three character templates based on ρ will be assigned 90% of total weights (i.e. the combined weights of the top three categories will be 0.9), while the rest of the categories will be assigned 10% of total weights (i.e. the combined weights of the rest of the categories will be 0.1). The relative weighting within each of these two grouping is linearly scaled using the score ρ. If one template out of the top

Table 5.1 Summary of CG optimization for HR image reconstruction

Initialization:

$\boldsymbol{p_0} = \boldsymbol{r_0}$, where $\boldsymbol{r_0} = \boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{f}}_0$.

Step 1. At the *k*-th iteration

Update the *k*-th iteration conjugate vector:

$$\boldsymbol{p_k} = \boldsymbol{r_k} + \beta_k \boldsymbol{p}_{k-1}, \text{ where } \beta_k = \frac{\boldsymbol{r_k}^T \boldsymbol{r_k}}{\boldsymbol{r}_{k-1}^T \boldsymbol{r}_{k-1}}. \quad (5.8)$$

Step 2. Update the (*k*+1)-th iteration HR image:

$$\hat{\boldsymbol{f}}_{k+1} = \hat{\boldsymbol{f}}_k + \alpha_k \boldsymbol{p_k}, \text{ where } \alpha_k = \frac{\boldsymbol{r_k}^T \boldsymbol{r_k}}{\boldsymbol{p_k}^T \boldsymbol{A} \boldsymbol{p_k}}. \quad (5.9)$$

Step 3. Update the (*k*+1)-th iteration residual:

$$\boldsymbol{r}_{k+1} = \boldsymbol{r_k} - \alpha_k \boldsymbol{A} \boldsymbol{p_k}. \quad (5.10)$$

Step 4. Repeat Steps 1-3 until convergence or a maximum number of iteration is reached.

three is dominant, then the soft weight assigned to this category of learning patches will be much higher than the other two categories. In this case, the weight coefficients can actually be real values in [0, 0.9]. The main rationale for this weight assignment scheme is that a misclassified character is typically grouped into one of the top three categories. For instance, a digit “6” may likely be classified as “6”, “5” or “S”. Thus, the HR patches originating from the top 3 characters templates should be given a large combined weight of 90%, while the rest of the categories will be considered as outliers and assigned a combined weight of 10%.

5.4.4 HR Image Reconstruction

After determining the soft learning weights $w_k^{[i,j]} \in [0,1]$, the HR image can be estimated by minimizing the proposed cost function in (5.6):

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left(\left\| \boldsymbol{g} - \boldsymbol{W}\boldsymbol{f} \right\|_2^2 + \lambda \sum_{i,j} \sum_{k \in \Omega[i,j]} w_k^{[i,j]} \left\| \boldsymbol{R}_{[i,j]} \boldsymbol{f} - \boldsymbol{x}_k^{[i,j]} \right\|_2^2 \right). \tag{5.11}$$

This is equivalent to solving for $\hat{\boldsymbol{f}}$ using the following linear equation:

$$\boldsymbol{A}\hat{\boldsymbol{f}} = \boldsymbol{b} \tag{5.12}$$

where $\boldsymbol{A} = \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{W} + \lambda \sum_{i,j} \sum_{k \in \Omega[i,j]} w_k^{[i,j]} (\boldsymbol{R}_{[i,j]}{}^T \boldsymbol{R}_{[i,j]})$ and $\boldsymbol{b} = \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{g} + \lambda \sum_{i,j} \sum_{k \in \Omega[i,j]} w_k^{[i,j]} (\boldsymbol{R}_{[i,j]}{}^T \boldsymbol{x}_{[i,j]})$.

In order to solve the optimization problem in (5.12), conjugate gradient (CG) optimization is employed. CG utilizes conjugate direction instead of local gradient to search for the minima. Therefore, it can achieve faster convergence when compared with steepest descent method [28]. It also requires less storage requirement and computation complexity when compared with Quasi-Newton method. The mathematical formulation of the HR image reconstruction is derived in Table 5.1.

5.4.5 Summary of the Proposed Method

The SR algorithm using the soft learning prior is summarized in Table 5.2.

Table 5.2 Summary of the proposed method

Initialization:

a. Initialize estimated image $\boldsymbol{f}_0$ using bi-cubic interpolation.

b. Assign $w_k^{[i,j]} = 1$ for all i, j and k.

1. Extract the overlapped LR patches $\boldsymbol{g}_{[i,j]}$ of sized $n \times n$ from the observed LR image $\boldsymbol{g}$ at location [i , j].
2. Find the corresponding set of $m \times m$ HR patches $\boldsymbol{x}_{[i,j]}$ with respect to the current LR patch $\boldsymbol{g}_{[i,j]}$.
3. Construct the matrix $\boldsymbol{R}_{[i,j]}$ that extract an $m \times m$ block at location [i , j].
4. At the i-th iteration:
   - Perform image enhancement and segmentation.
   - Calculate the cross-correlation score ρ.
   - Perform soft weight estimation $w_k^{[i,j]}$ of learning patches based on ρ.
   - Employ conjugate gradient optimization to estimate the HR image $\hat{\boldsymbol{f}}$ by solving the soft-learning cost function:

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left(\left\| \boldsymbol{g} - \boldsymbol{W}\boldsymbol{f} \right\|_2^2 + \lambda \sum_{i,j} \sum_{k \in \Omega[i,j]} w_k^{[i,j]} \left\| \boldsymbol{R}_{[i,j]} \boldsymbol{f} - \boldsymbol{x}_k^{[i,j]} \right\|_2^2 \right).$$

5. Repeat step 4 until convergence or a maximum number of iterations is reached.

5.5 Experimental Results

In this section, we will demonstrate the performance of the proposed method using both simulated and real-life images. In the patch database, the patch dimension of the LR and HR patches are chosen to be 4×4 and 13×13 respectively. Other patch dimensions can also be chosen. The LR patches are obtained by applying uniform blur on the HR patches, followed by a decimation factor D of 4, and additive white noise.

5.5.1 HR Reconstruction of License Images

In the first part of the experiment, the license plate image of the top right image (white Toyota) in Figure 5.3 is selected as the test image. To simulate the LR image, the HR image in Figure 5.3 is blurred by uniform blur with support of 5×5. A decimation factor of 4 is chosen and the images are degraded by additive white Gaussian noise (AWGN) to produce a signal-to-noise ratio (SNR) at 35 dB. The proposed algorithm is then run on the LR image. The result is compared with three other methods: bi-cubic interpolation, regularized interpolation method in [18] and binary learning-based method in [21]. The reconstructed HR images using the four methods are given in Figure 5.4.

Figure 5.3 Vehicle license plate images

The initial LR image and the scaled-up version of it are given in Figure 5.4(a) and (b), respectively. The result obtained using the proposed method is given in Figure 5.4(f). Comparing the result with the original HR image in Figure 5.3, it is clear that the proposed algorithm has reconstructed the HR image effectively. The results obtained using the bi-cubic interpolation, the regularized interpolation method [18], and the binary learning-based method [21] are given in Figure 5.4(c)-(e). It is observed that the bi-cubic interpolation method suffers from significant blurring, and the regularized interpolation method produces ragged edges which are unsatisfactory. The enlarged regions of the reconstructed HR images using the method in [21] and the proposed method are given in Figure 5.4(g) and (h), respectively. From the figures, it can be seen the proposed method can offer superior reconstruction. Different from the conventional

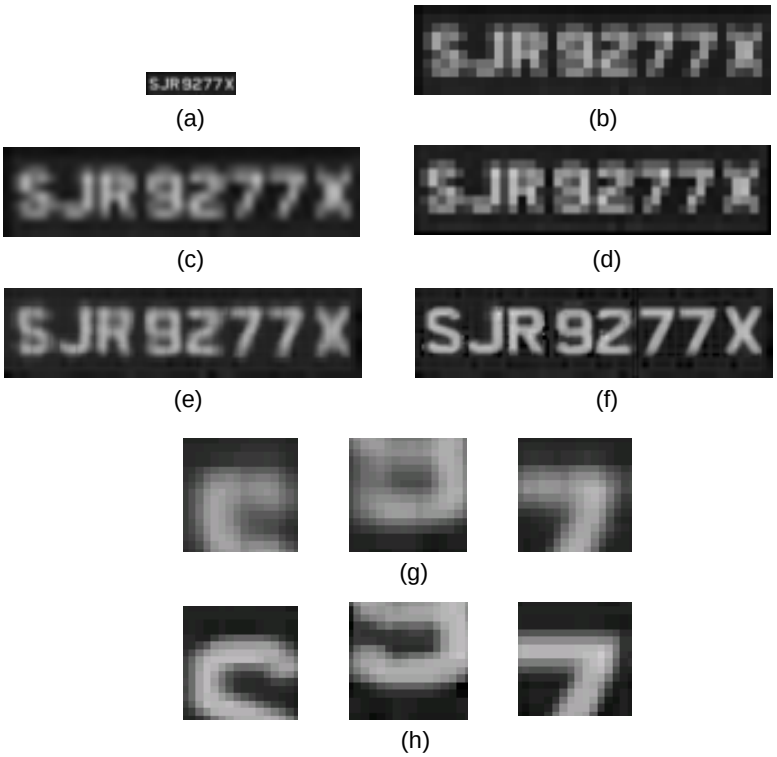


Figure 5.4 SR for license plate image. (a) LR image, (b) The scaled-up LR image, (c) Reconstructed image using the bi-cubic interpolation, (d) Reconstructed image using the regularized interpolation method in [18], (e) Reconstructed image using binary learning-based method in [21], (f) Reconstructed image using the proposed method, (g) Selected enlarged regions of (e), (h) Selected enlarged regions of (f).

methods [21] that utilize a hard-decision strategy for training patches, the proposed method adopts a soft learning prior which first estimates the importance of different learning samples, and then utilizes this information to improve the subsequent HR license plate image reconstruction. Comparison demonstrates that the proposed soft learning framework is better than the binary weighting scheme when performing VLP reconstruction.

Table 5.3 PSNR of HR image reconstruction at different noise levels

Honda

SNR	PSNR			
	Bicubic interploation	Regularized interpolation	Binary learning-based method	The proposed method
45	13.950	15.107	18.787	**19.837**
35	13.939	15.033	18.753	**19.767**
25	13.932	15.008	18.560	**19.674**

White Toyota

SNR	PSNR			
	Bicubic interploation	Regularized interpolation	Binary learning-based method	The proposed method
45	15.018	17.686	20.943	**22.622**
35	15.015	17.677	20.907	**22.481**
25	14.978	17.256	20.467	**21.703**

Hyundai

SNR	PSNR			
	Bicubic interploation	Regularized interpolation	Binary learning-based method	The proposed method
45	16.350	18.779	21.566	**22.924**
35	16.330	18.681	21.525	**22.888**
25	16.326	18.641	21.402	**22.519**

Black Toyota

SNR	PSNR			
	Bicubic interploation	Regularized interpolation	Binary learning-based method	The proposed method
45	14.556	16.744	20.498	**22.086**
35	14.548	16.646	20.476	**21.779**
25	14.544	16.611	19.831	**21.083**

5.5.2 Additional HR License Image Reconstruction Results

In order to illustrate the effectiveness of the proposed method to handle LR images degraded by different amount of noise, additive noise is added to the test images in Figure 5.3 to produce four different sets of LR images with different signal-to-noise ratios (SNRs) at 45, 35 and 25dB. The same experimental setup as in Section 5.5.1 was used to conduct the experiments. The peak SNR (PSNR) of the reconstructed HR images obtained using the four methods are given in Table 5.3. From the table, it can be seen that the PSNR of reconstructed image using the proposed method outperforms all the other three methods under different noise levels. The objective performance measures reconfirm the subjective evaluation of the reconstructed images.

We also perform experiments on a dataset consisting of 30 license plate images acquired using camera. The results of the proposed method in terms of PSNR of the reconstructed HR images are given in Table 5.4. The proposed method is compared with bicubic interpolation,

Table 5.4 Average PSNR of license plate image dataset

Bicubic interpolation	Regularized interpolation [18]	Binary learning-based method [21]	The proposed method
16.5	18.7	20.7	21.3

regularized interpolation [18] and binary learning-based method [21]. From the table, it can be observed that the PSNR of the reconstructed images using the proposed method outperforms the other methods. This demonstrates the effectiveness of the proposed soft learning method in handling vehicle license plate super-resolution reconstruction.

To further demonstrate the proposed method for HR reconstruction, we have provided the intermediate HR results of 2 LR license plate images in Figure 5.5-5.6. In Figure 5.5, the LR image is given in Figure 5.5(a). The initialized HR image after scaling up by a factor of 4 is shown in Figure 5.5(b). In iteration 2, the proposed soft learning method is performed on the image and it is observed that the HR result improves, as shown in Figure 5.5(c). In iteration 3, the HR result is given in Figure 5.5(d). The selected regions of HR images in Figure 5.5(b) and 5.5(d) are enlarged in Figure 5.5(e) and 5.5(f), respectively, for closer examination. It can be observed that the HR result by the proposed method in Figure 5.5(f) contains more detailed information and less artifacts (marked in red) when compared with the results in Figure 5.5(e). Similar observations can also be made for the next car plate image given in Figure 5.6.

Next, we will discuss the computational time of the proposed method. The binary learning-based method [21] is used as the benchmark for comparison. The experimental environments are given as follows: Windows XP, MATLAB R2010b, CPU Core Q9400-2.67GHz, and 3.25 GB RAM. The total computational time is 9.8s for our proposed method, 8.2s for binary learning-based method [21]. The computational cost of the proposed method is a little larger, due to the

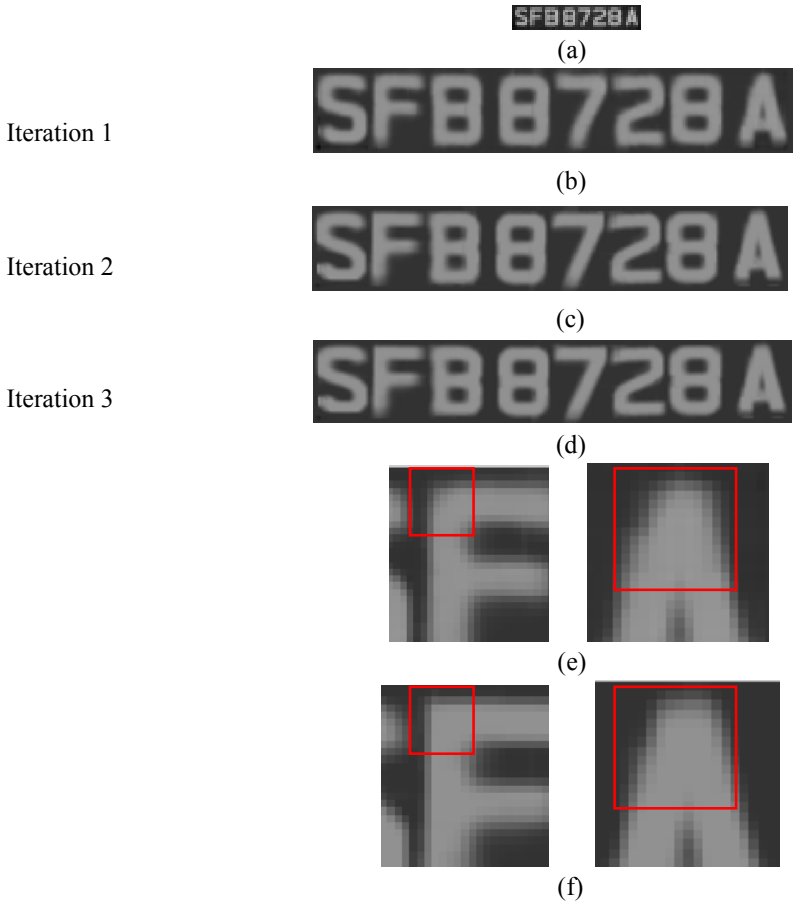


Figure 5.5 HR results for "SFB8728A" car plate image: (a) original LR image; (b)-(d) HR images of the first 3 iterations; (e)-(f) selected enlarged regions of (b) and (d), respectively.

additional calculation of soft weighting. As the current algorithm is implemented in MATLAB, the computational time can be further reduced if it is implemented in compiling language such as C.

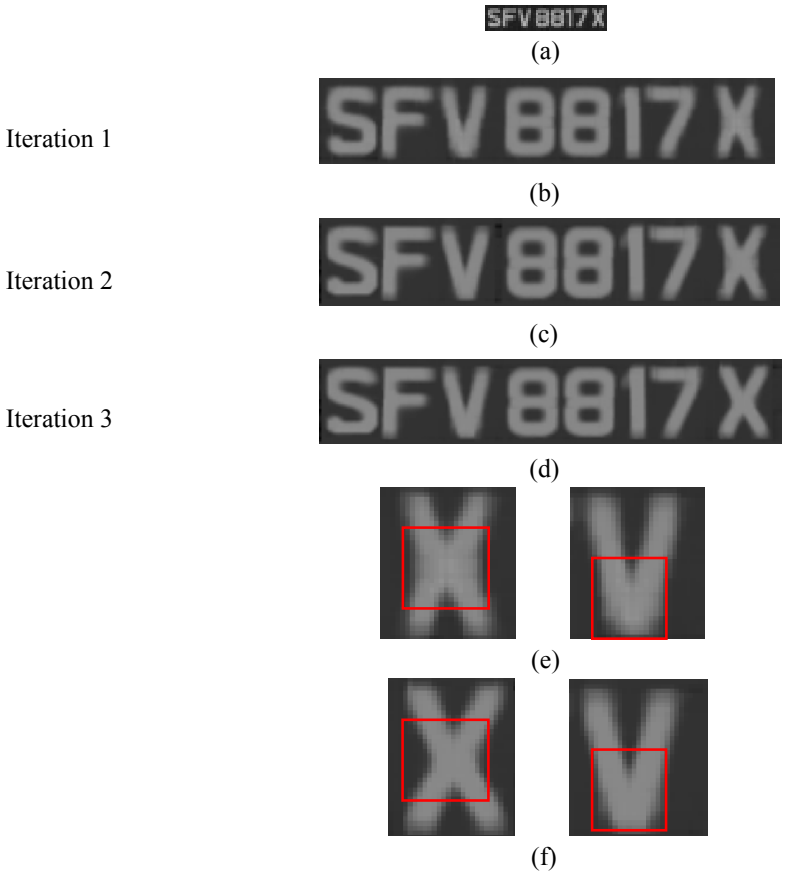


Figure 5.6 HR results for "SFV8817X" car plate image: (a) original LR image; (b)-(d) HR images of the first 3 iterations; (e)-(f) selected enlarged regions of (b) and (d), respectively.

5.5.3 Real-life License Plate HR Reconstruction

In this section, we will conduct HR reconstruction of real-life license plate images. Two VLP images are captured using a digital camera, as given in Figure 5.7 and 5.9. The size of the reconstructed HR images is set to be four times that of the captured license plate images. The reconstructed HR images using the proposed method are given in Figure 5.8 and 5.10. It can be seen from the figures that the proposed method has achieved satisfactory HR reconstruction where the overall clarity and sharpness of the license plates has been recovered. Significant amount of details has also been restored near the edges. When compared with the regularized interpolation method in [18] and binary learning-based method in [21], it is clear that the proposed approach is superior in handling real-life VLP SR reconstruction. The enlarged regions of images obtained using the proposed method also show that it manages to achieve better performance than the binary learning-based method in [21].

Figure 5.7 Car with license plate number ‘SFJ1848R’

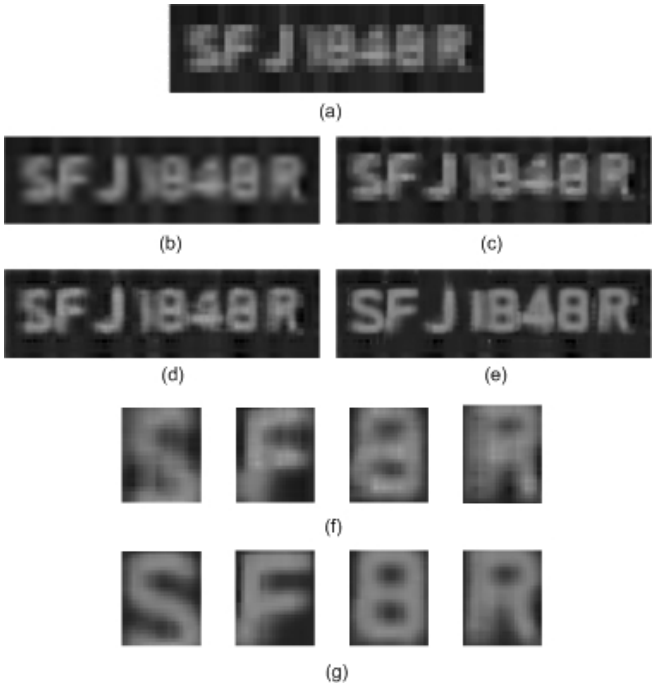


Figure 5.8 SR for first real-life license plate. (a) The scaled-up LR image, (b) Reconstructed image using the bi-cubic interpolation, (c) Reconstructed image using the regularized interpolation method in [18], (d) Reconstructed image using the binary learning-based method in [21], (e) Reconstructed image using the proposed method, (f) Selected enlarged regions of (d), (g) Selected enlarged regions of (e).

Figure 5.9 Car with plate number ‘SFN1518Y’

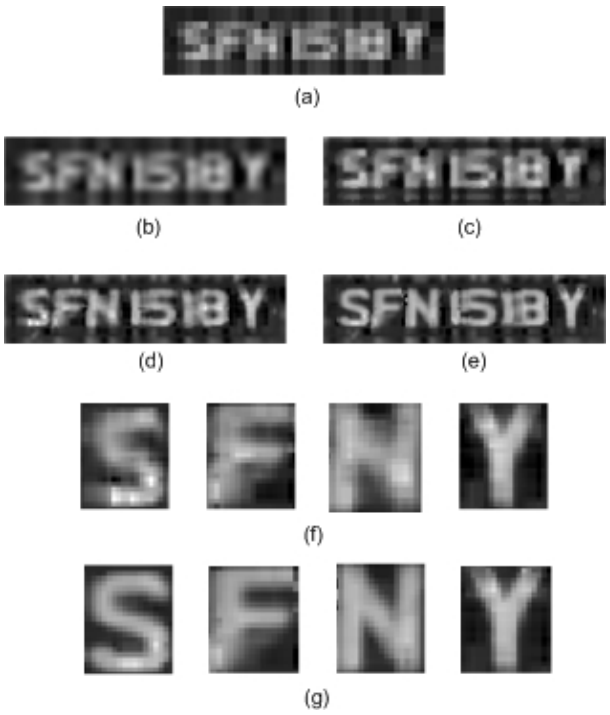


Figure 5.10 SR for second real-life license plate. (a) The scaled-up LR image, (b) Reconstructed image using the bi-cubic interpolation, (c) Reconstructed image using the regularized interpolation method in [18], (d) Reconstructed image using the binary learning-based method in [21], (e) Reconstructed image using the proposed method, (f) Selected enlarged regions of (d), (g) Selected enlarged regions of (e).

5.6 Summary

This chapter presents a new framework using a soft learning approach in single frame VLP super-resolution. As opposed to the conventional binary learning-based techniques, the method estimates the importance of the learning patches through the relevance scores obtained from the character prior information extraction. These are then incorporated as a soft prior into a new cost function. The iterative framework of HR reconstruction enables the quality of the image SR to be improved progressively. Experimental results show that the proposed method is effective in achieving good HR license plate reconstruction. The proposed method is suitable for applications including video surveillance systems of large parking areas such as in airport or shopping mall car parks. For example, if there is a car theft or explosion, the proposed algorithm can be employed to reconstruct a HR license image from the storage database to provide forensic evidence.

In this chapter, the proposed method is focused on single frame SR. In SR applications where vehicles are moving towards overheads video surveillance camera, the multi-frame SR method using a motion model consisting of translation, rotation and zooming in Chapter 4 can be used. Therefore, one possible direction is to integrate the learning-based SR in Chapter 5 with the multi-frame SR in Chapter 4 to address the issue.

Chapter 6

Joint Super-resolution and Demosaicing for Color Image Reconstruction

6.1 Introduction

In the previous chapters, we have developed three SR algorithms to reconstruct a high-resolution (HR) image from multiple or a single captured low-resolution (LR) images. Similar to most of the conventional SR methods [134], [62], [23], [28], [131], [154], [146], [137], [43], [121], [72], they are designed for gray-scale images, which can be regarded as gray-scale SR. In real-life digital imaging applications, an interesting problem is to reconstruct a HR color image from a sequence of LR color images, or namely, HR color image reconstruction. It is well-known that each pixel of a color image represents three separate data measurements for respective RGB color channels, as shown in Figure 6.1(a). In practice, however, many commercial digital cameras utilize a single CCD equipped with a CFA (Color Filter Array) due to production cost. The captured mosaiced color image and subsequent demosaiced color image are shown in Figure 6.1(b) and (c), respectively. In such cases with CFA, a process for HR color image reconstruction can be divided into two stages: (i) color demosaicing to estimate the missing color information at each pixel of the LR images followed by (ii) HR reconstruction performed on the restored full color images to reconstruct a HR color image.

(a) (b) (c)

Figure 6.1 Illustration of color image capturing process: (a) real-life scene; (b) captured mosaiced color image; (c) demosaiced full color image.

Numerous demosaicing techniques have been proposed to estimate the full resolution color image based on the mosaiced color images. The basic demosaicing method is the linear interpolation [48], [111], [7] applied to each channel of the color image. However, it tends to introduce color artifacts, especially in regions containing edges. Therefore, more sophisticated methods [65], [66], [94], [78], [142] have been developed to suppress the color artifacts based on the correlation between the G (green) channel with the R (red) channel and B (blue) channel. Currently, the color-ratio rule is broadly used to integrate inter-channel correlation during color demosaicking, which assumes that the color ratio (R/G or B/G) is smooth within a local image region. In [78], Lukac *et al*. developed a sequential demosaicing methods based on the color-ratio rule. In their work, the luminance (G) information of color image is estimated first and the missing chrominance (R or B) pixel values will be subsequently calculated from the interpolated R/G or B/G ratio, respectively. Similarly, the color-difference rule assumes that the color difference of R–G and B–G has small variation in a local region. Recently, an applications based on color-difference rule was proposed by Li in [142]. After obtaining the demosaiced LR color images, conventional gray-scale SR methods can be applied on RGB channels to calculate the

respective HR images. Then, the final color image is obtained by combining the HR images of three color channels. Although the two-stage direct process could produce HR color image, this process tends to be suboptimal as the preprocessing of demosaicing is performed on the mosaiced color image with low resolution. The aliasing effects in the LR images may degrade the performance of demosaicing and introduce color artifacts, which could be further propagated to the subsequent HR reconstruction. Moreover, the conventional SR methods treat a color image as the combination of three separate color channels of gray-scale images. They neglect color correlation that may lead to inconsistency of three separate color channels during HR reconstruction.

There have been several techniques to address the color SR problems in the literature [71], [36], [25], [6], [53]. It is noted that color artifacts often occur in the regions consisting of high frequency details, such as edges. Therefore, total variation (TV) technique using the color correlation is a good candidate due to its edge-preserving property. The underlying principle is that the regularization of each color channel benefits from the information of the other two channels. Based on this idea, Chan *et al*. extended the classical scalar TV norm to the vector-valued color image in their work [6]. Later, they developed a nonlocal regularizer based on Mumford-Shah model in [53]. These intensity-based methods, however, may fail to constrain iso-luminance edges where there is a "jump" in color. Therefore, the recent developed method [25] applied color correction technique [59] to keep the color consistency in a local region. This approach works well in smooth region but may blur the edges as it does not take full advantage of the edge direction information.

In this chapter, we propose a new framework to jointly perform color demosaicing and image SR from the mosaiced LR color images directly. Different from conventional two-stage process of color demosaicing followed by SR, the proposed method combines color SR and demosaicing into a single process. A color correction term is incorporated into the proposed cost

function to ensure consistency of different color channels and alleviate color artifacts. Further, an adaptive edge-directed bilateral total variation (BTV) regularization term is developed to preserve the high frequency information in the reconstructed HR image. An iterative framework is developed that enables the reconstructed HR image to be improved progressively. Experimental results show the proposed method is effective to perform color SR.

The organization of this chapter is outlined as follows. Section 6.2 presents the model description and problem formulation. In Section 6.3, the new proposed cost function is introduced. The details of the iterative scheme are explained in Section 6.4. Experimental results are provided in Section 6.5 and conclusions are given in Section 6.6.

6.2 Problem Formulation

In this work, we adopt a motion model consisting of translation, rotation and zooming to describe the relationship of LR images. The generation of the *k*th ($1 \le k \le N$) color mosaiced LR image $g_k(u,v,i)$ (i=*R,G,B*) represented in the RGB color space is shown in Figure 6.2(a). Based on the image acquisition process, the *i*th (i=*R,G,B*) channel of LR image $g_{ki}(u,v)$ of size $M_g \times N_g$ can be modeled by zooming the *i*th channel of the original color HR image $f_i(x,y)$ of size $M_f \times N_f$ by a zooming factor l_k, rotating it by an angle θ_k, shifting it by a translational vector $[s_{xk}, s_{yk}]$, blurring the zoomed, rotated and shifted HR image $f_{ki}(x,y)$ by a point-spread function (PSF) $h_k \otimes h_c$ and down-sampling the result by a downsampling operator $\downarrow_\rho$. Then, the LR full color image is subsampled by the Bayer filter. Finally, the mosaiced LR image is degraded by additive white Gaussian noise (AWGN) $n_{ki}(u,v)$. Thus, the degradation process can be modeled as

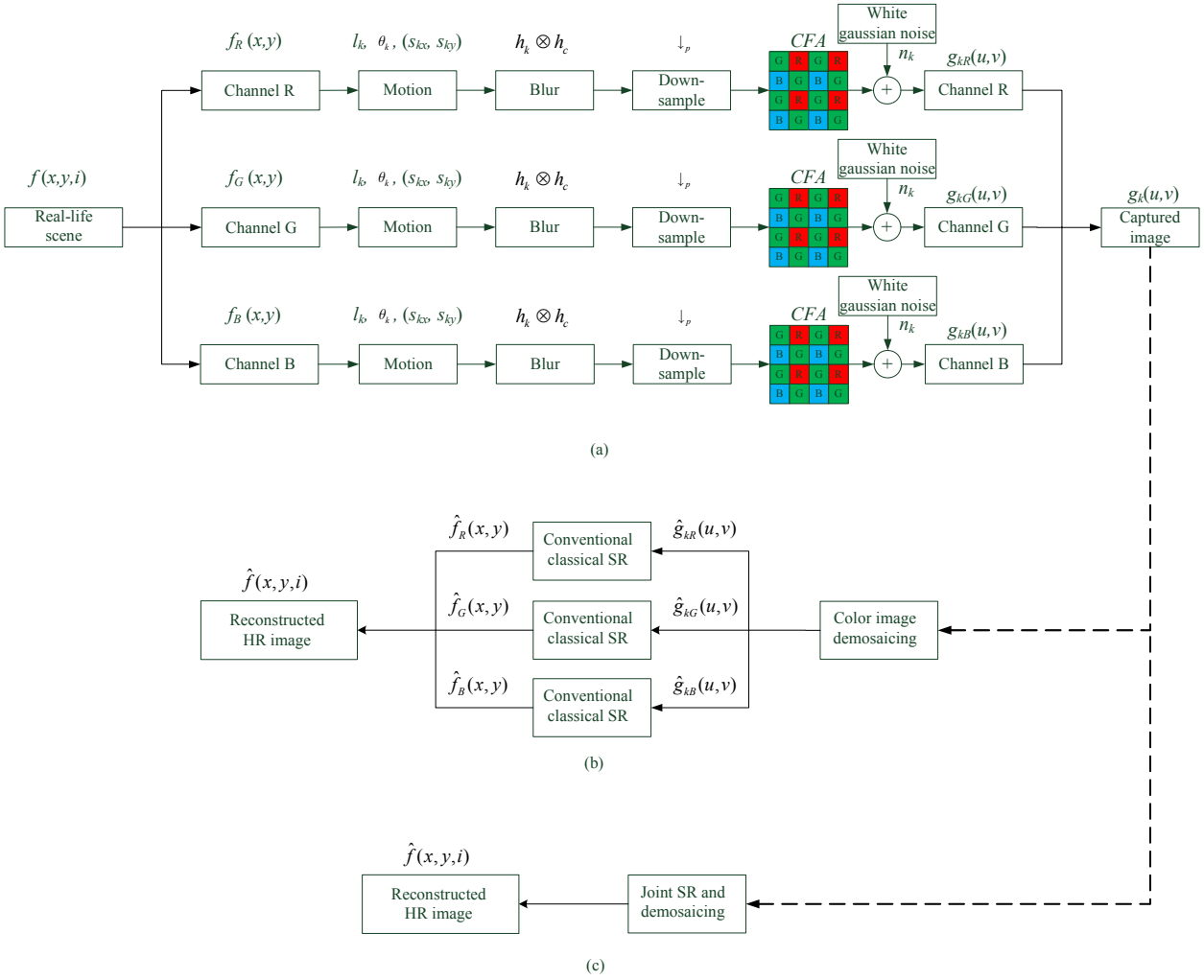


Figure 6.2 Color image capturing and SR reconstruction processes: (a) the generation of LR CFA images; (b) conventional SR processing; (c) the proposed joint SR and demosaicing approach.

$$\begin{aligned} g_{ki}(u,v) &= \left(f_i(l_k x\cos\theta_k - l_k y\sin\theta_k + s_{xk}, l_k x\sin\theta_k + l_k y\cos\theta_k + s_{yk}) \otimes h_k \otimes h_c\right)\downarrow_\rho + n_{ki} \\ &= \left(f_{ki}(x,y) \otimes h_k \otimes h_c\right)\downarrow_\rho + n_{ki}(u,v) \end{aligned} \tag{6.1}$$

where $\otimes$ is the 2D convolution operator. h_k represents the blur operator related to camera lens while h_c represents the blur operator due to a square-region integration of light intensity. Here, h_c is denoted as a $\rho \times \rho$ size uniform PSF.

As in our work, we assume the lens condition (h_k) is known and identical for all LR images. Then using the matrix-vector form, the imaging process in (6.1) can be expressed as

$$\boldsymbol{g}_k = \boldsymbol{D}_c \boldsymbol{D} \boldsymbol{H}_k \boldsymbol{S}(\boldsymbol{\alpha}_k) \boldsymbol{f} + \boldsymbol{n}_k \tag{6.2}$$

where $\boldsymbol{f} = [\boldsymbol{f}_R{}^T, \boldsymbol{f}_G{}^T, \boldsymbol{f}_B{}^T]^T$, $\boldsymbol{g}_k = [\boldsymbol{g}_{kR}{}^T, \boldsymbol{g}_{kG}{}^T, \boldsymbol{g}_{kB}{}^T]^T$ and $\boldsymbol{n}_k = [\boldsymbol{n}_{kR}{}^T, \boldsymbol{n}_{kG}{}^T, \boldsymbol{n}_{kB}{}^T]^T$ are the vectors which represent the discrete, concatenated and lexicographically ordered f_i, g_{ki} and n_{ki}, respectively. The matrix $\boldsymbol{D}_c$ represents the CFA operator and matrix $\boldsymbol{D}$ denotes the identical down-sampling operator for all LR images. $\boldsymbol{H}_k (1 \le k \le N)$ denotes the corresponding matrix representing the blurring operator $h_k \otimes h_c$. The matrix $\boldsymbol{S}(\boldsymbol{\alpha}_k) (1 \le k \le N)$ denotes the geometric motion operator for the *k*th color image, where $\boldsymbol{\alpha}_k$ represents the motion parametric vector of the *k*th color image and $\boldsymbol{\alpha}_k = [l_k \cos\theta_k, l_k \sin\theta_k, s_{xk}, s_{yk}]^T$. In this chapter, the purpose of SR reconstruction is to fuse *N* LR mosaiced color observations to produce a HR full-color image $\boldsymbol{f}$ of the original scene.

Based on the generation of the mosaiced LR image $\boldsymbol{g}_k$, it is noted that each pixel represents the intensity of only one color channel. Conventional SR approaches on these raw LR images may not produce satisfactory results due to the missing pixels in each color channel. Therefore, we will introduce a new cost function to address this problem in the next section.

6.3 Proposed Cost Function

Generally, image SR on a sequence of color LR images captured by a single CCD equipped with a CFA filter can be considered as a two-stage process, as shown in Figure 6.2(b). In this process, a major problem is the color artifacts existing in the demosaiced full resolution color LR images $\hat{\boldsymbol{g}}_k$, which may lead to suboptimal result in the subsequent SR process. To address this problem, in this work, we will develop a uniform framework combining color demosaicing and color SR into a single process. The proposed cost function consists of three components of the data fidelity term of the original raw mosaiced LR images, an adaptive edge-directed regularization term and a color correction term.

Given the image formation model in (6.1), the estimated color HR image $\boldsymbol{f}$ can be obtained by minimizing the following cost function:

$$\boldsymbol{E}(\boldsymbol{f}) = \left\| \boldsymbol{g} - \boldsymbol{D}_c \boldsymbol{D}\boldsymbol{H}\boldsymbol{S}(\boldsymbol{a})\boldsymbol{f} \right\|^2 + \lambda \boldsymbol{T}(\boldsymbol{f}) + \beta \boldsymbol{C}(\boldsymbol{f}) \tag{6.3}$$

where $\boldsymbol{g} = [\boldsymbol{g}_1^T, \cdots, \boldsymbol{g}_N^T]^T$, $\boldsymbol{H} = [\boldsymbol{H}_1^T, \cdots, \boldsymbol{H}_N^T]^T$ and $\boldsymbol{S}(\boldsymbol{a}) = \left[\boldsymbol{S}(\boldsymbol{a}_1)^T, \boldsymbol{S}(\boldsymbol{a}_2)^T, \cdots, \boldsymbol{S}(\boldsymbol{a}_N)^T\right]^T$. The matrix $\boldsymbol{S}_k(\boldsymbol{a}_k)$ represents the motion operator consisting of zooming, rotation and translation. The first term in (6.3) is the data fidelity term, while the second term in (6.3) is the regularization functional that introduces stability into the estimate of HR image $\boldsymbol{f}$. Here, we adopt an edge-directed BTV regularization functional for the HR image. λ is the regularization parameter that serves to control the relative contribution of the regularization term. The third term is the color correction term utilized to suppress the color artifacts in the final HR image and β is the regularization parameter for it.

6.4 Iterative Framework for HR Color Image Reconstruction

6.4.1 Framework Overview

The overview of the proposed iterative color SR reconstruction framework is shown in Figure 6.3. It contains three major steps of adaptive edge-based calculation of regularization term, calculation of color correction term and color HR reconstruction. Firstly, the initial HR image $\boldsymbol{f}^{(0)}$ is obtained by the shift-and-add result of the demosaiced LR images after bilinear interpolation. To preserve the high frequency details of the desired HR image, the initial estimate $\boldsymbol{f}^{(0)}$ will then be passed into the edge-based regularization construction step, where the edge direction information of $\boldsymbol{f}^{(0)}$ is explored and employed to construct the adaptive regularization term. Furthermore, an improved color correction term using the estimated edge information is developed to induce consistency of different color channels. The HR reconstruction of joint color SR and demosaicing is conducted by minimizing the proposed cost function using preconditioned conjugate gradient (PCG) optimization. The reconstructed HR image will be utilized to construct the regularization and color correction terms in the next iteration. The iterative process will continue so that progressively improved color HR image will be obtained. The process will continue until convergence or a maximum number of iterations is reached.

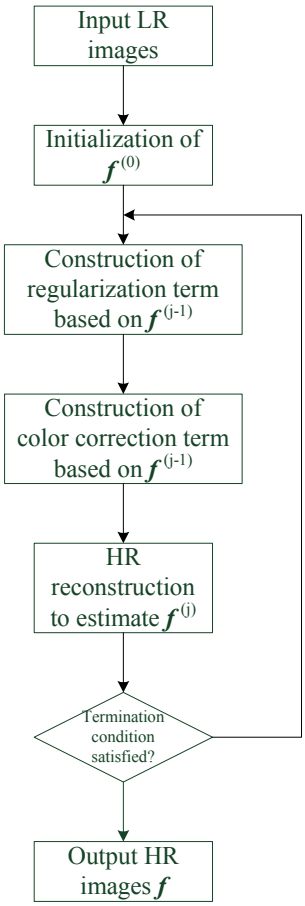


Figure 6.3 Overview of the proposed iterative framework.

6.4.2 Derivation of the Edge-directed Regularization Term

It is well-known that image SR is an ill-posed inverse problem that requires regularization. Various regularization techniques [39], [28], [119] have been developed and applied in different applications. One of the popular techniques is total variation (TV), which is often utilized to preserve the sharp features of the reconstructed HR image [28], [121], [6], [124], [11], [153], [118]. In an edge region, it is noted that the intensity gradients across the edge are larger than those along the edge. However, conventional TV penalizes intensity gradients in all directions equally. The equal weighting strategy often over-penalizes the pixels across the edges. In this work, we propose an adaptive edge-directed regularization term which takes this into consideration.

The proposed adaptive TV regularization term of the desired HR color image $\boldsymbol{f}$ can be expressed as

$$\begin{aligned} \boldsymbol{T}(\boldsymbol{f}) &= \int_{\Omega} \left(\omega_R \left| \nabla \boldsymbol{f}_R \right| + \omega_G \left| \nabla \boldsymbol{f}_G \right| + \omega_B \left| \nabla \boldsymbol{f}_B \right| \right) dxdy \\ &= \omega_R \int_{\Omega} \left| \nabla \boldsymbol{f}_R \right| dxdy + \omega_G \int_{\Omega} \left| \nabla \boldsymbol{f}_G \right| dxdy + \omega_B \int_{\Omega} \left| \nabla \boldsymbol{f}_B \right| dxdy \end{aligned} \tag{6.4}$$

where $\boldsymbol{f} = [\boldsymbol{f}_R{}^T, \boldsymbol{f}_G{}^T, \boldsymbol{f}_B{}^T]^T$. $(\omega_R, \omega_G, \omega_B)$ are the weighting coefficients for RGB channels, respectively. As the G channel contributes more intensity perceived by humans' eye than R/B channel, we select $(\omega_R, \omega_G, \omega_B)$ as (0.299, 0.587, 0.114) similar to the idea in [98]. Because of the incorporation of TV norm $\boldsymbol{T}(\boldsymbol{f})$, the minimization of (6.3) requires solving the nonlinear partial equations (PDEs). To alleviate this difficulty, a L_2 norm approximation for $\boldsymbol{T}(\boldsymbol{f})$ is developed. Selecting four directions of 0°, 45°, 90°, and 135° as the primary edge directions, $\int_{\Omega} \left| \nabla \boldsymbol{f}_i \right| dxdy$ (i=R, G and B) in (6.4) can be calculated as

$$\int_{\Omega} \left| \nabla \boldsymbol{f}_i \right| dxdy = \int_{\Omega} \frac{\left\| \nabla \boldsymbol{f}_i \right\|^2}{\left| \nabla \boldsymbol{f}_i \right|} dxdy = \sum_{x,y} \sum_{m=-1}^{1} \sum_{n=0}^{1} \frac{w_{m,n}}{\mu_{m,n}} (f_i(x+m, y+n) - f_i(x,y))^2 \tag{6.5}$$

where $\mu_{m,n}$ is defined as $\mu_{m,n} = \sqrt{m^2 + n^2} \sqrt{(f_i(x+m, y+n) - f_i(x,y))^2 + \gamma}$. $\gamma > 0$ is a small constant to ensure the term is non-zero. $w_{m,n}$ represents the weighting factor that indicates regularizations of the intensity gradients of the image $\boldsymbol{f}_i$ in four primary edge directions. An example of the adaptive weighting strategy for image pixels is shown in Figure 6.4. At each pixel location, the length of lines along each primary direction denotes the degree of smoothness regularization imposed in the corresponding direction. It is noted that the regularization of all 4 directions are almost identical for pixels in the smooth region while the regularization along the edge direction are significantly higher at pixels near the edge.

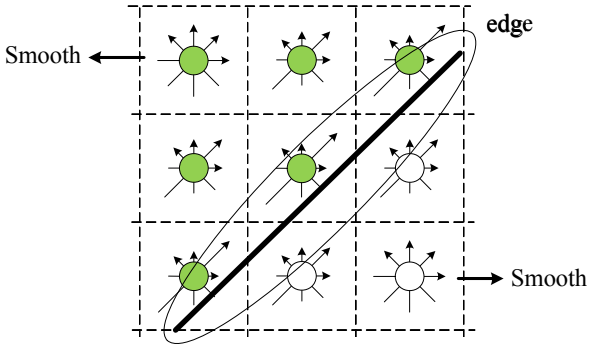


Figure 6.4 Illustration of the smoothness regularization for an edge in four directions.

We employ the normalized weighting strategy to estimate $w_{m,n}$ as follows:

$$w_{m,n} = \frac{\eta_{m,n}}{\frac{1}{4}\sum_{m=-1}^{1}\sum_{n=0}^{1}\eta_{m,n}} \tag{6.6}$$

where $\eta_{m,n} = \frac{\sqrt{m^2+n^2}}{\sqrt{(f_i(x+m,y+n)-f_i(x,y))^2+\gamma}}$ represents the reciprocal of intensity gradient of the HR image. It is noted that the intensity gradients of four primary edge directions are similar for pixels at smooth regions while the intensity gradient along the edge direction is lower than that across the edge direction for pixels at the edge regions. Therefore, the estimated $w_{m,n}$ of four directions are similar for pixels at smooth regions while $w_{m,n}$ along the edge direction is higher than that across the edge direction for pixels at the edge regions.

The fixed-point (FP) technique [14] is adopted to calculate the weights $w_{m,n}$ and $\mu_{m,n}$ by using the HR image $\boldsymbol{f}_i$ estimated in the previous iteration. Thus, the right-hand side of (6.5) can be expressed in a matrix-vector form as

$$\begin{aligned}
&\sum_{x,y}\sum_{m=-1}^{1}\sum_{n=0}^{1}\frac{w_{m,n}}{\mu_{m,n}}(f_i(x+m,y+n)-f_i(x,y))^2 \\
&=\sum_{m=-1}^{1}\sum_{n=0}^{1}\sum_{x,y}\frac{w_{m,n}}{\mu_{m,n}}(f_i(x+m,y+n)-f_i(x,y))^2 \\
&=\sum_{m=-1}^{1}\sum_{n=0}^{1}\boldsymbol{f}_i^T\boldsymbol{U}_{m,n}^T\boldsymbol{V}_{m,n}\boldsymbol{U}_{m,n}\boldsymbol{f}_i \\
&=\boldsymbol{f}_i^T\boldsymbol{B}_w\boldsymbol{f}_i
\end{aligned} \tag{6.7}$$

where matrix $\boldsymbol{U}_{m,n}$ represents the first-order derivative operator on the desired HR image. Matrix $\boldsymbol{V}_{m,n}=diag\{\boldsymbol{c}\}$ and vector $\boldsymbol{c}$ represents the lexicographically ordered $w_{m,n}/\mu_{m,n}$. $\boldsymbol{B}_w=\sum_{m=-1}^{1}\sum_{n=0}^{1}\boldsymbol{U}_{m,n}^T\boldsymbol{V}_{m,n}\boldsymbol{U}_{m,n}$. Finally, substituting (6.7) into (6.4), $\boldsymbol{T}(\boldsymbol{f})$ can be rewritten as

$$\boldsymbol{T}(\boldsymbol{f})=\boldsymbol{f}^T\boldsymbol{B}\boldsymbol{f}=diag\{\omega_R\boldsymbol{f}_R{}^T\boldsymbol{B}_w\boldsymbol{f}_R,\ \omega_G\boldsymbol{f}_G{}^T\boldsymbol{B}_w\boldsymbol{f}_G,\ \omega_B\boldsymbol{f}_B{}^T\boldsymbol{B}_w\boldsymbol{f}_B\} \tag{6.8}$$

where $\boldsymbol{f}=[\boldsymbol{f}_R{}^T,\boldsymbol{f}_G{}^T,\boldsymbol{f}_B{}^T]^T$ and $\boldsymbol{B}=diag\{\omega_R\boldsymbol{B}_w,\ \omega_G\boldsymbol{B}_w,\ \omega_B\boldsymbol{B}_w\}$.

6.4.3 Construction of the Color Correction Term

In this section, we will develop a color correction term to ensure the consistency of different color channels. The principle underlying our proposed color correction term is that the adjacent pixels in a local region of color image should have similar directions in 3D Euclidean space $\boldsymbol{R}^3$. Therefore, color artifacts can be minimized by inducing the color vectors to have similar directions within the neighborhood.

Let $P=[p_r,p_g,p_b]^T$ and $Q=[q_r,q_g,q_b]^T$ denote the color values of two adjacent pixels. The pixel-level color correction term for P and Q can be modeled as

$$\|P\times Q\|^2=\left[|P||Q|\sin(\vartheta)\right]^2=\left\|[p_gq_b-p_bq_g,\ p_bq_r-p_rq_b,\ p_rq_g-p_gq_r]^T\right\|^2 \tag{6.9}$$

where ϑ represents the angle between these two vectors and small angle indicates similar direction.

Although this assumption holds well with pixels located in smooth regions, it tends to over-penalize those located across the edge direction and lead to blurring effects around the edge regions. Unlike Farsiu's method [25] that treats the adjacent pixels located in all directions with equal emphasis, our color correction term takes the edge direction information into consideration. For instance, in a local region containing an edge, the adjacent pixels located along the edge direction will be given larger weight to ensure the color consistency along the edge direction. Combined with the edge weighting $w_{m,n}$ estimated in the previous section, the proposed color correction term can be expressed in a matrix-vector form as

$$\boldsymbol{C}(\boldsymbol{f}) = \sum_{m=-1}^{1}\sum_{n=0}^{1}\left(\left\|\boldsymbol{W}_{m,n}(\boldsymbol{f}_R \odot S_x^m S_y^n \boldsymbol{f}_G - \boldsymbol{f}_G \odot S_x^m S_y^n \boldsymbol{f}_R)\right\|^2 + \left\|\boldsymbol{W}_{m,n}(\boldsymbol{f}_G \odot S_x^m S_y^n \boldsymbol{f}_B - \boldsymbol{f}_B \odot S_x^m S_y^n \boldsymbol{f}_G)\right\|^2 + \left\|\boldsymbol{W}_{m,n}(\boldsymbol{f}_B \odot S_x^m S_y^n \boldsymbol{f}_R - \boldsymbol{f}_R \odot S_x^m S_y^n \boldsymbol{f}_B)\right\|^2\right) \tag{6.10}$$

where $\odot$ represents the element-by-element multiplication operator and $\boldsymbol{W}_{m,n}$ is a diagonal matrix consisting of lexicographically ordered edge weights $w_{m,n}$ described in Section 6.4.2.

Using the commutative rule of element multiplication operator, (6.10) can be rewritten as

$$\boldsymbol{C}(\boldsymbol{f}) = \sum_{m=-1}^{1}\sum_{n=0}^{1}\left(\left\|\boldsymbol{W}_{m,n}(S_x^m S_y^n \boldsymbol{f}_G \odot \boldsymbol{f}_R - \boldsymbol{f}_G \odot S_x^m S_y^n \boldsymbol{f}_R)\right\|^2 + \left\|\boldsymbol{W}_{m,n}(S_x^m S_y^n \boldsymbol{f}_B \odot \boldsymbol{f}_G - \boldsymbol{f}_B \odot S_x^m S_y^n \boldsymbol{f}_G)\right\|^2 + \left\|\boldsymbol{W}_{m,n}(S_x^m S_y^n \boldsymbol{f}_R \odot \boldsymbol{f}_B - \boldsymbol{f}_R \odot S_x^m S_y^n \boldsymbol{f}_B)\right\|^2\right) \tag{6.11}$$

We can substitute the element-by-element multiplication operator " $\odot$ " with the differentiable dot product by rearranging $\boldsymbol{f}_i\,(i = R, G, B)$ as diagonal matrix $\boldsymbol{F}_i\,(i = R, G, B)$

and $S_x^m S_y^n \boldsymbol{f}_i$ ($i = R, G, B$) as $\boldsymbol{F}_i^{m,n}$ ($i = R, G, B$), which is the diagonal form of shifted $\boldsymbol{F}_i (i = R, G, B)$ by m and n pixels in the horizontal and vertical directions. Then, (6.11) can be rewritten as

$$\begin{aligned}\boldsymbol{C}(\boldsymbol{f}) &= \sum_{m=-1}^{1}\sum_{n=0}^{1}\Big(\left\|\boldsymbol{W}_{m,n}(\boldsymbol{F}_G^{\,m,n}\boldsymbol{f}_R - \boldsymbol{F}_G S_x^m S_y^n \boldsymbol{f}_R)\right\|^2 + \left\|\boldsymbol{W}_{m,n}(\boldsymbol{F}_B^{\,m,n}\boldsymbol{f}_G - \boldsymbol{F}_B S_x^m S_y^n \boldsymbol{f}_G)\right\|^2 \\ &\qquad + \left\|\boldsymbol{W}_{m,n}(\boldsymbol{F}_R^{\,m,n}\boldsymbol{f}_B - \boldsymbol{F}_R S_x^m S_y^n \boldsymbol{f}_B)\right\|^2\Big) \\ &= \sum_{m=-1}^{1}\sum_{n=0}^{1}\left\|\begin{bmatrix}\boldsymbol{W}_{m,n}(\boldsymbol{F}_G^{\,m,n} - \boldsymbol{F}_G S_x^m S_y^n)\boldsymbol{f}_R \\ \boldsymbol{W}_{m,n}(\boldsymbol{F}_B^{\,m,n} - \boldsymbol{F}_B S_x^m S_y^n)\boldsymbol{f}_G \\ \boldsymbol{W}_{m,n}(\boldsymbol{F}_R^{\,m,n} - \boldsymbol{F}_R S_x^m S_y^n)\boldsymbol{f}_B\end{bmatrix}\right\|^2 \\ &= \sum_{m=-1}^{1}\sum_{n=0}^{1}\|\boldsymbol{\Lambda}\boldsymbol{f}\|^2\end{aligned} \tag{6.12}$$

where $\boldsymbol{\Lambda} = diag\left(\boldsymbol{W}_{m,n}(\boldsymbol{F}_G^{\,m,n} - \boldsymbol{F}_G S_x^m S_y^n), \boldsymbol{W}_{m,n}(\boldsymbol{F}_B^{\,m,n} - \boldsymbol{F}_B S_x^m S_y^n), \boldsymbol{W}_{m,n}(\boldsymbol{F}_R^{\,m,n} - \boldsymbol{F}_R S_x^m S_y^n)\right)$ and $\boldsymbol{f} = [\boldsymbol{f}_R^T, \boldsymbol{f}_G^T, \boldsymbol{f}_B^T]^T$.

6.4.4 Optimization Procedure

Incorporating the adaptive regularization term and color correction term into (6.3), the current HR image $\boldsymbol{f}^{(j)}$ can be estimated by minimizing the proposed cost function:

$$\hat{\boldsymbol{f}}^{(j)} = \arg\min_{\boldsymbol{f}}\left(\left\|\boldsymbol{g} - \boldsymbol{D}_c\boldsymbol{D}\boldsymbol{H}\boldsymbol{S}(\boldsymbol{a})\boldsymbol{f}\right\|_2^2 + \lambda\boldsymbol{f}^T\boldsymbol{B}\boldsymbol{f} + \beta\sum_{m=-1}^{1}\sum_{n=0}^{1}\boldsymbol{f}^T\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}\boldsymbol{f}\right). \tag{6.13}$$

This is equivalent to solving for $\hat{\boldsymbol{f}}^{(j)}$ using the following linear equation:

$$\boldsymbol{A}\hat{\boldsymbol{f}}^{(j)} = \boldsymbol{b} \tag{6.14}$$

where $\boldsymbol{A} = \boldsymbol{S}(\boldsymbol{a})^T\boldsymbol{H}^T\boldsymbol{D}^T\boldsymbol{D}_c{}^T\boldsymbol{D}_c\boldsymbol{D}\boldsymbol{H}\boldsymbol{S}(\boldsymbol{a}) + \lambda\boldsymbol{B} + \beta\sum_{m=-1}^{1}\sum_{n=0}^{1}\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}$ and $\boldsymbol{b} = \boldsymbol{S}(\boldsymbol{a})^T\boldsymbol{H}^T\boldsymbol{D}^T\boldsymbol{D}_c{}^T\boldsymbol{g}$.

In order to solve the optimization problem in (6.3), preconditioned conjugate gradient (PCG) optimization is employed. PCG utilizes conjugate direction instead of local gradient to search for the minima. Therefore, it can achieve faster convergence when compared with steepest descent method. It also requires less storage requirement and computation complexity when compared with Quasi-Newton method.

The mathematical formulation of the proposed joint image demosaicing and SR based on preconditioned conjugate gradient optimization is given as follows.

1) Initialize the residual vector $\boldsymbol{r}$ and conjugate vector $\boldsymbol{p}$

$$\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{f}^{(j-1)} \tag{6.15}$$

$$\boldsymbol{p}_0 = \boldsymbol{M}^{-1}\boldsymbol{r}_0 \tag{6.16}$$

where $\boldsymbol{M}$ is a symmetric positive-definite matrix to ensure fast convergence of PCG.

2) Update the $(i+1)$th iteration estimate vector $\boldsymbol{f}^{(j)}{}_{i+1}$ and residual vector $\boldsymbol{r}_{i+1}$

$$\boldsymbol{f}^{(j)}{}_{i+1} = \boldsymbol{f}^{(j)}{}_{i} + \xi_i \boldsymbol{p}_i \tag{6.17}$$

$$\boldsymbol{r}_{i+1} = \boldsymbol{r}_i - \xi_i \boldsymbol{A}\boldsymbol{p}_i \tag{6.18}$$

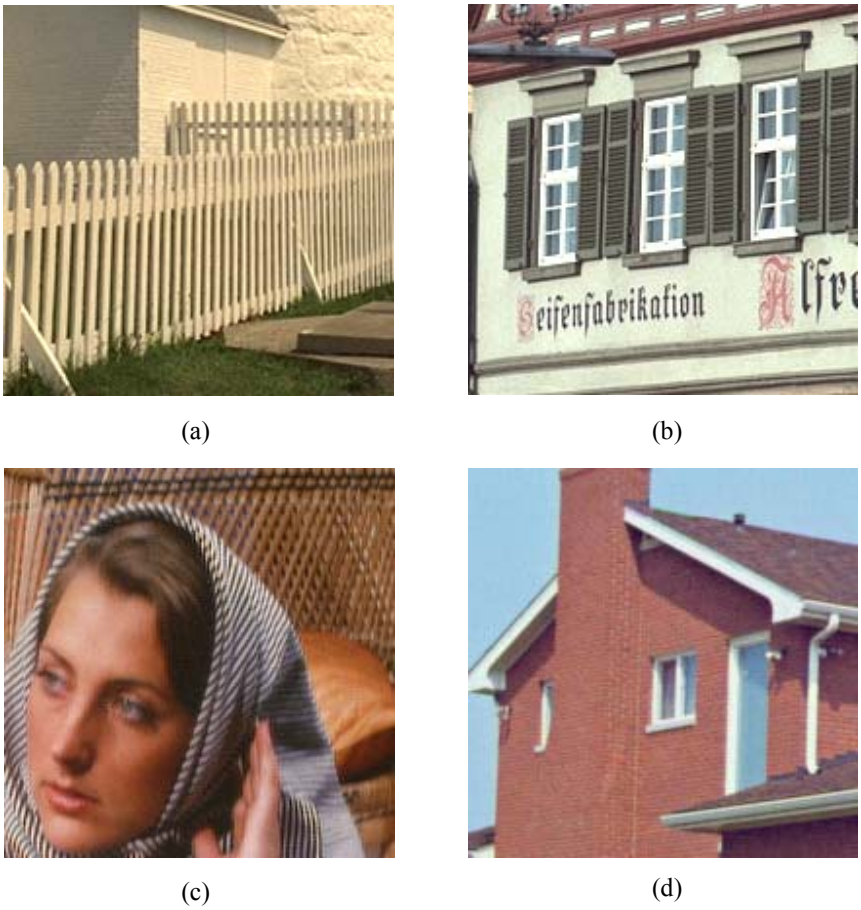
where $\xi_i = \dfrac{\boldsymbol{r}_i^T \boldsymbol{M}^{-1} \boldsymbol{r}_i}{\boldsymbol{p}_i^T \boldsymbol{A}\boldsymbol{p}_i}$.

3) Update the $(i+1)$th iteration conjugate vector

$$\boldsymbol{p}_{i+1} = \boldsymbol{M}^{-1}\boldsymbol{r}_{i+1} + \zeta_i \boldsymbol{p}_i \tag{6.19}$$

where $\zeta_i = \dfrac{\boldsymbol{r}_{i+1}^T \left(\boldsymbol{M}^{-1}\right)^T \boldsymbol{r}_{i+1}}{\boldsymbol{r}_i^T \left(\boldsymbol{M}^{-1}\right)^T \boldsymbol{r}_i}$.

4) Repeat steps 2) to 3) until convergence or a maximum number of iteration is reached.



(a) (b)

(c) (d)

Figure 6.5 4 Test images: (a) “Lighthouse” image; (b) “Window” image; (c) “Barbara” image; (d) “House” image.

6.5 Experimental Results

In our experiments, 4 benchmark images in Figure 6.5 are chosen as test images. They are chosen because they contain a significant amount of color texture, which can be used to evaluate different color SR methods. We conduct various experiments and compare the results obtained by the proposed method with other four methods, namely, two-stage method of image demosaicing in [142] followed by conventional gray-scale SR in [23], Color TV SR in [6],

Jung's method in [53] and Farsiu' method in [25]. To demonstrate the effectiveness of the methods, we use subjective visual evaluation along with the objective metrics PSNR.

6.5.1HR Reconstruction on the Natural Color Image

In this section, the "Window" image in Figure 6.5 is selected as the test image. The size of the test image is 200×200. Based on the LR imaging model in (6.1), 8 LR images were generated with zooming factors randomly selected from a uniform distribution over [0.9, 1.1], rotation angles randomly selected from a uniform distribution over [-5°, 5°], and translations randomly selected from a uniform distribution over [-3, 3] pixels. The first LR image was set as the reference. The shifted HR image was then blurred by a 3×3 Gaussian blur, followed by a down-sampling operator with the decimation factor of $\rho=2$. Subsequently, each LR image was subsampled by the Bayer filter to produce the CFA image. Finally, these LR images were degraded by additive white Gaussian noise (AWGN) to produce a signal-to-noise ratio (SNR) at 35dB.

In order to provide a fair comparison, we conducted various experiments to select the regularization parameter for our proposed method. We have selected the regularization parameters λ ranging from 1×10^{-6} to 0.1. The PSNR of the reconstructed images using our proposed method is given in Figure 6.6. From the figure, it can be seen that the best PSNR using our method is obtained at around $\lambda=1\times10^{-4}$. Next, similar experiments were repeated to select the best regularization parameters for other three comparative methods. The highest PSNRs are obtained at about $\lambda=1\times10^{-4}$ for Color TV SR method [6], Jung's method [53] and Farsiu's method [25], while $\lambda=1\times10^{-3}$ for two-stage approach [142]-[23]. For parameter β in our proposed method, we follow the idea in [25] to select β as 100λ. In the following experiments, for fair comparison, we will use the best regularization parameters for our method and the other comparative methods, respectively.

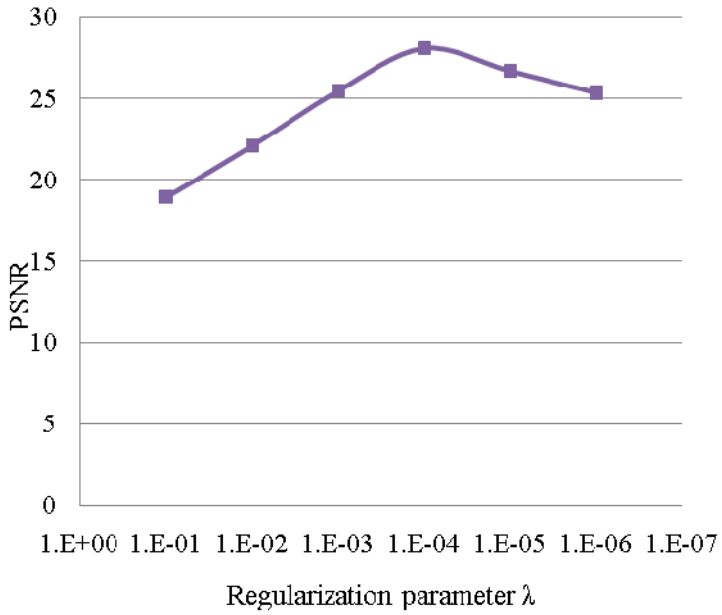


Figure 6.6 PSNR of reconstructed "Window" images against various regularization parameters.

4 samples of the CFA mosaiced LR images are shown in Figure 6.7(a). Firstly, a two-stage approach of demosaicing followed by SR was performed on the LR images. To estimate the full resolution demosaiced color image, Li's demosaicing method in [142] was adopted. The scaled-up version and the bicubic interpolated result of the demosaiced reference LR image are shown in Figure 6.7(b) and (c), respectively. SR method [23] was then performed to reconstruct HR images of each RGB channel and the final color image is given in Figure 6.7(d). It is noted that the result is unsatisfactory as the HR reconstruction suffers from the color artifacts arising from image demosaicing. Next, Color TV [6], Jung's method [53], Farsiu's method [25] and our proposed algorithm were performed on the raw LR images and the results are given in Figure 6.7(e)-(f), respectively. Comparison shows that the proposed method can achieve better HR reconstruction. The proposed method presents an iterative framework to simultaneously perform color demosaicing and image SR on the mosaiced LR color images. A color correction term is incorporated into the proposed cost function to induce consistency amongst different color channels and alleviate color artifacts. It is noted that color artifacts often occur in regions

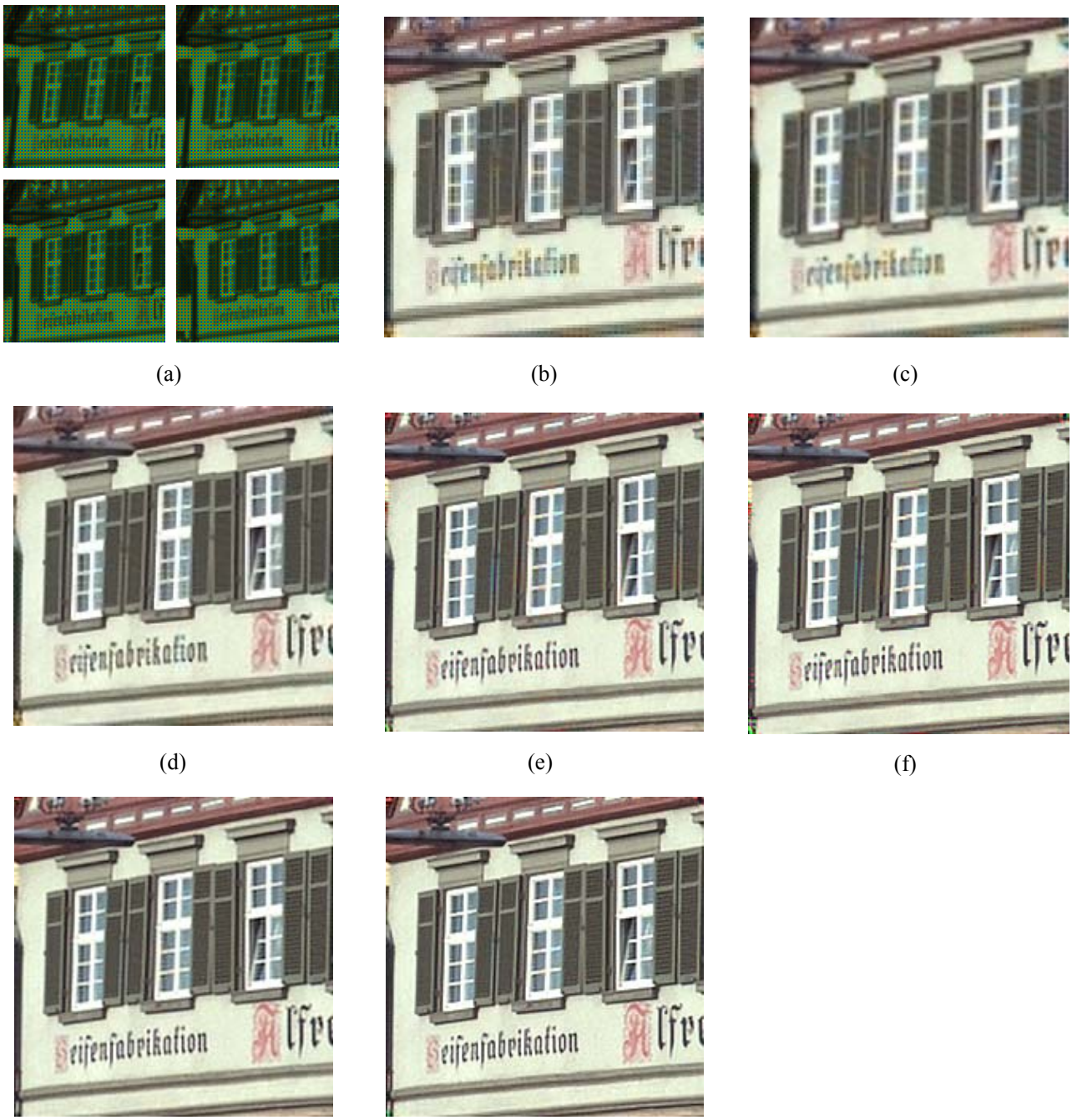
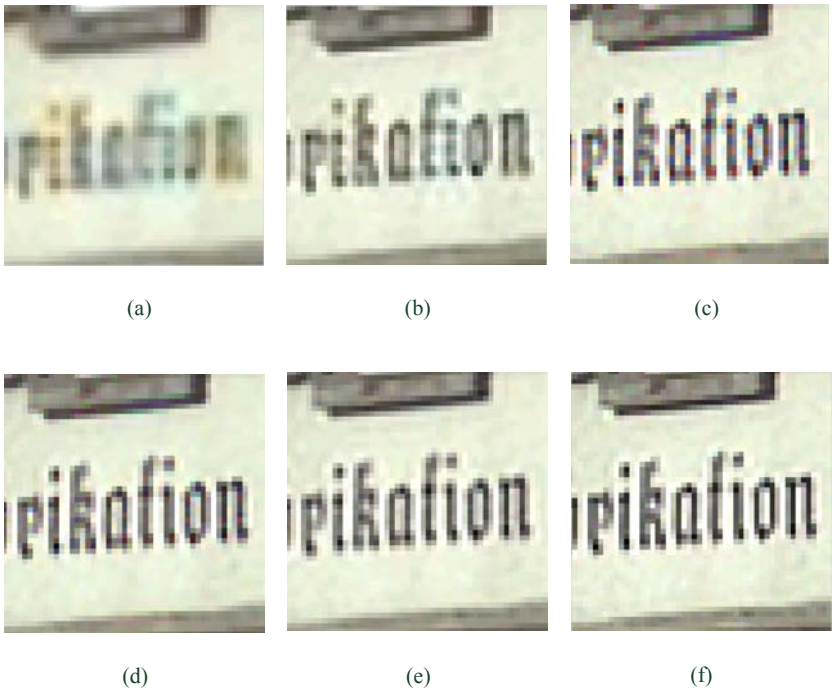


Figure 6.7 SR on "Window" image. (a) 4 samples of the mosaiced LR images, (b) The scaled-up version of the demosaiced reference LR image (1st LR image) using [142], (c) Bi-cubic interpolated result of (b), (d) Reconstructed result using two-stage approach of [142] followed by [23], (e) Reconstructed result using Color TV method [6], (f) Reconstructed result using Jung's method [53], (g) Reconstructed result using Farsiu's method [25], (h) Reconstructed result using our proposed method.

consisting of high frequency details, such as edges. Therefore, the proposed method utilizes the edge direction information to improve the HR reconstruction. This is different from conventional



(a) (b) (c)

(d) (e) (f)

Figure 6.8 Selected enlarged region of reconstructed "Window" image. (a)-(f) Selected regions of (c)-(h) in Figure 6.7.

Table 6.1 PSNR, SSIM and FSIM comparison using different methods.

	Two-stage approach	Color TV method [6]	Jung's method [53]	Farsiu's method [25]	Our proposed method
PSNR	25.0	26.1	27.1	27.6	28.0
SSIM	0.99994	0.99995	0.99996	0.99996	0.99997
FSIM	0.99854	0.99922	0.99936	0.99949	0.99952

joint color demosaicing and image SR approach [25]. An iterative framework is developed that enables the reconstructed HR image to be improved progressively. The selected region of the reconstructed HR images by the five methods is enlarged in Figure 6.8 for closer examination. The PSNRs, SSIMs and FSIMs of all the methods above are given in Table 6.1. From the table, it is observed that the proposed method outperforms the other methods. The objective performance measure confirms the effectiveness of the proposed method in handling color image SR.

6.5.2 Color Image SR against Various Noise Levels

We conducted Monte-Carlo simulations to demonstrate the effectiveness of the proposed method in handling LR images with different degrees of noise. The Monte-Carlo simulation was repeated 10 times for each set. Additive White Gaussian Noise (AWGN) was added to the test images in Figure 6.5 to produce four sets of LR images with different signal-to-noise ratios (SNRs) at 45, 35 and 25dB. The same experimental setup as in Section 6.5.1 was utilized to conduct the experiments. The PSNR of the reconstructed HR images obtained using our proposed method and the other four methods are given in Figure 6.9. From the figure, it can be seen that the PSNR of the reconstructed image using the proposed method outperforms all the other four methods under different noise levels. In all experiments, we obtained an average PSNR improvement of about 0.5dB, 1.7dB, 2.0dB and 2.7dB over the other four methods, respectively.

To further demonstrate the performance of the proposed method, subjective visual evaluation of reconstructed HR results by various SR methods at 35dB noise level are provided in Figure 6.10. Comparison shows that the proposed method can achieve better HR reconstruction when compared with other methods as the HR image reconstructed by the proposed method contains sharper edges (near the fence and headscarf regions) with less color artifacts. The reason why the proposed method can produce better results is because it performs joint image SR and color demosaicing progressively. Further, the textural information of the desired image is integrated into the color correction term and regularization term to improve the HR reconstruction. Both subjective evaluation and objective metrics confirm the effectiveness of the proposed method.

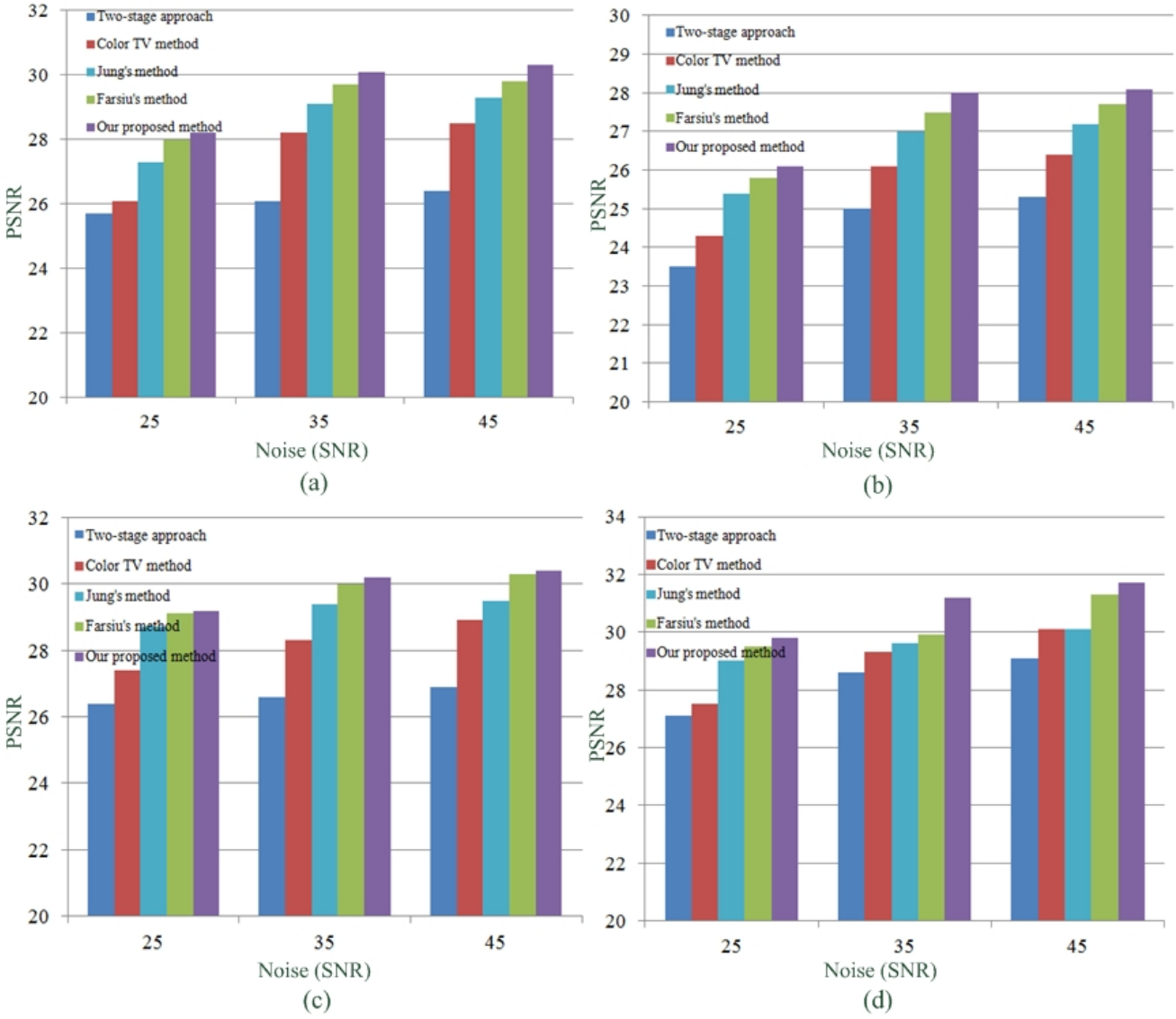


Figure 6.9 PSNR of HR images reconstruction at different noise levels. (a) “Lighthouse” image; (b) “Window” image; (c) “Barbara” image; (d) “House” image.

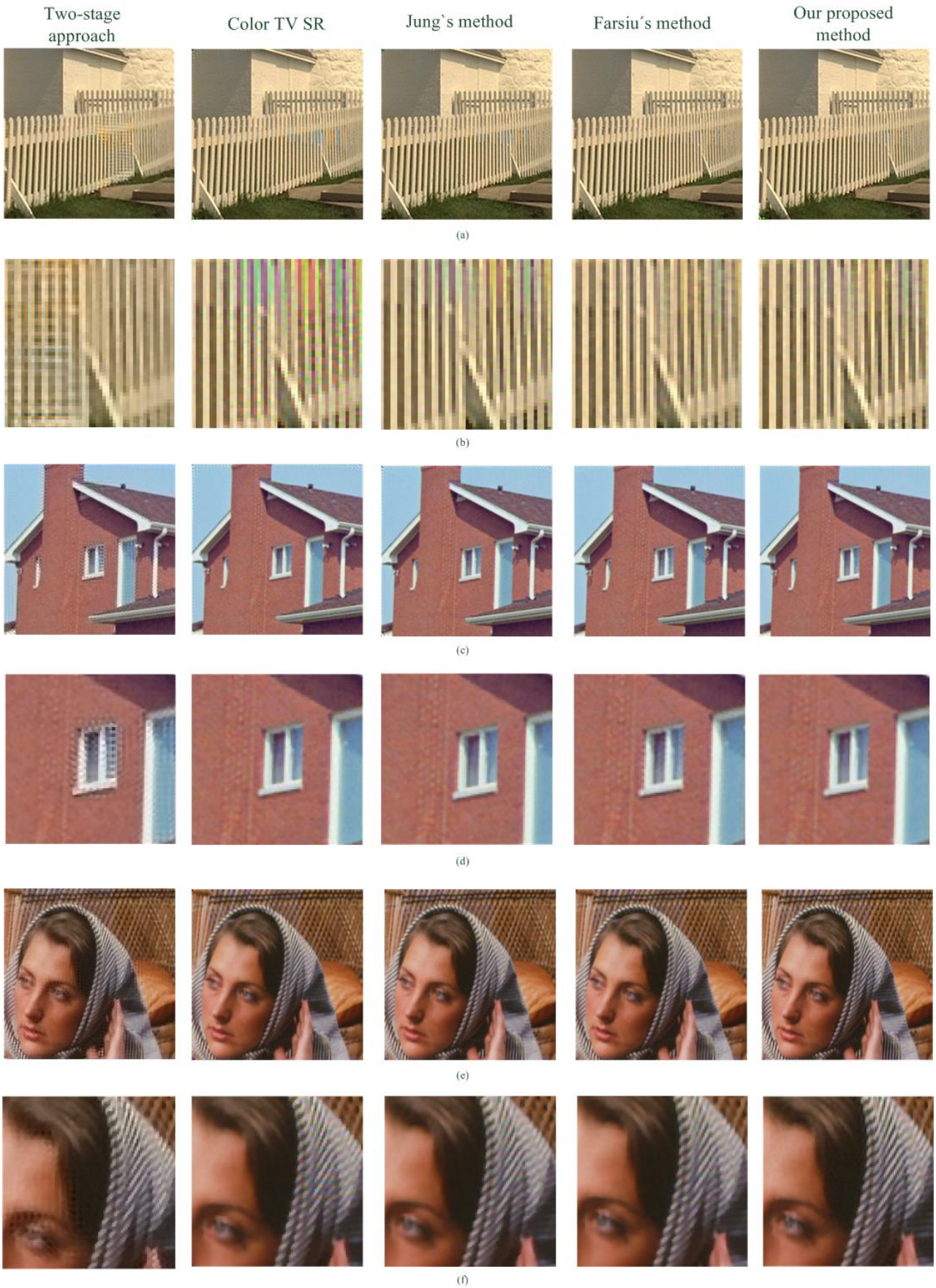

Two-stage approach
Color TV SR
Jung`s method
Farsiu´s method
Our proposed method
(a)
(b)
(c)
(d)
(e)
(f)

Figure 6.10 HR reconstructed results using various methods at 35dB noise level: (a) “Lighthouse” image; (b) Selected region of (a); (c) “House” image; (d) Selected region of (c); (e)"Barbara" image; (f) Selected region of (e).

Finally, we will use computational time to show the proposed method’s complexity. The two-stage approach of demosaicing and SR is used as the benchmark for comparison. The environments of simulation are given as follows: CPU Core Q9400-2.67GHz, 3.25 GB RAM, Windows XP OS (operation system) and MATLAB R2010b. The average computational time of test images with dimension 100×100 is 114s for our proposed method, 106s for Farsiu’s method [25], 95s for Jung’s method [53], 97s for Color TV method [6] and 126s for the two-stage approach of [142]-[23]. The computational time of our proposed method, Farsiu’s method, Jung’s method and Color TV SR method are comparable. Further, it should be pointed out that our proposed method is implemented in MATLAB, which is an interpreter programming platform whose processing speed is slower. The speed can be improved if it is implemented in a compiler language, for instance C.

6.6 Summary

In this chapter, we propose a new iterative framework for joint color image SR and demosaicing from a sequence of raw LR color images. Different from the two-stage process performing image demosaicing followed by SR, the proposed method combines the image demosaicing and SR into a single process. To suppress the color artifacts, a color correction term which employs the inter-channel correlation is developed. An edge-directed BTV regularization is adopted to preserve the high frequency information. The iterative scheme ensures the HR color image is improved progressively based on the edge direction information of the desired HR image. Experimental results demonstrate that the proposed method achieves satisfactory subjective visual evaluation along with the objective metrics.

A limitation of the proposed method is that it only estimates the gradient information along four primary directions. Hence, edges not lying along these directions may not be well characterized. A potential extension is to utilize more sophisticated techniques to exploit the edge information to improve the HR reconstruction.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In digital imaging applications, high resolution (HR) images with more details are required for computer vision and human viewing. Super resolution (SR) is a promising technique to tackle the problem of intrinsic hardware limitation, such as limited sensor pixels of the CCD chip. In this thesis, we have developed several SR algorithms to enhance the resolution of captured digital images. These algorithms can be classified into multi-frame SR, single-frame learning based SR, and color image SR.

7.1.1 Multi-frame Image SR

Multi-frame SR employs algorithms to combine a sequence of captured LR images to reconstruct a HR image of the original scene. The reconstruction process can be divided into two stages of image registration followed by HR reconstruction. Image registration serves to explore the sub-pixel level shifting among the LR images while HR reconstruction combine the unique information from each LR image to produce a HR image. Generally, accurate registration is a critical step for successful SR. However, many objective factors could degrade the performance of registration. For instance, the outliers existing in a LR image is a common obstacle for

registration to calculate the motion parameter of the LR image. Further, it would be propagated to the HR reconstruction result during the image fusion process. In Chapter 3, therefore, a L_1-norm nonlinear SR approach has been developed to alleviate the effects of outliers in the final HR reconstruction. The method involves development of an effective L_1-norm cost function, formulation of the SR problem into a joint optimization problem, and solving it using a linear programming (LP) framework. The algorithm enables image registration and HR reconstruction to be estimated simultaneously and improved progressively. Another important factor for registration is motion model. Generally speaking, the relationship of the captured LR images is described by using a motion model. The objective of registration is to explore the motion parameters based on the specific motion model. Therefore, an appropriate motion model would ensure the estimated motion parameters to reflect the true relationship of the LR images. In Chapter 4, we have developed a SR algorithm for more practical applications involving motion model consisting of translation, rotation and zooming. Different from most existing two-stage zooming SR methods, the proposed method combines HR reconstruction along with image registration into a single process and adopts an iterative framework to improve the estimated registration parameters and HR image progressively.

7.1.2 Single Frame Image SR

In contrast to multi-frame SR, single-frame SR involves HR reconstruction from a single LR image. This will make the SR reconstruction more ill-posed as many HR images could be mapped to the captured LR image. Therefore, strong regularization which could reflect the feature of the target image should be imposed to obtain a stable solution. An effective approach

is to learn the characteristic of the desired HR image from other images sharing the similar features. The learning-based SR technique is usually applied in domain specific applications, such as face SR and text SR. In Chapter 5, we have proposed a vehicle license plate (VLP) SR approach using training patches from an image database. Different from existing learning-based methods that utilize a hard binary decision strategy to determine whether the local patches are fully relevant or totally irrelevant, the proposed method incorporates soft local patches learned/reinforced based on the character prior information of the reconstructed VLP into HR license plate reconstruction. An iterative scheme is employed to ensure the license plate SR is improved progressively.

7.1.3 Color Image SR

In real-life imaging applications, color images can provide more visual information when compared with gray-scale images. It is well-known that a color image can be decomposed into three components representing different color channels. Therefore, a direct way to extend the existing SR algorithm for color SR applications is to perform separate HR reconstruction on different color channels first, and then combine the reconstructed HR images to produce a color image. Nevertheless, this process tends to introduce color artifacts as it neglects the consistency of different color channels. In Chapter 6, we have proposed a color SR method which takes into account the color correlation during the HR reconstruction process. Different from the two-stage process of performing image demosaicing followed by SR, the proposed method combines the image demosaicing and SR into a single process. To suppress the color artifacts, a color

correction term which employs the inter-channel correlation is developed. An edge-directed BTV regularization is adopted to preserve the high-frequency information.

7.2 Future Work

Image SR reconstruction is a promising research topic. Some potential research directions include example-based multi-frame SR, compressed video SR and medical image SR.

7.2.1 Example-based Multi-frame Super-Resolution

In this thesis, the proposed image SR methods can be classified into two families of methods: (i) the classical multi-frame SR (e.g. proposed methods in chapter 3 and 4) and (ii) example-based single-frame SR (e.g. proposed method in chapter 5). The classical SR process is to combine multi-frame LR images obtained at subpixel shifts to produce a HR image. The underlying HR information is assumed to be split across multi-frame LR images, implicitly found in the aliased form. However, the increase of resolution in HR reconstruction by this approach depends greatly on the number of captured LR images. On the other hand, the example-based SR can estimate the HR details from a single-frame LR image through learning correspondence between LR and HR patches from database. The ratio of HR image size to LR image size can be larger. Therefore, one potential extension of our work is to develop a unified framework which combines the multi-frame SR with example-based SR.

7.2.2 Compressed Video Super-resolution

In this thesis, we focus on the HR image construction, including gray-scale image and color image. Intuitively, we can extend our multi-frame SR algorithms directly to uncompressed video SR application as uncompressed video consists of a sequence of frames which can be considered as still images. However, video acquisition which involves compression procedure is much more complex than multi-frame images/raw video (e.g. uncompressed) acquisition. Therefore, it may not be applicable to apply current SR techniques designed for multi-frame images/raw videos to the compressed videos, especially for highly compressed ones. This is because the quantization error introduced in the quantization/compression process is often the dominant error when the compressed ratio is high. To address this issue, one potential direction for video SR is to investigate the quantization/compression error in compressed video SR process and develop video SR algorithms which incorporate quantization error estimation into SR reconstruction.

7.2.3 Medical Image Super-resolution

In medical imaging applications, HR images are often desirable to increase the accuracy of diagnoses. Various SR techniques have been proposed in recent years, including X-ray [80], magnetic resonance imaging (MRI) [34], [95] and positron emission tomography imaging system (PET) [58], [57]. Conventional medical image SR algorithms can be regarded as two-stage process of motion estimation and HR reconstruction. Therefore, one interesting extension of our work (e.g. proposed methods in chapter 3 and 4) is to develop a joint image registration and SR framework for the HR medical image reconstruction. However, the acquisition process of medical images (e.g. MRIs) is different from common photographic images [37]. A promising

direction is to adopt an appropriate motion model which can reflect the underlying principle of medical imaging acquisition and incorporate it into the joint framework proposed in this thesis to perform medical image SR.

List of Publications

[1]. Y. Tian and K.-H. Yap, "Joint Image Registration and Super-Resolution From Low-Resolution Images With Zooming Motion," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 23, pp. 1224-1234, 2013.

[2]. Y. Tian, K.-H. Yap, and Y. He, "Vehicle license plate super-resolution using soft learning prior," *Multimedia Tools and Applications,* vol. 60, pp. 519-535, 2011

[3]. K.-H. Yap, Yu He, Y. Tian, and Lap-Pui Chau, "A nonlinear L1-norm approach for joint image registration and super-resolution," *IEEE Signal Processing Letters,* vol. 16, pp. 981-984, 2009

[4]. Y. Tian and K.-H. Yap, "Joint Super-resolution and Demosaicing for Color Image Reconstruction," under review by *Signal Processing*

[5]. Y. Tian and K.-H. Yap, "Multi-frame Super-resolution from Observations with Zooming Motion," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Kyoto, Japan, March, 2012

[6]. Y. Tian, K.-H. Yap, and L. Chen, "L1-norm multi-frame super-resolution from images with zooming motion," in *IEEE 13th International Workshop on Multimedia Signal Processing*, 2011, pp. 1-6.

[7]. Y. Tian, K.-H. Yap, and Y. He, "High resolution vehicle license plate reconstruction using soft recognition learning," in *8th International Conference on Information, Communications and Signal Processing*, 2011, pp. 1-5.

[8]. Y. Tian and K.-H. Yap, “Iterative Joint Super-resolution and Demosaicing for Color Image Reconstruction”, submitted to *International Symposium on Intelligent Signal Processing and Communication Systems*, 2013

Bibliography

[1] *[online] http: //functions.wolfram.com/04.01.20.0001.01*.

[2] Aizawa, K., Komatsu, T., and Saito, T., "A scheme for acquiring very high resolution images using multiple cameras," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1992, pp. 289-292.

[3] Baboulaz, L. and Dragotti, P. L., "Exact Feature Extraction Using Finite Rate of Innovation Principles With an Application to Image Super-Resolution," *IEEE Transactions on Image Processing,* vol. 18, pp. 281-298, 2009.

[4] Baker, S. and Kanade, T., "Hallucinating faces," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 83-88.

[5] Baker, S. and Kanade, T., "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Machine Intell,* vol. 24, pp. 1167–1183, 2002.

[6] Blomgren, P. and Chan, T. F., "Color TV: total variation methods for restoration of vector-valued images," *IEEE Transactions on Image Processing,* vol. 7, pp. 304-309, 1998.

[7] Blu, T., Thevenaz, P., and Unser, M., "Linear interpolation revitalized," *IEEE Transactions on Image Processing,* vol. 13, pp. 710-719, 2004.

[8] Bo, L., Hong, C., Shiguang, S., and Xilin, C., "Locality preserving constraints for super-resolution with neighbor embedding," in *IEEE International Conference on Image Processing*, 2009, pp. 1189-1192.

[9] Bose, N. K. and Chappalli, M. B., "A second generation wavelet framework for superresolution with noise filtering," *International Journal of Imaging Systems and Technology,* vol. 14, pp. 84-89, 2004.

[10] Bose, N. K., Lertrattanapanich, S., and Chappalli, M. B., "Superresolution with second generation wavelets," *Signal Processing: Image Communication,* vol. 19, pp. 387-391, 2004.

[11] Bresson, X. and Chan, T. F., "Fast minimization of the vectorial total variation norm and applications to color image processing," *Inv. Problems Imag.,* vol. 2, pp. 455-484, 2008.

[12] Cain, S., Hardie, R. C., and Armstrong, E., *Restoration of aliased video sequences via a maximum-likelihood approach*. Monterey, CA, 1996.

[13] Chakrabati, A., Rajagopalan, A. N., and Chellappa, R., "Super-resolution of face images using kernel PCA-based prior," *IEEE Transaction on Multimedia,* vol. 9, pp. 888-892, 2007.

[14] Chan, T. F. and Chiu-Kwong, W., "Total variation blind deconvolution," *IEEE Transactions on Image Processing,* vol. 7, pp. 370-375, 1998.

[15] Chang, H., Yeung, D.-Y., and Xiong, Y., "Super-resolution through neighbor embedding," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, pp. I-275-I-282.

[16] Chang, T., Liu, T., and Chuang, J., "Direct energy minimization for super-resolution on nonlinear manifolds," *In Proc. European conf. computer vision,* pp. 281-294, 2006.

[17] Chappalli, M. B. and Bose, N. K., "Simultaneous noise filtering and superresolution with second-generation wavelets," *IEEE Signal Processing Letters,* vol. 12, pp. 772-775, 2005.

[18] Chen, L. and Yap, K.-H., "Regularized interpolation using kronecher product for still images," in *IEEE International Conference on image processing*, 2005, pp. 1014–1017.

[19] Corduneanu, A. and Platt, J. C., "Learning spatially-variable filters for super-resolution of text," in *IEEE International Conference on Image Processing*, 2005, pp. I-849-52.

[20] Dalley, G., Freeman, B., and Marks, J., "Single-frame text super-resolution: a Bayesian approach," in *IEEE International Conference on Image Processing*, 2004, pp. 3295-3298.

[21] Datsenko, D. and Elad, M., "Example-based single document image super-resolution: A global map approach with outlier rejection," *Multidimensional Systems and Signal Processing,* vol. 18, pp. 103-121, 2007.

[22] El-Khamy, S. E., Hadhoud, M. M., Dessouky, M. I., Salam, B. M.*, et al.*, "Efficient implementation of image interpolation as an inverse problem," *Digital Signal Process,* vol. 15, pp. 137–152, Mar. 2005.

[23] Elad, M. and Feuer, A., "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions on Image Processing,* vol. 6, pp. 1646-1658, 1997.

[24] Elad, M. and Hel-Or, Y., "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Transactions on Image Processing,* vol. 10, pp. 1187-1193, 2001.

[25] Farsiu, S., Elad, M., and Milanfar, P., "Multiframe demosaicing and super-resolution of color images," *IEEE Transactions on Image Processing,* vol. 15, pp. 141-159, 2006.

[26] Farsiu, S., Robinson, D., Elad, M., and Milanfar, P., "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology,* vol. 14, pp. 47-57, 2004.

[27] Farsiu, S., Robinson, D., Elad, M., and Milanfar, P., "Robust shift and add approach to superresolution," in *SPIE Applications of Digital Image Processing XXVI* 2003, pp. 121-130.

[28] Farsiu, S., Robinson, M. D., Elad, M., and Milanfar, P., "Fast and Robust Multiframe Super Resolution," *IEEE Trans. Image Processing,* vol. 13, pp. 1327-1344, 2004.

[29] Foroosh, H., Zerubia, J. B., and Berthod, M., "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing,* vol. 11, pp. 188-200, 2002.

[30] Freeman, W. T., Jones, T. R., and Pasztor, E. C., "Example based super-resolution," *IEEE Computer Graphics and Applications,* vol. 22, pp. 56-65, 2002.

[31] Freeman, W. T., Pasztor, E. C., and Carmichael, O. T., "Learning Low-Level Vision," *International Journal of Computer Vision,* vol. 40, pp. 25-47, 2000.

[32] Fu, H., Ng, M. K., Nikolova, M., and Barlow, J. L., "Efficient minimization methods of mixed l2-l1 and l1-l1 norms for image restoration," *SIAM Journal on Scientific Computing,* vol. 27, pp. 1881-1902, 2006.

[33] Geman, S. and Geman, D., "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. PAMI-6, pp. 721-741, 1984.

[34] Gholipour, A., Estroff, J. A., and Warfield, S. K., "Robust Super-Resolution Volume Reconstruction From Slice Acquisitions: Application to Fetal Brain MRI," *IEEE Transactions on Medical Imaging,* vol. 29, pp. 1739-1758, 2010.

[35] Golub, G. H. and Loan, C. F. V., "Matrix computations," *The Johns Hopkins University Press, Baltimore,* 1996.

[36] Gotoh, T. and Okutomi, M., "Direct super-resolution and registration using raw CFA images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, pp. II-600-II-607.

[37] Greenspan, H., "Super-Resolution in Medical Imaging," *The Computer Journal,* vol. 52, pp. 43-63, January 1, 2009.

[38] Gupta, M. R. and Chen, T., "Vector color filter array demosaicing," *In M. M. Blouke, J. Canosa, and N. Sampat, editors, Proc. SPIE,* vol. 4306, pp. 374–382, 2001.

[39] Hardie, R. C., Barnard, K. J., and Armstrong, E. E., "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing,* vol. 6, pp. 1621-1633, 1997.

[40] He, H. and Kondi, L. P., "An image super-resolution algorithm for different error levels per frame," *IEEE Transactions on Image Processing,* vol. 15, pp. 592-603, 2006.

[41] He, H. and Kondi, L. P., "Resolution enhancement of video sequences with simultaneous estimation of the regularization parameter," *Journal of Electronic Imaging,* vol. 13, pp. 586-596, 2004.

[42] He, Y., Yap, K.-H., Chen, L., and Chau, L.-P., "Joint Image Registration and Super-Resolution using Nonlinear Least Squares Method," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2007, pp. I-561-I-564.

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

[44] He, Y., Yap, K.-H., Chen, L., and Chau, L.-P., "A soft MAP framework for blind super-resolution image reconstruction," *Image and Vision Computing,* vol. 27, pp. 364-373, 2009.

[45] Hong, K. P., Paik, J. K., Kim, H. J., and Lee, C. H., "An edge-preserving image interpolation system for a digital camcorder," *IEEE Trans. Consumer Electronics,* vol. 42, pp. 279-284, Aug. 1996.

[46] Hwang, J. W. and Lee, H. S., "Adaptive image interpolation based on local gradient features," *IEEE Signal Processing Letters,* vol. 11, pp. 359-362, Mar. 2004.

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

[48] James E. Adams, J., "Interactions between color plane interpolation and other image processing functions in electronic photography," *Proceeding of SPIE,* vol. 2416, pp. 144-151, Feb,1995.

[49] Jiang, J., Hu, R., Han, Z., and Lu, T., "Efficient single image super-resolution via graph-constrained least squares regression," *Multimedia Tools and Applications,* pp. 1-24, 2013/06/30 2013.

[50] Jing, G., Shi, Y., Kong, D., Ding, W.*, et al.*, "Image super-resolution based on multi-space sparse representation," *Multimedia Tools and Applications,* pp. 1-15, 2012/01/01 2012.

[51] Joshi, M. V., Chaudhuri, S., and Panuganti, R., "A learning-based method for image super-resolution from zoomed observations," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 35, pp. 527-537, 2005.

[52] Joshi, M. V., Chaudhuri, S., and Panuganti, R., "Super-resolution imaging: use of zoom as a cue," *Image and Vision Computing,* vol. 22, pp. 1185-1196, 2004.

[53] Jung, M., Bresson, X., Chan, T. F., and Vese, L. A., "Nonlocal Mumford-Shah Regularizers for Color Image Restoration," *IEEE Transactions on Image Processing,* vol. 20, pp. 1583-1598, 2011.

[54] Kakarala, R. and Baharav, Z., "Adaptive demosaicing with the principal vector method," *IEEE Transactions on Consumer Electronics,* vol. 48, pp. 932-937, 2002.

[55] Kang, M. G. and Katsagelos, A. K., "Simultaneous iterative image restoration and evaluation of the regularization parameter," *IEEE Trans. Signal Process,* vol. 40, pp. 2329–2334, 1992.

[56] Kazemi, F. M., Samadi, S., Poorreza, H. R., and Akbarzadeh-T, M.-R., "Vehicle Recognition Using Curvelet Transform and SVM," in *Fourth International Conference on Information Technology*, 2007, pp. 516-521.

[57] Kennedy, J. A., Israel, O., Frenkel, A., Bar-Shalom, R.*, et al.*, "Improved image fusion in PET/CT using hybrid image reconstruction and super-resolution," *International Journal of Biomed Imaging,* vol. 2007, p. 10, 2007.

[58] Kennedy, J. A., Israel, O., Frenkel, A., Bar-Shalom, R.*, et al.*, "Super-resolution in PET imaging," *IEEE Transactions on Medical Imaging,* vol. 25, pp. 137-147, 2006.

[59] Keren, D. and Osadchy, M., "Restoring subsampled color images," *Machine Vision and Applications,* vol. 11, pp. 197-202, 1999.

[60] Kim, H. Y., "Binary operator design by k-nearest neighbor learning with application to image resolution increasing," *International Journal of Imaging Systems and Technology,* vol. 11, pp. 331-339, 2000.

[61] Kim, K. I., Franz, M., and Scholkopf, B., "Kernel hebbian algorithm for single-frame super-resolution," in *in proceeding of the European Conference on Computer Vision*, 2004, pp. 135-149.

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

[63] Kim, S. P. and Su, W. Y., "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Transaction on Image Processing,* vol. 2, pp. 534-539, 1993.

[64] Kim, S. P. and Su, W. Y., "Recursive high-resolution reconstruction of blurred multiframe images," in *International Conference on Acoustics, Speech, and Signal Processing*, 1991, pp. 2977-2980

[65] Kimmel, R., "Demosaicing: image reconstruction from color CCD samples," *IEEE Transactions on Image Processing,* vol. 8, pp. 1221-1228, 1999.

[66] Kuno, T., Sugiura, H., and Matoba, N., "New interpolation method using discriminated color correlation for digital still cameras," *IEEE Transactions on Consumer Electronics,* vol. 45, pp. 259-267, 1999.

[67] Lee, E. S. and Kang, M. G., "Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration," *IEEE Transactions on Image Processing,* vol. 12, pp. 826-837, 2003.

[68] Lertrattanapanich, S. and Bose, N. K., "High resolution image formation from low resolution frames using Delaunay triangulation," *IEEE Transaction on Image Processing,* vol. 11, 2002.

[69] Li, X., "Super-Resolution for Synthetic Zooming," *EURASIP Journal on Applied Signal Processing,* vol. 2006, pp. 1-12, 2006.

[70] Li, X. and Orchard, M. T., "New edge-directed interpolation," *IEEE Trans. Image Processing,* vol. 10, pp. 1521-1527, Oct. 2001.

[71] Li, Y.-R. and Dai, D.-Q., "Color Superresolution Reconstruction and Demosaicing Using Elastic Net and Tight Frame," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 55, pp. 3500-3512, 2008.

[72] Li, Y.-R., Dai, D.-Q., and Shen, L., "Multiframe Super-Resolution Reconstruction Using Sparse Directional Regularization," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 20, pp. 945-956, 2010.

[73] Lin, Z. and Shum, H.-Y., "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 26, pp. 83-97, 2004.

[74] Lin, Z. and Shum, H. Y., "Fundamental limits of reconstruction-based super-resolution algorithms under local translation," *IEEE Trans. Pattern Anal. Machine Intell,* vol. 26, pp. 83–97, 2004.

[75] Liu, W., Lin, D., and Tang, X., "Face hallucination through dual associative learning," in *IEEE International Conference on Image Processing*, 2005, pp. I-873-6.

[76] Long, Y. and Huang, Y., "Adaptive Demosaicking using Multiple Neural Networks," in *IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, 2006, pp. 353-357.

[77] Lu, W. and Tan, Y.-P., "Color filter array demosaicking: new method and performance measures," *IEEE Transactions on Image Processing,* vol. 12, pp. 1194-1210, 2003.

[78] Lukac, R., Martin, K., and Plataniotis, K. N., "Demosaicked image postprocessing using local color ratios," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 14, pp. 914-920, 2004.

[79] Lukac, R. and Plataniotis, K. N., "Demosaicking using Vector Spectral Model," in *IEEE International Conference on Multimedia and Expo*, 2006, pp. 1185-1188.

[80] Ma, Z. and Zhao, X., "Super-resolution from multiframe X-ray images," in *International Conference on Complex Medical Engineering*, 2009, pp. 1-6.

[81] Menon, D., Andriani, S., and Calvagno, G., "Demosaicing With Directional Filtering and a posteriori Decision," *IEEE Transactions on Image Processing,* vol. 16, pp. 132-141, 2007.

[82] Merino, M. T. and Nunez, J., "Super-Resolution of Remotely Sensed Images With Variable-Pixel Linear Reconstruction," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, pp. 1446-1457, 2007.

[83] Molina, R., Vega, M., Abad, J., and Katsaggelos, A. K., "Parameter estimation in Bayesian high-resolution image reconstruction with multisensors," *IEEE Transactions on Image Processing,* vol. 12, pp. 1655-1667, 2003.

[84] Muresan, D. D. and Parks, T. W., "Demosaicing using optimal recovery," *IEEE Transactions on Image Processing,* vol. 14, pp. 267-278, 2005.

[85] Ng, M. K., Shen, H., Chaudhuri, S., and Yau, A. C., "Zoom-based super-resolution reconstruction approach using prior total variation," *Opical Engineering,* vol. 46, Dec. 2007.

[86] Nguyen, N. and Milanfar, P., "A wavelet-based interpolation-restoration method for superresolution," *Circuits, Systems, and Signal Processing, Special issue on advanced signal and image reconstruction,* vol. 19, pp. 321-338, 2000.

[87] Nhat, N., Milanfar, P., and Golub, G., "A computationally efficient superresolution image reconstruction algorithm," *IEEE Transactions on Image Processing,* vol. 10, pp. 573-583, 2001.

[88] Ni, K. S. and Nguyen, T. Q., "Color Image Superresolution Based on a Stochastic Combinational Classification-Regression Algorithm," in *IEEE International Conference on Image Processing*, 2007, pp. II - 89-II - 92.

[89] Nijhuis, J. A. G., Ter Brugge, M. H., Helmholt, K. A., Pluim, J. P. W*., et al.*, "Car license plate recognition with neural networks and fuzzy logic," in *IEEE International Conference on Neural Networks*, 1995, pp. 2232-2236.

[90] Park, J.-S. and Lee, S.-W., "Enhancing low-resolution facial images using error back-projection for human identification at a distance," in *International Conference on Pattern Recognition*, 2004, pp. 346-349.

[91] Park, J., Kwon, Y., and Kim, J. H., "An example-based prior model for text image super-resolution," in *International Conference on Document Analysis and Recognition*, 2005, pp. 374-378.

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

[93] Patti, A. J. and Altunbasak, Y., "Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants," *IEEE Transactions on Image Processing,* vol. 10, pp. 179-186, 2001.

[94] Pei, S.-C. and Tam, I.-K., "Effective color interpolation in CCD color filter arrays using signal correlation," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 13, pp. 503-513, 2003.

[95] Peled, S. and Yeshurun, Y., "Superresolution in MRI: Application to human white matter fiber tract visualization by diffusion tensor imaging " *Magnetic Resonance in Medicine,* vol. 45, pp. 29-35, 2001.

[96] Peleg, S., Keren, D., and Schweitzer, L., "Improving image resolution by using subpixel motion " *Pattern Recognition Letters,* vol. 5, pp. 223-226, 1987.

[97] Pickup, L. C., Capel, D. P., Roberts, S. J., and Zisserman, A., "Bayesian Methods for Image Super-Resolution," *The Computer Journal,* vol. 52, pp. 101-113, 2009.

[98] Pratt, W. K., *Digital image processing*: John Wiely & Sons, INC., 2001.

[99] Prendergast, R. S. and Nguyen, T. Q., "Improving frequency domain super-resolution via undersampling model," in *IEEE International Conference on Image Processing*, 2005, pp. I-853.

[100] Prendergast, R. S. and Nguyen, T. Q., "Minimum Mean-Squared Error Reconstruction for Generalized Undersampling of Cyclostationary Processes," *IEEE Transactions on Signal Processing,* vol. 54, pp. 3237-3242, 2006.

[101] Protter, M., Elad, M., Takeda, H., and Milanfar, P., "Generalizing the Nonlocal-Means to Super-Resolution Reconstruction," *IEEE Transactions on Image Processing,* vol. 18, pp. 36-51, 2009.

[102] Pu, H., Wei, H., Dong-Feng, W., and Yong-Jie, Z., "Car license plate feature extraction and recognition based on multi-stage classifier," in *International Conference on Machine Learning and Cybernetics*, 2003, pp. 128-132.

[103] Ramponi, G., "Warped distance for space-variant linear image interpolation," *IEEE Trans. Image Processing,* vol. 8, pp. 629-639, May 1999.

[104] Reddy, B. S. and Chatterji, B. N., "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing,* vol. 5, pp. 1266-1271, 1996.

[105] Rhee, S. and Moon Gi, K., "DCT-based regularized algorithm for high-resolution image reconstruction," in *International Conference on Image Processing*, 1999, pp. 184-187.

[106] Robinson, D., Farsiu, S., and Milanfar, P., "Optimal Registration of Aliased Images Using Variable Projection with Applications to Superresolution," *The Computer Journal* vol. 52, pp. 31-42, 2009.

[107] Robinson, M. D., Toth, C. A., Lo, J. Y., and Farsiu, S., "Efficient Fourier-Wavelet Super-Resolution," *IEEE Transactions on Image Processing,* vol. 19, pp. 2669-2681, 2010.

[108] Rosenfeld, A. and Kak, C. A., "Digital Picture Processing," *2nd ed. London: Academic Press,* 1982.

[109] Roweis, S. T. and Saul, L. K., "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science,* vol. 290, pp. 2323-2326, December 22 2000.

[110] Sajjad, M., Ejaz, N., Mehmood, I., and Baik, S., "Digital image super-resolution using adaptive interpolation based on Gaussian function," *Multimedia Tools and Applications,* pp. 1-17, 2013/07/09 2013.

[111] Sakamoto, T., Nakanishi, C., and Hase, T., "Software pixel interpolation for digital still cameras suitable for a 32-bit MCU," *IEEE Transactions on Consumer Electronics,* vol. 44, pp. 1342-1352, 1998.

[112] Schultz, R. R. and Stevenson, R. L., "A Bayesian approach to image expansion for improved definition," *IEEE Transactions on Image Processing,* vol. 3, pp. 233-242, 1994.

[113] Schultz, R. R. and Stevenson, R. L., "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing,* vol. 5(6), pp. 996-1011, 1996.

[114] Shah, N. R. and Zakhor, A., "Resolution enhancement of color video sequences," *IEEE Transactions on Image Processing,* vol. 8, pp. 879-885, 1999.

[115] Shen, M. and Xue, P., "Super-resolution from observations with variable zooming ratios," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 2622-2625.

[116] Shyang-Lih, C., Li-Shien, C., Yun-Chung, C., and Sei-Wan, C., "Automatic license plate recognition," *IEEE Transactions on Intelligent Transportation Systems,* vol. 5, pp. 42-53, 2004.

[117] Smith, D. C., "Super-Resolution of Text Images through Neighbor Embedding," in *Proceedings of the IASTED International Conference on Signal and Image Processing*, 2011, pp. 19-26.

[118] Song, H., Qing, L., Wu, Y., and He, X., "Adaptive regularization-based space–time super-resolution reconstruction," *Signal Processing: Image Communication,* vol. 28, pp. 763-778, 2013.

[119] Sroubek, F., Cristobal, G., and Flusser, J., "A Unified Approach to Superresolution and Multichannel Blind Deconvolution," *IEEE Transactions on Image Processing,* vol. 16, pp. 2322-2332, 2007.

[120] Sroubek, F. and Flusser, J., "Multichannel blind deconvolution of spatially misaligned images," *IEEE Transactions on Image Processing,* vol. 14, pp. 874-883, 2005.

[121] Sroubek, F. and Flusser, J., "Multichannel blind iterative image restoration," *IEEE Transactions on Image Processing,* vol. 12, pp. 1094-1106, 2003.

[122] Stark, H. and Oskoui, P., "High-resolution image recovery from image-plane arrays, using convex projections," *Journal of the Optical Society of America A,* vol. 6, pp. 1715-1726, 1989.

[123] Stephenson, T. A. and Chen, T., "Adaptive Markov random fields for example-based super-resolution of faces," *EURASIP Journal on Applied Signal Processing,* vol. 2006, pp. 1-11, 2006.

[124] Strong, D. and Chan, T., "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse Problems,* vol. 19, pp. 165-187, 2003.

[125] Sun, J., Tao, H., and Shum, H., "Image Hallucination with Primal Sketch Priors," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003, pp. 729- 736.

[126] Suresh, K. V. and Rajagopalan, A. N., "A Discontinuity Adaptive Method for Super-Resolution of License Plates " *Graphics and Image Processing Computer Vision,* vol. 4338/2006, pp. 24-34, 2006.

[127] Tae Oh, B. and Jay Kuo, C. C., "Super-resolution texture synthesis using stochastic PAR/NL model," *Journal of Visual Communication and Image Representation,* vol. 23, pp. 995-1007, 2012.

[128] Tang, Y., Yuan, Y., Yan, P., and Li, X., "Greedy regression in sparse coding space for single-image super-resolution," *Journal of Visual Communication and Image Representation,* vol. 24, pp. 148-159, 2013.

[129] Tian, J. and Ma, K.-K., "Edge-adaptive super-resolution image reconstruction using a Markov chain Monte Carlo approach," in *6th International Conference on Information, Communications & Signal Processing*, 2007, pp. 1-5.

[130] Tian, J. and Ma, K.-K., "A MCMC approach for Bayesian super-resolution image reconstruction," in *IEEE International Conference on Image Processing*, 2005, pp. I-45-8.

[131] Tian, J. and Ma, K.-K., "Stochastic super-resolution image reconstruction," *Journal of Visual Communication and Image Representation,* vol. 21, pp. 232-244, 2010.

[132] Tian, J. and Ma, K.-K., "A survey on super-resolution imaging," *Signal, Image and Video Processing,* vol. 5, pp. 329-342, 2011.

[133] Tom, B. C. and Katsaggelos, A. K., "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Transactions on Image Processing,* vol. 10, pp. 278-287, 2001.

[134] Tsai, R. Y. and Huang, T. S., "Multiframe image restoration and registration," *Advances in Computer Vision and Image Processing,* vol. 1, pp. 317-319, 1984.

[135] Vandewalle, P., Sbaiz, L., Vandewalle, J., and Vetterli, M., "Super-Resolution From Unregistered and Totally Aliased Signals Using Subspace Methods," *IEEE Transactions on Signal Processing,* vol. 55, pp. 3687-3703, 2007.

[136] Vandewalle, P., usstrunk, S. S., and Vetterli, M., "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution," *EURASIP Journal on Applied Signal Processing,* vol. 2006, pp. 1-14, 2006.

[137] Wang, C., Xue, P., and W., L., "Improved Super-Resolution Reconstruction From Video," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 16, pp. 1411-1422, 2006.

[138] Wang, X. G. and Tang, X. O., "Hallucinating face by eigentransformation," *IEEE Transactions on System, Man, and Cybernetics, Part C: Applications and Reviews,* vol. 35, pp. 425-434, 2005.

[139] Woods, N. A. and Galatsanos, N. P., "Non-stationary approximate Bayesian super-resolution using a hierarchical prior model," in *IEEE International Conference on Image Processing*, 2005, pp. I-37-40.

[140] Woods, N. A., Galatsanos, N. P., and Katsaggelos, A. K., "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Transactions on Image Processing,* vol. 15, pp. 201-213, 2006.

[141] Wu, X. and Zhang, N., "Primary-consistent soft-decision color demosaicking for digital cameras (patent pending)," *IEEE Transactions on Image Processing,* vol. 13, pp. 1263-1274, 2004.

[142] Xin, L., "Demosaicing by successive approximation," *IEEE Transactions on Image Processing,* vol. 14, pp. 370-379, 2005.

[143] Xin, L. and Orchard, M. T., "New edge-directed interpolation," *IEEE Transactions on Image Processing,* vol. 10, pp. 1521-1527, 2001.

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

[145] Yang, J., Wright, J., Huang, T. S., and Ma, Y., "Image Super-Resolution Via Sparse Representation," *IEEE Transactions on Image Processing,* vol. 19, pp. 2861-2873, 2010.

[146] Yap, K.-H., He, Y., Tian, Y., and Chau, L.-P., "A nonlinear L1-norm approach for joint image registration and super-resolution," *IEEE Signal Processing Letters,* vol. 16, pp. 981-984, 2009.

[147] Yin, H., Li, S., and Fang, L., "Simultaneous image fusion and super-resolution using sparse representation," *Information Fusion,* vol. 14, pp. 229-240, 2013.

[148] Yu, W., "Colour demosaicking method using adaptive cubic convolution interpolation with sequential averaging," *IEE Proceedings -Vision, Image and Signal Processing,* vol. 153, pp. 666-676, 2006.

[149] Zhang, K., Gao, X., Tao, D., and Li, X., "Single Image Super-Resolution With Multiscale Similarity Learning," *IEEE Transactions on Neural Networks and Learning Systems,* vol. PP, pp. 1-1, 2013.

[150] Zhang, L. and Wu, X., "Color demosaicking via directional linear minimum mean square-error estimation," *IEEE Transactions on Image Processing,* vol. 14, pp. 2167-2178, 2005.

[151] Zhang, Y., "Solving large-scale linear programs by interior-point methods under the MATLAB environment," *Technical Report TR96-01,* vol. 1, Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD, 1995.

[152] Zhang, Y. and Zhang, C., "A new algorithm for character segmentation of license plate," in *IEEE International Conference on Intelligent Vehicles Symposium*, 2003, pp. 106–109.

[153] Zhou, L., Lu, X., and Yang, L., "A local structure adaptive super-resolution reconstruction method based on BTV regularization," *Multimedia Tools and Applications,* pp. 1-14, 2012/12/01 2012.

[154] Zibetti, M. V. W. and Mayer, J., "A Robust and Computationally Efficient Simultaneous Super-Resolution Scheme for Image Sequences," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 17, pp. 1288-1300, 2007.

[155] Zibetti, M. V. W., Mayer, J., and Bazan, F. S. V., "Determining the parameters in regularized super-resolution reconstruction," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 853-856.

[156] Zomet, A., Rav-Acha, A., and Peleg, S., "Robust super-resolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. I-645-I-650.

[157] Zou, H. and Hastie, T., "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 67, pp. 301-320, 2005.