Super-resolution reconstruction : analysis and applications

Shen, Minmin

2010

Shen, M. (2010). Super-resolution reconstruction : analysis and applications. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/51017>

<https://doi.org/10.32657/10356/51017>

Super-resolution Reconstruction: Analysis and Applications

SHEN MINMIN

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

2010

Statement of Originality

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

.....

Date

.....

SHEN MINMIN

To my parents and grandparents, for all your love and support.

Abstract

In many imaging applications, the acquisition of high resolution (HR) images/video is highly desirable, but may not be available due to various practical limitations. Even if the prices of HR sensors in many commercial digital cameras decrease, the transmission and processing of HR images/video could still be a problem. Besides, the limited number of sensors of many medical diagnostic systems or industrial non-destructive testing systems still constrains the resolution of acquired images. It is always desirable to develop some software schemes to enhance the resolution of the acquired images/video. Super-resolution reconstruction (SR) is a technique to yield one or a set of HR images from the low resolution (LR) observations.

In the first part of this thesis, we develop SR algorithms based on more general degradation models. Some existing multi-frame SR methods are constructed based on the assumption that the degradation process for each LR observation is identical. This assumption is widely accepted due to its simplicity, although it may lead to poor results with real data. For handling the real scenarios, we take into account more general assumptions: the LR observations are generated from the HR image through different degree of blurring or down-sampling with different magnification factors. For dealing with different out-of-focus blurs, we propose a new SR algorithm with better initial estimation and tighter constraint to improve the performance. For the case involving zoom motion, i.e. the LR images are captured with different magnification factors, we present a new linear model for the image capturing model and propose a MAP based algorithm.

The second part of this thesis deals with the application of SR as a post-processing of the acquisition and compression of images/video for improving the overall performance. From this point of view, the area of the applications of SR is broadened, and new SR algorithms should be proposed for particular requirements correspondingly. We propose to remain some video frame at HR during acquiring/encoding, and develop new SR algorithms to make use of these HR frames for restoring the whole video to HR. To utilize the information about high frequency component of the HR frames, we resort to example-based SR algorithms, which enhance the resolution of single LR image by adding HF details from a set of HR patches and cannot be applied directly to video enhancement. We address the problem of applying example-based algorithms on video by exploiting the inter-resolution correlation (the LR mapping between the down-samplings and their HR counterparts) and inter-frame correlation (the correlation among neighboring LR frames). We also explore its application in down-sampling based coding (DBC) and low power video acquisition.

Acknowledgement

First of all, I would like to express my at most gratitude to my supervisor, Associate Professor Xue Ping, for his supervision, advice, and guidance on this research and encouragement and support through the time of my study towards the Doctor of Philosophy degree. His knowledge, wisdom, modesty and sprit inspire and enrich my growth as a student and a researcher want to be.

I greatly acknowledge my fellow colleague Dr. Wang Ci for his advice and for providing the inspiring discussions and comments with his abundant knowledge to improve my works. Many thanks go in particular to Associate Professor Lin Weisi, for his thoughtful comments and encouragement on my research.

I would also like to express my appreciation to Assistant Professor Yap Kim Hui and his students Dr. He Yu and Dr. Chen Li for their kind help in my research.

Also, my special thanks go to my best friends Bao Man, Jin Feng, Ding Huijun, Luo Xuewen for bringing warm encouragements, joy and kind help to my life in Singapore. And I also would like to thank the technicians of the Media Technology Lab.

Finally, I am so grateful to my family for their support and understanding. My parents and grandparents cultivated my desire for knowledge and habit of reading since my childhood, and take pride in me for every little improvement I achieve. Their love and dedication are always the greatest encouragement for me.

Contents

Abstract	i
Acknowledgement	iii
List of Figures	vii
List of Tables	xi
List of Abbreviations	xii

1 Introduction

1.1 Motivation.....	2
1.2 Objectives.....	6
1.3 Major Contributions.....	7
1.4 Organization.....	8

2 Review of SR Methods

2.1 Multi-frame SR Methods	10
2.2 Example-Based Approach: Single-frame SR Methods	22

3 SR from Multiple Images Captured with Different Out-of-focus

Blur

3.1 Introduction.....	28
3.2 Problem Statement	30
3.3 MISO Framework	31
3.4 Proposed SR Framework.....	32
3.4.1 Construction of the Initial Estimation	34

3.4.2 Multi-channel Constraints	35
3.5 Experiments	36
3.6 Summary	44

4 SR from Multiple Down-sampled Images with Different Arbitrary Zooming Ratios

4.1 Introduction	45
4.2 Problem Statement	47
4.2.1 Image Capture Model without Zoom Motion	47
4.2.2 Image Capture Model with Zoom Motion.....	49
4.3 MAP Framework for Registration and Estimating the HR Image.....	51
4.3.1 MAP Framework.....	52
4.3.2 Estimation of Zooming Ratios M_k	52
4.3.3 Estimation of HR Image $\mathbf{x}$	53
4.4 Experiments	55
4.5 Summary	57

5 Applications of SR Techniques on Video Coding and Acquisition

5.1 Introduction	59
5.2 Down-sampling Based Video Coding.....	63
5.2.1 Background	63
5.2.2 Proposed DBC.....	65
5.2.3 Rate Distortion Analysis	66
5.3 Low-Power Video Acquisition.....	71
5.3.1 Background	71

5.3.2 Proposed Video Acquisition Scheme	72
5.4 Proposed SR Technique	74
5.4.1 System Model	75
5.4.2 Existing Example-based SR Algorithms	77
5.4.3 Proposed SR Algorithm	79
5.4.4 Implementation Details	82
5.5 Experiments	83
5.5.1 Down-sampling Based Coding.....	83
5.5.2 Low-power Video Acquisition.....	99
5.6 Summary	106
6 Conclusions and Recommendations	
6.1 Conclusions.....	108
6.1.1 SR Algorithms Based on More General Degradation Models	108
6.1.2 Applications of SR as Post-processing.....	110
6.2 Future Works.....	112
Author's Publications.....	114
Bibliography.....	116

List of Figures

Figure 2.1 Pictorial illustration of multi-frame SR for a typical digital imaging system. The degradation model is a mathematical description of the degradation process, by which each LR image can be obtained from the HR image. The inverse problem addresses the issue of estimating the original HR image.....	11
Figure 2.2 Basic premise for SR (reprinted from [18])	14
Figure 2.3 Block diagram showing raster-order per-patch processing. At each step, local low- and high-frequency details (in green and red, respectively) are used to search the training set for a new HF patch, which is added to the HF image (reprinted from [16])......	24
Figure 2.4 Patch centers, LR and HR patches, and HR patch overlap. The shaded HR patches have already been processed, and the shaded area overlapped by the current (unshaded) patch are used to enforce spatial consistency in the HR details (reprinted from [97])......	24
Figure 2.5 Steps of LLE: (1) Assign neighbors to each data point x_i . (2) Compute the weights w_{ij} that best linearly reconstruct x_i from its neighbors, solving the constrained least squares problem in (2.10). (3) Compute the low-dimensional embedding vectors best reconstructed by minimizing (2.11) (reprinted from [98])......	26
Figure 2.6 Neighbor embedding procedure applied to a LR patch for 3X magnification: (a) true HR patch; (b) input LR patch down-sampled from (a); (c) five nearest neighbor LR patches from the training images; (d) HR patches from the training images corresponding to the LR patches in (c); (e) target HR patch constructed from (d). (reprinted from [15])......	27
Figure 3.1 Flowchart of MISO framework.....	32
Figure 3.2 Flowchart of proposed framework.....	34
Figure 3.3 Test images from “ <i>F16</i> ”: (a) LR1 (b) LR2 (c) LR3 (d) LR4	37
Figure 3.4 Test images from “ <i>Hill</i> ”: (a) LR1 (b) LR2 (c) LR3 (d) LR4.....	37
Figure 3.5 Test images from “ <i>Man</i> ”: (a) LR1 (b) LR2 (c) LR3 (d) LR4.....	37
Figure 3.6 PSNR of the reconstructed HR image vs. number of iterations by MISO for four cases and the proposed method using tested images (a) “ <i>Hill</i> ” (b) “ <i>F16</i> ”(c) “ <i>Man</i> ”.	40

Figure 3.7 Part of SR reconstruction for “ <i>F16</i> ” by (a) Standard SR method—24.47dB. (b) MISO— 28.01dB. (c) Proposed method— 28.51dB. (d) Original image.....	41
Figure 3.8 Part of SR reconstruction for “ <i>Hill</i> ” by (a) Standard SR method — 25.42dB (b) MISO — 27.12dB (c) Proposed method—27.56dB (d) Original image.....	41
Figure 3.9 Part of SR reconstruction for “ <i>Man</i> ” by (a) Standard SR method — 24.9dB (b) MISO — 27.0dB (c) Proposed method—27.4dB (d) Original image	41
Figure 3.10 SR for real-life images. (a) Four LR images. (b) Bicubic interpolation of LR1. SR reconstructions by (c) Standard SR method, (d) MISO, (e) proposed method. (f)-(i) Selected enlarged region of (b)-(e), respectively.....	43
Figure 4.1 The LR images captured by a camera lens at different planes have different zooming ratios.....	46
Figure 4.2 Illustration of the relation between the LR pixels and the HR pixels in the HR pixel grids	51
Figure 4.3 Convergence of estimating M_4, M_8, M_{12}, M_{16} using(4.6).....	56
Figure 4.4 (a) Original image, the reconstructed image by (b) Li’s algorithm using 6 LR images (c) the proposed algorithm using 6 LR images, (d) bicubic interpolation, (e) Li’s algorithm using 16 LR images, (f) the proposed algorithm using 16 LR images.	57
Figure 5.1 The block diagram of the proposed video acquisition system.....	60
Figure 5.2 Block diagram of the proposed DBC encoder	65
Figure 5.3 Analytical rate distortion curves of the regular coding and the proposed DBC	70
Figure 5.4 Block diagram of the proposed uniform acquisition and SR on (a) VASI for saving power, (b) conventional sensor imager for producing high-frame-rate and HR frames.....	73
Figure 5.5 An example of compressing video from uniform acquisition using SVC..	74
Figure 5.6 Block diagram of the example-based SR algorithm in [15]	78
Figure 5.7 Block diagram of the proposed example-based SR algorithm	80
Figure 5.8 Average PSNR comparison of the regular coding (H.264), proposed coding and LLE method using only intra-frames for: (a) <i>Mother & daughter</i> (b) <i>Missa</i> (c) <i>Salesman</i>	85

- Figure 5.9 Mother & daughter sequence: (a) training image (intra-frame coded with QP=30), (b) original 30th frame and (c) its down-sampled version with quantization noise. 87
- Figure 5.10 *Mother & daughter* sequence: (a) Enlarged LR observation (b) SR result of LLE method (PSNR=33.9dB) (c) initial estimation (PSNR=34.5dB) (d) iterative correction (PSNR=34.9dB). 87
- Figure 5.11 Average PSNR comparison of the regular coding (H.264), proposed coding and LLE method using both intra-frame and inter-frame at intra-frame QP=40 for: (a) *Mother & daughter* (b) *Missa* (c) *Salesman*. 90
- Figure 5.12 Average PSNR comparison of the regular coding (H.264), proposed coding and LLE method using both intra-frame and inter-frame at intra-frame QP=30 for: (a) *Mother & daughter* (b) *Missa* (c) *Salesman*. 91
- Figure 5.13 PSNR comparison of the regular coding (H.264), proposed coding, Bicubic interpolation and LLE method in a GOP for sequence *Mother & daughter* at (a) intra-frame QP=40 (b) intra-frame QP=30. 92
- Figure 5.14 PSNR comparison of the regular coding (H.264), proposed coding, Bicubic interpolation and LLE method in a GOP for sequence *Salesman* at (a) intra-frame QP=40 (b) intra-frame QP=30. 93
- Figure 5.15 Average PSNR comparison for sequence *Foreman* with different GOP size of (a) (b) $P=50$ and (c) (d) $P=10$ 94
- Figure 5.16 Video inter-frame samples (a) Original *Missa* frame (b) Regular coding, PSNR=34.9dB. (c) Proposed coding, PRNR= 35.3dB. (d) Original *Salesman* frame (e) Regular coding, PSNR=28.3dB. (f) Proposed coding, PRNR= 29dB. (g) Original *Foreman* frame (h) Regular coding, PSNR=29.9dB. (i) Proposed coding, PSNR=31.4dB. 96
- Figure 5.17 PSNR comparison of the regular coding (H.264), proposed coding at different down-sampling factors in a GOP for sequence *Football* at (a) intra-frame QP=40 (b) intra-frame QP=30. 97
- Figure 5.18 Average PSNR obtained by different approaches at various key frame ratios on five test sequences: (a) *Foreman*, (b) *Hall*, (c) *Garden*, (d) *Container*, (e) *Coastguard*. 101
- Figure 5.19 Samples of reconstructed frames obtained by bicubic interpolation (left) and the proposed SR (middle), original frame (right) of: ((a)-(c)) *Foreman*, ((d)-(f)) *Hall*. 103
- Figure 5.20 Samples of reconstructed frames obtained by bicubic interpolation and the proposed SR, original frame of: ((a)-(c)) *Garden*, ((d)-(f)) *Container*, ((g)-(i)) *Coastguard*. 104

Figure 5.21 Reconstructed frames of noisy video sequence Mobile using (a) bicubic interpolation or super-resolved by (c) LLE method (d) the proposed SR with the help of the training image (e). A part of each sample frames is enlarged as (f)-(j), respectively. 106

List of Tables

Table 1 Estimation error and PSNR for four cases of test images.....	39
Table 2 The estimation error of $M_k, k = 2, \dots, 16$	56
Table 3 Comparison of power saving (measured by pixel number).....	102

List of Abbreviations

1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional
AM	Alternating minimization
AWGN	Additive white Gaussian noise
BSR	Blind super-resolution
CCD	Charged-coupled device
CG	Conjugate gradient
CLS	Constrained Least Squares
DBC	Down-sampling based coding
DCT	Discrete cosine transform
FOV	Field-of-view
GOP	Group of pictures
FR	Full resolution
HF	High frequency
HR	High resolution
JPEG	Joint photographic experts group
LLE	Locally linear embedding
LF	Low frequency
LR	Low resolution
MAP	Maximum a posteriori

MB	Macro block
MC	Motion compensation
MF	Magnification factor
MISO	Multi-input single-output
ML	Maximum likelihood
MRF	Markov random field
MSE	Mean square error
NLS	Nonlinear least squares
PDF	Probability density function
POCS	Projection-onto-convex-sets
PSF	Point spread function
PSNR	Peak signal-to-noise ratio
QP	Quantization parameter
RD	Relative displacement
ROI	Region of interest
SAR	Simultaneous autoregressive
SR	Super-resolution
SRC	Super-resolution constraint
SVC	Scalable video coding
TV	Total variation

Chapter 1

Introduction

There are two ways to achieve HR images: one belongs to hardware schemes and the other relies on software schemes. Although in many imaging system the acquisition for HR images becomes achievable, as the prices of HR sensors decrease, the transmission and processing of HR images may still suffer a high cost or be constrained by physical limits. Besides, in many medical diagnostic systems or in industrial non-destructive testing systems, the acquisition of HR images is still a problem. Therefore, it is promising to develop software schemes, i.e., some algorithms post-processing the LR images to achieve HR images. SR is such a technique for producing one or a set of HR images from one or a set of LR ones.

As a focus of research in image processing, SR techniques have been applied in many areas, of which some examples are [1]:

- Mobile LR imaging devices, such as mobile phones and portable computers embedded with LR cameras (e.g. web-cameras) [2];
- Multi-camera and multi-view recording in aerial or satellite imaging [3-5];
- Medical and biological imaging systems, where HR 2-D or 3-D images are always desired by combining different images obtained from the same object in different conditions (different time, viewing angles or energy levels) [5-7];

- Holographic and 3-D television imaging, where the data transmission limits the resolution of image sequences, thus the reconstruction of the HR 3-D scene from LR images becomes crucial;
- 3-D photography and surface modeling for 3-D scenes.

The studies on SR problem can be categorized into two levels: designing SR algorithms and the application of SR. The two levels are connected, although they have different focuses. The objective of introducing such methodologies is not only to reveal the importance of SR as an efficient tool to improve the quality of acquired images/video, but also to shed light on the inherent characteristics of the advanced signal processing techniques that manage to adapt to the specific requirement of acquisition and compression of images/video, providing a new standpoint in the evaluation of SR techniques. As the application of SR algorithms determines the characters of inputs (acquired images/video/bitstream), it restricts the designing of algorithms. This thesis will attack several issues in these two levels.

1.1 Motivation

In this thesis, our study is motivated by the cue on SR of its algorithm designing and applications, which will be elaborated in this section.

We will develop new algorithms based on more general assumptions. For different acquisition systems resulting in different degradation models describing the relation between the obtained LR data and desired HR image, numerous algorithms are constructed to address the SR problem. How to improve the performance of SR

algorithms with more general assumptions attracts many researchers' attentions. Based on a simple degradation model, which assumes pure translational motion and common space-invariant blur, many robust and fast SR algorithms have been developed. Unfortunately, these algorithms may not be successful if a more sophisticated model is required to describe the relation between the HR images and the LR observations. For instance, the auto focus function is not contained in some consumer level cameras (e.g. embedded cameras in hand phones and portable PCs). Therefore, blurring varies within the obtained LR images, since each of them may have different out-of-focus blur. Some existing RBAs randomly select one LR image and use its interpolation as the initial estimation. If the LR images suffer different degree of blurring, the selection of different referenced image as the initial estimation would affect the final SR result, since the reconstruction constraints are not a perfect convex set and the solution is not global minima. Therefore, a better initial estimation could be constructed to improve the performance of SR algorithms. Besides, tighter reconstruction constraints could restrict the indetermination of the SR solution.

Additionally, the condition of pure translational motion between the LR observations is not always followed in actual acquisition process, where slight zooming or rotation is almost unavoidable if the camera is hand-held. Taking zooming into account, a linear model is required to be constructed to represent the relation between the underlying HR image and LR images captured with arbitrary sampling lattices.

The other concern of this thesis is the application of SR. As a post-processing technique, SR is used for resolution enhancement of the obtained LR frames. An

interesting idea is to incorporate SR into the acquisition process or coding process to improve the over-all performance of the integrated acquisition or coding system. For this purpose, the area of applications of SR algorithms is broadened. For instance, a down-sampling based coding (DBC) has been designed for improving the video quality at low bit rate. In DBC schemes, the encoder is modified and deliberately reduces the resolution of encoded video frames, and SR could be applied at the decoder to restore the HR of decoded frames. According to the ways in which the encoder down-samples the video frames, new SR algorithms could be designed to fully make use of the encoded data to produce a HR video.

Besides, lossy image/video compression is common in network/wireless transmitting and recording. Due to the limitation on bandwidth or capacity of storage, a high compression ratio is usually required. Artifacts due to compression are then presented in the obtained images. How to reduce such artifacts is essential to both scientific and consumer communities. Some of the SR methods focus on the compressed images for reducing the artifacts caused by compression, in addition to the resolution enhancement. Therefore, the development of efficient SR algorithms is crucial to the quality enhancement of the compressed images as well as compressed videos.

In many art-of-state compression methods, such as MPEG and ITU family of standards, the discrete cosine transform (DCT) is utilized and the quantization of the DCT coefficients results in errors in the image representation as well as a loss of signal information [8]. A Bayesian framework was used to establish the maximum a posteriori (MAP) estimator for SR images. Utilizing the statistics of quantization noise,

the compression process was incorporated into the MAP model in the form of quantization constraint set to address the SR problem. However, these RBAs merely guarantee a solution satisfying the reconstruction constraints, while the high frequency (HF) components of the resultant image are generally not reconstructed well. Thus the improvement of these RBAs is limited, since the redundancy in the LR image sequence, which is exploited for reconstructing the HR image, is reduced by the compression procedure.

Example-based SR algorithms, originally proposed as single-frame SR approaches, give an inspiration for how to utilize the HR frames to improve the SR performance. These example-based SR algorithms treat the HR images in a very similar way: they construct a training set consisting of a large number of patch pairs from HR images and their down-sampled versions [9-15]. The fine details of the HR patches that correspond to different image region seen at a LR can be learned to predict the fine details in the tested LR image. Unfortunately, the results of directly applying the existing example-based SR algorithms in video enhancement are not plausible due to their intrinsic characteristic as single-frame SR approaches: the created HF details would lead to unpleasant subjective view in moving images (video). In addition, compression is almost unavoidable in acquired videos due to limited capacity of storage or transmitted bandwidth. The compression noise in the tested LR image is recognized as image content by the existing example-based SR algorithms and will be amplified in the reconstructed HR image. To address these two problems [16], the existing example-based algorithms need to be modified to meet the requirement of video enhancement.

1.2 Objectives

Motivated by the importance of SR, the works dedicated in this thesis focus in two respects of SR towards the following two objectives:

- **Design SR algorithms based on more general degradation**

models: to propose efficient methods to deal with the LR images with more relaxed assumptions. 1) To address the SR problem with different out-of-focus blur, a new framework with a better initial estimation and tighter constraints is established. 2) To represent the relation between the HR images and the LR images captured with arbitrary sampling lattices, a new degradation model is established. Based on this model, a MAP based SR algorithm is proposed.

- **Improve the efficiency of video acquisition using SR algorithms**

as post-processing: we generalized the definition of video acquisition and incorporate SR into the integrated video system, which can be used for DBC, low-power video acquisition and high-frame-rate HR video acquisition. To improve the over-all efficiency of the acquisition system, down-sampling is applied at some of the video frames during acquisition or before compression. The down-sampled frames are then restored back to HR by a SR process. In order to meet the requirement of video enhancement, a new SR algorithm is designed inspired by the example-based SR algorithms.

1.3 Major Contributions

In what follows in this thesis, we study the SR problem in terms of its methodology and application.

- We propose two SR algorithms to relax some generally adopted assumptions, such as: the relative motion is a pure translation, and the point spread function (PSF) and magnification factor (MF) are common and space invariant in all LR images. In Chapter 3, we take into account the different blur of each LR image. To improve the efficiency of SR, we design a new framework to obtain better initial estimation and tighter constraints to limit the indetermination of the SR solution. In Chapter 4, we present a new linear model to represent the relation between the underlying HR image and the LR images captured with arbitrary sampling lattices, assuming zooming exists when acquiring LR images. Based on this model, a MAP based SR algorithm is proposed.
- In Chapter 5, we propose an example-based SR algorithm which is applicable for enhancing the resolution of video frames. We also explore its application as a post-processing to improve the over-all performance of the generalized video acquisition system. Specifically, we design a DBC scheme with the proposed SR algorithm, and its advantage over regular coding at the low bit-rate range is analyzed. With the new SR algorithm, we also propose a low power video acquisition with VASI system. A similar approach can be used for capturing high-frame-rate and HR videos with conventional imaging systems.

1.4 Organization

The remainder of the thesis is organized as follows:

Chapter 2 presents the research background which gives an overview of SR algorithms from both multi-frame input and single-frame input points of view.

In Chapter 3, we propose a new algorithm to improve the performance of SR from images captured with different out-of-focus blur. A multiple reference-based fusion framework is proposed, where the SR outputs using different input LR image as referenced image are regarded as independent output channels. Different from the conventional multi-input single-output (MISO) framework, the dependency of initial estimation on the ultimate result is alleviated in the proposed framework. The improvement of the proposed algorithm is verified by testing simulated and real images.

In Chapter 4, an algorithm is proposed for reconstructing a HR image from LR images with different arbitrary zooming ratios. We overcome the difficulty of modeling an image capturing process including nonlinear zoom motion, and construct a linear system to relate the underlying HR image and the LR observations. Based on this system, a MAP based algorithm is proposed to estimate the zoom motion and the underlying HR image. Simulations verify the effectiveness of the proposed algorithm.

To address these two problems in using example-based SR algorithms in video enhancement, which is pointed out in [16], we propose an example-based algorithm for video enhancement purpose in Chapter 5. The potential applications of the proposed SR algorithm as a post-processing in the areas of video coding and video

acquisition are also explored. Experiments are constructed using various kinds of videos to present the effectiveness of the proposed video coding and video acquisition schemes

Finally, Chapter 6 summarizes the work in this thesis. Some future challenges in SR are also outlined and discussed.

Chapter 2

Review of SR Methods

Since the basic problem of multi-frame SR was first addressed by Tsai and Huang [17], many approaches have been developed to go beyond the camera's resolution limit. Reviews of the literature can be found in [1, 18-21]. According to the numbers of input LR images, these methods can be classified into Multi-frame SR methods and Single-frame SR methods. Multi-frame SR is often referred to a process of aggregating multiple LR images to form a HR one. This class of SR methods assumes that there exist some relative motions between the camera and the scene when capturing the LR images and the SR problem is treated as an inverse problem. Different to multi-frame SR approaches, some learning-based approaches have been proposed for enlarging single LR image, which are called example-based SR methods. As follows, we elaborate the representative approaches of these methods and present relative studies reported so far.

2.1 Multi-frame SR Methods

For multi-frame SR methods, each LR observation is assumed to be obtained from the HR image by a degradation process, which includes warping, blurring, and down-sampling. This degradation model describing the relation between the given LR

images and the estimated HR image serves as the SR constraints (SRCs). A pictorial illustration of the SRCs is shown in Figure 2.1.

Figure 2.1 Pictorial illustration of multi-frame SR for a typical digital imaging system. The degradation model is a mathematical description of the degradation process, by which each LR image can be obtained from the HR image. The inverse problem addresses the issue of estimating the original HR image.

The main idea behind the multi-frame SR methods is how to exploit the complementary information contained in each LR observation to construct a HR one.

These LR images may be obtained by [1]:

- A moving camera with fixed scene;
- A fixed camera but focusing on a slightly moving scene;
- Multiple fixed cameras focusing on a same fixed scene;
- Multiple fixed cameras but focusing on a slightly moving scene, etc.

Each LR image captures some unique characteristic of the original HR image, since the relative motion between the scene and the camera, the readout noise of the sensors, and possibly the optical lens characteristics vary through the time (such as the amount of out-of-focus blur). Multi-frame SR methods are to fuse the LR noisy

blurred images for producing a HR image, which has more HF component and less noise and blur effects compared with any of the LR observations.

The degradation model can be represented in matrix notation, i.e.

$$\mathbf{y}_k = \mathbf{W}_k \mathbf{x} + \mathbf{e}_k, k = 1, 2, \dots, N \quad (2.1)$$

The vector $\mathbf{y}_k$ consists of the k th LR image pixels in the lexicographical order, and $\mathbf{x}$ is the HR image. Assume that the desired HR image is of size $N_H \times N_H$, given N measured LR images, each of which is of size $N_L \times N_L$. The image capturing process is modeled by a degradation matrix $\mathbf{W}_k$ with an additive noise $\mathbf{e}_k$. The degradation matrix $\mathbf{W}_k$ includes the effect of blur $\mathbf{H}_k$, motion $\mathbf{F}_k$ and down-sampling $\mathbf{D}_k$:

$$\mathbf{W}_k = \mathbf{D}_k \mathbf{H}_k \mathbf{F}_k \quad (2.2)$$

The SRCs have been used in numerous SR methods to seek the reverse solution of (2.1). In the RBAs, SR problem is solved by finding the solution after denoising. Therefore, the SR result in (2.1) regresses to seek the solution

$$\mathbf{y} = \mathbf{W} \mathbf{x} + \mathbf{e} \quad (2.3)$$

where $\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix}$, $\mathbf{e} = \begin{pmatrix} \mathbf{e}_1^0 \\ \vdots \\ \mathbf{e}_k^0 \end{pmatrix}$ and $\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_k \end{pmatrix}$. The notation $\mathbf{e}_k^0$ denotes the error in

estimating the noise.

The premise in SR techniques is that the LR images are captured as different representations of the same scene (see Figure 2.2). The simplest model is assuming a translational movement between these images. Additionally, the blurring introduced by the integration of the LR sensor and lens, and the CCD noise are also considered

for a general degradation model. For compensating the degradation modeling operators, classical SR methods combine registration (movement compensation), interpolation (down-sampling compensation) and fusion to construct a HR image. Ur and Gross [22] performed a non-uniform interpolation on the shifted LR images, and then applied deblurring separately. This algorithm does not take additive noise into account, and it is unstable if the LR observations have close spatial positions or the number of observations is very large. Assuming that the blur is space invariant and common for all the LR images, the warps between them are a pure translational motion, and the additive noise is white, Elad and Hel-Or proposed a fast SR algorithm [23]. This algorithm uses a simple non-iterative fusion method, and the theoretic justification of such interpolation-fusion scheme is also provided in [23]. This special simplified SR problem can also be addressed by frequency domain methods such as [24, 25]. Unfortunately, such assumptions may not hold in actual situations, e.g. each LR images may suffer different degree of blurs, or the camera does not move parallel to the scene. Some early works do not have this restriction, but assume that the related parameters are exactly known [26]. Therefore, the SR problem regresses to the inversion of a large sparse matrix. In fact, the error is unavoidable in estimating those parameters, which makes these methods produce poor results when dealing with real images.

Figure 2.2 Basic premise for SR (reprinted from [18])

When considering a more general degradation model, such as described in (2.1), numerous approaches treat SR as an inverse problem. In these spatial domain-based multi-frame SR methods, a cost function is explicitly or implicitly (such as the Projection-onto-convex-sets (POCS) based methods [27-30]) defined to estimate the HR image $\mathbf{x}$. This type of cost function assures the fidelity of the final solution to the observed data, i.e.

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J(\mathbf{x}) = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{W}\mathbf{x}\|_{\beta} \quad (2.4)$$

where $\|\cdot\|_{\beta}$ denotes the L_{β} norm. With the additive noise $\mathbf{e}_k$ modeled as white Gaussian noise, the maximum likelihood (ML) methods seek the solution of (2.4) using a L_2 norm ($\beta = 2$) [23, 31, 32]. As shown in [33], $\beta = 2$ results in pixel-wise

mean and $\beta = 1$ results in pixel-wise median of the registered LR images in the fusion step. Therefore, the robust SR method proposed in [34] is in essential using L_1 norm minimization for (2.4).

In general, SR is an ill-posed problem either because it has infinite possible solutions or because it has great sensitivity to the noise. In the linear model, the ill-posedness results from inverting the effects of the matrix $\mathbf{W}$ in (2.3), which is an ill-conditioned or even singular matrix. To solve this problem, and to obtain an unique and stable solution, some form of regularization must be included in the cost function $J(\mathbf{x})$:

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{W}\mathbf{x}\|^2 + \alpha Y(\mathbf{x}) \quad (2.5)$$

The regularization term poses a penalty on the unknown $\mathbf{x}$ for choosing a stable solution. The regularization parameter α controls the strength of the penalty. The selection of α could be either accomplished manually or automatically using some method such as generalized cross-validation [35], L-curve [36] and other techniques [37]. Some SR algorithms simultaneously estimating the HR image and the regularization parameters have also been proposed [38] [39-42].

More importantly, a proper form of the regularization term should be constructed. In Constrained Least Squares (CLS) methods [39, 43-46], $\|\mathbf{C}\mathbf{x}\|^2$ is used for $Y(\mathbf{x})$, where the operator $\mathbf{C}$ is generally a high-pass filter. This type of regularization often penalizes HF energy of the solution, thus results a smooth HR image. For producing sharper HR images, the edge-preserved regularization is considered,

$$Y(\mathbf{x}) = \alpha \sum_{c \in S} \phi(\Phi \mathbf{x}) \quad (2.6)$$

where ϕ is a potential function that depends only on the pixel values located within clique c , and S denotes the set of cliques. Typically, the elements of the sequence $\{\Phi \mathbf{x}\}$ are either first- or second- order differences neighboring pixels from $\mathbf{x}$. The function ϕ is assumed to be smooth, convex, and edge-preserving (i.e. $\phi(t) < t^2$ as $|t| \rightarrow \infty$). Examples of such functions can be found in [47-50]:

$$\begin{aligned} \phi(t) &= t^\beta, 1 < \beta \leq 2, \\ \phi(t) &= \sqrt{\beta + t^2}, \\ \phi(t) &= \begin{cases} t^2 / 2 & \text{if } |t| \leq \beta \\ \alpha |t| - \beta^2 / 2 & \text{if } |t| > \beta \end{cases} \end{aligned} \quad (2.7)$$

From the statistical perspective, a Bayesian approach provides a flexible way to model a priori knowledge concerning the solution. The MAP estimator of $\mathbf{x}$ seeks the estimate $\hat{\mathbf{x}}$ that maximize the a posterior probability

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \Pr(\mathbf{x} | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) \quad (2.8)$$

Taking the logarithmic function and applying Bayes' theorem to the conditional probability, the MAP optimization problem can be rewritten as

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{\ln \Pr(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N | \mathbf{x}) + \ln \Pr(\mathbf{x})\} \quad (2.9)$$

where $\ln \Pr(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N | \mathbf{x})$ is the log-likelihood function and $\Pr(\mathbf{x})$ is the a prior probability density function (PDF) of $\mathbf{x}$. Many SR algorithms are constructed based on the MAP framework [1, 38, 51-57], since the Bayesian estimation framework is much richer. A major advantage of the Bayesian framework is the use of an edge-preserving image prior model. For instance, the total variation (TV) method [58], i.e.

$Y(\mathbf{x}) = \|\nabla \mathbf{x}\|_1$, is opted in many efficient MAP based SR algorithms such as [33, 59, 60].

The comparison of the strength and weakness of the aforementioned spatial domain-based SR methods is given as follows. It can be shown that if the Gaussian prior is used for a MAP estimator, the high frequency component is suppressed in any location of the image. In this sense, it is equal to the CLS estimator. Another special case of MAP estimation is the ML estimation, in which no prior term is included. Compared with the MAP-based methods, the POCS-based methods are easier to incorporate a priori constraints. But they have the weakness of non-uniqueness of solution and slow convergence.

Another class of MAP methods uses the regularization-like penalty terms derived from collections of training samples [12, 61-63]. These methods are also called reconstruction algorithm [61]. The term “reconstruction” was coined in [61] since the proposed algorithm includes the results of a set of recognition decisions as the prior term. The prior term enforces the condition that the gradient of the reconstructed image should be equal to the one of the best matching training image. In this way, the relation between the training images and their LR versions is learned and used for resolving the test image. Although the advantage of the reconstruction algorithms over the RBAs is shown in [61], they are often effective when applied to very specific scenarios (e.g. faces or text), since a large number of examples are required for the training database.

As the SR methods using motion cue belong to 2-D dense feature-matching technique, they assume that all the frames are captured at the same spatial resolution.

These methods fail when this assumption does not hold. To address this problem, some researchers work with zoom as a cue. In the area of computer vision, this cue has been studied in determination of depth [64-66], minimization of view democracies [67], and zoom tracking [68, 69]. Assuming that the zooming ratios between the obtained images are known, Joshi et al. [70] modeled the HR image either as a homogeneous Markov random field (MRF) or a simultaneous autoregressive (SAR) model, which can be chosen according to the available computation one can afford. The corresponding field parameters for the chosen model are learned from the most zoomed observation, and later used to super-resolve the least zoomed observation to obtain an image with largest scene and highest resolution among all the observations. Rajan and Chaudhuri proposed an approach to simultaneously estimate the HR image and the depth map at HR from LR defocused images [71]. This approach combines the two tasks of the super-resolved depth recovery and the super-resolved space varying image restoration through the interplay of two separate MRFs, which represent the depth map and the intensity field, respectively. For generating a HR image with arbitrary MF rather than limited to the spatial resolution of the given observations, a degradation model is required to be established to register the LR observations acquired with different zooming ratios. To circumvent the difficulty of constructing such a linear image capture model, Li [72] argued that all LR observations are linked by a line-geometric model under the assumptions of constant scene depth, small number of observations and zooming speed. This model alleviates the data inconsistency in fusing the differently zoomed LR images.

Having reviewed the existing SR methods which are often employed, we introduce some methods with the respective of the advanced issues within the SR area [18]:

- SR considering registration error,
- Blind SR image reconstruction,
- Computationally efficient SR algorithm.

In SR process, it is very crucial to accurately calibrate the sub-pixel displacements between the LR observations. However, registration is a very difficult task in itself. Taking into account the impact of unreliable initial registration, some SR algorithms are proposed. Considering the error caused by inaccurate registration in the degradation matrix $\mathbf{W}_k$, a class of SR methods employs iterative schemes based on alternating minimization (AM) to estimate the HR image and the motion parameters alternatively [51, 73]. Similar to these AM methods, Woods et al. [74] proposed algorithms for joint estimation of HR images and motion parameters using Bayesian and MAP formulation. To improve the performance of the AM methods, a nonlinear least squares (NLS) technique for simultaneous registration and SR is proposed in [60]. This method integrates registration and SR into a single estimation process, and employed an iterative scheme based on a NLS method to estimate the motion parameters and the HR image simultaneously.

Another approach to minimize the effect of the registration error is based on channel adaptive regularization [39, 41, 75]. The basic concept of channel adaptive regularization is that in estimating the HR image the LR images with a large amount of

registration error should contribute less than reliable LR images. It is assumed that the degree of registration error is different in each channel (LR image) and the regularization functional is applied to adaptively control the effect of registration error in each channel. For example, a regularization functional that is performed on the data consistency was proposed in [39] and the corresponding cost function is defined as

$$J(\mathbf{x}) = \sum_{k=1}^N \alpha_k(\mathbf{x}) \|\mathbf{y} - \mathbf{W}\mathbf{x}\|^2 + \|\mathbf{C}\mathbf{x}\|^2. \text{ The regularization functional } \alpha_k(\mathbf{x}) \text{ should have}$$

the properties as follows to reduce the effect of errors in SR process:

- $\alpha_k(\mathbf{x})$ is inversely proportional to $\|\mathbf{y} - \mathbf{W}\mathbf{x}^n\|^2$;
- $\alpha_k(\mathbf{x})$ is proportional to $\|\mathbf{C}\mathbf{x}^n\|^2$;
- $\alpha_k(\mathbf{x}) > 0$;
- $\alpha_k(\mathbf{x})$ considers the influence of the cross channel.

where $\mathbf{x}^n$ represents the partially reconstructed image at the n -th iteration.

Some SR methods which do not require explicit sub-pixel motion estimation have emerged in recent years [76-78]. In this class of methods, accurate motion information in sup-pixel precision is not a key issue, thus the effect of subsequent registration error on reconstructed HR images is avoided. To achieve this objective, i.e. accomplish SR on general image sequences while avoiding sub-pixel-accurate motion estimation, the algorithm in [76] measures the similarity of image patches across space and time extracting the motion information implicitly. Similarly, the approach in [77], which presents an extension of the block-matching 3-D filter

(BM3D) [79] for video SR, also avoids explicit motion estimation by classifying the image patches using block matching. As a specific extension of [76] for the purpose of video denoising and resolution enhancement, the algorithm in [78] was proposed based on multidimensional kernel regression. As this framework is fundamentally based upon the comparison of pixels in spatiotemporal neighborhood, it implicitly contains information about the local motion of the pixel across time, and thus capable of handling the fine-scale motion.

Besides assuming a priori known blurs, some researchers confront with a problem of blind super-resolution (BSR). For example, the SR techniques in [35, 80] are capable of handling parametric PSFs, which are modeled with one parameter. The first attempts for BSR with an arbitrary PSF appeared in [81, 82]. These methods represent the SR problem as the blind 2-D multi-input, multi-output system driven by poly-phase components of a band-limited signal. The same idea is adopted in [83], in which an extension of the multi-channel blind deconvolution (MBD) theory [84] is developed. This work constructed a cost function with the regularization carried out in both image and blur domains. To tackle the minimization of the regularized cost function, an AM approach consisting of two linear equations is used.

To apply the SR algorithms in practical situations, it is very important to develop efficient algorithms that reduce the computation cost. Besides the interpolation-based approaches [85-88] and some multi-frame SR methods designed for fast SR implementation mentioned earlier [23, 32, 33], the adaptive filtering approaches [89, 90] are also appropriate for this purpose.

2.2 Example-Based Approach: Single-frame SR

Methods

As the SR with multiple LR images from the same scene has been intensively studied, some single-frame SR (as coined in [16]) methods have emerged recently. As referred from the name, single frame SR produce a HR images from its single LR version. The related approaches include some image interpolation methods, such as bilinear interpolation, cubic convolution interpolation [91] and cubic spline interpolation [92]. These image interpolation methods are computationally efficient, but suffer from blurring edges and image details. Recently, some researchers advocated the approach on edge-guided interpolation [93-95].

As proposed in [16], the goal of example-based SR is to estimate missing HR details that isn't present in the original image, which cannot be made visible by simple sharpening. The basic idea of learning based SR algorithms is to learn the fine details that correspond to different image regions seen at a LR and then use those learned relation to predict the fine details in other images [16]. The same idea is used in [9, 11, 14, 61] to address the single-frame SR problem, while the relation between the pairs of training images is exploited from different perspectives. For example, a framework called "image analogies" is proposed in [11], which involves two stages. These two stages include:

- Design phase: presents training data comprises of pairs of images, with one image purported to be a "filter version" of the other;

- Application phase: the learned filter is applied to the testing image in order to create an “analogous” filtered result.

Such analogies between images could be used to describe the relation between a HR image and its down-sampled version, thus this training-based method could effectively super-resolve the testing image given some LR and HR pairs for its small portion.

Similarly, Freeman et al. [16] explored two different SR approaches to exploit the neighborhood relation between the pairs of HR and LR image patches. First, given a collection of HR images, a training set is generated by properly processing the HR images corresponding to a degradation process. The spirit of training set generation is similar to design phase in [11]. For a testing image desired to be enlarged, a naïve approach to make use of the training set patches is as follows: break the image into patches, and look up the missing HR detail. Unfortunately, this approach does not work because local patch information alone is insufficient for SR: the selected LR patches look fairly similar to the testing LR patch; however, the corresponding HR patches do not. To address this issue, a Markov network is used to probabilistically model the relation between patch pairs, and between neighboring HR patches [9, 13, 96]. Later, they developed a one-pass algorithm using the same local relation information as the Markov network in [16]. In this algorithm, the input image is divided into low-frequency (LF) patches that are traversed in raster-scan order. As illustrated in Figure 2.3, each HF patch at a time is selected by a nearest neighbor search from the training set based on the local LF details and adjacent, previously determined HF patches. To enforce the compatibility of neighboring high-frequency patches, the predicted patches are overlapped at their borders (see Figure 2.4).

Figure 2.3 Block diagram showing raster-order per-patch processing. At each step, local low- and high-frequency details (in green and red, respectively) are used to search the training set for a new HF patch, which is added to the HF image (reprinted from [16]).

Figure 2.4 Patch centers, LR and HR patches, and HR patch overlap. The shaded HR patches have already been processed, and the shaded area overlapped by the current (unshaded) patch are used to enforce spatial consistency in the HR details (reprinted from [97]).

Inspired by locally linear embedding (LLE) method [98], another approach for searching the HF patches in the training set is proposed in [15]. LLE is an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs. As shown in Figure 2.5, it is initially designed for compact representations of high-dimensional data. The data point in high-dimension can be mapped to the manifold of low-dimension using the following three steps [98]:

- 1) Assign neighbors to each data point X_i using the searching methods such as K nearest neighbors.
- 2) Compute the weights W_{ij} that best linearly reconstruct X_i from its neighbors, solving the constrained least squares problem

$$\varepsilon_1(W) = \sum_i \left\| X_i - \sum_j W_{ij} X_j \right\|^2 \quad (2.10)$$

- 3) Compute the low-dimensional embedding vectors best reconstructed by minimizing

$$\varepsilon_2(Y) = \sum_i \left\| Y_i - \sum_j W_{ij} Y_j \right\|^2 \quad (2.11)$$

Figure 2.5 Steps of LLE: (1) Assign neighbors to each data point X_i , (2) Compute the weights w_{ij} that best linearly reconstruct X_i from its neighbors, solving the constrained least squares problem in (2.10). (3) Compute the low-dimensional embedding vectors best reconstructed by minimizing (2.11) (reprinted from [98]).

LLE provides an effective approach to predict the data point in high-dimension X_i , given its neighbors X_j and their embedded manifold coordinates in low-dimension, i.e. Y_i and Y_j , since it constructs a neighborhood-preserving mapping. Based on this idea, Chang et al. [15] proposed an example based SR algorithm, whose spirit is that small image patches in the LR and HR images form manifolds with similar local geometry in two distinct feature spaces. In other words, the local geometry of the desired HR testing patch can be approximate by a linear combination of its neighbors. The weights determining the contribution of each neighbor to the

reconstructed HR patch are the same at LR. An example of the neighbor embedding procedure is illustrated in Figure 2.6. The reconstructed HR image is fused with all the estimated HR patches with overlapping, which is the same as in Figure 2.4. These example-based approaches produce plausible visual results, but they are usually not applicable for real-time processing, since the training database they require is usually huge and thus the training process is computationally costly.

Figure 2.6 Neighbor embedding procedure applied to a LR patch for 3X magnification: (a) true HR patch; (b) input LR patch down-sampled from (a); (c) five nearest neighbor LR patches from the training images; (d) HR patches from the training images corresponding to the LR patches in (c); (e) target HR patch constructed from (d). (reprinted from [15]).

Chapter 3

SR from Multiple Images Captured with Different Out-of-focus Blur

3.1 Introduction

Image SR is a process of combining multiple aliased LR images to construct a HR one. Typically, each LR image is assumed to be obtained from the HR image by a degradation process, which includes warping, blurring, and down-sampling. The degradation process for each LR image is different due to its different parameter. For example, each LR observation can suffer different out-of-focus blur. This degradation model describing the relation between the given LR images and the estimated HR image serves as the SR constraints (SRCs).

Under the SRCs, numerous SR methods have been developed to seek the inverse solution. The SR problem is ill-posed in that there are countless solutions with regards of the SRCs. A POCS algorithm is first formulated by Stark and Oskoui [99]. In the POCS-based SR methods, the space of HR image is intersected with a set of convex constraint sets. The residual error between the down-sampling of the current estimated HR image and the LR observation is back-projected to satisfy the SRCs [28-

30]. Another class of SR algorithms utilizes a Bayesian framework to incorporate image prior into the SR constraints. The MAP estimator is found as the solution [26, 51, 53]. Using the depth related defocus as the cue, Rajan and Chaudhuri [71] proposed an algorithm combining the super-resolved depth recovery and the super-resolved space varying image restoration. The depth and the image are both modeled as MRF and the HR fields is recovered by a MAP method. There is no relative motion between the scene and camera involved in the image capturing model, thus they did away the registration problem.

The precise estimation of degradation parameters, such as registration and the estimation of the imaging system's PSF are very important to the SR performance. However, the estimation error is unavoidable in real applications. Some SR algorithms adaptively adjust the channel (LR image) weights to tackle the inaccurate estimation of the registration and PSF parameters, and expect the optimization methods would give a global minimum solution. In these SR methods, one of the input LR image is randomly selected as the referenced image, and interpolated into the dimension of the HR image as the initial estimation. An iterative process, such as the steepest descent algorithm, is used to produce the final SR image [23, 39, 41]. We refer this scheme as a multi-input single-output (MISO) framework.

However, because the SRCs are not a perfect convex set and the optimization approaches may be trapped at local minima, the solution is sensitive to the initial estimation. As the reconstructed images initialized by interpolating different referenced image may be biased in a different way from the true HR image, we use a multiple reference-based fusion framework to describe the SR problem, where the SR

outputs using different input LR image as referenced image are regarded as independent output channels. In this framework, all the SR output channels are used to refine the reconstructed image.

The rest of this chapter is organized as follows. The problem formulation of image SR taking account of different blurs is introduced in Section 3.2. The MISO framework is described in Section 3.3. An algorithm under the proposed framework is developed in Section 3.4. Experiments constructed on synthetic and real-life images are presented and discussed in Section 3.5. Finally, Section 3.6 gives a brief summary.

3.2 Problem Statement

We follow the degradation model in (2.1), i.e. $\mathbf{y}_k = \mathbf{W}_k \mathbf{x} + \mathbf{e}_k, k = 1, 2, \dots, N$. More specifically, some common assumptions are made in this chapter for addressing the specific SR problem with differently out-of-focus blur.

1. $\mathbf{e}_k$ is modeled as white Gaussian noise (AWGN), and takes the form

$$\mathbf{e}_k \sim N(0, \mathbf{K}_k) \quad (3.1)$$

where $\mathbf{K}_k$ is the covariance matrix of the noise in the k -th frame.

2. All obtained images are assumed to be captured in plane parallel to the camera lens. Then all down-sampling matrices are assumed to be the same, i.e. $\forall k, \mathbf{D}_k = \mathbf{D}$.

3. In this work, the blur $\mathbf{H}_k$ is dominated by the out-of-focus blur, and their PSFs $\mathbf{B}_k$ are approximated as:

$$\mathbf{B}_k(x) = \begin{cases} \frac{1}{\pi R_k^2} & |x| \leq R_k \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

where R_k is the radius of the PSF $\mathbf{B}_k$, and the matrix $\mathbf{W}_k$ stands for the integrated degradation process to include the effects such as blur $\mathbf{H}_k$, motion $\mathbf{F}_k$ and down-sampling $\mathbf{D}$, i.e.

$$\mathbf{W}_k = \mathbf{D}\mathbf{H}_k\mathbf{F}_k \quad (3.3)$$

3.3 MISO Framwork

In the MISO framework, the maximum likelihood (ML) estimate of the SR result meets the following least squares constraint:

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} L(\mathbf{x}) \quad (3.4)$$

where $L(\mathbf{x})$ is the log-likelihood of the conditional probability density function of the LR observations

$$L(\mathbf{x}) = \left\{ \sum_{k=1}^N [\mathbf{y}_k - \mathbf{W}_k \mathbf{x}]^T \mathbf{K}_k^{-1} [\mathbf{y}_k - \mathbf{W}_k \mathbf{x}] \right\} \quad (3.5)$$

Let us consider the m -th LR images $\mathbf{y}_m$ is used as the referenced image, and is up-sampled to $\tilde{\mathbf{x}}_m^0$ be the initial estimate. Combined with the effect of correction, the final SR result is $\tilde{\mathbf{x}}_m = \tilde{\mathbf{x}}_m^0 + \delta \mathbf{x}$, where $\delta \mathbf{x}$ is the correction part.

As shown in Figure 3.1, the MISO framework seeks the HR estimate by an iteratively procedure:

- 1) Initialization: the m -th LR observation $\mathbf{y}_m$ is randomly selected as the referenced

image, and up-sampled as the initial estimation $\tilde{\mathbf{x}}_m^0$. The switch in Figure 3.1 closes at $n = 0$ only.

- 2) Correction: $\tilde{\mathbf{x}}_m^{n-1}$ is iteratively degraded by (2.1) and compared with the other observed LR frames $\mathbf{y}_k, k \neq m$, the differences are fed back to correct the current estimation $\tilde{\mathbf{x}}_m^{n-1}$ to generate a new estimation $\tilde{\mathbf{x}}_m^n$.

Figure 3.1 Flowchart of MISO framework

Here each LR observation suffers different degree of blurring, and this is a more realistic assumption than common blurring for all LR images. In this case, SR becomes more difficult as the number of the unknown parameters of the SRCs increases. In this chapter, a new framework is proposed to improve the SR performance by providing a better initial estimation and imposing constraints on the SRCs.

3.4 Proposed SR Framework

By MISO model, the SR estimation $\tilde{\mathbf{x}}_m$ is iteratively projected onto the convex set consisting of N least squares constraints. The cost function (3.5) can be reiterated as

$$L(\mathbf{x}) = \sum_{k=1}^N [\mathbf{y}_k - \mathbf{W}_k \mathbf{x}_m]^T \mathbf{K}_k^{-1} [\mathbf{y}_k - \mathbf{W}_k \mathbf{x}_m] \quad (3.6)$$

where $\mathbf{x}_m$ is the shifted ideal HR image satisfying $\mathbf{y}_m = \mathbf{D}\mathbf{H}_m \mathbf{x}_m$.

The constraint sets in (3.6) are quite loose, and hardly produces satisfactory result. The SR image by this model is still not satisfactory although the mean square error (MSE) of its degradation and the observed LR image set is converged. Some existing methods break the SR performance bound by images prior knowledge [61]. However, such knowledge is not always available.

The proposed framework is introduced here, which improves the performance without prior knowledge requirements. We regard an SR process with a different referenced image as an independent output channel. The ideal SR images on different output channels should be identical, apart from different shifts. Thus, the reconstruction distortion should be measured in all output channels. Motivated by this idea, the cost function of the SR image (3.6) is changed to

$$L(\mathbf{x}) = \sum_{m=1}^N \sum_{k=1}^N [\mathbf{y}_k - \mathbf{W}_{m,k} \mathbf{x}_m]^T \mathbf{K}_m^{-1} [\mathbf{y}_k - \mathbf{W}_{m,k} \mathbf{x}_m] \quad (3.7)$$

where $\mathbf{W}_{m,k}$ is the degradation matrix to describe the contribution of the k -th HR image to the m -th LR image and defined as $\mathbf{W}_{m,k} = \mathbf{D}\mathbf{H}_k \mathbf{F}_{m,k}$. The warping matrix $\mathbf{F}_{m,k}$ describes the translational motion between $\mathbf{x}_m$ and $\mathbf{x}_k$.

For clarity, we use $\mathbf{y}_1$ as the referenced image. The ideal HR image is then aligned to its up-sampling, i.e. $\mathbf{x} = \mathbf{x}_1$. The SR output of the m -th channel is a shifted HR estimate $\tilde{\mathbf{x}}_m$, $m \neq 1$. The displacement vector between the current frame $\tilde{\mathbf{x}}_m$ and the referenced frame $\mathbf{x}_1$ is denoted as $\mathbf{v}_{1m}$. Then $\tilde{\mathbf{x}}_m$ is written as

$$\tilde{\mathbf{x}}_m = \mathbf{F}_{1,m} \tilde{\mathbf{x}}_1 \quad (3.8)$$

The ML estimate $\bar{\mathbf{x}}$ is the solution of

$$\bar{\mathbf{x}} = \arg \min L(\tilde{\mathbf{x}}) \quad (3.9)$$

3.4.1 Construction of the Initial Estimation

The SR estimate of each channel is $\tilde{\mathbf{x}}_m$, $m=1,\dots,N$, as shown in Figure 3.2. And they are ones of the uncounted SR solutions. The LR frames captured from natural scene usually are highly correlated, since each LR image corresponds to the same visual content. However, due to different degradation (e.g. different out-of-focus PSFs, different translations), these LR images reflect different parts of the HR image, which have different energy distribution and information recovery by up-sampling. Better SR images can be obtained through constructing better initial estimation.

Figure 3.2 Flowchart of proposed framework

First, the initial SR estimation should have the minimum error with the original

HR image $\mathbf{x}$, but actually $\mathbf{x}$ is unavailable in actual SR reconstruction. The information loss of image down-sampling is highly related to the sampling position. The extremes, which are the pixels with largest gradient, are the most important ones for information recovery. In contrast, the samples on the flat area are useless for HR detail recovery. The up-sampling operator acts as a smoothing filter and does not create HF details. Therefore, the LR image with more HF energy in its up-sampling reflects the one with the less destructive PSF, and is in favor of the SR process.

The pixel of initial SR estimate $\tilde{\mathbf{x}}^0$ is fused from corresponding pixel in $\tilde{\mathbf{x}}_m$. For example, it is selected as

$$\tilde{\mathbf{x}}^0(i, j) = \mathbf{M}(\mathbf{F}_{m,1}\tilde{\mathbf{x}}_m(i, j), m = 1, \dots, N) \quad (3.10)$$

where $\mathbf{M}(\cdot)$ is an operator that selects the pixel with the maximum offset from the centroid of 3×3 adjacent HD pixels of all aligned images $\mathbf{F}_{m,1}\tilde{\mathbf{x}}_m$, $m = 1, \dots, N$. This selection produces much higher frequency information in the initial SR estimate. It may introduce more noise, however, experimental results show that if SNR is not very low the subsequent correction iteration effectively reduce it and produce good result.

3.4.2 Multi-channel Constraints

The switches in Figure 3.2 close at $n = 0$ only. The initialized estimation $\tilde{\mathbf{x}}_1^0, \dots, \tilde{\mathbf{x}}_N^0$ are then inputted to each channel for correction. At the n -th iteration, the correcting procedure in each channel is the same as MISO framework: the shifted SR estimate $\mathbf{F}_{k,m}\tilde{\mathbf{x}}_m^{n-1}$ is degraded and compared with each LR image $\mathbf{y}_k$, $k \neq m$, the difference is fed back to $\tilde{\mathbf{x}}_m^{n-1}$ for its update. The updated images $\tilde{\mathbf{x}}_m^n, m = 1, \dots, N$ are then fused to

form the new SR estimate $\bar{\mathbf{x}}''$. Above process executes iteratively and fast converges to a SR image.

Intuitively, the proposed framework outperforms the MISO framework is due to the two factors: it provides an initial estimate closer to the true HR image, so that the optimization procedure would not lead to a sub-optimal solution. Moreover, the multi-channel constraints are tighter than the SRCs of MISO framework, which would reduce the uncertainty of the SR solution.

3.5 Experiments

Experiments are constructed using synthetic and real-life images to evaluate the proposed proposed framework. The proposed method is also compared with two other SR algorithms, one of which does not take different error levels per frame into account similar to [34] (we call it standard SR method), and the other is the MISO framework similar to [41]. For the synthetic experiments, we use two 256×256 “*F16*” and “*Hill*” images. Four LR images are generated with down-sampling ratio of 2. The registration parameters for the four LR images are global translations $[0,0],[0,1],[1,0],[1,1]$ in the HR grid, respectively. We have also tested “*Man*” image with different subpixel shifts. To test the performance of our algorithm on SR of LR images with different out-of-focus blurring, the PSF are circular averaging filter with different radius $R_1 = 2, R_2 = 2.5, R_3 = 3, R_4 = 3.5$, respectively. The LR images are further degraded by AWGN to produce a signal-to-noise ratio (SNR) at 40 dB, as shown in Figure 3.3, Figure 3.4 and Figure 3.5.

Figure 3.3 Test images from “F16”: (a) LR1 (b) LR2 (c) LR3 (d) LR4

Figure 3.4 Test images from “Hill”: (a) LR1 (b) LR2 (c) LR3 (d) LR4

Figure 3.5 Test images from “Man”: (a) LR1 (b) LR2 (c) LR3 (d) LR4

The first experiment is constructed to verify the influence of initial estimation on SR performance. To get the estimation of registration and PSF parameters, we use the algorithms in [100] and [101] respectively. These algorithms are used in the simulation for both synthetic and real-life test images because of their efficiency and robustness. The accurate and estimated values of the registration and PSF parameters

are listed in Table 1 . As the MISO framework randomly select one LR image as the referenced image, we compare the MISO framework with different referenced LR image for fair comparison. For example, Case 1 corresponds to MISO framework with the first LR image (LR1) as reference image. In the experiment, the algorithms are considered to be converged if the relative change of the SR image goes below 10^{-6} . The Peak signal-to-noise ratio (PSNR) of reconstructed HR images against the number of iterations is plotted in Figure 3.6. These curves indicate the improvement of the SR output at each iterate. It is observed that using MISO with different initialization, the PSNR difference of the reconstructed images is quite large. For the PSNR difference of the reconstructed images between Case 1 and Case 4, it is 3.1 dB and 6.9 dB for “*Hill*” and “*F16*”, respectively. It is noted that this is not due to the error in estimating registration and the PSF for each LR observation, since they are not significantly different, and it is reported in [41] that this MISO is robust to the estimation error. Therefore, the PSNR difference indicates that the initial estimate imposes significant influence on the final SR result. An inappropriate initial estimation, e.g. the up-sampling of LR2 for “*Hill*” image, results in a sub-optimal solution, which does not gain much PSNR improvement during the iterative correction procedure. In the contrast, a better initial estimation, which is provided in proposed framework, not only contains more high-frequency details in the beginning, but also benefits the subsequent correction procedure.

Table 1 Estimation error and PSNR for four cases of test images

Test images	Registration parameters				Est. PSF's Radius	Accurate PSF's Radius	PSNR of SR using the corresponding LRI as ref.
	Vertical		Horizontal				
	Est.	Accurate	Est.	Accurate			
<i>Hill Case1</i>	0	0	0	0	2	2	29.8
<i>HillCase2</i>	-0.24	0	0.50	1	2	2.5	27.8
<i>Hill Case3</i>	0.51	1	0.64	0	3	3	27.7
<i>Hill Case4</i>	0.51	1	0.53	1	3	3.5	26.7
<i>F16 Case1</i>	0	0	0	0	2.5	2	32.6
<i>F16 Case2</i>	0.23	0	0.50	1	3	2.5	28
<i>F16 Case3</i>	0.51	1	0.67	0	3	3	27.7
<i>F16 Case4</i>	0.52	1	0.53	1	3	3.5	25.7
<i>Man Case1</i>	0	0	0	0	2	2	29.9
<i>Man Case2</i>	0.25	0.13	0.41	0.69	2.5	2.5	28.1
<i>Man Case3</i>	0.43	0.54	0.47	0.73	3	3	27.9
<i>Man Case4</i>	0.59	0.23	0.45	0.87	3	3.5	27.8

Figure 3.6 PSNR of the reconstructed HR image vs. number of iterations by MISO for four cases and the proposed method using tested images (a) "Hill" (b) "F16" (c) "Man".

To show the subjective quality improvement produced by the proposed framework, a part of the reconstructed images (100×100 pixels) of "Hill", "F16" and "Man" using different methods are shown in Figure 3.7, Figure 3.8 and Figure 3.9, respectively. The standard SR method does not take into account the destructive PSF (lens's PSF), thus blurring is actually not reduced in the reconstructed image. The reconstructions by MISO are obtained using LR1 (the sharpest observation) as reference image. The proposed method produces the sharpest reconstruction among the three methods.

Figure 3.7 Part of SR reconstruction for "F16" by (a) Standard SR method—24.47dB. (b) MISO—28.01dB. (c) Proposed method—28.51dB. (d) Original image

Figure 3.8 Part of SR reconstruction for "Hill" by (a) Standard SR method — 25.42dB (b) MISO — 27.12dB (c) Proposed method—27.56dB (d) Original image

Figure 3.9 Part of SR reconstruction for "Man" by (a) Standard SR method — 24.9dB (b) MISO — 27.0dB (c) Proposed method—27.4dB (d) Original image

The real-life experiment uses four "book" images captured with deliberately introduced different out-of-focus blur and different translations. The LR images are shown in Figure 3.10 (a). The top left LR image (LR1) suffers the least destructive

blur and its bicubic interpolation is shown in (b). The magnification is chosen as 2 so that the estimated HR image would double the resolution of the LR images. The out-of-focus blurs and translations are estimated using [100, 101], respectively, just as in synthetic experiments. The reconstructed images using the bi-cubic interpolation and three HR methods are provided in (b)-(e), respectively. It is seen the proposed multiple reference-based fusion method recovers the out-of-focus blur effectively. A region of these reconstructions are enlarged in (f)-(i) for a closer comparison. The two letters “p” and “l” are only seen separated in the reconstructed HR image by the proposed framework, which are not clearly observed in other super-resolved results. This comparison verifies the efficacy of the proposed method for real-life image SR.

Figure 3.10 SR for real-life images. (a) Four LR images. (b) Bicubic interpolation of LR1. SR reconstructions by (c) Standard SR method, (d) MISO, (e) proposed method. (f)-(i) Selected enlarged region of (b)-(e), respectively.

3.6 Summary

This chapter presents a new multiple reference-based fusion method. The proposed method is robust to different blurred LR inputs, and improves the SR performance by the construction of a better initial estimation and tighter SRCs. Experimental results show the validity of the proposed framework with different PSF in each LR observation (channel) for both simulated and real-life images. We will continue to explore a further analysis in our future work.

Chapter 4

SR from Multiple Down-sampled Images with Different Arbitrary Zooming Ratios

4.1 Introduction

SR is a process to reconstruct an HR image by fusing multiple LR observations. Typically, these LR observations are modeled as being sub-sampled from the HR image before it is blurred caused by the optical system and shifted with sub-pixel precision. The new information contained in each LR observation can be exploited to estimate the HR image. The key issue here is the accurate registration.

In recent years, various approaches have been developed to fulfill the registration as well as the HR image estimation [60, 102]. Most of these approaches assume the LR images are all captured at the same plane parallel to the camera lens, i.e. at the same sampling lattices. There are only sub-pixel relative displacements between LR images. In practice, however, the plane at which each LR image is captured may have slightly different distance to the camera lens, as shown in Figure 4.1. In other

words, the LR images have different zooming ratios. As some techniques for detecting zoom motion between two images have been reported [68, 69], few works have addressed the SR problem considering zooming due to the difficulty to construct a linear image capturing model to relate the HR image and its LR observations. Chaudhuri and Manjunath addressed the SR problem using zoom cue [103]. In their work, the SR image is modeled as a MRF and they used a MAP estimation method to derive the cost function. The MRF parameters are learnt from the given observations using a maximum pseudo-likelihood (MPL) estimator. This MAP-MRF approach produces good results, but is computationally demanding. For reducing the computation they also proposed an alternative prior, which uses the simultaneous autoregressive (SAR) model to model the SR image. Li [72] attempted to circumvent this problem and argued that all LR observations are linked by a line-geometric model under the assumptions of constant scene depth, small number of LR observations and zooming speed. This model alleviates the data inconsistency among the LR images introduced by zoom motion. The LR images are interpolated and fused to a new image, and then deblurring is applied separately. However, the approach of two disjoint stages will result in a suboptimal solution.

Figure 4.1 The LR images captured by a camera lens at different planes have different zooming ratios

In this chapter, we present a new model for jointly fusing LR images and deblurring. This model can overcome the difficulty of modeling the image capturing process including nonlinear zoom motion, and provide a linear system to relate the HR image and the LR observations. Under our model, the SR problem can be fulfilled by many well-established SR techniques, such as ML, MAP, and POCS. These techniques enable us to tackle image fusion and deblurring simultaneously, thus improve the SR performance. Due to the convenience of MAP based methods to incorporate the prior knowledge of image as regularization, we design a MAP based algorithm to reconstruct the HR image. In the proposed algorithm, the zooming ratios are estimated using optical flow calculation, and then the MAP estimation of HR image is computed.

The rest of this chapter is organized as follows: The problem formulation of image SR with different zooming ratios is introduced in Section 4.2. In Section 4.3, we propose a MAP based framework for registration and estimating HR image. Experimental results on simulated images are presented and discussed in Section 4.4. Finally, a brief conclusion is given in section 4.5.

4.2 Problem Statement

4.2.1 Image Capture Model without Zoom Motion

Let y_k be the k -th ($1 \leq k \leq N$) acquired LR image of size $N_L \times N_L$, and is lexicographically ordered. x denotes the original image. Most of the existing SR approaches model the acquisition of the LR images by rotating, shifting, blurring and

decimating the HR image. If shifting is considered, the parameters $\mathbf{v}_k$ needed to be estimated for registration are horizontal and vertical shifts in HR grids. Then the k th LR image $\mathbf{y}_k$ and the HR image $\mathbf{x}$ is related by the model in (2.1):

$$\mathbf{y}_k = \mathbf{D}_k \mathbf{H}_k \mathbf{F}_k(\mathbf{s}_k) \mathbf{x} + \mathbf{e}_k, k = 1, 2, \dots, N \quad (4.1)$$

where $\mathbf{D}_k$ are the down-sampling matrices, $\mathbf{H}_k$ are the blurring matrices, and $\mathbf{F}_k$ are the warping matrices with translational motion $\mathbf{v}_k$, $\mathbf{e}_k$ is the additive zero mean Gaussian noise. As the LR images are assumed to be captured with the same integer zooming ratio, the decimation process is the same, i.e. $\forall_k, \mathbf{D}_k = \mathbf{D}$. For registration, one of the LR images is selected as the referenced image. Each LR images can be interpolated to the HR grids by uniform pixel interpolation, and aligned to each other by the estimated shift $\mathbf{v}_k$.

However, if the zooming ratio for each LR image is different and not an integer, the image capturing model in (4.1) is not applicable due to the following two reasons. First, the fractional part of the zooming ratio makes it hard to decimate the HR image in integer pixel grids. Second, it is not possible to align $\mathbf{y}_k$ with different zooming ratios to the referenced image by $\mathbf{v}_k$, since the motion vectors are not uniformly distributed. Li [72] attempted to circumvent the difficulty of constructing a linear image capturing model, however, the SR performance is limited by the simple line-geometric model.

4.2.2 Image Capture Model with Zoom Motion

In this chapter, we present a new model to simulate the image capturing process with zoom motion under two assumptions: 1) The ROI has constant scene depth. 2) The PSF of sensors is a box function described as

$$\mathbf{B}(x, y) = \begin{cases} M^{-2} & |x| \leq M/2, |y| \leq M/2 \\ 0 & \text{otherwise} \end{cases} \quad (4.2)$$

As the sensor PSF represents the effect of the spatial integration of light intensity over a square region to simulate the image acquisition of the sensors, it is defined in LR pixel size. Thus it has a parameter of M when it is redefined in HR pixel grids. The imaging system's PSF is the convolution of the sensor PSF and the lens PSF. Here we only consider the sensor PSF and assume the imaging system is a pinhole camera to simplify the analysis, since the lens PSF could vary with M (the focal length is changed), and is too complicated to be represented as a function of M . Moreover, the reconstructed HR image can be deconvolved to reduce the effect of the lens PSF by some deblurring technique [104].

Based on the two assumptions, we formulate a linear system relating the m -th LR pixels from the k -th observation and the HR image

$$\mathbf{y}_k(m) = \sum_{n=1}^{N_H \times N_H} w_{k,m,n}(\mathbf{s}_k, M_k) \mathbf{x}(n) + \mathbf{e}_k(m) \quad (4.3)$$

where $N_H \times N_H$ is the size of $\mathbf{x}$ for $n=1,2,\dots,N_H \times N_H$, M_k is the zooming ratio for $\mathbf{y}_k$ and $k=1,2,\dots,K$. $\mathbf{e}_k$ represents the additive noise, and $\mathbf{s}_k = [\mathbf{s}_{kx}, \mathbf{s}_{ky}]$ is the

coordinate of the center of the zoom operation. The weight $w_{k,m,n}$ represents the portion of the n -th HR pixel comprising the m -th pixel of the k -th LR image.

Each LR image is divided into four parts with the zooming center as the “origin” (i.e. $\mathbf{s}_k = [0,0]$), as shown in Figure 4.2. The dash line denotes HR grid, and the grey square is a pixel of $\mathbf{y}_k$ nearest to the origin. With the origin, we define the coordinate of each LR pixel and HR pixel at the HR pixel grids. For example, the coordinate of the furthest vertex of the LR pixel from the origin is $(-M_k, M_k)$. We denote this LR pixel as $\mathbf{y}_k(-M_k, M_k)$, and the four HR pixels it covers are $\mathbf{x}(-1,1)$, $\mathbf{x}(-2,1)$, $\mathbf{x}(-1,2)$ and $\mathbf{x}(-2,2)$, respectively. The weight $w_{k,m,n}$ of each HR pixel is measured following [104] as the area covered by the LR pixel. For example, the LR pixel (the grey square) in Figure 4.2 covers the whole HR pixel $(-1,1)$, thus the area of $\mathbf{x}(-1,1)$ is 1, and factor M^2 is used for normalization to ensure the $\sum_{n=1}^{N_H \times N_H} w_{k,m,n} = 1$.

Then we have

$$\begin{aligned} \mathbf{y}_k(-M, M) = & \frac{1}{M^2} \mathbf{x}(-1, 1) + \frac{M-1}{M^2} \mathbf{x}(-2, 1) \\ & + \frac{M-1}{M^2} \mathbf{x}(-1, 2) + \frac{(M-1)^2}{M^2} \mathbf{x}(-2, 2) \end{aligned} \quad (4.4)$$

If the zooming center is the center of the LR images, the k -th LR image $\mathbf{y}_k$ comprises of consecutive LR pixels $\mathbf{y}_k(\pm pM_k, \pm pM_k)$, $p = 1, 2, \dots, N_L/2$, and $N_H = \lceil M_k N_L \rceil$, where $\lceil a \rceil$ denotes the nearest integer greater than a . Then we can rewrite (4.3) in matrix notation:

$$\mathbf{y}_k = \mathbf{W}_k(\mathbf{s}_k, M_k) \mathbf{x} + \mathbf{e}_k \quad (4.5)$$

where $\mathbf{W}_k$ is a matrix of size $N_H^2 \times N_L^2$ with element (m, n) as $w_{k,m,n}$.

Figure 4.2 Illustration of the relation between the LR pixels and the HR pixels in the HR pixel grids

4.3 MAP Framework for Registration and

Estimating the HR Image

Based on the image capturing model we have constructed, the SR problem can be readily solved by estimating the registration parameters M_k and $\mathbf{s}_k$ for each $\mathbf{y}_k$, as well as estimating the HR image. The estimation of $\mathbf{s}_k$ can be fulfilled by detecting the feature points using existing techniques. Thus we would not investigate it and simply assume that the zooming center is at the center of each LR image, i.e. $\mathbf{s}_k = \mathbf{s}_0, k = 1, 2, \dots, N$. Then $\mathbf{W}_k(\mathbf{s}_0, M_k)$ in (4.5) is denoted as $\hat{\mathbf{W}}_k(M_k)$ in the following statements.

4.3.1 MAP Framework

We wish to form a MAP estimate for $\mathbf{x}$ and M_k jointly, i.e. to minimize the following cost function with respect to $\mathbf{x}$ and M_k simultaneously

$$L(\mathbf{x}, M_k) = \frac{1}{2} \sum_{k=1}^K \alpha_k (\mathbf{y}_k - \dot{\mathbf{W}}_k(M_k) \mathbf{x})^T (\mathbf{y}_k - \dot{\mathbf{W}}_k(M_k) \mathbf{x}) + \frac{\lambda}{2} \mathbf{x}^T \mathbf{T} \mathbf{x} \quad (4.6)$$

where α denotes the regularization parameter, and $\mathbf{x}^T \mathbf{T} \mathbf{x}$ is the regularization term using TV technique. α_k is the channel weight we propose to tackle the deblurring problem for different PSF of $\mathbf{y}_k$.

However, unlike the translational motion parameters can be simultaneously or alternatively corrected by the current estimation of the HR image [60, 102], it is a challenge to update the estimation of M_k along with $\mathbf{x}$, due to the fact that (4.6) is not readily differentiable with respect to M_k . In addition, it is difficult to represent the Jacobian of $\dot{\mathbf{W}}_k(M_k) \mathbf{x}$ with respect to M_k , since the minimization problem of (4.6) is nonlinear with respect to M_k and the linear approximation is not applicable. Therefore, we have to estimate M_k first, and then minimize (4.6) to obtain the final estimation of $\mathbf{x}$.

4.3.2 Estimation of Zooming Ratios M_k

As M_k is not readily to be directly detected by existing registration methods, we propose an algorithm to estimate M_k iteratively. Assuming the first image $\mathbf{y}_1$ has the highest resolution among the LR images, it is bilinear interpolated to be the referenced

image. M_1 is used as the initial guess of M_k , i.e. $M_k^0 = M_1$. At the t -th iteration, optical flow (OF) calculation is used to compute pixel displacements between y_k and the down-sampled referenced image by M_k^t . The pixel displacements will have a “blow in” pattern if $M_k^t > M_k$, or a “blow out” pattern if $M_k^t < M_k$. Then M_k^{t+1} is updated as

$$M_k^{t+1} = M_k^t - \eta \sum_{m=1}^{N_L \times N_L} u^t(m) \cdot \vec{r} \quad (4.7)$$

where $u^t(m)$ is the pixel displacement, which is equivalent to motion vector here, η is a scalar for tuning the step size. $\vec{r}$ is a vector denoting the direction of $u^t(m)$, i.e., its value is set to $+1$ if $u^t(m)$ points inward to the center, or -1 otherwise. Therefore, the

sum $\sum_{m=1}^{N_L \times N_L} u^t(m) \cdot \vec{r}$ converges to zero when M_k^t approaches the actual M_k . The

iteration terminates if $\sum_{m=1}^{N_L \times N_L} u^t(m) \cdot \vec{r}$ falls below a threshold or the number of iterations achieves a pre-defined maximum.

4.3.3 Estimation of HR Image $\mathbf{x}$

Given M_k estimated as $\hat{M}_k$, the HR image can be obtained by minimizing

$$L(\mathbf{x}) = \frac{1}{2} \sum_{k=1}^K v_k(\mathbf{x})(\mathbf{y}_k - \mathbf{W}_k(\hat{M}_k)\mathbf{x})^T (\mathbf{y}_k - \mathbf{W}_k(\hat{M}_k)\mathbf{x}) + \frac{\lambda}{2} \mathbf{x}^T \mathbf{T} \mathbf{x} \quad (4.8)$$

The channel weight v_k is determined by the difference $\mathbf{y}_k - \dot{\mathbf{W}}_k(M_k)\mathbf{x}$ as well as the PSF. According to (4.2), the LR image with a larger M_k should have a smaller weight because its PSF introduce more blurring. Then we have

$$v_k(\mathbf{x}) = \frac{1}{M_k \|\mathbf{y}_k - \dot{\mathbf{W}}_k(M_k)\mathbf{x}\|_2} \quad (4.9)$$

Minimizing (4.8) is equivalent to solve the equality $\nabla_{\mathbf{x}} L(\mathbf{x}) = 0$. The gradient of the function $L(\mathbf{x})$ is given by the partial derivatives of it with respect to each pixel [39]:

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}) &= \sum_{k=1}^K \{ \nabla_{\mathbf{x}} v_k(\mathbf{x}) (\mathbf{y}_k - \dot{\mathbf{W}}_k(\hat{M}_k)\mathbf{x})^T (\mathbf{y}_k - \dot{\mathbf{W}}_k(\hat{M}_k)\mathbf{x}) \\ &\quad + v_k(\mathbf{x}) \dot{\mathbf{W}}_k^T(\hat{M}_k) (\dot{\mathbf{W}}_k(\hat{M}_k)\mathbf{x} - \mathbf{y}_k) \} + \lambda \mathbf{T}\mathbf{x} \\ &\approx \sum_{k=1}^K v_k(\mathbf{x}) \dot{\mathbf{W}}_k^T(\hat{M}_k) (\dot{\mathbf{W}}_k(\hat{M}_k)\mathbf{x} - \mathbf{y}_k) + \lambda \mathbf{T}\mathbf{x} \end{aligned} \quad (4.10)$$

The approximation $\nabla_{\mathbf{x}} v_k(\mathbf{x}) \approx 0$ is also used in (17) of [39], where the channel weight is also a function of $\mathbf{x}$. It is reasonable because the channel weight $v_k(\mathbf{x})$ is determined by the difference $\mathbf{y}_k - \dot{\mathbf{W}}_k(M_k)\mathbf{x}$ as well as the PSF to ensure that the LR image with a larger M_k should have a smaller weight because its PSF introduce more blurring. For this purpose, it is computed using (4.9) with the current estimates of M_k and $\dot{\mathbf{W}}_k$.

The solution of (4.10) is given by

$$\mathbf{x} = \left(\sum_{k=1}^K \alpha_k(\mathbf{x}) \dot{\mathbf{W}}_k^T(\hat{M}_k) \dot{\mathbf{W}}_k(\hat{M}_k) + \lambda \mathbf{T} \right)^{-1} \left(\sum_{k=1}^K \alpha_k(\mathbf{x}) \dot{\mathbf{W}}_k^T(\hat{M}_k) \mathbf{y}_k \right) \quad (4.11)$$

The inversion of the matrix $\sum_{k=1}^K \alpha_k(\mathbf{x}) \dot{\mathbf{W}}_k^T(\hat{M}_k) \dot{\mathbf{W}}_k(\hat{M}_k) + \lambda \mathbf{T}$ is computationally intensive. Thus we employ the conjugate gradient (CG) optimization to compute the solution due to its fast convergence.

4.4 Experiments

In this simulation, the 384×384 “*building*” image is used as tested image. 16 LR images are generated for simulation, each of which has the size 40×40 . M_1 is set as 2, and $M_k, k = 2, 3, \dots, 16$ are randomly selected between 2.0 and 2.5, and the values are ascending sorted, i.e. $M_1 (= 2) < M_k < 2.5$. To simulate the image capturing process, we fix the zooming center as (274, 270), and cut a part of the tested image with this center and of the size $\lceil 40M_k \times 40M_k \rceil$ as the k -th ideal HR image. Then each ideal HR image is down-sampled using (4.5) for generating $\mathbf{y}_k$. For evaluating the SR performance, we let the first ideal image as $\mathbf{x}$.

$M_k, k = 2, 3, \dots, 16$ are estimated using (4.7) by setting $\eta = 10^{-6}$. We considered the algorithm had converged if $\sum_{m=1}^{N_L \times N_L} u(m) < 150$ and set the maximum iteration number to 200. It is observed that M'_k in (4.7) converge around 150~180 iterations. To show the convergence of the algorithm of estimating M_k , the difference between the estimated M_4, M_8, M_{12}, M_{16} and accurate one at each iteration are plotted in Figure 4.3. The estimation error of $M_k, k = 2, \dots, 16$ is listed in Table 2. We can see that the

estimation error of each zooming ratio is close to each other, and the average value is 0.078.

Figure 4.3 Convergence of estimating M_4, M_8, M_{12}, M_{16} using (4.6)

Table 2 The estimation error of $M_k, k = 2, \dots, 16$

$M_2 \sim M_6$	0.076	0.078	0.078	0.077	0.078
$M_7 \sim M_{11}$	0.079	0.074	0.078	0.075	0.075
$M_{12} \sim M_{16}$	0.076	0.078	0.078	0.082	0.082

Using the estimated zooming ratios $\hat{M}_k$, the reconstructed image is obtained from (4.11) by CG method. We also implement the SR algorithm proposed in [72], which uses a simple line-geometric model to simulate the relation between the LR observations.

The reconstructed HR images by three methods are shown in Figure 4.4. As the two different algorithms are based on different observation models, for a fair comparison, we select six LR images for simulating the SR result in [72] to follow its two assumptions: constant zooming speed and small number of input LR images. Using the same LR images, the reconstructed HR image by the proposed algorithm is also given. It can be seen that (b) presents a better resolution than (d) at the left window frame, as well as the horizontal texture in the middle jalousies which is not

recognizable in (d). The proposed algorithm (c) produces the best image quality, where the left window and the texture are clearer. Some noise in the border of (c) is observed, which is due to the ill-posedness of SR with insufficient LR observations. This noise can be eliminated when the number of input LR images is further increased as in (f). On the contrary, no improvement is observed in (e), as the assumption of small number of inputs in [72] is no longer held.

Figure 4.4 (a) Original image, the reconstructed image by (b) Li's algorithm using 6 LR images (c) the proposed algorithm using 6 LR images, (d) bicubic interpolation, (e) Li's algorithm using 16 LR images, (f) the proposed algorithm using 16 LR images.

4.5 Summary

In this chapter, a linear system is presented to relate the LR observations captured with different zooming ratios and the HR image. Based on this system model, we

developed a technique to estimate the zooming ratios and construct a MAP based SR algorithm. Under the proposed framework, the reconstructed HR image presents a better image quality compared to some existing algorithm. Experiments verify our image capturing model and show the efficiency of the proposed SR approach.

As the proposed framework is appropriate and is validated by simulation, the algorithm used for estimating zooming ratios is simple and its performance strongly depends on the optical flow calculation for computing the pixel displacement $u(m)$ in (4.6). If noise is present, optical flow calculation would not be accurate, and the algorithm for estimating zooming ratios does not work well. Besides, the zooming center $\mathbf{s}_k$ in (4.2) needs to be estimated using some technique such as detection of feature points. In the future, we will address these issues to prepare the work in Chapter 4 for real implementation.

Chapter 5

Applications of SR Techniques on Video Coding and Acquisition

5.1 Introduction

For some imaging systems, the limitation of resolution is not determined by the density of sensors, but other factors that restrict the acquisition of all frames at full resolution (FR). For example, the battery power is a bottle neck in mobile devices. The variable acuity sensor imaging (VASI) system is capable to turn off some of the sensors during acquiring each frame. A good idea with such a system to save power is to sacrifice the FR of some video frames during acquisition, and restore them afterwards using SR techniques. Another restriction for capturing FR video is the acquisition at high-frame-rate. For some consumer level digital cameras with the movie function, it is possible to achieve high-frame-rate to reveal fast motions at the cost of a reduction of image resolution. In this case, the proposed scheme could be used to overcome the difficulty of obtaining high-frame-rate and HR videos.

On the other hand, SR can not only be used to ameliorate the undesirable resolution reduction by imaging systems, but also be applied as a post-processing in

video compression, combined with intentional down-sampling before compression to improve the overall quality of the reconstructed video. At a certain low bit rate range, such a down-sampling coding (DBC) produces better image quality compared with a regular coding, where no down-sampling and post-processing are applied.

Figure 5.1 The block diagram of the proposed video acquisition system

In the applications mentioned above, SR serves as a part of the video acquisition system, as shown in Figure 5.1. In this acquisition system, since it is possible to selectively down-sample some video frames at LR during capturing (for low power acquisition or high-frame-rate video acquisition), or before compression (for better reconstruction quality) while remaining the others at HR, the extra information of the HR frames can be exploited to restore the acquired LR frames during SR process. The three blocks (capture video, compress video and post-processing) can be regarded as a generalized acquisition system, then a trade-off between image quality and the efficiency of this system is achieved by tuning the ratios of HR frames or down-sampling factor (equivalent to MF) of the LR frames [105]. For specific applications, our work on the integrated video acquisition/coding

scheme includes: 1) DBC and 2) low-power video acquisition and high-frame-rate and HR video acquisition.

Apparently, the over-all performance of the integrated video acquisition/coding scheme strongly depends on the efficacy of the SR techniques employed. While down-sampling helps to improve the efficiency of the integrated system, it will eliminate HF details, which are hardly reproduced by a simple pixel interpolation. Therefore, we propose an example-based SR algorithm to restore the HF components of the down-sampled frames.

RBAs make use of the correlation of multiple aliased LR observations to produce a HR image, among which the MAP estimator of the HR image is commonly used [51, 53, 55, 57]. Some works have been proposed for estimating HR videos from their compressed LR observations utilizing Bayesian framework to incorporate the quantization constrain set [59, 106, 107]. Using the compressed bitstream as the input, Segall et al. [106] developed an algorithm to exploit the information from the compressed bitstream, knowledge of the encoder structure and models for the HR intensities and displacement values. Assuming the quantization noise is dominant, the compression error is modeled following Gaussian distribution, and the covariance matrix can be derived from the compressed bit-stream. Their work is improved concerning that the “useful” pixels in the motion compensated LR observations should be detected for SR [59, 107]. However, the SR gains decrease when quantization error increases since the redundancy between LR observations is reduced. Inspired by [61], we propose a new example-based algorithm to improve the SR performance. The basic idea of example-based SR algorithms is to construct a training set consisting of a large

number of patch pairs from HR images and their down-sampled versions. The resolution of the tested LR image can then be enhanced by adding HF details, which are obtained by a nearest neighbor searching in the training set. This is called a training process in this chapter.

Brandi et al. [108] used the example-based SR algorithm in [16] to restore the down-sampled frames. However, as pointed in [16], there are two major problems in applying the existing example-based SR algorithms to video enhancement. First, the compression noise in the LR image is recognized as image content and will be amplified in the corresponding HR reconstruction. Second, the created HF details would lead to unpleasant subjective view in moving images. In order to solve these problems, we propose a new example-based SR algorithm to exploit the LR mapping between the samples and their HR counterparts (referred as “*inter-resolution correlation*”) as well as the correlation among neighboring LR frames (referred as “*inter-frame correlation*”):

- Imposing constraints of nearest neighbor searching region.
- Rectifying the “unreal” pixels using *inter-resolution correlation* and *inter-frame correlations*.

The proposed example-based SR algorithm comprises of two parts: 1) The training set is constructed similar to [15]. The HF component of each patch in the input LR frame is found from the training set with some constraints considering the influence of compression noise. 2) The output of the training process is iteratively corrected with the LR counterpart and accurately registered pixels of the training image. The final SR output converges to an HR image without apparent made up

artifacts.

The rest of this chapter is organized as follows. The DBC scheme with SR technique is elaborated in Section 5.2. Section 5.3 introduces the low-power video acquisition scheme with SR technique. The proposed example-based SR algorithm is developed in Section 5.4. Experiments are given in Section 5.5. Section 5.6 gives the summary.

5.2 Down-sampling Based Video Coding

In this section, some related work of DBC is provided as a background. Then a new DBC is proposed, where a SR technique is employed to restore the down-sampled frames to their original frame size. The performance improvement of the proposed DBC is analyzed at low bit rates.

5.2.1 Background

In many applications, such as video surveillance and mobile multimedia, there is a need to compress video signals with a low bit rate due to the limited bandwidth available. Due to its energy compactness, DCT is widely used in video compression standards to convert a frame into a statistically uncorrelated data set. The error of lossy compression comes from the quantization of the DCT coefficients. Coding frames with a low bit rate (corresponds to a coarse quantizer) would result in the compression noise shown up in the image, including block effect, the loss of sharpness and ringing

artifacts.

If frames in a video are sparsely sampled before compression, more bits can be reserved to each DCT coefficient at the same bit rate, as a result, less quantization error is introduced and consequently compression loss is alleviated. It was shown on still images [109] that at low bit rate an image down-sampled before Joint Photographic Experts Group (JPEG) compression visually beats the original image compressed via JPEG with the same number of bits. H. A. Ilgin et al. [110] and Viet-Anh Nguyen et al. [111] applied the down-sampling strategy for video coding on DCT domain. This DBC truncates the higher frequency DCT coefficients before quantization, thus more bits are allocated in DC and lower frequency AC coefficients. This low frequency (LF) preferred bits allocating strategy leads to a better performance over H.263 in the intra-frame coding. However, the improvement decreases when coding inter-frames, since less information is contained in the prediction residual. Therefore, when compared to the H.264/AVC codec, which possesses more powerful motion compensation (MC) techniques such as variable block-size MC and quarter-sample-accurate MC, the down-sampling strategy only performs better if intra-frames are coded with a low bit rate. D. Barreto et al. [112] proposed a DBC scheme by down-sampling of the HR video frames before compression and then utilizing SR for generating HR ones at the decoder. Side information defining the down-sampling process is needed to be transmitted to the decoder to help the SR process. This scheme uses a pre-process (including segmentation and down-sampling) and post-processing (SR procedure) to improve the quality of compressed video at low bitrates.

5.2.2 Proposed DBC

We propose a different framework of DBC with SR technique, in which no side information is required to be extracted and transmitted. The encoder of the proposed DBC is shown in Figure 5.2. The intra-frames are coded at their original frame size, and their reconstructions are down-sampled and stored in the frame memories. For the inter-frames, they are down-sampled before compression. Thus all the macro blocks (MB) of the inter-frames are down-sampled, and the MC is applied at LR. At the decoder, the proposed SR is applied to restore the LR inter-frames to their original frame size with the help of the intra-frames.

As the number of DCT coefficients is reduced in such a DBC, the bit budget of each coefficient will increase, thus reduce the quantization error. Therefore, this DBC produces better quality of reconstructed frames compared with regular coding.

Figure 5.2 Block diagram of the proposed DBC encoder

5.2.3 Rate Distortion Analysis

To justify the performance of the proposed DBC, we present a rate distortion analysis from a mathematical view, which relates the MSE of the coded frames with the down-sampling factor and bit budget.

It has been shown in [113] why the compression efficiency is improved by down-sampling, JPEG compression and interpolation at low bit rate. As ITU video compression standards use block based DCT, which is the same as JPEG compression, we extend the analytical model in [113] to describe the procedure of the proposed DBC and explain its advantage at low bit rate.

In [113], assumptions are made for the analytical model: the given images are on the unit square $[0,1] \times [0,1]$, and an image $\{\mathbf{x}_m(x,y)\}$ is a 2-D random process, which follows $E[\mathbf{x}_m(x,y)] = 0$, and its autocorrelation is $\Re_f(\tau_x, \tau_y) = E[\mathbf{x}_m(x,y)\mathbf{x}_m(x+\tau_x, y+\tau_y)] = r_0^2 e^{-\alpha_x|\tau_x| - \alpha_y|\tau_y|}$, where τ_x, τ_y is the Euclidean distance between (x,y) and $(x+\tau_x, y+\tau_y)$ in x, y direction, respectively. $r_0^2, \alpha_x, \alpha_y$ are constants determined by image statistics. In the block based DCT, $\mathbf{x}_m(x,y)$ is sliced into $U \times U$ pieces (assume each block is square) and over each one a subset Ω is used for the basis functions (i.e., $\Omega \subset \{(p,q) | p,q = 0,1,2,\dots\}$). The DCT coefficients of the piece (p,q) are represented as $\mathbf{x}_{pq}$, and $\mathbf{x}_{pq}^Q$ denotes the quantized DCT coefficients with b_{pq} bits.

As each 8×8 block of intra-frame is intra coded, we treat the compression process in the same way as JPEG. Therefore its MSE due to quantization is computed as in [113]:

$$E[(x_{pq} - x_{pq}^Q)^2] = r_0^2 \left\{ 1 - \sum_{p,q \in \Omega} (2 - \delta_p) B\left(\frac{\alpha_x}{U}; p, p\right) \times (2 - \delta_q) B\left(\frac{\alpha_y}{U}; q, q\right) \left[1 - \frac{Z}{2^{2b_{pq}}}\right] \right\} \quad p, q = 0, \dots, 7 \quad (5.1)$$

where B is a function of parameters A, p, q :

$$B(A; p, q) = (1 + \delta_{p-or-q}) \delta_{|p-q|} \frac{A}{A^2 + p^2 \pi^2} - [1 + (-1)^{p+q}] \frac{A^2}{(A^2 + (p\pi)^2)(A^2 + (q\pi)^2)} \frac{(2 - e^{-A}[(-1)^p + (-1)^q])}{2} \quad (5.2)$$

Z is a constant in the range $[1, 3]$, its value is chosen according to different bits allocation policies in [113] and its value would not affect the following analysis. Here we set $Z = 3$. δ_w is the impulse function

$$\delta_w = \begin{cases} 1, & \text{if } w \equiv 0 \\ 0, & \text{otherwise} \end{cases}. \text{ Note that } U \text{ is related to the down-sampling factor (i.e. MF}$$

in the previous sections) M as $M = \frac{U \times \text{block size}}{\text{image size}}$.

By (5.1), the MSE of each block in intra-frame due to quantization can be predicted with given M and b_{pq} , and the image statistics α_x, α_y can be estimated following [113].

Now we proceed to analyze the MSE of the blocks in inter-frames (P and B frames) due to quantization. MBs of P frames can be intra coded, motion-

compensated predictive coded or skipped, and the MBs of B frames are predicted from the ones of forward, backward frames. Thus these MBs are classified to two catalogues: intra coded macro blocks and motion-compensated predictive coded ones. In the first type, the MSE are calculated by (5.1), just as the ones in intra-frame. We will focus on the second type of MBs.

We denote prediction residual of motion-compensated predicted pixel as $\mathbf{n}_{l,k}(x, y) = \mathbf{x}_l(x, y) - \mathbf{x}_k(x + d_x, y + d_y)$, where $\mathbf{x}_l$ is the current coded frame and $\mathbf{x}_k$ is the referenced frame, where the motion estimation is used to search the prediction macro blocks, and d_x, d_y are the motion components at the direction x, y , respectively.

To simplify our analysis, we assume there is only local translation in the video and thus the prediction residual is due to registration error. Assuming the correct motion vector is (d'_x, d'_y) , we can obtain $\mathbf{x}_l(x, y) = \mathbf{x}_k(x + d'_x, y + d'_y)$. Then the auto-correlation of $\mathbf{n}_{l,k}(x, y)$ is

$$\begin{aligned}
 & \Re_{\mathbf{n}_{l,k}}\{x, y; x + \tau_x, y + \tau_y\} \\
 &= E\{[\mathbf{x}_l(x, y) - \mathbf{x}_k(x + d_x, y + d_y)][\mathbf{x}_l(x + \tau_x, y + \tau_y) - \mathbf{x}_k(x + \tau_x + d_x, y + \tau_y + d_y)]\} \\
 &= E\{[\mathbf{x}_k(x', y') - \mathbf{x}_k(x' + v_x, y' + v_y)][\mathbf{x}_k(x' + \tau_x, y' + \tau_y) - \mathbf{x}_k(x' + \tau_x + v_x, y' + \tau_y + v_y)]\} \\
 &= \Re_f(x', y'; x' + \tau_x, y' + \tau_y) - \Re_f(x' + v_x, y' + v_y; x' + \tau_x, y' + \tau_y) \\
 &\quad - \Re_f(x', y'; x' + \tau_x + v_x, y' + \tau_y + v_y) + \Re_f(x' + v_x, y' + v_y; x' + \tau_x + v_x, y' + \tau_y + v_y) \\
 &= 2r_0^2 e^{-\alpha_x|\tau_x| - \alpha_y|\tau_y|} - r_0^2 e^{-\alpha_x|\tau_x - v_x| - \alpha_y|\tau_y - v_y|} - r_0^2 e^{-\alpha_x|\tau_x + v_x| - \alpha_y|\tau_y + v_y|}
 \end{aligned} \tag{5.3}$$

where $(x', y') = (x + d'_x, y + d'_y)$, and $(v_x, v_y) = (d_x, d_y) - (d'_x, d'_y)$ is the registration error.

Then we have $E[\mathbf{n}_{l,k}^2(x, y)] = 2r_0^2(1 - e^{-\alpha_x|v_x| - \alpha_y|v_y|})$.

This approximation of the second order statistics of $\mathbf{n}_{l,k}(x, y)$ is valid since the variance of prediction residual is smaller than the original image. For example, we set the registration error is 0.08 pixel as in [104], and the size of image is 512×512 , thus $v_x = v_y = \frac{0.08}{512}$ (the scaling is due to the assumption that images are on the unit square

$[0,1] \times [0,1]$). We also set $\alpha_x = \alpha_y = 150$ as [113], then we get $2r_0^2(1 - e^{-\alpha_x|v_x| - \alpha_y|v_y|}) = 0.1r_0^2$.

Similar to (5.1), we can get the relation between MSE of motion-compensated predicted 8×8 blocks, down-sampling factor and bits:

$$E[(n_{l,k}^{pq} - \hat{n}_{l,k}^{pq})^2] = r_0^2 (2 - e^{-\alpha_x|v_x| - \alpha_y|v_y|} - e^{\alpha_x|v_x| + \alpha_y|v_y|}) \{1 - \sum_{p,q \in \Omega} (2 - \delta_p) B(\frac{\alpha_x}{M}; p, p) (2 - \delta_q) B(\frac{\alpha_y}{M}; q, q) [1 - \frac{Z}{2^{2b'_{pq}}}] \} \\ p, q = 0, \dots, 7 \quad (5.4)$$

where $n_{l,k}^{pq}$ are the DCT coefficients of piece (p, q) of $\mathbf{n}_{l,k}$, $\hat{n}_{l,k}^{pq}$ are the ones after quantization, b'_{pq} is the number of bits allocated in the coefficient $\hat{n}_{l,k}^{pq}$.

Using (5.1) and (5.4), we can now calculate the PSNR of coded video frames: $M = 2$ in (5.4) and $M = 1$ in (5.1) for the proposed DBC; $M = 1$ in (5.1) and (5.4) for the regular coding (with no down-sampling applied). The other parameters are set in [113]: the image size is 352×288 , $\alpha_x = \alpha_y = 150$ and $Z = 3$; the number of bits used is: $0.3B_{total}$ for coding intra-blocks and $0.7B_{total}$ for inter-blocks (i.e.

$$\sum_{p,q \in \Omega} b_{pq} = 0.3B_{total}, \sum_{p,q \in \Omega} b'_{pq} = 0.7B_{total} \text{), and the bit allocation strategy}$$

follows [113]. Additionally, we set the registration error to 0.08 pixels as in [104]. The curves of PSNRs vs. bit rates from the proposed DBC and the regular coding are presented in Figure 5.3, which are denoted by a dot curve and a solid curve, respectively. At the low bit rate range, the dot curve is higher than the solid one. It is because for smaller image size, more bits can be allocated at each DCT coefficient, thus favorable to restore the HR image. The PSNR saturates after a certain bit rate, because limited information is available from the down-sampled frames, and the information loss due to down-sampling cannot be recovered even no quantization error is present. As a result, the PSNR curve saturate because higher Kbps cannot recover the information loss induced by down-sampling. A crossing point is clearly observed in the figure.

Figure 5.3 Analytical rate distortion curves of the regular coding and the proposed DBC

This rate distortion analysis predicts that higher PSNR of the decoded frames from the proposed DBC than the regular coding could be obtained to restore the HR in a certain low bit rate range. However, we cannot quantitatively estimate the optimal

down-sampling for a given video. This is because the statistics of compression error in video coding is much more complicated than image coding.

5.3 Low-Power Video Acquisition

Based on the basic idea of integrated video acquisition scheme described in Figure 5.1, we propose a video acquisition scheme by selectively acquiring some video frames at FR, while the others at LR. This scheme could not only be implemented in the VASI system for saving power, but also be applied in conventional sensors for producing high-frame-rate and HR videos. Before elaborating our scheme, a brief review of the background for these two applications is given as follows.

5.3.1 Background

Video acquisition has become a common feature in most mobile devices, such as mobile phones and digital cameras, where the battery power is still a bottleneck. How to reduce the power consumption during video acquisition with little effect on the quality of the acquired video attracts much attention in industrial and scientific communities. With the VASI system [114], which is capable to turn on/off each sensor when capturing each frame, a low-power content-based video acquisition system is proposed in [115]. It detects the region of high spatial or temporal activity, within which all the sensors are turned on. The sensors in the other region are turned off to reduce the power. This acquisition system requires a pre-analyzing to detect the spatially and temporally active regions in current acquired frame. An additional buffer

is needed to keep the first two FR frames.

Nowadays, some consumer-level digital cameras with the movie function can achieve high-frame-rates to reveal fast motions. However, the resolution is usually sacrificed as a cost of such high-frame-rates. For example, current cameras work either at a high-frame-rate (1000fps) with a low-resolution (224×56), or at a high-resolution (1280×720) with a low-frame-rate (30fps). The incompatibility of high-resolution and high-frame-rate is due to the limited pixel transfer rate. The conventional sensor imagers at dilution modes throw away pixel data in exchange for higher-frame-rates to maintain total field-of-view (FOV). For producing very high-frame-rate movie functions, the dual-sensor camera proposed in [116] consists of a HR sensor with a low-frame-rate and a LR sensor with a high-frame-rate. Two sequences with different spatial-temporal frequencies are recorded from the same FOV, and thus doubling the memory requirements.

5.3.2 Proposed Video Acquisition Scheme

Motivated by the low-power acquisition with VASI system in [115], we propose a uniform acquisition supported by SR technique, where no pre-analysis is needed (Figure 5.4(a)). The acquired video consists of groups of equal number frames. Within each group of pictures (GOP), one frame is captured at FR as the key frame while others are at LR with only a quarter of the sensors turned on. A SR technique is proposed to enhance the whole acquired video to FR and restore the missing HF details. As SR can be applied in the post-processing, our framework consumes much less power in mobile devices while maintaining a reasonable good video quality.

As the uniform acquisition does not need the content-based analysis to determine which sensor should be turned on/off, this framework can be implemented on conventional sensors for producing high-frame-rate and HR videos. With the help of our scheme, the high-frame-rate can be maintained by reducing data for a majority of frames (LR frames) while remaining FR for key frames (Figure 5.4 (b)). The whole video sequence is enhanced to HR using SR technique during post-processing.

Figure 5.4 Block diagram of the proposed uniform acquisition and SR on (a) VASI for saving power, (b) conventional sensor imager for producing high-frame-rate and HR frames

Moreover, the proposed uniform acquisition indicates power saving not only for sensors, but also in the subsequent video processing such as compression. In prevalent high-dimension video and digital cameras, each video frame or interlaced field of a digital video sequence is separately compressed by intra-frame compression techniques such as Motion JPEG (M-JPEG) and intra-frame MPEG2. Therefore, the video frames are treated as a sequence of still images during compression. Naturally,

processing on smaller frames would consume less power than on FR ones because less computation is needed for DCT and motion estimation. It was reported in [109] that compression on down-sampled images produces better image quality than directly compression at low bit rates. So the uniform acquisition actually benefits the video compression not only in power saving but also quality. Since many state-of-art video compression standards such as H.264 has spatially scalable extension (scalable video coding), the uniform acquired video can be readily compressed using such techniques. An example of using scalable video coding (SVC) for coding the uniform acquired video is similar to the one in our earlier work [117], as presented in Figure 5.5. The key frames are coded at the enhancement layer while the LR ones are at the base layer.

Figure 5.5 An example of compressing video from uniform acquisition using SVC

5.4 Proposed SR Technique

To restore the video acquired by the generalized acquisition system in Figure 5.1, which comprises of down-sampled frames interlaced with some HR frames, we

propose a new example-based SR algorithm in this section. The proposed SR makes use of both *inter-resolution correlation* and *inter-frame correlation* to restore the down-sampled frames to HR, and solves the two problems in applying existing example-based SR algorithms in video enhancement: 1) the compression noise in the LR image is recognized as data and amplified in the reconstructed HR frame; 2) the created HF details would scintillate in moving images.

5.4.1 System Model

We use the well known image degradation model to simulate the relation between the acquired FR frames $\mathbf{x}_n$ and their down-sampled version $\mathbf{y}_n$:

$$\mathbf{y}_n = \mathbf{W}\mathbf{x}_n + \mathbf{e}_n, n = 1, \dots, P \quad (5.5)$$

where $\mathbf{W}$ is a degradation matrix incorporating blurring and down-sampling operators, and $\mathbf{e}_n$ is the acquisition error, P is the GOP size .

Pictures in each GOP typically contain similar image content. Using the first frame of the GOP as a intra-frame and the rest as inter-frames, the relation between HR frames are described as

$$\mathbf{x}_n = \mathbf{C}(\mathbf{d}_{n,1})\mathbf{x}_1 + \mathbf{e}_{n,1}, n = 2, \dots, P \quad (5.6)$$

where $\mathbf{C}(\mathbf{d}_{n,1})$ is the two dimensional matrix describing the displacement, $\mathbf{d}_{n,1}$ is the displacement that relates the pixel at the n -th inter-frame to the pixel at the intra-frame, and $\mathbf{e}_{n,1}$ is the registration error.

The *inter-resolution correlation* and the *inter-frame correlation* of the acquired frames by the proposed acquisition scheme in Figure 5.4 can be described by

(5.5) and (5.6), respectively. For the video frames compressed using the proposed DBC in Figure 5.2, compression noise should be taken into account. We analyze the *inter-resolution correlation* and the *inter-frame correlation* of the compressed video as follows.

According to the coding process of Figure 5.2, the intra-frame $\mathbf{x}_1$ is divided into blocks, DCT transformed, quantized and entropy coded into a bit stream. The reconstructed frame is down-sampled as $\mathbf{y}_1^Q$ serving as a reference frame. For each block of the down-sampled inter-frames $\mathbf{y}_n (n = 2, \dots, P)$, its most similar block is searched from the reference frame $\mathbf{y}_1^Q$ and then subtracted from the current block. The prediction residual is coded.

After decoding, the observed intra-frame $\mathbf{y}_1^Q$ is expressed by

$$\mathbf{x}_1^Q = \mathbf{T}^{-1} \mathbf{Q}(\mathbf{T}(\mathbf{x}_1)) \quad (5.7)$$

and LR inter-frame $\mathbf{y}_n^Q$ is

$$\mathbf{y}_n^Q = \mathbf{T}^{-1} \mathbf{Q}[\mathbf{T}(\mathbf{y}_n - \mathbf{C}(\mathbf{v}_{1,n}) \mathbf{y}_1^Q)] + \mathbf{C}(\mathbf{v}_{1,n}) \mathbf{y}_1^Q, n = 2, \dots, P \quad (5.8)$$

where $\mathbf{y}_1^Q = \mathbf{W} \mathbf{x}_1^Q$, $\mathbf{T}$ denotes DCT transformation, and $\mathbf{Q}$ is the quantization operator, the motion vector $\mathbf{v}_{1,n}$ is the spatial offset between the prediction block and the current block.

Considering the quantization noise, we rewrite (5.5) and (5.6) as:

$$\mathbf{y}_n^Q = \mathbf{W} \mathbf{x}_n + \mathbf{e}'_n, n = 1, \dots, P \quad (5.9)$$

$$\text{and} \quad \mathbf{x}_n = \mathbf{C}(\mathbf{d}_{n,1}) \mathbf{x}_1^Q + \mathbf{e}'_{n,1}, n = 2, \dots, P \quad (5.10)$$

with $\mathbf{e}'_n = \mathbf{e}_n + \mathbf{e}_n^Q$, $\mathbf{e}'_{n,l} = \mathbf{e}_{n,l} + \mathbf{e}_{n,l}^Q$, and $\mathbf{e}_n^Q, \mathbf{e}_{n,l}^Q$ are the equivalent quantization noise in spatial domain. It should be noted that while $\mathbf{e}_{n,l}^Q$ is due to quantization of the intra-frame, whose representation in frequency domain within a block is $\mathbf{x}_{pq} - \mathbf{x}_{pq}^Q$ of (5.1), $\mathbf{e}_n^Q$ comes from quantization of both prediction residual in (5.8) and intra-frame in (5.7) and its statistics is much more complex.

In what follows, the input HR and LR images by the DBC scheme or low-power acquisition are denoted as $\mathbf{x}_l^Q$ and $\mathbf{y}_n^Q$, respectively. For the case of uncompressed images, we simply treat them with $\mathbf{e}_n^Q = 0, \mathbf{e}_{n,l}^Q = 0$.

5.4.2 Existing Example-based SR Algorithms

The example-based SR methods make use of a training set of images in a very similar way [9-14]. The training images are treated as “examples” where the HF details of a tested LR image are learnt. In the example-based SR algorithms, the tested LR image and HR training images are subdivided into patches in raster-scan order, and the patch pairs form a training set. For each patch of the tested LR image, its HR counterpart is searched by a training process. The found HR patches are then fused to a HR image.

W.T. Freeman et al. proposed an example-based SR approach using the Markov network to probabilistically model the relation between HR and LR patches and between neighboring HR patches is proposed in [9]. They also developed a one-pass algorithm that simplify the iterative solution in [16]. The results of the one-pass algorithm are nearly as good as the iterative solution to the Markov network.

To be more flexible, a SR approach is proposed based on LLE in [15]. It

provides an explicit way to extract features from the HR and LR patches, which is elaborated in Figure 5.6. The LR frames $\mathbf{y}_n$, the training images $\mathbf{x}_r$ and its LR counterpart $\mathbf{y}_r$ are divided into patches. We denote the sets of image patches corresponding to $\mathbf{y}_n$, $\mathbf{y}_r$ and $\mathbf{x}_r$ as $\{\mathbf{y}_n^t\}_{t=1}^{N_b}$, $\{\mathbf{y}_r^s\}_{s=1}^{N_b}$ and $\{\mathbf{x}_r^s\}_{s=1}^{N_b}$, respectively, where N_b is the number of the patches in each image. Each patch is represented as its feature vector, which is composed by the first and the second order gradients of its center pixel. We use $\mathbf{Y}_n^t$, $\mathbf{Y}_r^s$ and $\mathbf{X}_r^s$ to denote the feature vectors.

Figure 5.6 Block diagram of the example-based SR algorithm in [15]

For each LR patch $\mathbf{Y}_n^t$, its HR counterpart will be constructed as follows: The LR patch is compared to all patches $\{\mathbf{Y}_r^s\}_{s=1}^{N_b}$. Then the K nearest neighbors of $\mathbf{Y}_n^t$ are selected as candidate patches $\{\mathbf{Y}_r^s | s \in C'\}$, where C' is a set in which the distance of $\mathbf{Y}_r^s$ and $\mathbf{Y}_n^t$ is the smallest in the feature space. The HR counterparts $\{\mathbf{x}_r^s | s \in C'\}$ are subtracted by its mean value, multiplied with a weight $w_k, k = 1, 2, \dots, K$ and added to estimate the local geometry of the HR patch $\mathbf{x}_n^t$. Then the estimation is added to the mean value of $\mathbf{Y}_n^t$ for the reconstructed HR patch $\mathbf{x}_n^t$. The reconstructed HR patches

$\{\mathbf{x}_n^t\}_{t=1}^{N_b}$ are fused into a HR image $\mathbf{x}_n$ ' using an overlapping between adjacent patches.

5.4.3 Proposed SR Algorithm

If the LR observation $\mathbf{y}_n$ is not contaminated with compression noise, the SR approach in [15] (referred as “LLE method”) produces a plausible result with details well restored. However, the low-level training set alone is not enough to distinguish the compression noise from correct image data since it interprets the artifacts as image data and enhances them [16]. Based on the fact that the training image has strong correlation with the LR image (i.e., use $\mathbf{x}_l$ as $\mathbf{x}_r$), we impose constraints on the searching range of the nearest neighbors to address this problem. Besides, we expect the made-up artifacts would not scintillate in the video sequence. For this purpose, we iteratively compare the result of the training process with its LR counterpart and the warped training image to rectify the “unreal” created image details.

Our proposed SR algorithm comprises of two parts (see Figure 5.7): the initial estimation and iterative correction. In details, the algorithm differs from LLE method as follows. Hereby we denote the sets of image patches corresponding to $\mathbf{y}_n^Q$, $\mathbf{y}_l^Q$ and $\mathbf{x}_l^Q$ as $\{\mathbf{y}_n^t\}_{t=1}^{N_b}$, $\{\mathbf{y}_l^s\}_{s=1}^{N_b}$ and $\{\mathbf{x}_l^s\}_{s=1}^{N_b}$, respectively. The corresponding feature vector for each patch is denoted as $\mathbf{Y}_n^t$, $\mathbf{Y}_l^s$ and $\mathbf{X}_l^s$.

Figure 5.7 Block diagram of the proposed example-based SR algorithm

1. For SR in video enhancement in our work, the value of K used in [15] should be set as 1, since the training image is highly correlated to the LR frames. Furthermore, to avoid introducing unreal HF details in the SR reconstruction, we set a threshold to select the nearest neighbors, i.e., only $\{x_1^s \mid s \in C_1 : \rho(Y_n^t, Y_1^s) < \zeta\}$ are selected, where $\rho()$ is a cost function, measured by L1 norm of the difference of the features, ζ is a threshold. The difference $\rho(Y_n^t, Y_r^s)$ larger than ζ indicates that the patch we are seeking does not exist in x_1^Q , and then the bilinear interpolation of y_n^t is used as the estimation of x_n^t .
2. Moreover, an image is modeled as a Markov random field (MRF), that is to say, the probability distribution of a patch on the SR image is conditionally independent of all but the neighborhood of the patch. Assuming that the consecutive frames in the same GOP is strongly temporally correlated, a constraint is imposed that t, s must belong to the same neighborhood: $s \in C_2 : \{t - w \leq s \leq t + w\}$, where w is the size of neighborhood.

Intuitively, the constraint is introduced because the artifact due to quantization may make the similarity between x'_n and its corresponding LR version y'_n distorted, thus the selection from all patches in Y_n may not work. The imposed constraint C_2 prevents the patches x_1^s which satisfy that $\rho(Y'_n, Y_1^s)$ is small while $\rho(X_n^t, X_1^s)$ is large being chosen, where $\rho(X_n^t, X_1^s)$ represent the difference between HR patches in vector space. Then the result of the training process is expressed by

$$\tilde{x}_n = \text{ArgMin} \left[\sum_{t=1}^{N_b} \sum_{s \in C_1 \cap C_2} \rho(Y'_n, Y_1^s) \right] \quad (5.11)$$

3. Iterative correction is applied using *inter-resolution correlation* and *inter-frame correlation* to rectify the made up artifacts produced by the training process.

The *inter-resolution correlation* between the training result $\tilde{x}_n$ and the LR observation y_n^Q is

$$\tilde{x}_n' = \text{ArgMin}_{\tilde{x}_n} [\|y_n^Q, W_n \tilde{x}_n\|^2] \quad (5.12)$$

The correction process using (5.12) does not guarantee $\tilde{x}_n'$ would approach to the ground truth x_n due to the lack of the prior knowledge of e_n in (5.9). But it removes the made-up artifacts by avoiding the down-sampling of $\tilde{x}_n'$ diverge from y_n^Q .

Pixels of the warped intra-frame are also used to rectify the artifact in SR result. It exploits the inter-frame relation to restore the current frame:

$$\tilde{x}_n'' = \text{ArgMin}_{\tilde{x}_n'} [\|\tilde{x}_n', S_n C(d_{n,1}) x_1^Q\|^2] \quad (5.13)$$

To prevent the error due to inaccurate registered pixel, the inaccurate HR

displacements are detected using the optical flow (OF) calculation, as introduced in [59]. The OF at each point is calculated based on the assumption that the brightness of a particular point in the pattern is constant, i.e.

$$\frac{\partial \mathbf{x}_1^Q}{\partial x} u + \frac{\partial \mathbf{x}_1^Q}{\partial y} v + \frac{\partial \mathbf{x}_1^Q}{\partial t} = 0 \quad (5.14)$$

where u, v are pixel displacements in the x and y directions, and $\frac{\partial \mathbf{x}_1^Q}{\partial t}$ is the temporal brightness derivative between $\hat{\mathbf{x}}_n$ and $\mathbf{x}_1^Q$. If the forward and backward displacements of the corresponding pixel in two frames do not match, i.e. $\mathbf{d}_{n,1} \neq -\mathbf{d}_{1,n}$, it is regarded as inaccurate and excluded by a mask, which is a matrix denoted as S_n comprises of “0” and “1”.

It is well known that SR is an ill-posed problem, thus regularization is applied to ensure a stable solution $\hat{\mathbf{x}}_n$. The TV method is used for the regularization [58]. Adding the TV regularization term $\lambda_3 \cdot (\|\nabla \tilde{\mathbf{x}}_n\|)$ and combining (5.11), (5.12) and (5.13), the SR reconstruction is defined as the following minimization problem:

$$\hat{\mathbf{x}}_n = \underset{\tilde{\mathbf{x}}_n}{\text{ArgMin}} [\lambda_1^2 \cdot \|\mathbf{y}_n^Q, \mathbf{W} \tilde{\mathbf{x}}_n\|^2 + \lambda_2^2 \cdot \|\tilde{\mathbf{x}}_n, \mathbf{S}_n \mathbf{C}(\mathbf{d}_{n,1}) \mathbf{x}_1^Q\|^2 + \lambda_3^2 \cdot \|\nabla \tilde{\mathbf{x}}_n\|] \quad (5.15)$$

where $\lambda_1, \lambda_2, \lambda_3$ are the weighting scalars and ∇ is the gradient operator.

5.4.4 Implementation Details

As presented in Figure 5.7, the implementation of the proposed SR algorithm is as follows:

1. Nearest-neighbor Searching: For each vector $\mathbf{Y}_n^t$, its most similar vector $\mathbf{Y}_1^s$ with

minimum $\rho(Y_n^t, Y_1^s)$ in a window $[s-w \leq t \leq s+w]$ is found under constraint C_2 .

We use a threshold θ to decide whether y_n^t contains local geometry needed to be restored or not, which is measured by its variance. Combined with the constraint C_1 ,

x_n^t is constructed by adding the average of y_n^t and the HF details of y_{n-1}^s if

$\rho(Y_n^t, Y_1^s) < \zeta$ and $E\|y_n^t - \mu\|^2 > \theta$, where μ is the average of pixels in y_n^t . Otherwise,

the bilinear interpolation of y_n^t is used as the estimation of x_n^t .

2. HR frame reconstruction: To enforce the local compatibility and smoothness of the fused HR image $\tilde{\mathbf{x}}_n$, patch searching and motion compensation are applied on patches with an overlap of one pixel between adjacent patches.
3. Iterative correction: The fused HR frame $\tilde{\mathbf{x}}_n$ is then iteratively corrected using steepest descent algorithm to solve the minimization problem in (5.15).

$$\begin{aligned}\hat{\mathbf{x}}_n^{(0)} &= \tilde{\mathbf{x}}_n \\ \hat{\mathbf{x}}_n^{(j+1)} &= \hat{\mathbf{x}}_n^{(j)} - [\lambda_1 (\mathbf{W}_n \hat{\mathbf{x}}_n^{(j)} - \mathbf{y}_n) + \lambda_2 (\hat{\mathbf{x}}_n^{(j)} - \mathbf{S}_n \mathbf{C}(\mathbf{d}_{n,l}) \mathbf{x}_1^{\mathcal{Q}})] \\ &\quad + \lambda_3 \cdot \nabla \hat{\mathbf{x}}_n^{(j)}\end{aligned}\tag{5.16}$$

where $\hat{\mathbf{x}}_n^{(j)}$ is the SR estimate at the j -th iterate.

5.5 Experiments

5.5.1 Down-sampling Based Coding

In this subsection, we construct a set of experiments for examining the effectiveness of the proposed DBC in Section 5.2. In these experiments, we use gray-level CIF video sequences with frame size 352×288 and the down-sampling factor $M = 2$ for inter-

frames. The H264/AVC reference software (Joint Model version 8.6) is referred as the “regular coding”. For comparison, the test sequences are encoded with approximately the same bit rates using the regular coding and our proposed scheme. The SR results of the down-sampled frames using the LLE method are also provided to show the efficacy of the proposed SR. To verify the performance improvement of the proposed scheme at low bit rates, we use coarse quantization parameters (QP) to obtain rate-distortion curves shown below.

For the implementation of the proposed SR, the parameters are set as follows: the searching window size w is 8, the LR patch size is 3×3 as in [15], and $\theta = 5$ for all the sequences, the threshold ζ is set as a larger value for the sequences with low motion and a smaller value for the ones with higher motion. In the experiments, we empirically fixed $\zeta = 5000$ for sequences “*Missa*”, “*Mother & daughter*” and “*Salesman*”, and $\zeta = 1000$ for “*Foreman*” and “*Football*”. For iterative correction, the weighting scalars $\lambda_1, \lambda_2, \lambda_3$ are all set to 0.1 for all the sequences. The HR displacements $\mathbf{d}_{n,l}$ in (5.6) is estimated using the code of [118]. The iterations stops when the relative change $\|\bar{\mathbf{x}}_n^{(j+1)} - \bar{\mathbf{x}}_n^{(j)}\|^2 / \|\bar{\mathbf{x}}_n^{(j)}\|^2$ falls below 10^{-6} . The number of needed iterative steps varies in the SR process of each frame, and it does not exceed 5 steps.

In the first set of experiments in Section 5.5.1, we use only intra-frames in the test sequence. The first frame of each GOP is coded with the original frame size and used as the training image at the decoder, while the others are down-sampled. Figure 5.8 shows the PSNR vs. bit rate for different test sequences: “*Mother & daughter*”, “*Missa*” and “*Salesman*”. It is observed that the proposed coding scheme outperforms

the H.264 as well as the SR results using LLE method over a wide range of bit rates. The improvements vary with different sequences, since the image statistics is different [113].

Figure 5.8 Average PSNR comparison of the regular coding (H.264), proposed coding and LLE method using only intra-frames for: (a) *Mother & daughter* (b) *Missa* (c) *Salesman*.

In the second set of experiments, the test sequences include both intra-frame and inter-frame. Frame rate is 30 frames per second and each GOP consists of one I frame and 49 P frames, as set in [110]. While the first set of experiments verified the better performance of DBC over H.264 at low bit rates, these experiments show the

advantage of the proposed coding scheme at the aspect of exploiting the *inter-resolution correlation* and *inter-frame correlation*.

We compare the reconstructed HR frame obtained by the proposed SR algorithm and LLE method using the sequence “*Mother & daughter*”. The training image (intra-frame coded with QP=30), a sample of the down-sampled inter-frames (30th frame coded with QP=30) and its original version are shown in Figure 5.9. It can be seen that although the training image suffers certain compression loss, it maintains sufficient HF details to restore the down-sampled inter-frames. Figure 5.10 (a) is the enlarged LR image. First we use LLE method to get the SR reconstruction. Our experiments show that setting $K = 1$ gives the best result, since the training image used is highly correlated to the LR frames. We set $K = 1$ for LLE methods in all our experiments. It is shown that the compression noise in the LR observation (Figure 5.10 (a)) is enhanced by the LLE method, and the resultant artifacts are observed on the human faces and the mother’s hair in Figure 5.10 (b). The reconstructed HR frame by the proposed SR algorithm is obtained following the three steps described in Section 5.4.3. Figure 5.10 (c) shows the interim output after step 1 and 2, and it can be seen that the made-up artifact around the mother’s eyes is much suppressed. Finally, Figure 5.10 (d) shows the ultimate SR result after four iterations of step 3. The unreal pixels in the area of mother’s eyes and hair are rectified by the iterative correction. It can also be seen that the compression noise at the edges (as the highlighted circle region) is suppressed by regularization operator $\lambda_3^2 \cdot \|\nabla \tilde{\mathbf{x}}_n\|$ used in (5.15).

Figure 5.9 Mother & daughter sequence: (a) training image (intra-frame coded with QP=30), (b) original 30th frame and (c) its down-sampled version with quantization noise.

Figure 5.10 Mother & daughter sequence: (a) Enlarged LR observation (b) SR result of LLE method (PSNR=33.9dB) (c) initial estimation (PSNR=34.5dB) (d) iterative correction (PSNR=34.9dB).

The quality of decoded video sequence depends on many factors, one of which is the bit rate of the intra-frames. The DBC in [110] outperforms H.264 only if intra-frames are encoded with a low bit rate. The proposed coding scheme breaks this limitation by exploiting the *inter-resolution correlation* and *inter-frame correlation* to restore the down-sampled frames. As a result, the performance of the proposed coding scheme also benefits from an intra-frame with better quality. The average PSNR improvements when intra-frames are coded with both low bit rate (intra-frame QP=40) and moderate bit rate (intra-frame QP=30) are shown in Figure 5.11 and Figure 5.12, respectively. For inter-frames, various QP's between 30 to 40 are used in the proposed coding, and the QP's in regular coding are manually set to obtain approximate equal bit rate (each inter-frame is coded with equal QP). Typically the range is [38, 50]. For example, to compare the performance of two coding methods at the bit rate of 10.8kbps of sequence “*Mother & daughter*” (shown in Figure 5.13(a)), the QP for coding inter-frames in regular coding is 50, and the one in the proposed coding is 40. And the intra-frames are both coded with QP=40. It is seen at a wide range of bit rate, the average PSNR of inter-frames from the proposed coding is higher than H.264 with the same intra-frames.

Besides the average PSNR improvements, the constant PSNR improvements of every inter-frame when intra-frames are coded with both low bit rate and moderate bit rate are observed in Figure 5.13 and Figure 5.14. The advantage of the proposed coding for inter-frames comes from down-sampling and the SR process. By down-sampling, less quantization error is imposed but high frequency information is lost. The tradeoff is represented by the PSNR of bicubic interpolation of the down-sampled

frames. It is observed that for the sequence “*Mother & daughter*”, the PSNR of bicubic interpolation is higher than H.264 when the prediction residual between current frame and the intra-frame increases by a certain degree (when the frame number is larger than 20). As this sequence does not contain many edges or texture, the improvements of the proposed coding mainly results from down-sampling. And the PSNR gain of the proposed coding over the bicubic interpolation is produced by the iterative correction and regularization of the proposed SR process. For the sequence “*Salesman*”, the shelf in the background comprises of many edges, thus the lost of HF information by down-sampling makes the PSNR of the bicubic interpolation constantly lower than H.264. The improvement of the proposed coding over H.264 mainly depends on the SR process distilling HF information from the intra-frame.

Figure 5.11 Average PSNR comparison of the regular coding (H.264), proposed coding and LLE method using both intra-frame and inter-frame at intra-frame QP=40 for: (a) *Mother & daughter* (b) *Missa* (c) *Salesman*.

Figure 5.12 Average PSNR comparison of the regular coding (H.264), proposed coding and LLE method using both intra-frame and inter-frame at intra-frame QP=30 for: (a) *Mother & daughter* (b) *Missa* (c) *Salesman*.

Bit Rate $\approx 10.8\text{kb/sec}$ (Intraframe QP=40)

(a)

Bit Rate $\approx 23.3\text{kb/sec}$ (Intraframe QP=30)

(b)

Figure 5.13 PSNR comparison of the regular coding (H.264), proposed coding, Bicubic interpolation and LLE method in a GOP for sequence *Mother & daughter* at (a) intra-frame QP=40 (b) intra-frame QP=30.

Bit Rate ≈ 20.6 kb/sec (Intraframe QP=40)

Bit Rate ≈ 48.1 kb/sec (Intraframe QP=30)

Figure 5.14 PSNR comparison of the regular coding (H.264), proposed coding, Bicubic interpolation and LLE method in a GOP for sequence *Salesman* at (a) intra-frame QP=40 (b) intra-frame QP=30.

Experiments are also done with different size of GOP using “*Foreman*” sequence, as shown in Figure 5.15. The average PSNR of the frames coded with GOP size $P=50$ are shown in (a) (b), and (c) (d) presents the cases of $P=10$. The averaged PSNRs for the two coding schemes increase with shorter GOP size, since both of them

benefit from the increased temporal redundancy. The proposed coding outperforms the regular coding at low bit rates. As the bit-rate increases, the efficiency of the proposed coding decreases. This verifies the analysis in Section 5.2.3. It shows that advantage of coding schemes results from its effective utilization of limited bits to preserve useful information about encoded frames when the number of bits is small. As the down-sampled inter-frames contain less information than full frames, the performance of the proposed coding will saturate as the available bits increases, and the extra bits will be wasted.

Figure 5.15 Average PSNR comparison for sequence *Foreman* with different GOP size of (a) (b) $P=50$ and (c) (d) $P=10$

Figure 5.16 shows the comparison of sample video inter-frames obtained by H.264 and the proposed coding scheme. It can be seen that the block effect is much alleviated in the sample frames from the proposed coding than the ones from H.264, e.g. the edge of the woman's face in "*Missa*" and the right eye of the man in "*Foreman*". As the "*Salesman*" sequence contains moderate motion, the quantization error on prediction residual is more visible. The proposed coding well preserves meaningful details in the reconstructed frames. For example, one cannot tell from (e) whether the salesman is holding something on his hands, while a box can easily be distinguished from the hands in (f). The improvement of face restoration can also be observed. The better visual quality of inter-frames indicates the training process in the proposed SR algorithm makes a good use of the motion-related temporal dependency.

Figure 5.16 Video inter-frame samples (a) Original *Missa* frame (b) Regular coding, PSNR=34.9dB. (c) Proposed coding, PRNR= 35.3dB. (d) Original *Salesman* frame (e) Regular coding, PSNR=28.3dB. (f) Proposed coding, PRNR= 29dB. (g) Original *Foreman* frame (h) Regular coding, PSNR=29.9dB. (i) Proposed coding, PSNR=31.4dB.

The proposed coding with different down-sampling factors is also examined using sequence “*Football*”, which has high motion. The inter-frame QP of the proposed coding at down-sampling factors $M = 2$ is constantly set to 40, and intra-frames are coded with both low bit rate (intra-frame QP=40) or moderate bit rate (intra-frame QP=30). The inter-frame QPs for other two coding methods are corresponding set for approximately equal bit rates. PSNR results are shown in Figure

5.17. The proposed coding achieves certain PSNR gain at $M = 2$. But the advantage of the proposed coding over regular coding is not seen at $M = 4$, since the MC is applied at LR and the registration error increases significantly when M becomes larger.

$P = 50$, Bit Rate $\approx 108.0\text{kb/sec}$ (Intraframe QP=40)

(a)

$P = 50$, Bit Rate $\approx 145.8\text{kb/sec}$ (Intraframe QP=30)

(b)

Figure 5.17 PSNR comparison of the regular coding (H.264), proposed coding at different down-sampling factors in a GOP for sequence *Football* at (a) intra-frame QP=40 (b) intra-frame QP=30.

It should be pointed out that the performance improvement of the proposed method is achieved at the cost of a moderate increase of computational complexity. For the encoder of the proposed coding, only the down-sampling procedure is added for decreasing the frame size of inter-frames. The increase of complexity comes mainly from the SR algorithm, which can be regarded as a post-process at the decoder. As shown in Fig 5.7, the SR algorithm includes (a) Nearest neighbor searching, (b) HR frame reconstruction (image fusion), (c) OF calculation & HPDP outlier detection, (d) EIHPD and (e) Iterative correction. As the exact number of arithmetic operations of implementing the proposed SR algorithm is difficult to determine, we follow [112] to use computational time to show the complexity of the proposed coding. The SR algorithms are run using Matlab v7.6.0 R2008a under Windows XP Professional on an Intel Core 2 Quad Q9400 processor at 2.66 GHz with 3.25G of RAM. The encoding/decoding is done by H.264 reference software. Computational time is calculated as the mean time for all the sequences and the bitrates considered normalized by the number of frames in each sequence with GOP=50. For the regular coding, the encoding/decoding time was 6.39s. For the proposed coding the computational time consists of two parts: encoding/decoding and SR. The encoding/decoding time was equal to 1.98s (including down-sampling procedure). It is noted that the image size of inter-frames in the proposed coding is a quarter of the ones in regular coding, thus the time consumed in encoding/decoding is reduced. The total running time of the SR process consumes about 15.8s including Module (a) 5.69s (b) 2.76s (c) 7.20s (d) 0.0025s (e) 0.17s. As a post-processing, the running time of the

SR process is comparable with the one in [112]. Finally, the total time of the proposed coding is around 17.8s.

5.5.2 Low-power Video Acquisition

For simulating the power saving of the proposed video acquisition in Section 5.3 and the quality of the reconstructed video, we have tested five gray-level CIF video sequences consisting of high motion, high detail, low motion, or low detail regions as full-resolution frames. The LR frames are obtained by blurring the FR frames with a 4×4 averaging filter and subsequently decimating by $M = 2$. They are enhanced to HR by the proposed SR algorithm. The convergence criterion is the same as in Section 5.5.1. The SR results using the LLE method in [15] are presented as a comparison, where the reconstructed video comprises of original key frames and super-resolved frames.

To evaluate the objective quality of the video frames, we follow [115] using frame-to-frame and average PSNR with respect to original full-resolution video sequences. The average PSNR of the video frames from the proposed uniform acquisition at various ratios of key frame (called as key frame ratio) vs. total frames in a GOP is shown in Figure 5.18, as compared with results from content-based acquisition [115], LLE method [15] and coarse sampling (capturing a video at LR for all the frames). The video from the proposed framework achieves higher PSNR as compared with the LLE method and coarse sampling in all key frame ratios, and outperform the content-based acquisition from 0.2 to 0.4 onwards depending on the

sequences tested. It is seen that the PSNR improvement over [115] is up to 3 dB, which verifies the ability of SR technique to reduce registration error (i.e. error due to inaccurate motion estimation) and restore the HF details in the reconstructed frames. As the key frame ratio decreases, the proposed framework remains a significant PSNR gain over the coarse sampling. It is noted that, when the key frame ratio decreases to 0.1 (i.e. only one FR frame is needed when the other 9 frames of each group are coarsely sampled), the PSNR gain over coarse sampling is up to 8.9 dB.

Figure 5.18 Average PSNR obtained by different approaches at various key frame ratios on five test sequences: (a) *Foreman*, (b) *Hall*, (c) *Garden*, (d) *Container*, (e) *Coastguard*.

For power measurements, we follow [115] to assume the power is proportional to the sampling solution. The comparison of the uniform acquisition and the content-based acquisition in [115] is listed in Table 3. It is shown that proposed framework consume less power than [115] at all key frame ratios, and the power saving

improvement over [115] is up to 38.1% when the key frame ratio is 0.1. In this case, the power saving, as well as the total number of pixels transferred, is close to the coarse sampling. It means that a significant PSNR gain (8.9dB) can be achieved over the coarse sampling while maintaining a high-frame-rate.

Table 3 Comparison of power saving (measured by pixel number)

Video sequence	Uniform acquisition (10 ⁶)			Full resolution (10 ⁶)	Using [120] (10 ⁶)	% Power saving improvement over[120]		
	Key frame ratio					Key frame ratio		
	0.5	0.2	0.1			0.5	0.2	0.1
<i>Container</i> (300 f)	19	12.2	9.9	30.4	19.8	1.5	24	31.5
<i>Hall</i> (300f)	19	12.2	9.9	30.4	19.5	1.5	24	31.5
<i>Coastguard</i> (300 f)	19	12.2	9.9	30.4	20.3	4.1	26.1	33.6
<i>Garden</i> (115f)	6.1	3.9	3.2	9.7	6.9	8.1	30.6	38.1
<i>Foreman</i> (300f)	19	12.2	9.9	30.4	20.1	3.7	26.2	33.7

As for subjective video quality, the following two problems are solved by the proposed SR algorithm in applying example-based SR algorithms: 1) the created high-frequency details degrade the subjective quality of the reconstructed video, and 2) introducing the noise in the training image to the reconstructed HF frame. The samples of reconstructed frames when the key frame ratio is 0.5 are shown in Figure 5.19 and Figure 5.20. For comparison, the bi-cubic interpolation of the LR frames are also shown, which is regarded to remain the LR but increases frame size for display. The resolution of the SR results is significantly enhanced compared to the bicubic interpolation, and the image details are well restored. The difference between the super-resolved frames by the proposed SR and the LLE method is not apparent when

viewing the individual frame. This is because without noise, the reconstruction error of the LLE method is mainly due to sub-pixel registration error, which is not sensitive to viewers. However, it would degrade the subjective quality of the reconstructed video with uneven motion appeared, since the LLE method intends to reconstruct frames close to the referred key frame rather than the real ones.

Figure 5.19 Samples of reconstructed frames obtained by bicubic interpolation (left) and the proposed SR (middle), original frame (right) of: ((a)-(c)) *Foreman*, ((d)-(f)) *Hall*.

Figure 5.20 Samples of reconstructed frames obtained by bicubic interpolation and the proposed SR, original frame of: ((a)-(c)) *Garden*, ((d)-(f)) *Container*, ((g)-(i)) *Coastguard*.

The second experiment is constructed to show that the proposed SR technique is able to avoid recognizing the noise in the training image as image content and adding it to the reconstruction. To simulate the acquisition noise such as high ISO noise, an AWGN at the level of SNR=25dB is added to all the frames of the sequence *Mobile*. The reconstructed frames by the bi-cubic interpolation, the LLE method and the proposed SR are presented in Figure 5.21, as well as the key frame used as training image. For a better comparison, part of each frame in a square is enlarged as (f)-(j).

Enhancing the LR frame (a) is a tough task since both (a) and the training image (e) are rather noisy. Such noise imposes difficulties for block searching and recognizing correct data in the LLE method. As a result, the inaccurate high-frequency details are added to the reconstruction, which is shown in (g). In contrast, it can be seen in (j) that the number “6” can be better recognized, and the noise is alleviated. This simulation indicates that the proposed SR technique achieves not only higher PSNR, but also significant subjective quality improvement, with noise suppressed and resolution enhanced.

Figure 5.21 Reconstructed frames of noisy video sequence Mobile using (a) bicubic interpolation or super-resolved by (c) LLE method (d) the proposed SR with the help of the training image (e). A part of each sample frames is enlarged as (f)-(j), respectively.

5.6 Summary

We proposed a generalized video acquisition system, which integrates a down-sampling operator for some of the acquired video frames during capturing or before compression and SR technique for restoring the acquired LR frame to HR. The

advantage of the proposed acquisition system could be summarized as follows:

- For video compression, the proposed framework (in the form of DBC) could achieve high PSNR than the regular coding at a certain low bit rate range.
- For applications with VASI-like sensors, the proposed framework could achieve better video quality than the content-based acquisition [115], while it saves more power and the acquired video is applicable for most prevalent video compression techniques. No new hardware is required for this application.
- For applications with conventional sensors, the proposed framework is applicable at the high-frame-rate mode. It enhances the resolution and quality of the acquired video, and produces high-frame-rate and HR video. The proposed high-frame-rate and HR video acquisition can be implemented by predefining the capturing mode as remaining the first frame at full resolution and the others at LR by throwing away pixel data at LR mode at each second.

Another contribution of this chapter is to develop a new SR technique. The existing example-based SR algorithms are not applicable for video enhancement because 1) they cannot handle the compression noise and 2) the created image details would deteriorate the subjective quality. These problems have been addressed by the proposed SR algorithm exploiting the *inter-resolution correlation* and *inter-frame correlation* of video frames. Experimental results show that the proposed SR algorithm outperforms the method in [15] in processing video frames.

Chapter 6

Conclusions and Recommendations

6.1 Conclusions

In many computer vision applications, the resolution of acquired images is limited by many factors related to the imaging system, such as the density of sensors and the characteristics of lens. This thesis investigates the SR techniques for HR image and video construction. In the following subsections, we summarize the work of the thesis at two respects: algorithm and applications.

6.1.1 SR Algorithms Based on More General Degradation

Models

This thesis has addressed the SR from differently degraded images: 1) images with different degree of out-of-focus blur and 2) images with different zooming ratios.

Many successful SR algorithms are constructed based on an ideal assumption: the input LR images have common PSF. Then one of the captured LR images is

randomly selected as the referenced image and interpolated to higher dimension as the initial SR estimation. This initial SR estimation is then degraded and compared with the other LR images for correction. This MISO framework, however, is sensitive to the initial estimation, and the solution may not converge to a global minimum. In other words, the performance of MISO framework is not stable when the input images have different PSFs. This issue is addressed in the thesis by designing a multiple reference-based fusion framework with a better initial estimation and tighter constraints to limit the indetermination of the SR solution. In this framework, the SR outputs using different input LR images as referenced image are regarded as independent output channels, and all the SR output channels are used to refine the reconstructed image. Experimental results have verified the advantage of the proposed algorithm in both subjective and objective terms.

Another issue arising from the SR problem is the registration of the LR images. The problem becomes more difficult when the LR images are not acquired at the same plane parallel to the camera lens. Few works have addressed the SR problem considering zooming due to the difficulty to construct a linear image capture model. In this thesis, we investigate this issue by constructing a linear system relating the underlying HR image and the LR observations with two parameters, which are the coordinate of the centre of the zoom operation and the zooming ratio. A MAP based SR method was established under this image capture model for estimating the zoom ratios and the HR image. The validity of this algorithm has been verified by simulated images.

6.1.2 Applications of SR as Post-processing

We explored the applications of SR techniques by integrating it into a generalized video acquisition system. Some of the acquired video frames are deliberately down-sampled for particular purposes, which are then super-resolved to original size during post-processing. Specifically, the proposed applications using SR techniques as post-processing include: DBC, low-power acquisition, and high-frame-rate and HR video acquisition.

It has been reported that over-sampling a still image before compression may further degrade the image quality. Similarly, down-sampling before video compression in low bit rate video coding may alleviate the blocking effect and improve PSNR of the decoded frames. When the number of DCT coefficients is reduced in such a DBC, the bit budget of each coefficient will increase, thus the quantization error is reduced. We proposed a DBC scheme in this thesis, where a SR technique is employed to restore the down-sampled frames to HR. We also analyzed the MSE of the coded frames by DBC, justified the advantage of DBC over regular coding (with no down-sampling). The effectiveness of the proposed DBC scheme was verified by experiments.

Besides, power optimization in video acquisition is a critical issue in mobile devices. With the VASI, which is capable to turn on/off each sensor during capturing each frame, we developed a low-power acquisition scheme. Different from some low-power content-based acquisition, the proposed acquisition scheme does not require a pre-analysis to detect the ROI. It divides video sequences into groups with equal number of frames. The first frame of each GOP is captured at the FR and the rest are at

the LR. A SR technique is employed afterwards to reconstruct LR frames to HR. Experiments showed that the proposed scheme further reduces the power consumption effectively with some video quality improvement compared to some existing low-power acquisition.

In addition, the proposed scheme can overcome the pixel transfer rate limit to obtain HR video at high-frame-rate. In some conventional techniques, the frame rate is increased at the cost of reducing the resolution of acquired video frames. The proposed scheme provides a new approach to increase the frame rate while retaining the HR of the acquired video. Moreover, the quality of acquired video can be enhanced by SR techniques.

For super-resolving the down-sampled video frames of the generalized video acquisition system, we proposed an example-based SR technique in this thesis to make full use of the interlaced HR frames (the first frame of each GOP). As the existing example-based SR algorithms are designed for single-frame SR, they are not readily applicable for video enhancement because 1) they cannot handle the compression noise and 2) the created image details would deteriorate the subjective quality in moving frames. The proposed SR algorithm has addressed these problems by exploiting the inter-resolution and inter-frame correlation of the obtained video frames. Experimental results have shown that the proposed SR algorithm is effective for video enhancement.

6.2 Future Works

Based on the research in this thesis, we recommend the potential directions for future research in this section.

- Joint image registration and SR from arbitrarily zoomed images

Studies in this thesis have established an image capture model, under which a MAP based SR algorithm is constructed. One possible direction for future research is to develop an approach to jointly estimate the registration parameters and the underlying HR image based on this model. The success of fulfilling registration and SR in a joint fashion has been shown in the work considering translational motion [60, 102], where the simultaneously or alternatively corrected estimation of registration parameters could significantly improve the quality of the ultimate reconstructed HR image, as exactly accurate registration is unavailable. Although we provided an effective approach for estimating the registration parameters in advance of the estimation of HR image in the presence of zoom motions, it is expected that it would improve the SR performance to develop a method for jointly registering the LR images and reconstructing HR image.

- SR from color video frames

In this thesis, SR problem has been investigated mainly for grayscale videos. So a possible extension of the current work is to the SR from color videos. A color image is typically represented by combining three separate monochromatic images. Although the current SR algorithms can be extended in principle to color images by applying it to each color channel in a full RGB field individually, however, this is not

the case for some advanced video coding schemes, which uses a sampling structure in which the chroma component (Cb and Cr components) has one fourth of the number of samples than the Y component. Besides, some color digital cameras using a single CCD also suffer from color mosaicking [119], for which demosaicking is required to interpolate the other colors from neighboring samples. Merely using demosaicking to pre-process the decoded frames prior to SR may lead to suboptimal results. It is expected a joint framework by merging SR and demosaicing into one single process like [120, 121] would be a good approach. But it should be noted that this issue cannot be addressed by applying the existing methods, since they are designed only for CFA demosaicking using the Bayer pattern. For a compressed video, the missing information of the chroma component due to compression should also be considered. Therefore, we believe that the future research in this area will be valuable for SR from color video frames.

- SR from all LR video frames

We proposed in this thesis some applications of SR with some video frames retaining their original frame size and more HF information, and correspondingly a SR technique that can readily be implemented for these applications. The basic idea behind the proposed SR technique, which is developing example-based SR algorithms for video enhancement, should be extended to other learning based SR algorithms from a video with no HR training images. Then these SR algorithms can be applied in more general situations, where the HR frames may not be available.

Author's publications

Journal Papers

Minmin Shen, Ci Wang, Ping Xue, and Weisi Lin, "Performance of Reconstruction-based Super-resolution with Regularization," *Journal of Visual Communication and Image Representation*, vol. 21, issue 7, pp. 640-650, 2010.

Minmin Shen, Ping Xue, "Low-Power Video Acquisition with Super-Resolution Reconstruction for Mobile Devices", *IEEE Trans. Consumer Electronics*, vol.56, no.4, pp.2520-2528, Nov. 2010.

Minmin Shen, Ping Xue, and Ci Wang, "Down-Sampling Based Video Coding Using Super-resolution Technique," will appear on *IEEE Trans. Circuits and Systems for Video Technology*, vol. 21, issue 6, Jun. 2011.

Minmin Shen, Ci Wang, and Ping Xue, "Super-resolution for Images Captured with Different Out-of-focus Blur", resubmitted to *Journal of Visual Communication and Image Representation*.

Conference Papers

Minmin Shen and Ping Xue, "A Fast algorithm for rain detection and removal from videos", will appear on *IEEE International Workshop on Advances in Automated Multimedia Surveillance for Public Safety*, 2011.

Minmin Shen and Ping Xue, "Super-Resolution from Observations with Variable Zooming Ratios", *IEEE International Symposium on Circuits and Systems (ISCAS)*,

May 30-Jun. 2, 2010, Paris, France.

Minmin Shen, Ping Xue and Ci Wang, “Down-Sampling Based Video Coding with Super-Resolution Technique”, *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 30- Jun.2, 2010, Paris, France.

Minmin Shen and Ping Xue, “Low-power uniform video acquisition with super-resolution reconstruction”, *7th International Conference on Information, Communications and Signal Processing*, Dec. 8-10, 2009, Macao, China.

Minmin Shen, Ping Xue, Ci Wang, “The Influence of Regularization Parameter on Error Bound in Super-Resolution Reconstruction”, *Pacific-Rim Conference on Multimedia*, Dec. 9-13, 2008, Taiwan, China.

Minmin Shen, Ping Xue, Ci Wang, “A Novel Scalable Video Coding Scheme Using Super Resolution Techniques,” *IEEE International Workshop on Multimedia Signal Processing*, Oct. 8-10, 2008, Cairns, Queensland, Australia.

Minmin Shen, Ci Wang, Ping Xue and Weisi Lin, “Initial Image Selection and Its Influence on Super-Resolution Reconstruction,” *IEEE Workshop on Signal Processing Systems*, Oct. 17-19, 2007, Shanghai, China.

Ci Wang, Ping Xue and Weisi Lin, **Minmin Shen**, “Image Super-Resolution Framework with Multi-Channel Constraints,” *IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2-5, 2007, Beijing, China.

Bibliography

- [1] A. Mohammad-Djafari, "Super-Resolution: A Short Review, A New Method Based on Hidden Markov Modeling of HR Image and Future Challenges " *The computer journal*, vol. 52, pp. 126-141 Jan., 2009.
- [2] P. M. Q. Aguiar and J. M. F. Moura, "Three-dimensional modeling from two-dimensional video," *IEEE Trans. Image Process*, vol. 10, pp. 1541-1551 2001.
- [3] Y. Sato and K. Ikeuchi, "Recovering shape and reflectance properties from a sequence of range and color images," in *IEEE/SICE/RSJ Int. Conf. Multisensor Fusion and Integration for Intelligent Systems*, Dec. 1996, pp. 493-500.
- [4] M. Daum and G. Dudek, "On 3-D surface reconstruction using shape from shadows " in *IEEE Conf. Computer Vision and Pattern Recognition*, 1998, pp. 461-468
- [5] X. Xu and R. M. Narayanan, "Three-dimensional interferometric ISAR imaging for target scattering diagnosis and modeling " *IEEE Trans. Image Process*, vol. 10, pp. 1094-1102, 2001.
- [6] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: interpolation methods in medical image processing " *IEEE Trans. Medical Imaging*, vol. 18, pp. 1049 - 1075, 1999.
- [7] A. Schatzberg and A. J. Devaney, "Super-resolution in Diffraction Tomography," *Inverse Problems*, vol. 8, pp. 149-164, 1992.
- [8] M. A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, pp. 27-38, Jan. 2005.
- [9] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *International Journal of Computer Vision*, vol. 40, pp. 25-47, Oct. 2000.
- [10] J. Sun, N. N. Zheng, H. Tao, and H. Y. Shum, "Image hallucination with primal sketch priors," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Madison, WI, USA, 18-20 Jun. 2003, pp. 729-736.
- [11] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image analogies," in *Proc. Twenty-Eighth Annual Conf. Computer Graphics and Interactive Techniques (SIGGRAPH 2001)*, Los Angeles, CA, USA, 12-17 Aug. 2001, pp. 327-340.
- [12] C. B. Atkins, C. A. Bouman, and J. P. Allebach, "Tree based resolution synthesis," in *Proc. Conf. Image Processing, Image Quality and Image Capture Systems* Savannah, GA, USA, 25-28 Apr. 1999, pp. 25-28.
- [13] W. T. Freeman and E. C. Pasztor, "Learning low-level vision," in *Proc. IEEE Int. Conf. Computer Vision*, Kerkyra, Greece, 20-27 Sep. 1999, pp. 1182-1189.
- [14] C. Liu, H. Y. Shum, and C. S. Zhang, "A two A two-step approach to hallucinating faces: global parametric model and local nonparametric model,"

- in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Kauai, HI, USA, 8-14 Dec. 2001, pp. 192-198.
- [15] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Washington, USA, Jun. 2004, pp. I-275-I-282.
 - [16] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example based super-resolution," *IEEE Computer Graphics and Applications*, pp. 22(2):56-65, 2002.
 - [17] R. Y. Tsai and T. S. Huang, "Multiple frame image restoration and registration," *Advances in Computer Vision and Image Processing*, Greenwich, CT: JAI Press Inc., pp. 317-339, 1984.
 - [18] S. C. Park, M. K. Park, and M. G. Kang, "Superresolution Image Reconstruction - A Technical Overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 21-36, May 2003.
 - [19] J. D. v. Ouwerkerk, "Image super-resolution survey," *J. Image and Vision Computing*, vol. 24, pp. 1039-1052, 1 Oct. 2006.
 - [20] M. K. Ng and N.K. Bose, "Mathematical Analysis of Super-Resolution Methodology," *IEEE Signal Processing Magazine*, pp. 62-74, May 2003.
 - [21] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *Int. J. Imag. Syst. Technol.*, vol. 14, pp. 47-57, Oct. 2004.
 - [22] H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," *CVGIP: Graphical Models and Image Processing*, vol. 54, pp. 181-186, Mar. 1992.
 - [23] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Processing*, vol. 10, pp. 1187-1193, Aug. 2001.
 - [24] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1013-1027, Jun. 1990.
 - [25] N. K. Bose, H. Kim, and B. Zhou, "Performance analysis of the TLS algorithm for image reconstruction from a sequence of undersampled noisy and blurred frames," in *IEEE Int. Conf. Image Processing*, Austin, TX, Nov. 1994, pp. 571-575.
 - [26] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646-1658, Dec. 1997.
 - [27] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lowerresolution image sequences and space varying image restoration," in *IEEE ICASP*, San Francisco, CA., Mar. 1992, pp. 169-172.
 - [28] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Process.*, vol. 6, pp. 1064-1076, Aug. 1997.
 - [29] P. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust, object-based high-resolution image reconstruction from low-resolution video," *IEEE Trans. Image Process.*, vol. 6, pp. 1446-1451, Oct. 1997.
 - [30] A. J. Patti and Y. Altunbasak, "Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-

- order interpolants," *IEEE Trans. Image Process.*, vol. 10, pp. 179-186, Jan 2001.
- [31] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes " in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Minneapolis, MN, Apr. 1993, pp. 269 - 272.
 - [32] N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Processing*, vol. 10, pp. 573-583, Apr. 2001.
 - [33] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Processing*, vol. 13, pp. 1327-1344, Oct. 2004.
 - [34] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," *IEEE Conf. Comp. Vis. Pattern Recognit.*, vol. 1, pp. I-645 - I-650, 2001.
 - [35] N. Nguyen, P. Milanfar, and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement " *IEEE Trans. Image Process*, vol. 10, pp. 1299 - 1308 2001.
 - [36] S. Oraintara, W. C. Karl, D. A. Castanon, and T. Q. Nguyen, "A method for choosing the regularization parameter in generalized Tikhonov regularized linear inverse problems," in *IEEE Int. Conf. Image Processing*, 2000, pp. 93-96.
 - [37] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Image Processing*, vol. 1, pp. 322-336, Jul. 1992.
 - [38] H. He and L. P. Kondi, "Resolution enhancement of video sequences with simultaneous estimation of the regularization parameter," *J. Electron. Imag.*, vol. 13, pp. 586-596, Jul. 2004.
 - [39] E. S. Lee and M. G. Kang, "Regularized Adaptive High-Resolution Image Reconstruction Considering Inaccurate Subpixel Registration," *IEEE Trans. Image Processing*, vol. 12, pp. 826 -837, Jul. 2003.
 - [40] M. G. Kang and A. K. Katsaggelos, "Simultaneous multichannel image restoration and estimation of the regularization parameters," *IEEE Trans. Image Processing*, vol. 6, pp. 774-778, May 1997.
 - [41] H. He and L. P. Kondi, "An image super-resolution algorithm for different error levels per frame," *IEEE Trans. Image Processing*, vol. 15, pp. 592-603, Mar. 2006.
 - [42] N. K. Bose, S. Lertrattanapanich, and J. Koo, "Advances in superresolution using L-curve," in *Int. Symp. Circuits and Systems*, 2001, pp. 433-436.
 - [43] M. C. Hong, M. G. Kang, and A. K. Katsaggelos, "A regularized multi-channel restoration approach for globally optimal high resolution video sequence," in *SPIE VCIP*, San Jose, CA, Feb. 1997, pp. 1306-1317.
 - [44] M. C. Hong, M. G. Kang, and A. K. Katsaggelos, "An iterative weighted regularized algorithm for improving the resolution of video sequence," in *IEEE Int. Conf. Image Processing*, 1997, pp. 474-477.
 - [45] M. G. Kang, "Generalized multichannel image deconvolution approach and its applications," *Opt. Eng.*, vol. 37, pp. 2953-2964, Nov. 1998.

- [46] R. C. Hardie, K. J. Barnard, J. G. Bogner, E. E. Armstrong, and E. A. Watson, "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," *Opt. Eng.*, vol. 37, pp. 247-260, Jan. 1998.
- [47] G. Aubert and L. Vese, "A Variational Method in Image Recovery," *SIAM J. Numer. Anal.*, vol. 34, pp. 1948-1979, 1997.
- [48] M. Black and A. Rangarajan, "On the Unification of Line Processes, Outlier Rejection, and Robust statistics with Applications to early vision," *Int. J. Comput. Vision*, vol. 19, pp. 57-91, 1996.
- [49] C. Bouman and K. Sauer, "A Generalized Gaussian Image Model for Edge-preserving MAP Estimation," *IEEE Trans. Image Process*, vol. 2, pp. 296-310, Jul. 1993.
- [50] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic Edge-preserving Regularization in Computer Imaging," *IEEE Trans. Image Process*, vol. 6, pp. 298-311, Feb. 1997.
- [51] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled image," *IEEE Trans. Image Processing*, vol. 6, pp. 1621-1633, Dec. 1997.
- [52] R. R. Schultz and R. L. Stevenson, "A bayesian approach to image expansion for improved definition," *IEEE Trans. Image Processing*, vol. 3, pp. 233-242, May 1994.
- [53] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequence," *IEEE Trans. Image Processing*, vol. 5, pp. 996-1011, Jun. 1996.
- [54] R. R. Schultz and R. L. Stevenson, "Improved definition video frame enhancement," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, pp. 2169-2171, May 1995.
- [55] L. P. Kondi, D. Scribner, and J. Schuler, "A comparison of digital image resolution enhancement techniques," in *Proc. SPIE AeroSense Conf.*, 2002, pp. 220-229.
- [56] H. He and L. P. Kondi, "Resolution enhancement of video sequences with simultaneous estimation of the regularization parameter," *Proc. SPIE Electron. Imag.*, vol. 5022, pp. 1123-1133, Jan. 2003.
- [57] H. He and L. P. Kondi, "MAP based resolution enhancement of video sequences using a Huber Markov random field image prior model," in *IEEE Int. Conf. Image Processing*, 2003, pp. 933-936.
- [58] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259-268, Nov. 1992.
- [59] C. Wang, P. Xue, and W. S. Lin, "Improved Super-Resolution Reconstruction from Video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, pp. 1411-1422, Nov. 2006.
- [60] Y. He, K. H. Yap, L. Chen, and L. P. Chao, "A Nonlinear Least Square Technique for Simultaneous Image Registration and Super-Resolution," *IEEE Trans. Image Process.*, vol. 16, pp. 2830-2841, Nov. 2007.
- [61] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 1167-1183, Sep. 2002.

- [62] E. Haber and L. Tenorio, "Learning regularization functionals-a supervised training approach," *Inverse Problems*, vol. 19, pp. 611- 626, 2003.
- [63] S. C. Zhu and D. Mumford, "Prior learning and Gibbs reaction-diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, pp. 1236 -1250, 1997.
- [64] C. Delherm, J. M. Lavest, M. Dhome, and J. T. Lapreste, "Dense reconstruction by zooming," in *Fourth Eur. Conf. Compt. Vision*, Apr. 1996, pp. 427-438.
- [65] J. M. Lavest, G.Rives, and M. Dhome, "Three Dimensional Reconstruction by Zooming," *IEEE Trans. Robotics Autom.*, vol. 9, pp. 196-207, Apr. 1993.
- [66] J. Ma and S. I. Olsen, "Depth zoom from zooming," *J. Ameri. Opt. Soc.*, vol. 7, pp. 1883-1890, Oct. 1990.
- [67] D. Wilkes, S. J. Dickinson, and J. K. Tsotsos, "A quantitative analysis of view degeneracy and its use for active focal length control," in *IEEE Int. Conf. Computer Vision*, 1995, pp. 938-944.
- [68] R. Jin, Y. Qi, and A. Hauptmann, "A probabilistic model for camera zoom detection," *Proceedings of IEEE 16th International Conference on Pattern Recognition*, vol. 3, pp. 859–862, Aug. 2002.
- [69] Y. Seo and K. S. Hong, "Theory and practice on the self-calibration of a rotating and zooming camera from two views," *IEE Proc. Vision, Image and Signal Processing*, vol. 148, pp. 166 - 172 Jun. 2001.
- [70] M. V. Joshi, S. Chaudhuri, and R. Panuganti, "A learning-based method for image super-resolution from zoomed observations," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 35, pp. 527-537, Jun. 2005.
- [71] D. Rajan and S. Chaudhuri, "Simultaneous Estimation of Super-Resolved Scene and Depth Map from Low Resolution Defocused Observations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, pp. 1102-1117, 2003.
- [72] X. Li, "Super-Resolution for Synthetic Zooming," *J. on Applied Signal Processing*, vol. 2006, pp. 1-12, 2006.
- [73] M. K. Ng, J. Koo, and N. K. Bose, "Constrained total least-squares computations for high-resolution image reconstruction with multisensors," *J. Imag. Sci. Technol.*, vol. 12, pp. 35-42, Jan. 2002.
- [74] N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Trans. Image Process*, vol. 15, pp. 201-213, Jan. 2006.
- [75] M. Park, E. Lee, J. Park, M. G. Kang, and J. Kim, "DCT-based high-resolution image reconstruction considering the inaccurate sub-pixel motion information," *SPIE Optical Engineering*, vol. 41, pp. 370-380, Feb. 2002.
- [76] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the non-local-means to super-resolution reconstruction," *IEEE Trans. Image Processing*, vol. 16, pp. 36-51, Jan. Jan. 2009.
- [77] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Image and video super-resolution via spatially adaptive block-matching filtering," in *Int.Workshop on Local and Non-Local Approximation in Image Processing* Lausanne, Switzerland, Aug. 2008.

- [78] H. Takeda, P. Milanfar, M. Protter, and M. Elad, "Super-resolution without explicit subpixel motion estimation," *IEEE Trans. Image Processing*, vol. 18, pp. 1958-1975, Sep. 2009.
- [79] K. Dabov, A. Foi, V. Katkovnic, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, pp. 2080-2095, Aug. 2007.
- [80] N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "EM-based simultaneous registration, restoration, and interpolation of super-resolved images," in *IEEE Int. Conf. Image Processing*, 2003, pp. 303-306.
- [81] W. Wirawan, P. Duhamel, and H. Maitre, "Multi-channel high resolution blind image restoration," in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing* 1999, pp. 3229-3232.
- [82] A. E. Yagle, "Blind superresolution from undersampled blurred measurements," in *Advanced signal processing algorithms, architectures, and implementations XIII* 2003, pp. 299-309.
- [83] F. Sroubek and J. Flusser, "Resolution enhancement via probabilistic deconvolution of multiple degraded images," *Pattern Recognition Letters*, vol. 27, pp. 287-293, Mar. 2006.
- [84] F. Sroubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *IEEE Trans. Image Process.*, vol. 14, pp. 874-883, Jul. 2005.
- [85] T. Komatsu, T. Igarashi, K. Aizawa, and T. Saito, "Very high resolution imaging scheme with multiple different-aperture cameras," *Signal Processing: Image Commun.*, vol. 5, pp. 511-526, Dec. 1993.
- [86] M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda, "Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames," *IEEE Trans. Instrum. Meas.*, vol. 49, pp. 915-923, Oct. 2000.
- [87] N. R. Shah and A. Zakhor, "Resolution enhancement of color video sequences," *IEEE Trans. Image Processing*, vol. 8, pp. 879-885, Jun. 1999.
- [88] N. Nguyen and P. Milanfar, "An efficient wavelet-based algorithm for image superresolution," *Proc. Int. Conf. Image Processing*, vol. 2, pp. 351-354, 2000.
- [89] M. Elad and A. Feuer, "Superresolution restoration of an image sequence: adaptive filtering approach," *IEEE Trans. Image Processing*, vol. 8, pp. 387-395, Mar. 1999.
- [90] M. Elad and A. Feuer, "Super-resolution reconstruction of image sequences," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 817-834, Sept. 1999.
- [91] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 1153-1160, Dec. 1981.
- [92] H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 508-517, Dec. 1978.
- [93] X. Li and M. T. Orchard, "New Edge-Directed Interpolation," *IEEE Trans. Image Process.*, vol. 10, pp. 1521-1527, Oct. 2001.

- [94] L. Zhang and X. Wu, "Image interpolation via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, pp. 2226-2238, Aug. 2006.
- [95] X. Zhang and X. Wu, "Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Processing*, vol. 17, pp. 887-896, Jun. 2008.
- [96] W. T. Freeman and E. C. Pasztor, "Makov Networks for Superresolution," in *34th Ann. Conf. Information Sciences and Systems*, Dept. Electrical Eng., Princeton Univ., 2000.
- [97] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-Based Super-Resolution," MERL, Mitsubishi Electric Research Labs., 201 Broadway Cambridge, MA 02139 TR-2001-30, Aug. 2001.
- [98] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, pp. 2323-2326, 22 Dec. 2000.
- [99] H. Stark and P. Oskoui, "High resolution image recovery from image- plane arrays, using convex projections," *J. Opt. Soc. Am. A*, vol. 6, pp. 1715-1726, 1989.
- [100] P. Vandewalle, S. Susstrunk, and A. Vetterli, "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution," *EURASIP J. Applied Signal Processing*, 2006.
- [101] L. Chen and K.-H. Yap, "Efficient discrete spatial techniques for blur support identification in blind image deconvolution," *IEEE Trans. Signal Processing*, vol. 54, pp. 1557-1562, April 2006.
- [102] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE Trans. Image Process.*, vol. 6, pp. 1621-1633, Dec. 1997.
- [103] S. Chaudhuri and J. Manjunath, *Motion-free Super-resolution*. NY: Springer Science+Business Media, 2005.
- [104] Z. Lin and H. Y. Shum, "Fundamental limits on reconstruction-based superresolution algorithms under local translation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, pp. 83-97, Jan. 2004.
- [105] S. H. Bae, T. N. Pappas, and B. H. Juang, "Subjective Evaluation of Spatial Resolution and Quantization Noise Tradeoffs " *IEEE Trans. Image Process*, vol. 18, pp. 495 - 508 Mar. 2009.
- [106] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Trans. Image Processing*, vol. 13, pp. 898-911, Jul. 2004.
- [107] L. D. Alvarez, J. Mateos, R. Molina, and A. K. Katsaggelos, "High resolution images from compressed low resolution video: Motion estimation and observable pixels," *Int. J. Imaging Syst. Technol.*, vol. 14, pp. 58-66, Aug. 2004.
- [108] F. Brandi, Q. R.D, and D. Mukherjee, "Super resolution of video using key frames," in *IEEE Int. Symp. Circuits and Systems(ISCAS)*, 18-21 May 2008.
- [109] X. J. Zhang and X. L. Wu, " Can Lower Resolution Be Better?," in *Data Compression Conf. Snowbird*, UT, Mar. 2008, pp. 302-311.
- [110] H. A. Ilgin and L. F. Chaparro, "Low Bit Rate Video Coding Using DCT Based Fast Decimation/Interpolation and Embedded Zerotree Coding," *IEEE*

- Trans. Circuits and Systems for Video Technology*, vol. 17, pp. 833-844, Jul. 2007.
- [111] V. A. Nguyen, Y. P. Tan, and W. S. Lin, "Adaptive downsampling/upsampling for better video compression at low bit rate," in *IEEE International Symposium on Circuits and Systems (ISCAS)* May 2008.
 - [112] D. Barreto, L. D. Alvarez, R. Molina, A. K. Katsaggelos, and G. M. Callicó, "Region-Based Super-Resolution for Compression," *Multidimensional Systems and Signal Processing*, vol. 18, pp. 59-81, Sep. 2007.
 - [113] A. M. Bruckstein, M. Elad, and R. Kimmel, "Down-Scaling for Better Transform Compression," *IEEE Trans. Image Processing*, vol. 12, pp. 1132-1144, Sep. 2003.
 - [114] J. P. Curzan, C. R. Baxter, and M. A. Massi, "Variable acuity imager with dynamically steerable, programmable superpixels," in *SPIE Conf. Aerosense Infrared Technology and Applications*, 2002.
 - [115] S. Banerjee, "Low Power Content-Based Video Acquisition for Super-resolution Enhancement," *IEEE Trans. Multimedia*, vol. 11, pp. 455-464, Apr. 2009.
 - [116] H. Nagahara, A. Hoshikawa, T. Shigemoto, Y. Iwai, M. Yachida, and H. Tanaka, "Dual-Sensor Camera for Acquiring Image Sequences with Different Spatio-temporal Resolution," in *IEEE Conf. Advanced Video and Signal Based Surveillance*, Sept. 2005, pp. 15-16.
 - [117] M. M. Shen, P. Xue, and C. Wang, "A novel scalable video coding scheme using super resolution techniques," in *IEEE workshop, Multimedia Signal Processing (MMSP)*, 8-10 Oct. 2008, pp. 196-199.
 - [118] A. S. Ogale and Y. Aloimonos, "A roadmap to the integration of early visual modules," *International Journal of Computer Vision*, vol. 72, pp. 9 -25, April 2007.
 - [119] W. Lu and Y. P. Tan, "Color filter array demosaicking: new method and performance measures," *IEEE Trans. Image Process*, vol. 12, pp. 1194-1210, Oct. 2003.
 - [120] T. Gotoh and M. Okutomi, "Direct super-resolution and registration using raw CFA images," in *Int. Conf. Computer Vision and Pattern Recognition*, Washington, DC, Jul. 2004.
 - [121] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Trans. Image Process*, vol. 15, pp. 141-159, Jan. 2006.