Robust Multiframe Super-Resolution with Adaptive Norm Choice Using Difference Curvature Based BTV Regularization

by

Xiaohong Liu

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements for the M.A.Sc. degree in
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

Multi-frame image super-resolution focuses on reconstructing a high-resolution image from a set of low-resolution images with high similarity. Since super-resolution is an ill-posted problem, regularization techniques are widely used to constrain the minimization function. Combining image prior knowledge with fidelity model, Bayesian-based methods can effectively solve this ill-posed problem, which makes this kind of methods more popular than other methods.

Our proposed model is based on maximum a posteriori probability (MAP) estimation. In this thesis, we propose a novel initialization method based on median operator to initialize our estimated high-resolution image. For the fidelity term in our proposed algorithm, the half-quadratic estimation is used to choose error norm adaptively instead of using fixed L_1 or L_2 norm. Furthermore, for our regularization term, we propose a novel regularization method based on Difference Curvature (DC) and Bilateral Total Variation (BTV) to suppress mixed noises and preserve image edges simultaneously. In our experimental results, synthetic data and real data are both tested to demonstrate the superiority of our proposed method in terms of clearer texture and less noise over other state-of-the-art methods.

Acknowledgements

I would like to express my gratitude to my supervisor Prof. Jiying Zhao for his full support, expert guidance, understanding and encouragement throughout my study and research. Without his incredible patience and timely wisdom and counsel, my thesis work would have been frustrating and overwhelming pursuit.

I would like to thank all my lab colleagues, especially Dr. Wenyi Wang, Lei Chen and Jun Hu for helping me with my academic research.

Finally, I would like to thank my girlfriend Dan Zhou, my parents and brother for their unconditional love and support during the last two years. I would not have been able to complete this thesis without their continuous love and encouragement.

This thesis is dedicated to my family and the perfect girl who stays with me even if across the whole world.

Table of contents

List of tables

List of figures

Nomenclature

Abbreviations

BEP	Bilateral Edge-Preserving
BTV	Bilateral Total Variation
CG	Conjugate Gradient
DC	Difference Curvature
DCBTV	Difference Curvature Based Bilateral Total Variation
DFT	Discrete Fourier Transform
EM	Expectation-Maximization
FGWs	First-Generation Wavelets
GMRF	Gaussian Markov Random Field
HR	High-Resolution
IBP	Iterative Back-Projection
IRW	Iteratively Re-Weighted Method
LR	Low-Resolution
LSI	Linear Space Invariant
LSV	Linear Space Variant

MAP	Maximum a Posteriori Probability
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
MRF	Markov Random Fields
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NL	Non-Local
PAR	Piecewise Auto-Regressive
PDF	Probability Density Function
POCS	Projection onto Convex Sets
PSF	Point Spread Function
PSNR	Peak Signal-to-Noise Ratio
PVE	Partial Volume Effect
RS	Remote Sensing
RGB	A Cartesian coordinate notation by defining R, G, and B as three primary colour lights – Red, Green, and Blue respectively
SAR	Simultaneous Autoregression
SCG	Scaled Conjugate Gradient
SD	Steepest Decent
SGWs	Second-Generation Wavelets
SR	Super-Resolution
SSIM	Structural Similarity
TV	Total Variation

Chapter 1

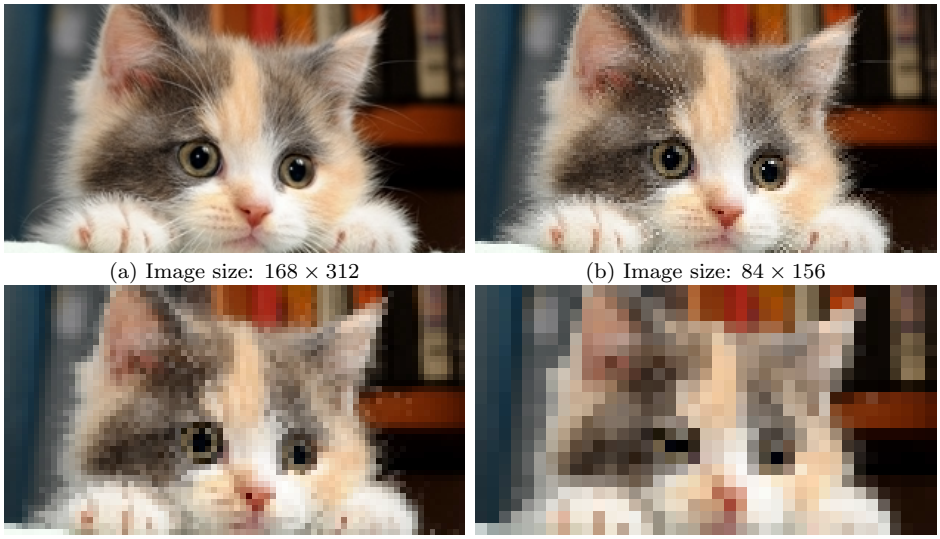
Introduction

Super-resolution (SR) is a method to increase the image resolution without modifying the sensor of camera. Different from single image super-resolution, multi-frame super-resolution focuses on reconstructing a high-resolution image from a set of similar low-resolution images. In 1984, Tsai *et al.* first proposed a method to solve this problem based on frequency domain [1], which is easy to implement and computationally efficient. However, frequency domain based multi-frame SR methods usually have serious visual artifacts. Due to this drawback, the spatial domain algorithms become more and more popular [2, 3] in recent years. Since super-resolution is an ill-posted problem, regularization techniques are widely used to constrain the minimization function. Bayesian-based methods can easily combine regularization techniques with image observation models, which makes this kind of methods have better performance than others.

1.1 Image super-resolution

In digital cameras, a pixel is captured by a corresponding detector sensor which is usually based on CCD (Charge-Coupled Device) and CMOS (Complementary Metal-Oxide Semiconductor). In general, to increase pixels of an image, hardware-based methods and software-based methods are usually used. For hardware-based methods, they directly increase the size of CCD array of cameras so that more pixels will be captured while taking photos. However, in practical application, reducing the sensor pixel size to increase the resolution of acquired photo will decrease the light intensity for each pixel [4, 5]. And the lower size of CCD increases the manufacturing difficulties as well. Due to the drawbacks of hardware based methods, many software methods are proposed as an effective and economical way to improve image quality.

Image super-resolution (SR) technique is one of the software-based methods, which has been studied over the last two decades. The goal of image super-resolution is to reconstruct a high-resolution (HR) image from a low-resolution (LR) image or a LR sequence. According to the number of input images, super-resolution methods can be sorted as single-image super-resolution, multi-frame image super-resolution and video super-resolution. For single-image super resolution, only one LR image is used as the reference to reconstruct the corresponding HR image. Different from single-image super-resolution, multi-frame super-resolution enhances an image from some LR frames with high similarity. And video super-resolution can be regarded as an iterative processing of multi-frame super-resolution. Figure 1.1 shows the same scene image with different resolutions. Compared with low-resolution images, the high-resolution one has adequate detail information to give people a better viewing experience.



(a) Image size: 168×312 (b) Image size: 84×156

(c) Image size: 42×78 (d) Image size: 21×39

Figure 1.1: The original image and its down-sampling images.

1.2 Multi-frame super-resolution

In this thesis, the problem of multi-frame super-resolution is addressed. Multi-frame super resolution is a technique which reconstructs a high-resolution image from a set of similar low-resolution images without increasing the size of pixel array. Due to this property, multi-frame super-resolution technique is usually used in fields of satellite Remote Sensing (RS), security surveillance and medical imaging. For medical imaging, one example is the Magnetic Resonance Imaging (MRI). Due to the Partial Volume Effect (PVE), the small objects or regions are lost because of the limited image resolution [6, 7]. In general, the differences between two low-resolution images are caused by sub-pixel movement, different focus points and blurring effect. In practical applications, the irregular pixel movement and unknown blurring directly influence the super-resolved result and limit the performance of simple observation model [8]. Figure 1.2 shows an example of recovering a high-resolution image from a

related low-resolution sequence by using our proposed algorithm. From this example, the super-resolved result can capture more detail information than any one of the input low-resolution images.

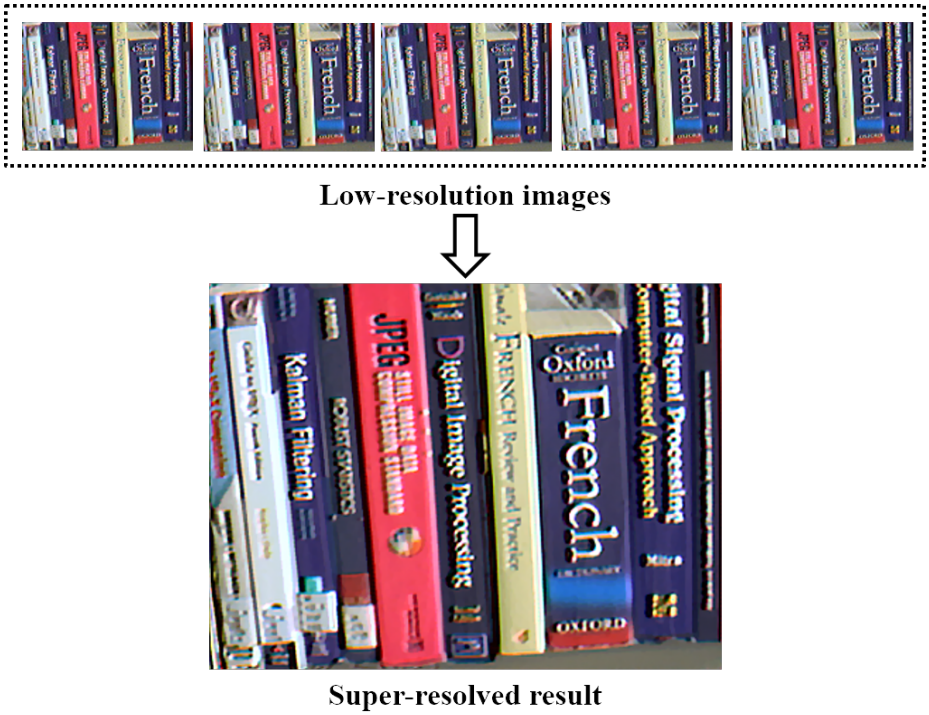


Figure 1.2: An example of multi-frame super-resolution for real data.

Different from single-image super-resolution, multi-frame super-resolution has more detail information hiding in low-resolution sequences. Therefore, how to accurately extract the image texture existed in different LR frames is a vital challenge to reconstruct an HR image with good quality. In this case, image registration and blur identification should be taken into consideration. Image registration is used to estimate the displacements among LR images, which directly influences the quality of final output. For blur identification, point-spread function is usually used to model

the blur kernel. Moreover, for the reconstruction step, if the low-resolution images have non-redundant information, the ill-posed nature of SR problem can be over-determined by adding more low-resolution images to the objective function. The pixels in low-resolution images can be aligned on a high-resolution grid according to the sub-pixel shifting with respect to the reference low-resolution image. Figure 1.3 illustrates this idea to reconstruct a high-resolution image from a set of related low-resolution images.

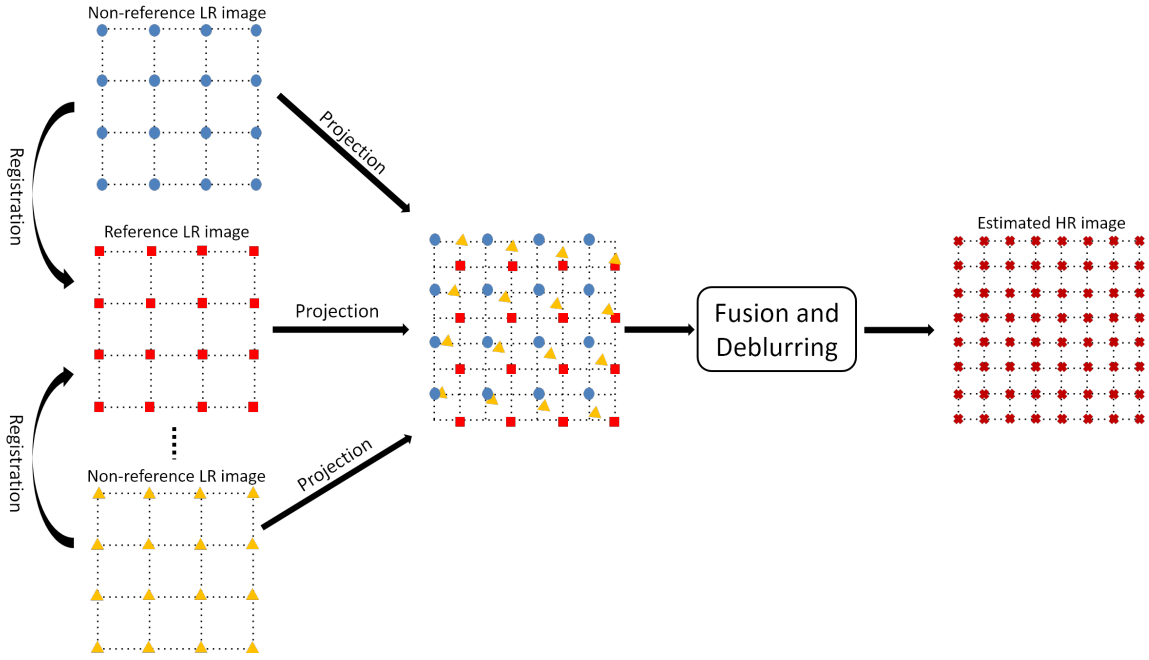


Figure 1.3: An example of recovering a high-resolution image from a sequence of non-redundant low-resolution images.

However, in practical applications, the low-resolution images are not always non-redundant, which limits the performance of this simple image reconstruction model. Besides, the blur kernel of a camera is usually unknown, which makes the multi-frame image super-resolution more challenging.

1.3 Thesis contribution

In our proposed multi-frame super-resolution algorithm, there are two main contributions to effectively improve the quality of final estimated HR image.

The first contribution is proposing a novel initial method based on median operator. The initializations of other traditional multi-frame SR methods are directly generated by bilinear or bicubic interpolations from the reference LR frame. In this case, the detail information in non-reference LR frames is not considered, which limits the quality of the initial high-resolution image. Different from the traditional methods, our proposed initialization uses the image registration technique to align LR images referenced by the target LR image. After alignment, the median operator is used to generate a composed LR image which is outlier-insensitive. Finally, we use the bicubic interpolation to obtain our initial HR image. Compared with the other traditional initial algorithms, our method can eliminate most of the outliers existing in LR images, which makes our initial HR image have better visual quality and objective quality.

The second contribution in this thesis is proposing a novel regularization method based on Difference Curvature (DC). The normal regularization methods suppress image noise and edges simultaneously. Due to the drawbacks of traditional regularization methods, an adaptive matrix W_D is used in our proposed regularization method. Each element in W_D represents a weight to adjust the relevant value in the Bilateral Total Variation (BTV) regularization. Therefore, the performance of the BTV regularization is controlled by this adaptive matrix based on difference curvature algorithms. The experimental results demonstrate that our proposed method suppresses more image noise and has clearer edges with the comparison of other

state-of-the-art methods.

1.4 Thesis structure

In this thesis, we focus on multi-frame image super-resolution and propose our novel method. In our proposed framework, a novel initialization method is used based on median operator to obtain a good initial high-resolution image. The fidelity term in our method is based on half-quadratic estimation, which adaptively determines the error norm. Moreover, we propose a difference curvature based BTV regularization to regularize our objective function. Compared with some existing state-of-the-art multi-frame super-resolution methods, the results of our proposed algorithm have better quality in visual comparison and higher values in image assessment matrices. The rest of the thesis is organized as follows. Chapter 2 introduces the basic concepts and ideas in multi-frame super-resolution field. Chapter 3 illustrates some specific multi-frame super-resolution algorithms in detail. Chapter 4 presents our proposed multi-frame image super-resolution algorithm. Chapter 5 demonstrates our experimental results and Chapter 6 concludes the thesis.

Chapter 2

Related work

2.1 Observation model

Observation model formulates the relationship between a high-resolution frame and a sequence of low-resolution frames with high similarity. Therefore an accurate observation model is vital for any multi-frame super-resolution algorithm. According to the study of camera sensor, some assumptions have been made to describe the observation model, which directly affect the performance of the final result.

In general, low-resolution frames can be regarded as the related high-resolution frame going through the geometric motion operator, blurring operator, down-sampling operator and noise adding operator successively. The geometric motion operator represents the displacement between two images caused by camera motion. The simplest motion in image sequences is translation. It moves image pixels in horizontal and vertical directions according to a motion vector. However, in practical applications, the motion operator model should also consider affine motion and non-rigid motion that

make the model hard to establish.

The blurring operators describe the impact of motion blur, out-of-focus blur and sensor blur which obeys the standard Gaussian distribution in most situations. Each of them can be considered as a Point Spread Function (PSF) to represent the response of optical system. The PSF kernel is usually modeled as Linear Space Invariant (LSI) or Linear Space Variant (LSV) and represents the spatial averaging between the high-resolution image and the corresponding low-resolution image.

Figure 2.1 shows a low-resoluton sensor PSF based on spatial averaging with the down-sampling factor of 2. H_i represent four high-resolution pixels where $i = 0, 1, 2, 3$ and L_0 stands for corresponding low-resolution pixel. As shown in the figure, the value of L_0 is the weighted sum of high-resolution pixels, which equals to $\frac{\sum H_i}{4}, i = 0, 1, 2, 3$.

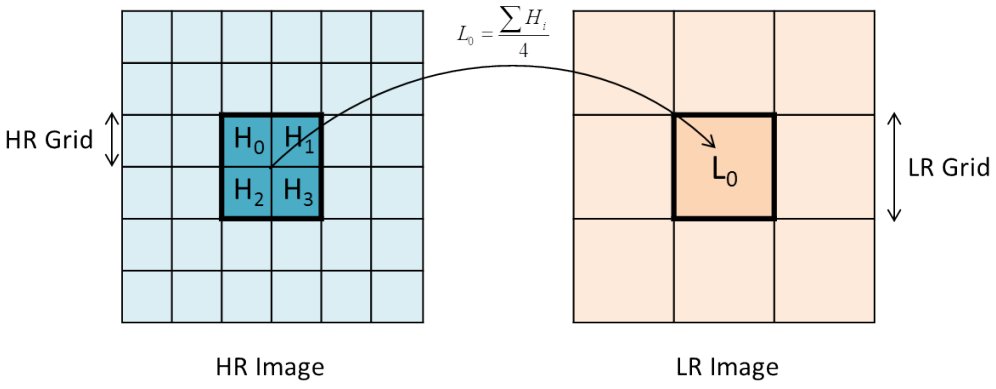


Figure 2.1: A low-resolution sensor PSF based on spatial averaging.

In the field of multi-frame super-resolution, the PSF kernel is usually assumed to be known. Nowadays, unknown PSF kernels are considered in some blind image super-resolution methods, which makes this problem more challenging. Figure 2.2

shows the *Butterfly* images blurred by Gaussian blur, motion blur and out-of-focus blur respectively. The lower left corner of every figure shows the corresponding PSF kernel. As shown below, the shape and size of PSF kernels determine the blurring degree.

(a) Original image (b) With Gaussian blur (c) With motion blur (d) With out-of-focus blur

Figure 2.2: *Butterfly* image with different blurring kernels.

In addition, adding noise in generated low-resolution sequences is a very vital part to simulate the electronic noise of digital camera. According to their properties, noises can be classified as Gaussian noise, *Salt&Pepper* noise and so on. Gaussian noise caused by taking a photo in poor illumination or high temperature condition is the most commonly considered noise in image processing. *Salt&Pepper* noise is another widely used noise caused by broken sensors or transition errors. Different from some denoising methods which consider high derivation of the noise, multi-frame super-resolution methods use nature images which have slight noise in general. Figure 2.3 shows the *Butterfly* images with Gaussian noise and *Salt&Pepper* noise.

To establish a good observation model, all degradative operators introduced above should be considered. The expression is formulated as

$$Y_k = DB_kM_kX + n_k, \tag{2.1}$$

(a) Original image (b) With Gaussian noise (c) With *Salt&Pepper* noise

Figure 2.3: *Butterfly* image with different noises.

where X is the HR frame and expressed in lexicographic order as $X = [x_1, x_2, ..., x_N]^T$, where N is the total number of pixels in HR frame which equals to $rm \times rn$ and r is the downsampling factor. Therefore, the size of X is $rmrn \times 1$. Similar to the definition of X, $Y_k = [y_{k,1}, y_{k,2}, ..., y_{k,L}]^T$, which represents the kth LR frame with the size of $mn \times 1$, where $k = 1, 2, ..., K$. K is the number of LR frames and $L = m \times n$ is the Lth pixels of the LR frame. M_k represents the geometric motion matrix between HR frame and kth LR frame with the size of $rmrn \times rmrn$. B_k is the blur matrix for the kth LR frame with the size of $rmrn \times rmrn$ and D is the downsampling matrix with the size of $mn \times rmrn$. For practical situation, the noise should be taken into consideration. n_k represents the noise added into the kth LR frame with the size of $mn \times 1$. Figure 2.4 shows a general structure of the observation model.

In order to simplify the Equ. (2.1), DB_kM_k can be regarded as a system matrix $W^k \in \mathbb{R}^{L \times N}$ proposed in [9]. As mentioned above, each low resolution image pixel can be generated by a weighted sum of the corresponding pixels in the high resolution image. In W^k, each row represents one of the low resolution pixels in lexicographic order. Similar to the definition of each row, each column of W^k represents one of the high resolution pixels in lexicographic order as well. For each low resolution pixel, the

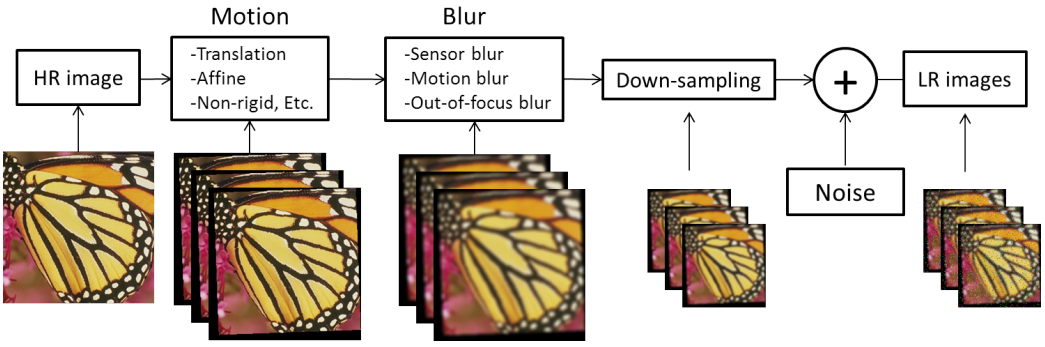


Figure 2.4: A general observation model for multi-frame super-resolution.

W^k has non-zero value in the location of related high resolution pixels. That makes W^k is a sparse matrix because the out of kernel pixels should not have significant weights for the center pixel.

For the simple pixel displacement, the W^k can be formalized with an isotropic 2D Gaussian PSF as

$$W_{ij}^k = \frac{\bar{W}_{ij}^k}{\sum_j' \bar{W}_{ij'}^k}, \tag{2.2}$$

$$\bar{W}_{ij}^k = \exp\left\{-\frac{\|v_j - u_i'\|_2^2}{2\sigma^2}\right\}, \tag{2.3}$$

where u_i' is the center kernel pixel of the corresponding high resolution image. And v_j is one of the pixels surrounding u_i' in the same kernel. The σ is the variance of the point spread function. This variance is measured in terms of low resolution pixels because for high resolution pixels, different zoom factor will affect the measurement

of PSF variance [10]. Thus, u_i' can be formulated as follows:

$$u_i' = H(\theta)u_i, \tag{2.4}$$

where u_i is the center kernel pixel of the corresponding low resolution image. H is the homographic motion matrix with the size of 3×3 parameterized by vector θ. And u_i and u_i' are in homogeneous coordinates rather than Cartesian coordinates. Equ. (2.4) represents the mapping relationship between the low resolution images and the estimated high resolution image.

By combining the different transformations as a united system matrix W^k, the Equ. (2.1) can be rewritten as follows:

$$Y_k = W^k X + n_k, \tag{2.5}$$

which simplifies the original expression and makes the observation model clearer.

2.2 The basic framework of multi-frame SR

The basic framework of multi-frame super-resolution contains fidelity term and regularization term. Fidelity term is used to keep the fidelity between the HR frame and the corresponding LR frames. Regularization term aims at regularizing the minimization function. For the fidelity term, M-estimation is usually used to minimize the observation residual. The definition of M-estimation has been written as follows:

$$\hat{X} = arg\min_X \sum_{k=1}^{K} \rho(DB_k M_k X - Y_k), \tag{2.6}$$

and the implicit equation of M-estimation is:

$$\sum_{k=1}^{K} \psi(DB_k M_k X - Y_k) = 0, \tag{2.7}$$

where ρ measures the error between the estimated HR frame and given LR frames. And $\psi(DB_k M_k X - Y_k) = \frac{\partial}{\partial X}\rho(DB_k M_k X - Y_k)$ is the derivative function of ρ with respect to X. If function ρ uses Lp norm for the residual, the M-estimation can be written as

$$\hat{X} = arg\min_{X} \sum_{k=1}^{K} \|DB_k M_k X - Y_k\|_p^p, \tag{2.8}$$

The regularization term is used to constrain the model so that the reconstructed image will easily reach a robust state. The traditional framework of multi-frame super-resolution can be expressed as follows:

$$\hat{X} = arg\min_{X} \left\{ \sum_{k=1}^{K} \|DB_k M_k X - Y_k\|_p^p + \lambda \Upsilon(X) \right\}, \tag{2.9}$$

where $\Upsilon(X)$ is the regularization term with respect to X. λ is the trade-off parameter between these two terms and p represents the choice of Lp norm.

Since the noise in observation model usually obeys the Gaussian distribution, choosing L_2 norm for fidelity term can obtain a good super-resolved result. But in practical applications, the observation model suffers various noise and errors due to inaccurate estimation of registration and blurring kernel. Farsiu *et al.* firstly used L_1 norm rather than L_2 norm in fidelity term and achieved better results than L_2 norm [11].

However, although the L_1 norm is robust for outliers, it may introduce more observation errors than L_2 norm when the estimation of images is accurate. The drawbacks of fixed norms motivated researchers to combine the advantages of L_1 and L_2 norms. Nowadays, M-estimators such as Huber function [12], Leclerc function and Lorentzian function have been proposed as the replacements of fixed norms. These robust M-estimators behave like L_2 norm when the observation errors are small and maintain the properties of L_1 norm to suppress the large errors such as outliers. Besides, robust M-estimators can be derivable at every point. Therefore, the traditional optimization methods such as steepest decent or conjugate gradient guarantee the convergence of the minimization function [12].

Huber estimator is one of the most widely used estimators which are less sensitive than quadratic estimator in outliers [13]. The expression of Huber estimator is formulated as follows:

$$\rho_{Huber}(x,T) = \begin{cases} 0.5x^2, & |x| \leq T \\ T(|x| - 0.5T), & |x| > T. \end{cases} \tag{2.10}$$

where x represents the observation errors and T is the threshold at which the function changes from quadratic to linear. For small observation errors, the Huber function behaves like quadratic function. However, if the parameter x is beyond the threshold, the Huber function performs like linear to decrease the penalty for large errors. The derivative of Huber function is formulated as

$$\psi_{Huber}(x,T) = \begin{cases} x, & |x| \leq T \\ T \cdot sign(x), & |x| > T. \end{cases} \tag{2.11}$$

where the *sign* function represents the sign of variable x. The derivative of Huber function is linear when the observation errors are small. For the errors larger than threshold T, the Huber function is constant to suppress the outliers.

Figure 2.5 compares the Huber estimator and its derivative function with L_1 and L_2 norm. The threshold T is set to 1.

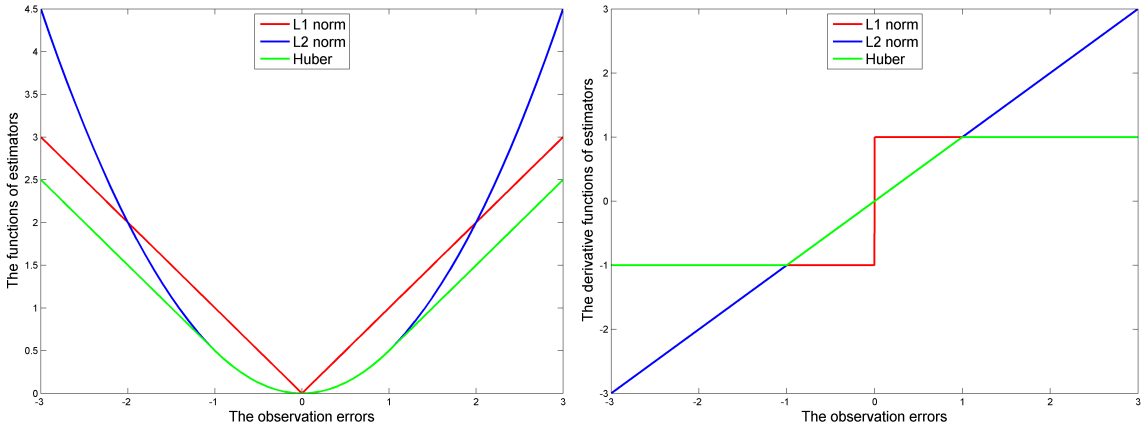


Figure 2.5: Comparing Huber estimator and its derivative function with L_1 and L_2 norm.

As shown above, the green line represents the derivative of Huber estimator. When the observation errors are small, it behaves like L_2 norm to decrease the penalty of image edges. For large errors, Huber estimator is similar to L_1 norm to suppress the large errors such as isolated noise and mis-registration.

The Leclerc estimator is expressed as follows:

$$\rho_{Lec}(x, T) = 1 - e^{-\frac{x^2}{T^2}}, \tag{2.12}$$

where T is the threshold to control the behavior of Leclerc estimator and the derivative

function of Leclerc estimator is expressed as

$$\psi_{Lec}(x,T) = \frac{2x}{T^2} e^{-\frac{x^2}{T^2}}, \tag{2.13}$$

Another popular robust M-estimator which needs to be introduced is Lorentzian estimator. The expression of Lorentzian estimator is

$$\rho_{Lor}(x,T) = \log\left(1 + \frac{x^2}{2T^2}\right), \tag{2.14}$$

and the derivative function of Lorentzian estimator is expressed as follows:

$$\psi_{Lor}(x,T) = \frac{2x}{2T^2 + x^2}, \tag{2.15}$$

Some estimators are not differentiable, which makes the iteration process numerically unstable. For example, $\rho(x) = |x|$ based on L_1 norm is not differentiable at zero point and Huber estimator expressed in Equ. (2.10) is not differentiable at threshold T since it's a piecewise function. Leclerc and Lorentzian estimators combine the advantages of L_1 norm and L_2 norm as well. However, different from Huber estimator, Leclerc and Lorentzian estimators are robust, differentiable and strictly convex. Therefore, simple optimization methods can also be used to optimize the minimization function.

Figure 2.6 compares the Leclerc and Lorentzian estimators and their derivative functions with L_1 norm and L_2 norm. The threshould T in Leclerc and Lorentzian is set to be 1 as well.

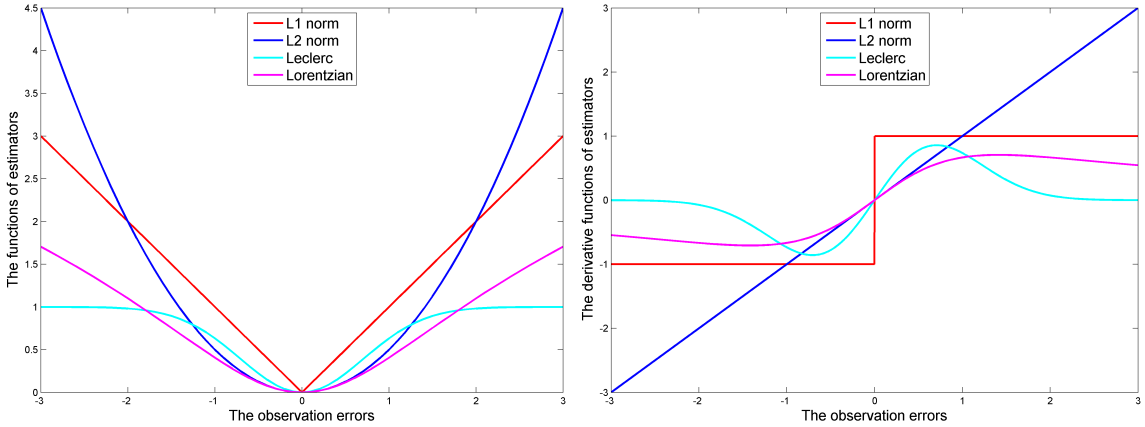


Figure 2.6: Comparing Leclerc and Lorentzian estimators and their derivative functions with L_1 and L_2 norm.

2.3 Regularization methods of super-resolution

Super-resolution is regarded as an ill-posed problem or an inverse problem [14]. Different from forward problems which start with the associated parameters and then calculate the related result, inverse problems start with the final results and then calculate the associated parameters. For example, the reconstruction of the observation model is a forward problem because it generates the low-resolution sequences from given high-resolution image. As a typical inverse problem, super-resolution methods infer the high-resolution image from the observed low-resolution images.

Super-resolution problem is not only an inverse problem but also an ill-posed problem which makes it harder to solve. The definition of ill-posed problem was first proposed by Hadamard [15]. For a well-posed problem, it should satisfy three properties which are 1) a solution exists 2) the solution is unique 3) the solution depends continuously on the initial conditions. A problem which is not a well-posed problem is called an ill-posed problem. For an ill-posed problem, since the number of

unknown information is more than the known information, the number of solutions are infinite. Therefore, the regularization methods are introduced to compensate the lost information and as a prior of the desired high-resolution image. Typically, the regularization terms involve additional assumptions such as gradient correlation or smooth matrix. In this section, some most commonly used regularizations will be introduced in detail.

2.3.1 Tikhonov regularization

Tikhonov regularization [14, 16, 17] is one of the most widely used regularizations. The Tikhonov cost function is expressed as follows:

$$\Upsilon_T(X) = \|\Gamma X\|_2^2, \tag{2.16}$$

where Γ is a high-pass filter such as Laplacian filter or simple identity matrix. Laplacian filter represents the second-order derivative of the image X. For simple identity matrix, Equ. (2.16) limits the total energy of X. Since Laplacian filter is a second-order derivative of the image X, it has good performance to suppress the high-frequency noise and artifacts. The Laplacian function for image X is defined as

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}, \tag{2.17}$$

for horizontal direction,

$$\frac{\partial^2 f}{\partial x^2} = f(x+1,y) + f(x-1,y) - 2f(x,y), \tag{2.18}$$

for vertical direction,

$$\frac{\partial^2 f}{\partial y^2} = f(x, y+1) + f(x, y-1) - 2f(x, y). \tag{2.19}$$

By substituting $\frac{\partial^2 f}{\partial x^2}$ and $\frac{\partial^2 f}{\partial y^2}$, the Laplacian function for image X is expressed as

$$\nabla^2 f(x, y) = f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) - 4f(x, y). \tag{2.20}$$

Besides, if the Equ. (2.20) is represented in terms of Laplacian blur kernel, the Γ in Equ. (2.16) can be formulated as

$$\Gamma_{kernel} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{2.21}$$

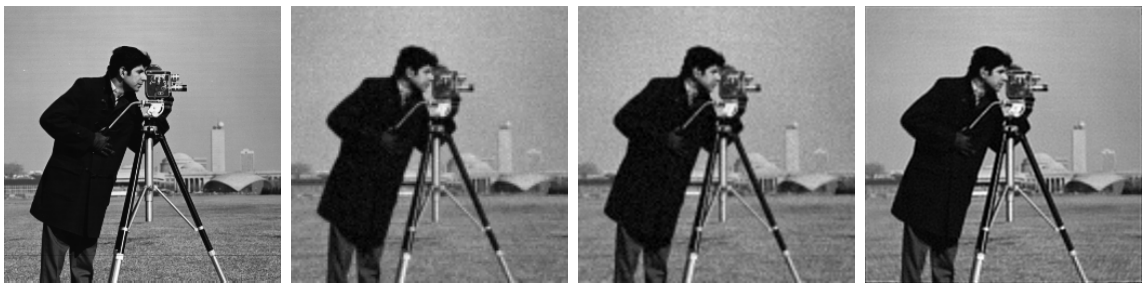
If the Laplacian kernel considers the pixels of 8-adjacent connections rather than 4-adjacent connections, the Laplacian function is expressed as

$$\begin{aligned} \nabla^2 f(x, y) = {} & f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) \\ & + f(x-1, y-1) + f(x+1, y-1) + f(x-1, y+1) \\ & + f(x+1, y+1) - 8f(x, y). \end{aligned} \tag{2.22}$$

The corresponding Laplacian blur kernel is represented as

$$\Gamma_{kernel} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{2.23}$$

Tikhonov regularization is easy to implement and has low computational complexity. It suppresses image noise by penalizing the high-frequency information. However, image edges also appear as high-frequency components. Therefore, the final result of Tikhonov regularization is hard to have sharp edges. Figure 2.7 shows the original image and the Tikhonov regularized image with the zoom factor of 2.



(a) Original HR image (b) LR image (c) Bicubic interpolation (d) Tikhonov regularized image

Figure 2.7: Super-resolution image with Tikhonov regularization.

2.3.2 Total Variation (TV) regularization

The Total Variation (TV) regularization model was first proposed by Rudin *et al.* [18] for image denoising and deblurring. Different from Tikhonov regularization, the total variation penalizes the total changes in the image so that it can preserve edges and detailed information of the estimated HR image. For the high-resolution image

X, the isotropic total variation model is defined as follows:

$$\Upsilon_{TV}^{iso}(X) = \sum_{i,j} \sqrt{|X_{i+1,j} - X_{i,j}|^2 + |X_{i,j+1} - X_{i,j}|^2}, \tag{2.24}$$

except for isotropic total variation model, the anisotropic model is sometimes used since it is easier to minimize. The expression of anisotropic TV regularization is

$$\begin{aligned} \Upsilon_{TV}^{aniso}(X) &= \sum_{i,j} \sqrt{|X_{i+1,j} - X_{i,j}|^2} + \sqrt{|X_{i,j+1} - X_{i,j}|^2} \\ &= \sum_{i,j} |X_{i+1,j} - X_{i,j}| + |X_{i,j+1} - X_{i,j}|. \end{aligned} \tag{2.25}$$

In general, the Total Variation regularization describes the relationship between the center pixel and its surrounding pixels and the final results of this regularization have clear edges. However, in some situations, it will introduce *staircase* artifacts in the flat region of the SR image. These artifacts can be suppressed by adjusting trade-off parameter λ, but the texture of the SR image will be blurred at the same time.

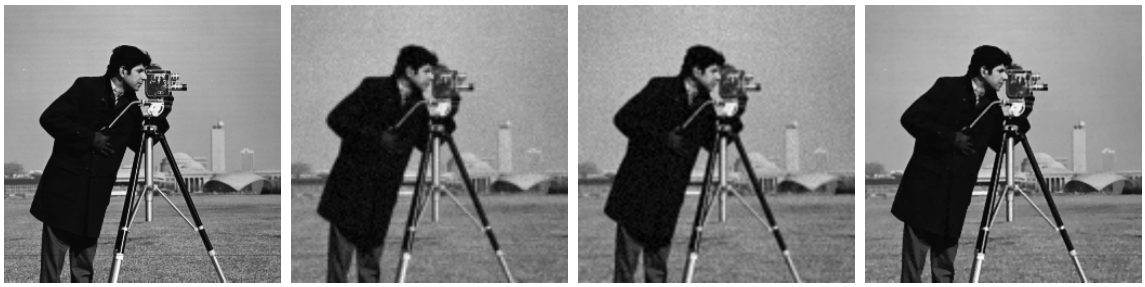
Based on the total variation regularization, Farsiu *et al.* [11] proposed a related method named the Bilateral Total Variation (BTV), which combines the bilateral filter and the TV regularization. BTV regularization is also computationally cheap and easy to implement. The formula of BTV regularization is expressed as follows:

$$\Upsilon_{BTV}(X) = \sum_{l=-P}^{P} \sum_{m=0}^{P} \alpha^{|m|+|l|} \left\| X - S_x^l S_y^m X \right\|_1, \tag{2.26}$$

where S_x^l shifts X by l pixels in horizontal direction and S_y^m shifts X by m pixels

in vertical direction. α is a scaled weight with the range of $0 < \alpha < 1$ and P is a control parameter. It controls the decaying effect to the summation of the BTV regularization.

Different from Tikhonov regularization that uses L_2 norm [14], BTV regularization term suppresses image noise by penalizing the first-order gradient magnitudes. Moreover, the BTV regularization term is based on L_1 norm to handle the outliers. Figure 2.8 shows the original image and the BTV regularized image with the zoom factor of 2.



(a) Original HR image (b) LR image (c) Bicubic interpolation (d) BTV regularized image

Figure 2.8: Super-resolution image with BTV regularization.

2.4 Quality assessment

In image processing, quality assessment is a significant field to numerically reflect the superiority of algorithms. Compared with visual quality which has strong subjective factors by different people, quality assessment methods evaluate the images more objectively. Nowadays, many quality assessment methods are proposed to measure the processed images with the ground truth. In the field of multi-frame super-resolution, researchers usually use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity

(SSIM) to assess the methods since they are efficient and easy to implement. In our experiments result, PSNR and SSIM are used as well to evaluate the quality of our recovered images and then compare the results with other methods.

The Peak Signal-to-Noise Ratio (PSNR) is the most widely used quality metric in image processing. The PSNR can be easily defined by the Mean Squared Error (MSE). Given a degraded monochrome image Y and its ground truth X, MSE is defined as

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [X(i,j) - Y(i,j)]^2, \tag{2.27}$$

where degraded monochrome image Y and its ground truth X have the same size $m \times n$. From the expression, MSE mainly reflects the similarity of two different images. Based on MSE, the PSNR in dB is defined as follows:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right), \tag{2.28}$$

where MAX represents the largest possible pixel value of the image. Generally, 8-bits are used to represent one pixel value. That means the maximum value in this image is 255. Besides evaluating the quality of monochrome images, PSNR can also assess the quality of color images. For an RGB color image, R, G, B channels are evaluated individually and then calculate their average number as the final assessment. From the definition, it's easy to find higher PSNR values to reflect the better quality of recovered images.

Besides PSNR, Structural Similarity (SSIM) is another popular method for image assessment proposed by Wang *et al.* [19]. Different from MSE and PSNR which

estimate the *absolute errors* between ground truth and degraded image, the SSIM is a perception-based model to search perceived changes in image structure. By incorporating the luminance, contrast and structure comparison, the most important structural information is included, which makes SSIM similar to the human visual perception. The expression of SSIM can be formulated as follows:

$$SSIM(X,Y) = l(X,Y)^{\alpha} \cdot c(X,Y)^{\beta} \cdot s(X,Y)^{\gamma}, \tag{2.29}$$

where α, β and γ are the weights of SSIM and set to be 1 in general. $l(X,Y)$, $c(X,Y)$, $s(X,Y)$ represent the luminance, contrast and structure comparison respectively which are formulated respectively as

$$l(X,Y) = \frac{2\mu_X\mu_Y + c_1}{\mu_X^2 + \mu_Y^2 + c_1}, \tag{2.30}$$

$$c(X,Y) = \frac{2\sigma_X\sigma_Y + c_2}{\sigma_X^2 + \sigma_Y^2 + c_2}, \tag{2.31}$$

$$s(X,Y) = \frac{\sigma_{XY} + c_3}{\sigma_X\sigma_Y + c_3}, \tag{2.32}$$

where μ_X and μ_Y are the mean values of X and Y respectively. σ_X and σ_Y represent the variances of X and Y respectively. σ_{XY} stands for covariance of X and Y. The parameter c_1, c_2 and c_3 stabilize the division with some small denominators which are expressed as

$$c_1 = (K_1 L)^2, \quad c_2 = (K_2 L)^2, \quad c_3 = c_2/2, \tag{2.33}$$

where K_1 is set to be 0.01 and K_2 is 0.03 by default. L is the range of pixel values. Due to regular images usually represented by 8 bits per pixel, L is equal to $2^8 - 1 = 255$.

By substituting Equ. (2.30), Equ. (2.31) and Equ. (2.32) into Equ. (2.29), SSIM is formulated as follows:

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)}. \tag{2.34}$$

In general, the Structural Similarity (SSIM) is more consistent with visual quality than PSNR or MSE. For evaluating the image quality, the recovered image has better visual quality when SSIM value is higher.

2.5 Summary

In this chapter, we firstly introduced the observation model which establishes the relationship between a high-resolution image and a sequence of low-resolution images. Then, the point spread function and other processing operators are introduced in detail. After that, the M-estimator and the basic framework of Multi-frame Super-resolution are introduced as two most important parts in this chapter. Besides, due to the ill-posedness of multi-frame super-resolution, the impact of some regularization methods such as Tikhonov regularization and Total Variation (TV) family are demonstrated. Finally, the image quality assessment methods used in this thesis are briefly introduced.

Chapter 3

Literature review

3.1 Frequency domain methods

Frequency domain super-resolution method was first proposed by Tsai and Huang [20], which uses frequency domain to recover lost high frequency information according to the Fourier shift theorem and the aliasing relationship between high-resolution image and corresponding low-resolution images. If x is used to represent the HR image and y_k is used to stand for corresponding kth LR image, the 2D Discrete Fourier Transform (DFT) of x and y_k is expressed as

$$Y_k(u,v) = \frac{1}{r^2}\sum_{p=0}^{r-1}\sum_{q=0}^{r-1} X[(u-Mp)_{Mr},(v-Nq)_{Nr}], \tag{3.1}$$

where $(\cdot)_S$ is the modulo operation and r is the down-sampling factor. M and N are the numbers of low-resolution image pixels in horizontal and vertical directions. X and Y_k represent the DFT of HR image x and kth LR image y_k respectively. As intro-

duced in Chapter 2, every low-resolution image can be regarded as the related high-resolution image going through the geometric motion, blurring and down-sampling operator successively [21]. For the algorithm of Tsai and Huang, the geometric motion is simply considered as the translation motion. Therefore, the DFT function of HR image x and the intermediate HR image x_k is formulated as

$$X_k(u,v) = X(u,v) \cdot e^{-j\frac{2\pi\alpha_k}{Mr}u} e^{-j\frac{2\pi\beta_k}{Nr}v}, \tag{3.2}$$

where α_k and β_k are the shifted pixels in horizontal and vertical directions of HR image x. By Substituting Equ. (3.2) into Equ. (3.1), the 2D Discrete Fourier transform of x and y_k is

$$\begin{aligned} Y_k(u,v) = \frac{1}{r^2} \sum_{p=0}^{r-1} \sum_{q=0}^{r-1} \Big\{ & X[(u-Mp)_{Mr}, (v-Nq)_{Nr}] \\ & \times e^{-j\frac{2\pi\alpha_k}{Mr}(u-Mp)_{Mr}} e^{-j\frac{2\pi\beta_k}{Nr}(v-Nq)_{Nr}} \Big\}. \end{aligned} \tag{3.3}$$

If the number of LR images is large, the Equ. (3.3) is overdetermined. Therefore, the lost DFT coefficients of estimated HR image can be solved. After that, the HR image x can be calculated by solving the inverse DFT function. Following the work of Tsai and Huang, Kim *et al.* [22] considered the effects of sensor blur and noises introduced by observation model which made the framework more practical. The expression of their algorithm can be formulated as

$$\begin{aligned} Y_k(u,v) = \frac{1}{r^2} \sum_{p=0}^{r-1} \sum_{q=0}^{r-1} \Big\{ & X[(u-Mp)_{Mr}, (v-Nq)_{Nr}] \times H[(u-Mp)_{Mr}, (v-Nq)_{Nr}] \\ & \times e^{-j\frac{2\pi\alpha_k}{Mr}(u-Mp)_{Mr}} e^{-j\frac{2\pi\beta_k}{Nr}(v-Nq)_{Nr}} \Big\} + S(u,v), \end{aligned} \tag{3.4}$$

where H is the DFT of a blur operator and S(u,v) represents the DFT of image noise.

Different from other reserchers who considered registation and deblurring steps seperately, Tom and Katsaggelos [23] combined all sub-problems together and using Expectation Maximization algorithm to solve it. Figure 3.1 shows the result of Tom and Katsaggelos with a synthetic data sequence. The down-sampling factor is 2 and the LR sequence contains 4 frames with noise. As shown in Figure 3.1, the estimated HR image has severe visible artifacts and unsharp edges.

(a) Degraded LR images (b) Estimated HR image

Figure 3.1: Example of multi-frame super-resolution in frequency domain.

Besides the methods mentioned above, wavelet based methods have been used to solve SR problem. Similar to Fourier-domain methods, Nguyen and Milanfar [24] utilized the wavelet coefficients to represent the corresponding low-resolution images. Afterwards, the desired high-resolution image is estimated by relating the LR coefficients to HR coefficients. This method is regarded as the First-Generation Wavelets (FGWs). However, the first-generation wavelets are defined on infinite do-

mains, which limits their performance to deal with finite boundaries. Bose *et al.* [25] proposed a method based on the Second-Generation Wavelets (SGWs), which can process arbitrary boundary conditions under bounded domains. Different from the first-generation wavelets only processing the simple motions such as translation and rotation, the second-generation wavelets allow more general camera motions due to less assumptions about sampling structure.

With the help of Fast Fourier Transform (FFT), the implementation of Frequency domain based SR methods is fast and convenient. However, the popular prior information is hard to be applied in the frequency domain based methods and the estimated SR images usually have bad visual quality and present high frequency artifacts.

3.2 Spatial domain methods

3.2.1 Iterative Back-Projection (IBP) based SR methods

The Iterative Back-Projection (IBP) methods are based on spatial domain. Compared with frequency domain methods, spatial domain methods can handle the image noise more easily and by considering sub-pixel motions among low-resolution images, the image registration is more accurate [26].

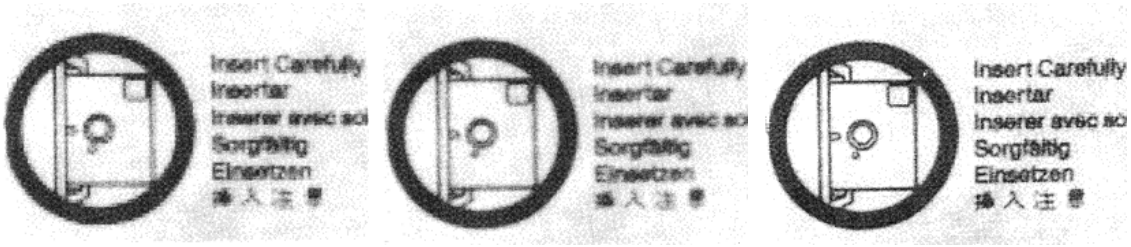
The basic idea of IBP methods is to calculate the residual error between the estimated HR image and the observed LR images. Afterwards, the estimated HR image is updated by projecting the residual error to the previous HR image. If the updated HR image has smaller residual error compared with the previous estimated HR image, the new HR image has better quality than the previous one. The expression of IBP

based SR reconstruction can be formulated as

$$X_{i+1} = X_i + H_f \sum_{k=0}^{K} (W X_i - Y_k), \tag{3.5}$$

where X_i represents the ith iteration of the HR image. Y_k stands for the kth LR image with the total number of K. W is the system matrix and H_f is the projection matrix to project the residual error which is $(W X_i - Y_k)$ to the previous HR image.

Based on this idea, S. Peleg *et.al.* [27] first proposed an Iterative Back-Projection algorithm to reconstruct the HR image. Irani *et.al.* [28] improved the IBP algorithm by back-projecting the residual error between the observed LR images and the simulated LR images. Then the difference of this residual error was projected back to update the previous estimated HR image. Figure 3.2 shows one of the estimated HR images taken from Irani's thesis.



(a) One of the LR images (b) Initial guess (c) Estimated HR image

Figure 3.2: An example of Iterative Back-Projection methods.

The Iterative Back-Projection method has some drawbacks. Due to the ill-posed nature of super-resolution problem, the IBP can not guarantee the stability of final results. Besides, the projection matrix H_f is hard to construct. These drawbacks limit the performance of the IBP based super-resolution methods.

3.2.2 Projection onto Convex Sets (POCS) based SR methods

Projection onto Convex Sets (POCS) is a classic super-resolution method in spatial domain based on set-theoretic [29]. It was first proposed by Stark and Oskoui [30]. In their method, the super-resolution image was constrained by the intersection of many convex sets including data fidelity constraint, finite energy constraint and so on. Therefore, the final estimated image is in the intersection of all constraint sets $C_j(j = 1 \cdots S)$. If P_j is the projection of the estimated HR image X onto set C_j, the POCS method can be formulated as

$$X_{i+1} = P_S P_{S-1} \cdots P_1 X_i, \tag{3.6}$$

where X_i represents the ith iteration of estimated HR image.

The observation model describes the relationship between the estimated HR image and the corresponding LR images. This relationship is used as a fidelity constraint to ensure data consistency. Therefore, the convex set of data consistency constraint is expressed as follows:

$$C_{m,n;K} = \left\{ X(k,l) : Y_K(m,n) = \sum_{k,l} W_K(m,n;k,l) \cdot X(k,l) \right\}, \tag{3.7}$$

where Y_K is the Kth LR image and $W_K(m,n;k,l)$ is the system matrix. Based on the convex set of fidelity constraint, the iterative updated high-resolution image can

be expressed as

$$X_{i+1}(k,l) = X_i(k,l) + \frac{r(m,n) \cdot W_K(m,n;k,l)}{\sum_{k,l} W_K^2(m,n;k,l)}, \tag{3.8}$$

where

$$r(m,n) = Y_K(m,n) - \sum_{k,l} W_K(m,n;k,l) \cdot X_i(k,l). \tag{3.9}$$

Tekalp *et al.* [31] proposed an extended POCS method which considered the image noise introduced by observation model. Their convex constraint was defined as

$$C_{m,n;K} = \{X(k,l) : |r(m,n)| \leq \delta)\}, \tag{3.10}$$

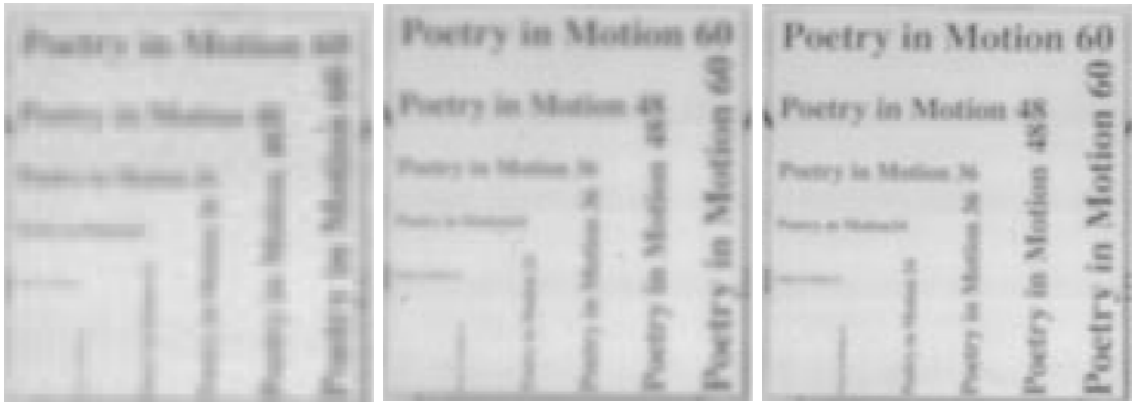
where δ represents the confidence of the observation model. The projection $X_{i+1}(k,l) = P_{m,n;K}\{X_i(k,l)\}$ onto $C_{m,n;K}$ is defined as [32]

$$X_{i+1}(k,l) = X_i(k,l) + \begin{cases} \frac{[r(m,n)-\delta]\cdot W_K(m,n;k,l)}{\sum_{k,l} W_K^2(m,n;k,l)}, & r(m,n) > \delta \\ 0, & -\delta < r(m,n) < \delta \\ \frac{[r(m,n)+\delta]\cdot W_K(m,n;k,l)}{\sum_{k,l} W_K^2(m,n;k,l)}, & r(m,n) < -\delta. \end{cases} \tag{3.11}$$

Besides the fidelity constraint, the other constraints such as positivity, finite energy, limited support can be used to improve the quality of estimated HR image as well.

Patti *et al.* [33, 34] improved the previous POCS methods by considering the family of homographies for the image registration. Their method also uses edge

adaptive constraints to decrease the ringing noise. Elad and Feuer [14] proposed a hybrid method combining the maximum likelihood (ML), Maximum a posteriori (MAP) and POCS together to optimize the final result. The advantage of POCS methods is easy to incorporate the prior knowledge as constraint conditions. But the initial guess of HR image directly influences the result of POCS methods, which makes POCS based SR methods unstable. Figure 3.3 shows an example of POCS super-resolution methods taken from Patti *et al.* [34].



(a) One of the LR images (b) Bilinear interpolation (c) POCS result

Figure 3.3: An example of POCS methods.

3.3 Stochastic methods

Stochastic based SR methods use stochastic functions to estimate the most probable HR image with the observed LR images. The most commonly used theories in stochastic SR methods are Maximum Likelihood (ML) estimation and Maximum a Posteriori (MAP) estimation. They both use the conditional probability density function (PDF) to estimate the desired HR image. The difference between them is

the MAP estimation reforms the traditional ML estimation by introducing an image prior. In general, stochastic methods are widely used to solve the problems of image restoration [35–37], super-resolution [38–40], interpolation [41, 42] and so on.

3.3.1 Maximum Likelihood (ML) estimation

For the Maximum Likelihood estimation, the prior information is not considered, In this case, the estimated HR image X can be recovered by maximizing the conditional probability density function under the observed LR images Y_k, which can be formulated as

$$\hat{X}_{ML} = arg\max_{X} P(Y_1, Y_2 \cdots Y_K|X), \tag{3.12}$$

where K is the total number of LR images. $P(Y_1, Y_2 \cdots Y_K|X)$ is the joint conditional PDF of Y_k given X. In general, every LR image is observed independently. Therefore, Equ. (3.12) can be rewritten as

$$P(Y_1, Y_2 \cdots Y_K|X) = P(Y_1|X)P(Y_2|X) \cdots P(Y_K|X), \tag{3.13}$$

the formula of Maximum Likelihood estimation is expressed as below:

$$\hat{X}_{ML} = arg\max_{X} \prod_{k=1}^{K} P(Y_k|X). \tag{3.14}$$

The ML estimation is modeled by the PSF of the observation error mentioned in Chapter 2. If the image noise is assumed to be independent and identically distributed

(i.i.d.) Gaussian noise with zero mean, it can be formalized as

$$P(n_k) = \frac{1}{(2\pi\sigma_g^2)^{N/2}} \cdot \exp\left(-\frac{n_k^T n_k}{2\sigma_g^2}\right), \tag{3.15}$$

where n_k represents the noise for kth image with the size of $N \times 1$ in column order. N stands for the total number of kth LR image pixels and σ_g is the standard derivation of the noise. According to the observation model, the noise n_k can be expressed as

$$n_k = Y_k - W_k X. \tag{3.16}$$

Therefore, Equ.(3.15) can be rewritten as follows:

$$P(Y_k|X) = \frac{1}{(2\pi\sigma_g^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma_g^2}(Y_k - W_k X)^T (Y_k - W_k X)\right\}, \tag{3.17}$$

where $P(Y_k|X)$ represents the conditional PDF of the low-resolution image Y_k given X. To consider the impact of every low-resolution image, the desired HR image is formulated as

$$\hat{X}_{ML} = arg\max_X \left\{\prod_{k=1}^{K} \frac{1}{(2\pi\sigma_g^2)^{N/2}} \exp\left[-\frac{1}{2\sigma_g^2}(Y_k - W_k X)^T (Y_k - W_k X)\right]\right\}. \tag{3.18}$$

In order to simplify Equ. (3.18), the logarithm operation is used. Due to the monotonic properties of logarithm, the solution of ML estimation is the same as previous expression. However, the minus sign in front of every probability part makes the maximization problem become minimization problem, which is formulated as

follows:

$$\hat{X}_{ML} = arg\min_{X}\left\{\sum_{k=1}^{K}\frac{1}{2\sigma_g^2}(Y_k - W_kX)^T(Y_k - W_kX) - K\ln\frac{1}{(2\pi\sigma_g^2)^{N/2}}\right\}. \quad (3.19)$$

The constant part doesn't impact the minimization equation. Therefore, the cost function of ML estimation can be expressed as

$$F_{cost} = \sum_{k=1}^{K}||W_kX - Y_k||_2^2\,. \quad (3.20)$$

As mentioned in Chapter 2, Equ. (3.20) also expresses the fidelity term based on L_2 norm. Now, if the image noise is independent and identically distributed Laplacian noise with zero mean, the probability density function is [43]

$$P(n_k) = \frac{1}{(2\sigma_l)^N}\cdot\exp\left(-\frac{|n_k|_1^1}{\sigma_l}\right), \quad (3.21)$$

where σ_l is the standard derivation of the Laplacian noise. Similar to the expression of ML estimation with Gaussian noise, the formula with Laplacian noise is

$$\hat{X}_{ML} = arg\max_{X}\prod_{k=1}^{K}\frac{1}{(2\sigma_l)^N}\exp\left\{-\frac{1}{\sigma_l}||Y_k - W_kX||_1^1\right\}. \quad (3.22)$$

By using logarithm operation, Equ. (3.22) becomes

$$\hat{X}_{ML} = arg\min_{X}\left\{\sum_{k=1}^{K}\frac{1}{\sigma_l}||Y_k - W_kX||_1^1 - K\ln\frac{1}{(2\sigma_l)^N}\right\}. \quad (3.23)$$

Same as Equ. (3.20), the cost function of ML estimation with Laplacian noise is

$$F_{cost} = \sum_{k=1}^{K} ||W_k X - Y_k||_1^1 . \tag{3.24}$$

Therefore, L_1 norm and L_2 norm are actually the ML estimation with the assumption of the noise in the Laplacian distribution and the Gaussian distribution respectively. ML estimation based SR methods can handle complicated displacements due to the properties of their observation model. However, their results may not be unique due to the ill-posed nature of SR problem.

3.3.2 Maximum a Posteriori (MAP) estimation

To solve the ill-posed nature of SR problem, prior knowledge is usually used as an effective method. Compared with ML estimation, Maximum a Posteriori (MAP) estimation combines the Bayes's theorem and prior knowledge together to estimate the HR image [44, 45]. The basic Bayes's theorem is expressed as

$$P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}. \tag{3.25}$$

According to Baye's theorem, if X stands for estimated HR image, Y_k is the kth LR image, the expression of MAP estimation can be formulated as

$$\begin{aligned} \hat{X}_{MAP} &= arg \max_X P(X|Y_1, Y_2 \cdots Y_K) \\ &= arg \max_X \frac{P(Y_1, Y_2 \cdots Y_K|X)P(X)}{P(Y_1, Y_2 \cdots Y_K)}, \end{aligned} \tag{3.26}$$

where K is the total number of LR images. Since $P(Y_1, Y_2 \cdots Y_K)$ has no influence

on the maximization function $\hat{X}_{MAP}$, the above MAP estimation can be rewritten as

$$\hat{X}_{MAP} = arg\max_{X} P(Y_1, Y_2 \cdots Y_K|X)P(X). \tag{3.27}$$

If each LR image Y_k is observed independently, Equ. (3.27) can be simplified to

$$\hat{X}_{MAP} = arg\max_{X} \prod_{k=1}^{K} P(Y_k|X) \cdot P(X), \tag{3.28}$$

where $P(X)$ represents the prior knowledge which reflects the properties of estimated HR image. Compared with Equ. (3.14), the only difference between ML and MAP estimations is that MAP takes the image prior information into consideration.

In general, Markov Random Fields (MRF) is widely used as image prior information incorporating with Bayesian based estimation [46]. Markov process is a stochastic process which can predict the future based only on its present state but have the same prediction with the full history states of the process. This process can be expressed as follows:

$$P_r(x_n|x_n - 1) = P(x_n|x_{n-1}, x_{n-2}, x_{n-3}, x_{n-4}, \cdots), \tag{3.29}$$

where n is the time index. For the nature image, MRF based methods are fairly true because the image pixels are only dependent on their neighbors in most situations. A general model of Markov Random Fields (MRF) characterized by Gibbs distribution [47] is

$$P_r(X) = \frac{1}{H} \exp\left\{ -\sum_{c \in C} V_c(X) \right\}, \tag{3.30}$$

where C is the set of cliques and H is the normalizing constant. $V_c(X)$ is the clique potential function. Cliques defines some adjacent pixel pairs between center pixel and its corresponding neighbors. Figure 3.4 shows four cardinal neighbors (left, right, up and down) of an arbitrary center pixel X_i. The right cliques is represented in the blue dash lines.

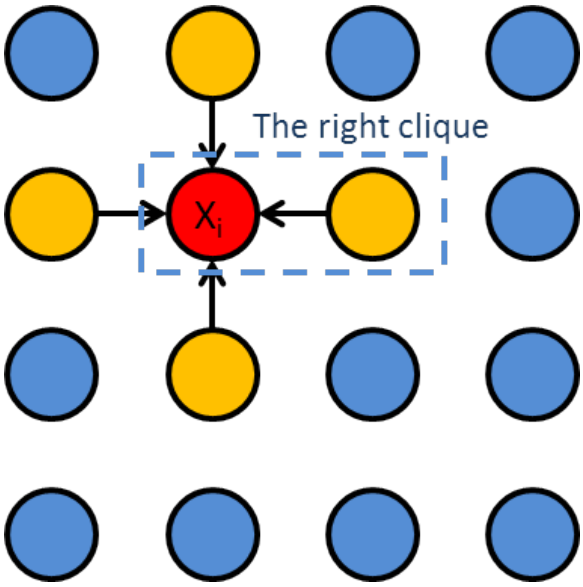


Figure 3.4: The cardinal neighbors of a center pixel X_i in high-resolution image.

Besides the MRF with Gibbs distribution, Markov Random Fields with Gaussian distribution is commonly utilized for prior knowledge as well [48, 49] which is expressed as

$$P_r(X) = \frac{1}{(2\pi)^{\frac{N}{2}}|C_X|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}X^T C_X^{-1} X\right\}, \tag{3.31}$$

where N is the total number of pixels in high-resolution image. C_X is the covariance matrix of HR image X with the size of $N \times N$. The exponential term can be further

calculated with a sum of products as

$$P_r(X) = \frac{1}{(2\pi)^{\frac{N}{2}}|C_X|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\lambda}\sum_{i=1}^{N} X^T d_i d_i^T X\right\}, \tag{3.32}$$

where λ is the tuning parameter to control the weight of the prior part. $d_i = [d_{i,1}, d_{i,2}, \cdots, d_{i,N}]^T$ is the coefficient vector of the corresponding HR image. Therefore, the GMRF can be rewritten as

$$P_r(X) = \frac{1}{(2\pi)^{\frac{N}{2}}|C_X|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\lambda}\sum_{i=1}^{N} \left(\sum_{i=j}^{N} d_{i,j} X_j\right)^2\right\}. \tag{3.33}$$

If the $(i,j)th$ element in covariance matrix C_X^{-1} is denoted as $C_{i,j}^{-1}$, the expression of this covariance matrix is

$$C_{i,j}^{-1} = \frac{1}{\lambda}\sum_{r=1}^{N} d_{r,i} d_{r,j}. \tag{3.34}$$

The coefficient elements in $d_{i,j}$ express the prior information of the relationship between adjacent pixel values in estimated HR image X. Each value is commonly selected according to their pixel similarities. Therefore, if $d_{i,j}$ is assumed as 2-D Laplacian matrix, the expression is

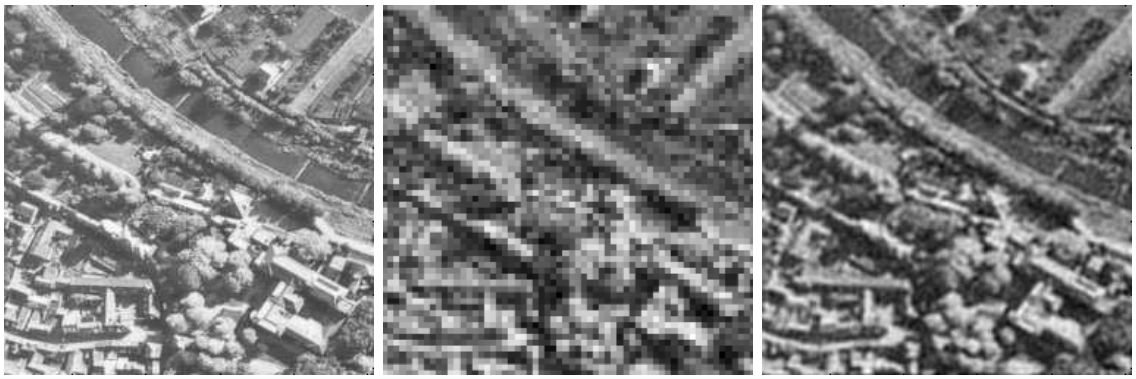
$$d_{i,j} = \begin{cases} 1, & \text{for } i = j \\ -1/4, & \text{for } j : X_j \text{ represents a cardinal neighbor of } X_i \\ 0, & \text{otherwise.} \end{cases} \tag{3.35}$$

With adding image prior information to the conditional probability density func-

tion (PDF), the MAP estimation with Gaussian Markov random field (GMRF) can be formulated as

$$\hat{X}_{MAP} = arg\min_{X}\left\{\sum_{k=1}^{K}\frac{1}{2\sigma_g^2}(Y_k - W_k X)^T(Y_k - W_k X) + \frac{1}{2}X^T C_X^{-1} X\right\}. \quad (3.36)$$

Figue 3.5 shows one example of GMRF methods. From left to right, the ground-truth image, low-resolution image and estimated *Aerial* image are shown respectively with zoom factor $r = 4$ captured from [49].



(a) Original *Aerial* image (b) One of the noisy low-resolution images (c) Estimated HR image

Figure 3.5: An example of GMRF based methods.

Besides the MAP estimation methods mentioned above, a Huber-Markov Random Field Model was proposed in [41, 50]. With the non-linear property of Huber function, it suppresses the unwished noise and applies fewer penalties to image edges. Tian and Ma [51] proposed a Markov Chain Monte Carlo (MCMC) method to generate samples from MAP estimation of unknown high-resolution image. Then, an updated MCMC method was proposed in [52] to preserve the image edges. Tae OH *et al.* [53] proposed a PAR/NL model whose parameters are calculated by a Non-Local (NL) training procedure. A Piecewise Auto-Regressive (PAR) process is used to

model the 2D random texture and the Expectation-Maximization (EM) method is used to optimize the object function. Recently, R. Molina *et al.* [54] utilized a simultaneously Spatial Auto-Regressive (SAR) model to make the estimated image stable and improve the visual quality.

Compared with ML estimation, MAP estimation can efficiently solve the ill-posed nature of SR problem with image prior information that makes MAP estimation better than ML estimation. However, the disadvantage of MAP estimation is the iteration procedure may cause the high computational cost.

3.4 Multi-frame SR methods with adaptive norm choice

Traditional multi-frame SR methods are based on MAP estimation with fixed norm choice, which can be expressed as

$$\hat{X} = arg\min_{X} \left\{ \sum_{k=1}^{K} \|DB_k M_k X - Y_k\|_p^p + \lambda \Upsilon(X) \right\}, \tag{3.37}$$

where $\|\cdot\|_p^p$ denotes the choise of Lp norm. K is the total number of LR images. $\Upsilon(X)$ is the regularization term. λ is the trade-off parameter between the fidelity term and the regularization term. The estimated HR image can be iteratively estimated by using the optimization methods such as Steepest Decent (SD), Conjugate Gradient (CG) [55], Scaled Conjugate Gradient (SCG) [56] and so on. Compared with other optimization methods, steepest descent is the simplest method to solve

this minimization problem [11]. The iterative function is

$$\hat{X}^{n+1} = \hat{X}^n - \beta \cdot \nabla f(\hat{X}^n), \tag{3.38}$$

where β is the step size in the gradient direction. $\hat{X}^n$ represents the nth iteration. $\nabla f(\hat{X}^n)$ stands for the first-order gradient of the minimization function with respect to estimated HR image X.

If the fidelity term of Equ. (3.37) chooses L_1 norm, the $\nabla f(\hat{X}^n)$ can be expressed as

$$\nabla f(\hat{X}^n) = \sum_{k=1}^{K} (DB_kM_k)^T \cdot sign(DB_kM_k\hat{X}^n - Y_k) + \lambda \Upsilon'(\hat{X}^n), \tag{3.39}$$

where $\Upsilon'(\hat{X}^n)$ is the gradient of the regularization term and $sign(\cdot)$ extracts the sign symbol of related parameters.

If the traditional SR framework chooses L_2 norm, the gradient function $\nabla f(\hat{X}^n)$ is

$$\nabla f(\hat{X}^n) = \sum_{k=1}^{K} 2(DB_kM_k)^T \cdot (DB_kM_k\hat{X}^n - Y_k) + \lambda \Upsilon'(\hat{X}^n). \tag{3.40}$$

Substituting Equ. (3.39) or Equ. (3.40) into Equ. (3.38), the HR image can be estimated after some iterations. However, according to the analysis of ML and MAP estimation, L_1 norm is chosen when the noise is assumed in the Laplacian distribution but if the noise fits Gaussian distribution, L_2 norm is a better choice. In practical situations, the noise is usually mixed and complicated. For example, the outliers in image can be fitted by Laplacian distribution but some other noises introduced by

the shooting environment are usually fitted by Gaussian distribution. Therefore, Yue *et al.* [57] proposed a locally adaptive L_1, L_2 norm to handle the images with mixed noise and outliers. In their method, the pixels with Gaussian noise is processed by L_1 norm. For the pixels of impulse noise or motion outliers, L_2 norm is chosen to handle these errors. The model can be formulated as

$$\begin{aligned}\hat{X} = arg\min_{X}\Big\{\sum_{k=1}^{K}\Big[w_k \cdot ||C_k(DB_kM_kX - Y_k)||_1^1 \\ + ||\tilde{C_k}(DB_kM_kX - Y_k)||_2^2\Big] + \lambda\Upsilon(X)\Big\},\end{aligned} \tag{3.41}$$

where C_k is a diagonal matrix for the kth LR image with non-zero values if the related pixels belong to outliers. $\tilde{C_k}$ equals to $I - C_k$ with the identity matrix I. w_k is a trade-off weight to handle the balance between L_1 norm and L_2 norm.

Although Yue's *et al.* method selects the norm adaptively according to the local pixel information, it causes the minimization function non-derivable. Therefore, traditional optimization methods can not guarantee its convergence to minimized point. Due to the non-linearity, Iteratively Reweighted Norm (IRN) [58, 59] is used to solve this nonlinear inverse problem, which increases the computational cost.

To solve this limitation, Zeng *et al.* [60] proposed a robust multi-frame SR based on half-quadratic estimation. This half-quadratic estimation is twice continuously differentiable and strictly convex. Therefore, the objective function can converge to a global minimizer by using optimization technique. Moreover, according to the properties of half-quadratic estimation, they proposed a Bilateral Edge-Preserving (BEP) regularizer based on traditional BTV regularization. The cost function of

their minimization model can be defined as

$$C(X) = \sum_{k=1}^{K} \alpha_k \sqrt{\alpha_k^2 + e_k^2} + \lambda \sum_{l=-P}^{P} \sum_{m=0}^{P} \sum_{i=1}^{N} \alpha^{|m|+|l|} \rho((X - S_x^l S_x^m X)[i], c), \quad (3.42)$$

where e_k is the observation error which equals to $(DB_kM_kX - Y_k)$. a_k is the threshold parameter for kth LR image, which is adaptively chosen by the averaged observation error. N is the total number of pixels in the estimated HR image. $\rho(x, c)$ is the half-quadratic estimation with threshold parameter c similar to the function in fidelity term.

This cost function can be minimized by simple Steepest Decent (SD) algorithm. However, it usually converges faster by using Conjugate Gradient (CG) optimization. The iterative function of CG algorithm is described as

$$\hat{X}_{n+1} = \hat{X}_n + \alpha_n P_n, \quad (3.43)$$

where α_n is the step size for the nth iteration. In order to find the optimal α_n, backtracking line search technique [61] is used to solve this problem. P_n is the CG vector for the nth iteration with the initial value being set to $P_0 = -\nabla C(X_0)$. The Iterative function of P_n is

$$P_{n+1} = -\nabla C(X_{n+1}) + \beta_n P_n, \quad (3.44)$$

where

$$\beta_n = \frac{\nabla C(X_{n+1})^T \nabla C(X_{n+1})}{\nabla C(X_n)^T \nabla C(X_n)}, \quad (3.45)$$

and $\nabla C(X)$ represents the gradient of $C(X)$ with respect to X.

3.5 Summary

Compared with other super-resolution methods mentioned in this chapter, the Bayesian-based SR methods can easily incorporate fidelity terms with prior knowledges to overcome the ill-posed nature of SR problem. Moreover, the quality of estimated HR image is gradually from coarse to fine due to the iteration procedure. That makes the relative parameters such as motion vectors and blur kernel updated in every iteration. Nowadays, Bayesian-based SR methods become the most widely used methods in the field of multi-frame image super-resolution. Compared with other SR methods, the final results of Bayesian-based SR methods have clearer texture and less noise.

Chapter 4

Proposed multi-frame super-resolution algorithm

In this chapter, we introduce our proposed algorithm in detail. For our initial setting of estimated HR image, we use a novel initial method based on median operator, which generates an outlier-insensitive HR image as our initial value. The traditional fidelity terms use fixed L_1 or L_2 norm to handle the observation errors. However, as mentioned in Chapter 3, L_1 or L_2 norm based traditional methods have limitations. In our proposed algorithm, the half-quadratic estimation is used to choose norm adaptively instead of fixing L_1 or L_2 norm. The traditional BTV regularization suppresses image noise by penalizing the first-order gradient magnitudes. However, since it can not distinguish image edges from noises, some edge information is suppressed as well. Due to this drawback of the traditional BTV regularization, our proposed BTV regularization method uses second-order gradient based on difference curvature which designates a corresponding weight to each pixel. In this case, if the pixel contains

edge information, the corresponding weight will become small. Otherwise, this weight will be large. Therefore, our proposed regularization method can both suppress noise and preserve image texture adaptively due to the corresponding weight imposed on different pixels.

4.1 The estimation of initial high-resolution image

Many multi-frame super-resolution uses optimization algorithm to minimize the objective function so that the final HR image is reconstructed when the function reaches the minimal point. Due to the iterative calculation, the estimated HR image is from coarse to fine. In this case, if the initial HR image is estimated well, the minimization function may reach a stable state very fast and the final recovered HR image will have good quality. In other words, the setting of initial value directly influences the quality of final result.

In traditional multi-frame super-resolution methods, the initial HR estimation X_0 is obtained by using bicubic or bilinear interpolation on the reference LR image. If the reference LR image has bad quality, the initial HR is poor as well. To solve this problem, we use a novel method based on image warping. At first, it warps every LR frame to the reference frame and then reconstructs a composed LR image by using median operator. Finally the initial HR image is obtained by using bilinear interpolation. Therefore, unlike traditional SR initial estimation only using information of the reference image, our proposed initialization method combines all detail information in LR sequence to estimate the initial HR image.

Image warping transforms an image from one plane to another plane based on

some mathematical functions [62]. If the transformation maps every point without changing the image color, it is called pure warping. In our novel initialization, every non-reference LR image is warped as the shape of the reference LR image. Therefore, pixels in the same location of registrated LR images can be considered as displaying the same detail information. If we choose the first LR frame Y_1 as our reference image, the warping procedure can be expressed as

$$Y_1 \approx P_w(Y_k), \quad k = 2, 3, 4, \cdots, K, \tag{4.1}$$

where $P_w(\cdot)$ is the projection function, which maps the non-reference LR images to the reference image. K is the total number of LR frames in sequence.

After the process of image warping, the LR images are aligned and the composed LR image should be generated from them to contain the novel information as much as possible. In general, mean operator and median operator are widely used to solve this problem. If the mean operator is chosen to generate the composed LR image and Y'_k is used to replace $P_w(Y_k)$ where $k = 2, 3, 4, \cdots, K$, the expression can be formulated as

$$Y_c = f_{mean}(Y_1, Y'_2, Y'_3, \cdots, Y'_K), \tag{4.2}$$

where f_{mean} stands for the pixel-wise mean operator and Y_c represents the composed LR image. The value of $(i, j)th$ pixel in composed image is calculated by

$$Y_c^{(i,j)} = \frac{Y_1^{(i,j)} + Y_2'^{(i,j)} + Y_3'^{(i,j)} + \cdots + Y_K'^{(i,j)}}{K}. \tag{4.3}$$

Different from mean operator, the median operator searches the median value of pixels in the same location of registrated LR images, which can be expressed as

$$Y_c = f_{med}(Y_1, Y_2', Y_3', \cdots, Y_K'), \tag{4.4}$$

where f_{med} represents the median operator. The $(i,j)th$ pixel in the composed LR image is the median value of all $(i,j)th$ pixels in relevant LR images. Figure 4.1 shows the composed LR image by using mean and median operator respectively.

(a) Mean operator (b) Median operator

Figure 4.1: Example of using mean and median operator to generate the composed LR image respectively.

Compared with these two operators, mean operator calculates the mean value of each LR pixel. However, due to the different distortion levels, some LR images contain more novel information than others. In this case, if we use the mean operator to reconstruct the composed LR image, the distorted pixels have the same contribution to the fine pixels, which may make the result undesirable. Different from mean

operator, the median operator extracts the median value of the corresponding pixels. Even though there are some distorted pixels, they barely affect the value of median one. Therefore, the median operator is more robust than mean operator to handle the outliers in the sequence of corresponding pixels. In consequence, we choose the median operator to generate our composed LR image rather than mean operator.

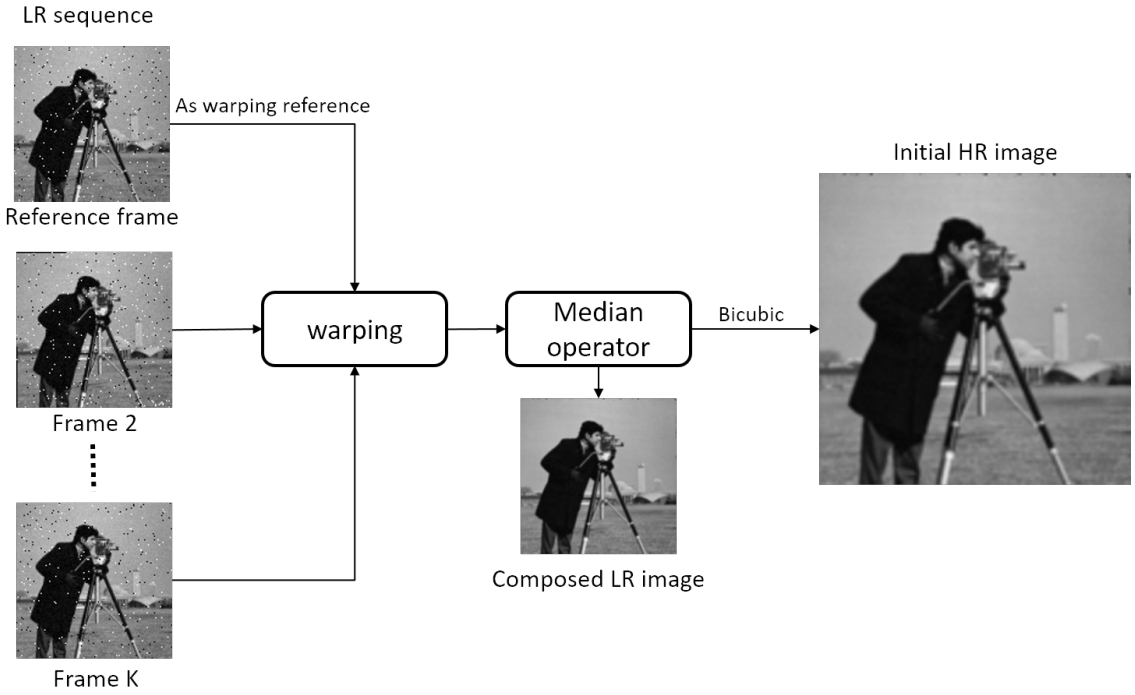


Figure 4.2: The framework of generating the initial HR image X_0.

Last but not least, by using bicubic interpolation, the initial HR image is generated from the composed LR image. The procedure of generating the initial HR image is shown in Figure 4.2. The isolated noise is added to the LR images. If we generate the initial HR image directly from the reference LR frame, the isolated noise still exists. Our initial algorithm can eliminate the outliers by searching the median value of all corresponding pixels. Therefore, our initial estimation has better quality due to the

combination of LR images. Moreover, a good initial estimation helps the convergence of minimization function as well.

4.2 Adaptive fidelity term based on half-quadratic estimation

The half-quadratic function was first proposed in [63] as a potential function, which combines the advantages of L_1 and L_2 norm. According to the parameter α, the half-quadratic function can reduce the effect of different kinds of errors such as large registration errors and small Gaussian errors. The half-quadratic estimation is defined as

$$f(x, \alpha) = \alpha\sqrt{\alpha^2 + x^2} - \alpha^2, \tag{4.5}$$

where α is a positive constant. For each LR frame, x represents the observation error which is $(DB_kM_kX - Y_k)$. The half-quadratic function has attributes of strictly convex and twice continuously differentiable so that any convex optimization algorithms can easily solve it.

For The first derivative of $f(x, \alpha)$, it behaves like linear proportion to small errors and then gradually behaves similarly to a constant. The first derivative function with respect to x is shown below:

$$f'(x, \alpha) = \frac{\alpha x}{\sqrt{\alpha^2 + x^2}}. \tag{4.6}$$

If α is set to be 1, the half-quadratic function and its derivative is shown in Figure 4.3.

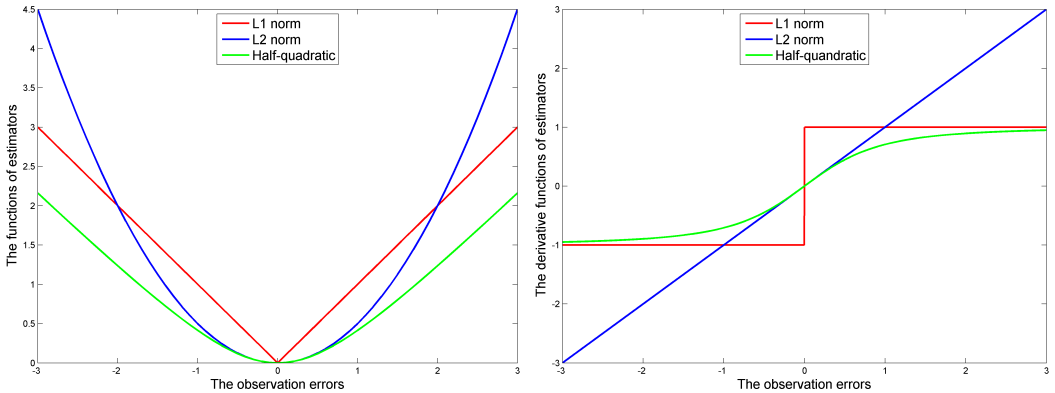


Figure 4.3: Error norms. (a) The norm functions of L_1, L_2 and $f(x, \alpha)$, (b) Their corresponding derivative norm functions.

As shown in Figure 4.3, the derivative of half-quadratic function performs like L_2 norm when the observation errors are small and then gradually performs like L_1 norm when the observation errors are large to suppress the outliers like isolated noise and mis-registration errors.

Compared with other M-estimators such as Leclerc and Lorentzian mentioned in Chapter 2, the half-quadratic estimation has the best performance. Figure 4.4 shows the norm functions and their derivative functions of L_1 norm, L_2 norm, Leclerc, Lorentzian and half-quadratic estimation respectively. The thresholds of Leclerc and Lorentzian are both set to be 1.

As shown in Figure 4.4, the traditional fixed norm function such as L_1 and L_2 norms can not adjust their output according to different observation errors. For the commonly used M-estimators, the Leclerc and Lorentzian fits L_1 and L_2 norm

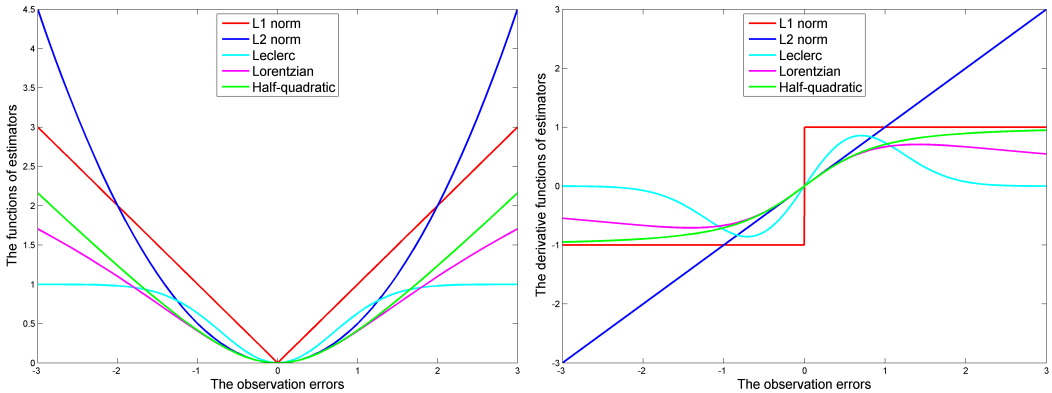


Figure 4.4: Error norms. (a) The norm functions of L_1, L_2, Leclerc, Lorentzian and half-quadratic estimation, (b) Their corresponding derivative norm functions.

adaptively with different input. However, they both have extreme points with the increasing of the observation error, which makes them nonmonotonic. Unlike the Leclerc and Lorentzian estimators, the half-quadratic estimation is monotonically increasing, which makes it robust to process observation errors. Moreover, the half-quadratic function fits L_1 and L_2 norm effectively. Combing the advantages of L_1 and L_2 norm, the half-quadratic function can handle mixed noises with Gaussian distribution and Laplacian distribution.

Using the half-quadratic function as an adaptive norm, the fidelity term of SR estimation is [60]

$$\hat{X} = arg\min_{X}\left\{\sum_{k=1}^{K}\alpha_k\sqrt{\alpha_k^2+\left(DB_kM_kX-Y_k\right)^2}-\alpha_k^2\right\}, \tag{4.7}$$

where the choice of parameter α_k is not fixed due to the different accuracy levels of each observed LR frame. As we all know, accurate registration of subpixel displacement and estimation of Point Spread Function (PSF) are difficult in real application

but directly affect the reconstruction of high-resolution images. In general, the observation errors consist of the blurring errors, the registration errors and the noise errors. For each LR frame, the accuracy level of the PSF estimation and the registration procedure may be different. The frame which has large residual error represents the inaccurate estimation, which should have less contribution to the final recovered HR image. In contrast, if the frame has small observation error, it should have more contribution to the estimated HR image. Therefore, the parameter α_k should be adaptively determined according to the observation error of each LR frame.

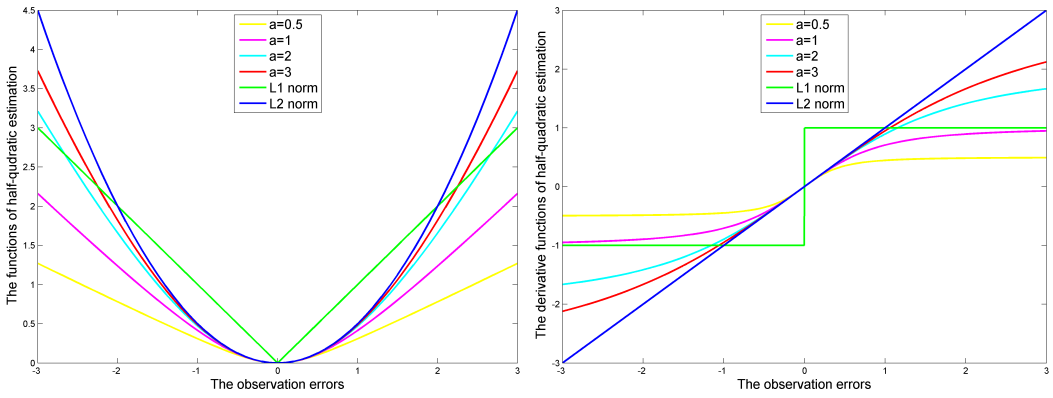


Figure 4.5: Error norms. (a) The half-quadratic estimation $f(x, \alpha)$ with different α values, (b) Their corresponding derivative functions.

Figure 4.5 shows the performance of the half-quadratic estimation with some different α values. When the parameter α tends to 0, the half-quadratic function performs like L_1 norm. Furthermore, with the increasing of parameter α, the adaptive error norm performs gradually close to L_2 norm. In order to define the accuracy level of each LR frame, the averaged observation error is used. It represents the averaged

error in each image pixel, which can be formulated as

$$E_k = \|DB_kM_kX_0 - Y_k\|_1/L, \tag{4.8}$$

where L means the total number of pixels in each LR frame. X_0 is the initial HR image.

In general, E_k has a small value when the LR frame doesn't have registration error or isolated error and accurately estimates the blurring kernel. In this case, the observation model approximately fits the Gaussian distribution. As analysis in Chapter 3, the half-quadratic function should behave like L_2 term. Therefore, the parameter a_k should be set to a large value. However, for those LR frames suffering from outliers and registration errors caused by inaccurate estimation, E_k has a large value. Due to the fine detail information being distorted terribly, the parameter a_k should be small to behave like L_1 term.

Based on above analysis, the parameter a_k should be a positive vector and inversely proportional to E_k. Considering the relationship between averaged observation error E_k and adaptive parameter a_k, the objective function is defined as

$$\alpha_k = f(E_k), \tag{4.9}$$

where $f(\cdot)$ is a monotonically decreasing function and should satisfy all conditions of parameter a_k

The choice of $f(\cdot)$ is not difficult since there are many kinds of monotonically decreasing functions satisfying the above conditions. Zeng *et al.* [60] proposed a

suitable function to calculate a_k from E_k which is defined as

$$\alpha_k = -\eta {E_k}^2 + \gamma, \tag{4.10}$$

where η should be large than 0 to decay the part of E_k^2 and make the function monotonically decreasing. The parameter η and γ are defined respectively as

$$\eta = \frac{\alpha_{max} - \alpha_{min}}{E_{max}^2 - E_{min}^2}, \tag{4.11}$$

$$\gamma = \frac{\alpha_{max} E_{max}^2 - \alpha_{min} E_{min}^2}{E_{max}^2 - E_{min}^2}, \tag{4.12}$$

where E_{max} and E_{min} are the maximum and minimum values of averaged observation error E_k respectively. α_{min} is the minimum value of a_k. Due to the positivity of a_k, it is set to be 10^{-5} in our experiment. a_{max} stands for the maximum value of a_k, which is set to be the same value as E_{max}.

Since the constant part of Equ. (4.7) only parallelly moves the function along the direction of Y-axis, it doesn't affect the performance of the fidelity term. To simplify this minimization function, α_k^2 can be deleted from the minimization function. Therefore, the fidelity term of our proposed algorithm can be rewritten as

$$\hat{X} = arg \min_X \sum_{k=1}^{K} \alpha_k \sqrt{\alpha_k^2 + (DB_k M_k X - Y_k)^2}. \tag{4.13}$$

4.3 Proposed difference curvature based BTV regularization

The traditional regularization term such as Tikhonov and TV can not distinguish edges from noises. Therefore, although the noise is eliminated, the texture information is suppressed as well, which limits the performance of these traditional regularization methods. Chen *et al.* [64] proposed a new edge indicator called difference curvature which can effectively distinguish the edges from image noise. Motivated by this property, we combine the traditional BTV regularization term with this new edge indicator to propose a novel regularization method, which can suppress image noise and preserve edges simultaneously. The definition of difference curvature is

$$D = ||I_{\eta\eta}| - |I_{\xi\xi}||, \tag{4.14}$$

where operator $|\cdot|$ denotes the absolute value. D represents the edge indicator based on difference curvature. The expressions of $I_{\eta\eta}$ and $I_{\xi\xi}$ can be formulated as follows respectively:

$$I_{\eta\eta} = \frac{I_x^2 I_{xx} + 2I_x I_y I_{xy} + I_y^2 I_{yy}}{I_x^2 + I_y^2}, \tag{4.15}$$

$$I_{\xi\xi} = \frac{I_y^2 I_{xx} - 2I_x I_y I_{xy} + I_x^2 I_{yy}}{I_x^2 + I_y^2}, \tag{4.16}$$

where I_x, I_y are the first-order derivatives of image I and I_{xx}, I_{xy} and I_{yy} are the second-order partial derivatives of image I. By using the central differences, these

derivatives can be formulated as

$$\begin{aligned} I_x(x,y) &= \frac{I(x+1,y) - I(x-1,y)}{2}, \\ I_y(x,y) &= \frac{I(x,y+1) - I(x,y-1)}{2}, \\ I_{xx}(x,y) &= I(x+1,y) - 2I(x,y) + I(x-1,y), \\ I_{yy}(x,y) &= I(x,y+1) - 2I(x,y) + I(x,y-1), \\ I_{xy}(x,y) &= \frac{I(x-1,y-1) + I(x+1,y+1) - I(x-1,y+1) - I(x+1,y-1)}{4}. \end{aligned} \tag{4.17}$$

Using the above functions to calculate the derivatives are troublesome. In our implementation, these partial derivatives are regarded as the convolution with the following corresponding kernels

$$\begin{aligned} h_x &= \begin{bmatrix} 0.5 & 0 & -0.5 \end{bmatrix}, \\ h_y &= \begin{bmatrix} 0.5 & 0 & -0.5 \end{bmatrix}^T, \\ h_{xx} &= \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}, \\ h_{yy} &= \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}^T, \\ h_{xy} &= \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{4.18}$$

This new indicator based on difference curvature combines the first-order and second-order gradients together to distinguish image edges from isolated noise. Table 4.1 shows the performance of $I_{\eta\eta}$, $I_{\xi\xi}$ and the Difference Curvature (DC) indicator in

various regions of a distorted image.

Table 4.1: The performance of difference curvature indicator in various areas

Various areas	$I_{\eta\eta}$	$I_{\xi\xi}$	D
Edge	Large	Small	Large
Flat	Small	Small	Small
Ramp	Small	Small	Small
Isolated noise	Large	Large	Small

In Table 4.1, $I_{\eta\eta}$, $I_{\xi\xi}$ and D are normalized within $[0, 1]$. 'Large' means that the parameter value is larger than 0.5. And 'Small' means that the parameter value is smaller than 0.1. The parameter values between 0.1 and 0.5 are not defined in our algorithm. In general, $I_{\eta\eta}$ has large value in isolated noise and edge regions. However, $I_{\xi\xi}$ only has large value in areas of isolated noise. Since the large values of isolated noise in $I_{\eta\eta}$ and $I_{\xi\xi}$ are in the same range, the new edge indicator D utilizes the difference between them. After subtracting $|I_{\xi\xi}|$ from $|I_{\eta\eta}|$, the indicator D only has large value in edge regions. For other regions, the values of D are all small. Therefore, D has good performance to distinguish edges from isolated noise. More results are given in Chapter 5.

In our proposed regularization term, D is used as a reference to control the performance of the adaptive weight matrix W_D which has the same size as the HR image. Every element in W_D represents a weight added in the corresponding location in BTV regularization term. For the noise areas, W_D should have large values to suppress those parts. But for the texture areas, W_D should have small values to preserve the detail information. Therefore, W_D can control the BTV parameters by giving them different penalty weights. From the above analysis, W_D should be inverse proportional to D. Therefore, the expression of adaptive weight matrix W_D is formulated

as

$$W_D = \frac{1}{w + \sqrt{\frac{D}{D_{max}}}}, \tag{4.19}$$

where w is a positive constant as a tuning parameter to adjust the extent of W_D. The default number of w is 0.5 in our experiment. D_{max} represents the maximum value of D. $\frac{D}{D_{max}}$ scales D to the range of 0 to 1. The square root operator is used to extend the difference between values, which efficiently divides the weight for noise regions and edge regions.

Combing the adaptive weight W_D and the traditional BTV model, we propose a novel Difference Curvature based BTV regularization method (DCBTV), which is expressed as

$$\Upsilon_D(X) = \sum_{l=-P}^{P} \sum_{m=-P}^{P} \alpha^{\|m\|+\|l\|} W_D \left\| X - S_x^l S_y^m X \right\|_1, \tag{4.20}$$

where P is a control parameter and α is a scaled weight with the range of $0 < \alpha < 1$, which controls the decaying effect to the summation.

Different from traditional regularization, the DCBTV model penalizes both first-order and second-order gradient magnitudes to suppress image noise and preserve edges at the same time according to the adaptive weight matrix based on difference curvature. Our experimental results also show that the DCBTV model can eliminate more image noise than other traditional regularization method.

4.4 The framework of our proposed algorithm

In our proposed multi-frame super-resolution framework, the fidelity term based on half-quadratic estimation and the novel regularization term based on difference curvature are combined to estimate the HR image from a sequence of corresponding LR images. Due to the improvement in fidelity and regularization parts, the results of our proposed method have better quality than other state-of-the-art methods. Figure 4.6 shows the whole framework of our proposed multi-frame super-resolution algorithm.

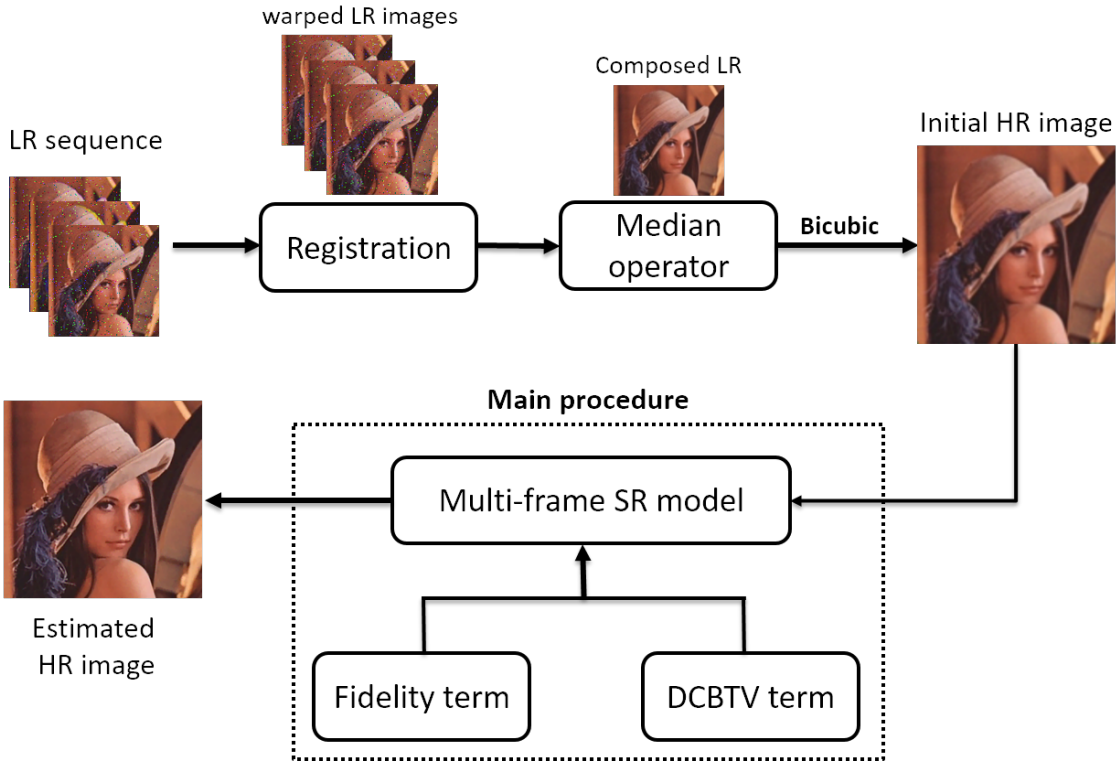


Figure 4.6: The framework of proposed multi-frame super-resolution algorithm.

The initial step of our proposed algorithm mainly generates a composed LR image as a reference to initialize the HR image X. The composed LR image has less

outliers and more detail information compared with any of the original LR images, which makes our initial HR image have better quality than traditional initial settings. The main process in our proposed algorithm contains the fidelity term introduced in Section 4.2 and the novel regularization based on difference curvature. The final estimated HR image is reconstructed when the objective function has the minimum value. Therefore, the recovered HR image can be formulated as

$$\begin{aligned}\hat{X} =& arg\min_{X}\sum_{k=1}^{K}\alpha_k\sqrt{\alpha_k^2+(DB_kM_kX-Y_k)^2}+\\ &\lambda\sum_{l=-P}^{P}\sum_{m=-P}^{P}\alpha^{|m|+|l|}W_D\left\|X-S_x^lS_y^mX\right\|_1,\end{aligned} \tag{4.21}$$

where λ is the trade-off parameter to control the balance between the fidelity term and regularization term.

Since the first-order gradient of objective minimization function $f(x)$ is needed to be calculated in the optimization step, the expression of the first-order derivative function with respect to X is calculated as

$$\begin{aligned}f'(X) =& \sum_{k=1}^{K}\frac{\alpha_k(DB_kM_k)^T(DB_kM_kX-Y_k)}{\sqrt{\alpha_k^2+(DB_kM_kX-Y_k)^2}}\\ &+\lambda\sum_{l=-P}^{P}\sum_{m=-P}^{P}\alpha^{|m|+|l|}W_D(I-S_y^{-m}S_x^{-l})sign(X-S_x^lS_y^mX),\end{aligned} \tag{4.22}$$

where $f'(X)$ denotes the derivative of $f(x)$. I is a identity matrix. S_x^l shifts image X by l pixels in the horizontal direction and S_y^m shifts X by m pixels in the vertical direction. S_x^{-l} and S_y^{-m} are the transposes of matrices S_x^l and S_y^m. They shift image X in the opposite directions as S_x^l and S_y^m do respectively.

In Chapter 2, we introduced a method to establish a system matrix W_k which unites the geometric motion matrix M_k, the blur matrix B_k and the down-sampling matrix D to simply describe the relationship between a sequence of LR images and their corresponding super-resolved HR image. The establishment of W_k is not complicated. Pickup *et al.* [9] proposed a method to generate this system matrix, which is used in our method as well.

By using the system matrix W_k, the Equ. (4.22) can be simplified as

$$\begin{aligned} f'(X) = & \sum_{k=1}^{K} \frac{\alpha_k W_k^T (W_k X - Y_k)}{\sqrt{\alpha_k^2 + (W_k X - Y_k)^2}} \\ & + \lambda \sum_{l=-P}^{P} \sum_{m=-P}^{P} \alpha^{|m|+|l|} W_D (I - S_y^{-m} S_x^{-l}) sign(X - S_x^l S_y^m X). \end{aligned} \tag{4.23}$$

The quality of estimated HR image is gradually improved with the increasing iterations. For multi-frame super-resolution problem, there are many optimization methods to solve it such as Steepest Decent (SD) [65], Conjugate Gradient (CG) [66] and so on. In our proposed algorithm, we use Scaled Conjugate Gradient (SCG) to solve the function expressed in Equ. (4.21). Compared with other optimization methods, the Scaled Conjugate Gradient (SCG) can adaptively adjust the step size according to the approximative speed in gradient direction [56]. Moreover, the convergence of SCG is faster than Steepest Decent and Conjugate Gradient due to the scaled step size as well.

4.5 Summary

In this chapter, we introduced our proposed multi-frame super-resolution method in detail. For the initial setting of the high-resolution image, we use a novel initial method based on median operator, which generates an outlier-insensitive HR image as our initial value. For the fidelity term, the half-quadratic estimation is used to make norm chosen adaptively instead of using fixed L_1 or L_2 norm. For the regularization part, a novel regularization method based on difference curvature is proposed to suppress the image noise and preserve the edges simultaneously according to local information.

Chapter 5

Experimental results

In this chapter, we mainly show the experimental results of our proposed multi-frame super-resolution algorithm. At first, we compare the quality of our initial HR image obtained by our novel initialization method with other traditional methods. Then the performance of difference curvature based indicator is shown in detail. Furtherfore, the synthetic low-resolution frames are generated by a high-resolution image and the real low-resolution frames are provided by Multi-Dimensional Signal Processing Research Group (MDSP) [67]. Compared with other 6 methods, the experimental results for synthetic data and real data both show that our proposed algorithm outperforms other methods.

5.1 Quality of the estimated initial HR image

In our proposed multi-frame super-resolution algorithm, we use a novel method to estimate the initial HR image. Different from the traditional methods only using information from the reference image, our proposed algorithm combines all detail information in LR frames to establish an HR image as our initial setting. As introduced in Chapter 4, the LR frames are first warped according to the shape of the reference LR image. Then, a median operator is used to eliminate isolate noise and reconstruct a composed LR image. At last, the initial HR image is obtained by using bicubic interpolation.

In our experiments, we first add Gaussian noise in our generated LR images. Our interpolated LR frames are obtained by applying some operators on the original HR image, which are affine motion operator, blurring operator and down-sampling operator. The affine motion operator decides the geometric motion matrix which generates different LR frames from an HR image. The blurring operator represents a blur kernel with the window size of 4 and standard deviation of 0.4. The down-sampling operator decimates the HR image with the zoom factor of 2.

Figure 5.1 shows the comparison of our initialization method and some widely used interpolation methods such as bilinear interpolation and bicubic interpolation with Gaussian noise. The mean and variance of additive Gaussian noise is 0 and 0.05 respectively. Since the traditional interpolation methods only use the detail information of the reference LR images, they have limited performance to suppress the Gaussian noise. Different from traditional interpolation methods, our composed LR image is the combination of all warped LR images, which makes it noise-insensitive. Therefore, our initial HR image has better visual quality and higher PSNR and SSIM

values than other traditional initial settings.

(a) Original Lena image (b) First frame of LR sequence (c) Bilinear interpolation

(d) Bicubic interpolation (e) Our initialization method

Figure 5.1: The comparison of different initial estimations with Gaussian noise.

Moreover, except for adding Gaussian noise in our test images, *Salt&Pepper* noise is taken into consideration as well to simulate broken sensors and transition errors. In our experiments, the variance of additive *Salt&Pepper* is 0.02. Figure 5.2 shows the results of different initialization methods with *Salt&Pepper* noise respectively. Similar to above analyses, the interpolation methods can not eliminate *Salt&Pepper* noise because they only use the information from the reference low-resolution image. However, for our initial estimation, the lost information caused by *Salt&Pepper* noise can be compensated by other low-resolution images, which makes our initial HR image

have ability to estimate most of the *Salt&Pepper* noise.

(a) Original Lena image (b) First frame of LR sequence (c) Bilinear interpolation

(d) Bicubic interpolation (e) Our initialization method

Figure 5.2: The comparison of different initial estimations with *Salt&Pepper* noise.

Except for visual comparison, we also use image quality assessment to evaluate our novel initialization method. Table 5.1 gives the PSNR (dB) and SSIM results of different initialization methods for *Lena* image with Gaussian noise and *Salt&Pepper* noise respectively.

From the table, our initial HR image has higher PSNR and SSIM than other results, which demonstrates that our initialization method has better performance. In general, multi-frame super-resolution methods use optimization algorithm to minimize the objective function. Due to the iterative process, the initial HR image directly

Table 5.1: PSNR (dB) and SSIM results of different initialization methods for *Lena* image with Gaussian noise and *Salt&Pepper* noise respectively

Method	*Salt&Pepper* noise		Gaussian noise	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
bilinear interpolation	24.2403	0.8661	26.4310	0.8931
bicubic interpolation	22.9268	0.8475	25.7472	0.8741
Our initialization	**28.4663**	**0.9379**	**28.2466**	**0.9327**

influences the quality of final estimated HR image. Therefore, this novel initialization method can effectively improve the final result of our proposed multi-frame super-resolution algorithm.

5.2 The performance of difference curvature based indicator

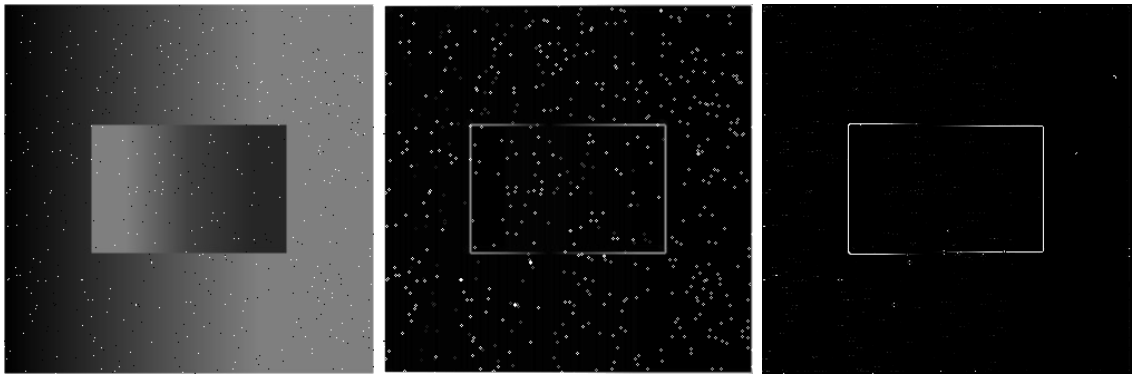
In this thesis, we propose a novel BTV regularization method based on difference curvature. Therefore, the indicator of difference curvature directly affects the performance of our regularization term. Compared with other traditional indicators such as gradient indicator, the Difference Curvature (DC) based indicator combines the first-order and second-order gradients together to distinguish image edges from isolated noise which is independent distributed. In our experiments, we add Gaussian noise and *Salt&Pepper* noise respectively to the test images to show the good performance of difference curvature based indicator. Besides, the traditional gradient indicator is

also generated as a comparison which is expressed as

$$G = \sqrt{I_x(x,y)^2 + I_y(x,y)^2}, \tag{5.1}$$

where $I_x(x,y)$ and $I_y(x,y)$ are the first-order partial derivatives of image I.

For the test image, it should contain ramp areas and isolated noise to make the distinction more complicated. Figure 5.3 shows the comparison of gradient indicator and difference curvature based indicator by adding *Salt&Pepper* noise to the test image. The variance of added *Salt&Pepper* noise is 0.005.



(a) Test image with *Salt&Pepper* noise (b) Processed by gradient indicator (c) Processed by difference curvature based indicator

Figure 5.3: Compare two indicators with *Salt&Pepper* noise.

The test image with the size of 330 × 330 is generated by ourselves. It contains two rectangle areas which have inverse color changing directions. As seen from left to right in Figure 5.3 (a), the inner rectangle is changing from gray to black and the outer rectangle is from black to gray which represents ramp regions mentioned in Table 4.1. For the gradient indicator, it can not distinguish image edges from the noise since the pixel values for both of them are in the same range. However, the

difference curvature based indicator suppresses most of the noise pixels and the edge pixels are still preserved at the same time. Figure 5.4 shows one line profile of the middle row in the test image, which demonstrates the performance of indicators to distinguish image edges from the noise.

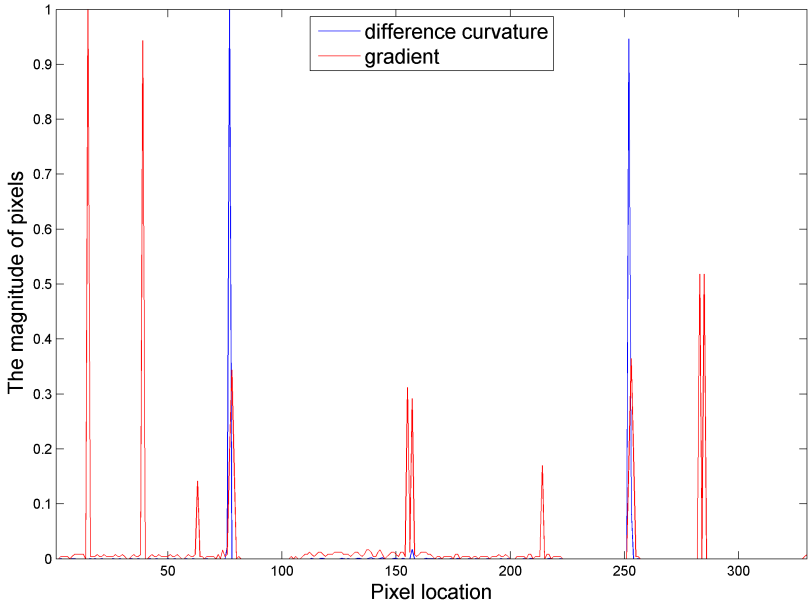
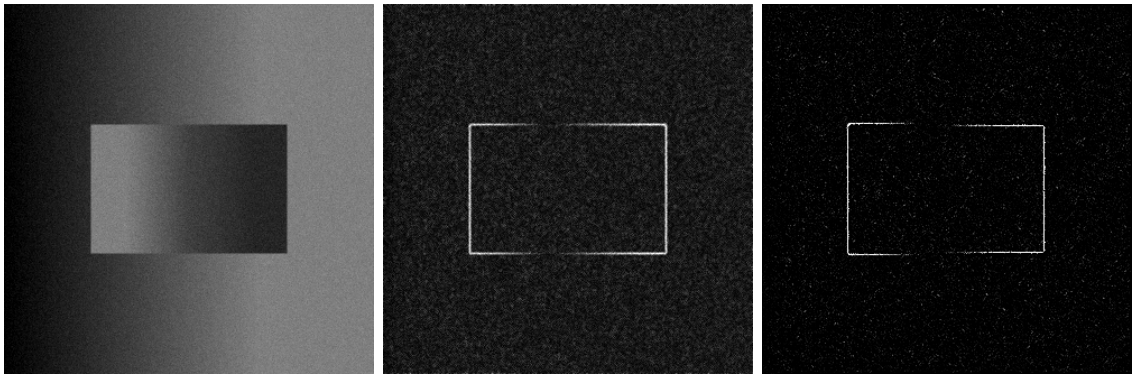


Figure 5.4: One line profile of the test image with *Salt&Pepper* noise.

As shown above, the red line represents the gradient indicator and the blue line stands for the difference curvature based indicator. The $78th$ pixel and $253th$ pixel in the middle row is the boundary of the inner rectangle and all pixels are normalized within $[0, 1]$. If the image noise is suppressed well, the one line profile should only have two peaks in the inner edges since the magnitude of edge pixels are much bigger than the ones in smooth areas. However, due to the effect of *Salt&Pepper* noise, the gradient indicator has eight peaks rather than just two. For the difference curvature based indicator, the noise is suppressed by the difference between $I_{\eta\eta}$ and $I_{\xi\xi}$. Therefore, there are only two peaks in the edges of inner rectangle.

Moreover, we use Gaussian noise to corrupt our test image as well. The mean and variance of the added Gaussian noise are 0 and 5×10^{-4} respectively and it is generated by Matlab built-in function *imnoise*. Figure 5.5 shows the comparison of the gradient indicator and the difference curvature based indicator. With the effect of Gaussian noise, the gradient indicator and the difference curvature based indicator both can highlight image edges from the Gaussian noise. However, the difference curvature based indicator suppresses the noise more efficiently. The Gaussian noise in Figure 5.5 (c) is almost eliminated compared with Figure 5.5 (b).



(a) Test image with Gaussian noise (b) Processed by gradient indicator (c) Processed by difference curvature based indicator

Figure 5.5: Compare two indicators with Gaussian noise.

Similar to the analysis of the test image with *Salt&Pepper* noise, one line profile of the middle row in the test image with Gaussian noise is shown in Figure 5.6. For the Gaussian noise, the gradient indicator and the difference curvature based indicator both have high magnitudes in image edges. However, the difference curvature based indicator has higher magnitudes in image edges and lower magnitudes in the location of Gaussian noise, which makes it have better performance of distinguishing image edges from Gaussian noise than the gradient indicator.

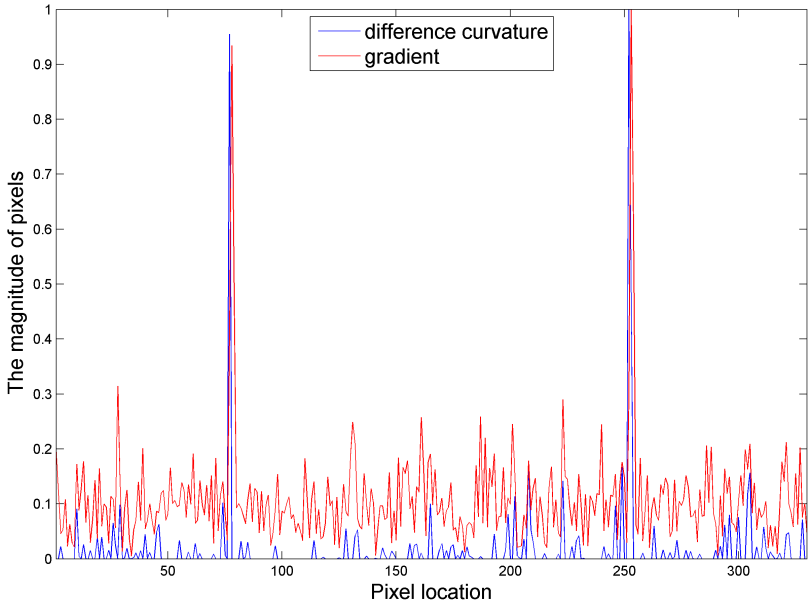


Figure 5.6: One line profile of the test image with Gaussian noise.

From the above analyses, difference curvature has good performance of distinguishing image edges from noises such as Gaussian noise and *Salt&Pepper* noise. Due to this property, it is used to make the traditional BTV regularization adaptive. For the pixel which contains edge information, the corresponding weight is small enough to preserve it. In contrast, if the pixel is in noise area, the adaptive weight is large enough to suppress it.

5.3 Proposed DCBTV based multi-frame super-resolution with adaptive norm choice

In this section, we use both artificially generated synthetic data and real data to test our proposed multi-frame super resolution, and compare its performance with

some other state-of-the-art algorithms. The synthetic data is created by using one HR image and the real data is provided by MDSP dataset [67]. Besides, the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) are used to measure the quality of our estimated HR images. In our experiments, the first frame of low-resolution sequence is chosen as our reference frame and the initial setting of high-resolution image is obtained by our novel initialization method. The intensity range of our test images is set to be $[0, 1]$. For color images, we use RGB model as our color model and apply our algorithm on all channels. The regularization parameter λ is selected to produce best qualities in visual comparison and image assessment matrices. The iterative number is selected by considering the balance between the computational complexity and the degree of convergence. Except for the nature images used in our experiments, our proposed algorithm can also be used for medical images. However, due to the space limitation, we don't demonstrate them in this thesis.

5.3.1 Experiments on simulated data

Our simulated data is generated to conduct a quantitative evaluation of our proposed algorithm since there is no ground truth image for real data. For each dataset, we generate 16 low-resolution frames from a high-resolution image. The displacement of every frame is simulated as a rigid motion. Therefore, the HR image is displaced by uniform distributed random translations and rotations. The range of random translations is from -2 to 2 pixels and the rotation angles are randomly changing from -1° to 1°. Then the displaced HR frames are blurred by a 4×4 Gaussian kernel with $\sigma = 0.4$ and subsampled with factor r. Gaussian noise and *Salt&Pepper*

noise are added in simulated LR sequence simultaneously as mixed noises to increase the difficulty of accurate estimation. The mean and variance of additive Gaussian noise is $\mu_G = 0$ and $\sigma_G = 0.02$ respectively. For *Salt&Pepper* noise, the variance is set to be $\sigma_{SP} = 0.02$ as well. The regularization parameter λ is a vital parameter to balance the fidelity term and the regularization term. In this case, λ is set to 3×10^{-4} in our experiments. Besides, the Scaled Conjugate Gradient (SCG) is used to minimize the objective function. For parameter settings in the SCG method, the termination criterion is set to be $\eta_t = 10^{-3}$. Since the minimization function usually converges within 25 iterations, the maximum iteration number is set to be $T_{max} = 25$. In general, the most commonly used down-sampling factor in multi-frame super-resolution algorithms is 2. Therefore, due to the space limitation, we only present the simulated images with $r = 2$ in this thesis. For the images with higher zoom factors such as 3 and 4, we present them in the experiments on real data. For motion estimation, the Enhanced Correlation Coefficient (ECC) [68, 69] is used to estimate the subpixel movement between two LR frames. In our experiments on the simulated data, we choose four widely used test images in super-resolution field to generate our simulated LR sequences which are *Cameraman*, *Lena*, *Butterfly* and *Hat*. Figure 5.7, Figure 5.8, Figure 5.9 and Figure 5.10 show them respectively and present the comparation of our proposed method with other state-of-the-art methods. From sub-figure (a) to (i), the ground truth image, the reference LR image, the result of bicubic interpolation, the result of L_2 + Tikhonov[14], the result of L_2 + BTV[11], the result of L_1 + BTV[11], the result of Bilateral Edge Preserving (BEP) method [60], the result of Iteratively Re-weighted (IRW) method [8] and our proposed algorithm are shown respectively.

(a) Ground truth (b) Reference LR image (c) Bicubic interpolation

(d) L_2 + Tikhonov (e) L_2 + BTV (f) L_1 + BTV

(g) BEP (h) IRW (i) Proposed

Figure 5.7: Visual comparison of multi-frame super-resolved results from different methods for *Cameraman* image with mixed noises.

(a) Ground truth (b) Reference LR image (c) Bicubic interpolation

(d) L_2 + Tikhonov (e) L_2 + BTV (f) L_1 + BTV

(g) BEP (h) IRW (i) Proposed

Figure 5.8: Visual comparison of multi-frame super-resolved results from different methods for *Lena* image with mixed noises.

(a) Ground truth (b) Reference LR image (c) Bicubic interpolation

(d) L_2 + Tikhonov (e) L_2 + BTV (f) L_1 + BTV

(g) BEP (h) IRW (i) Proposed

Figure 5.9: Visual comparison of multi-frame super-resolved results from different methods for *Hat* image with mixed noises.

(a) Ground truth (b) Reference LR image (c) Bicubic interpolation

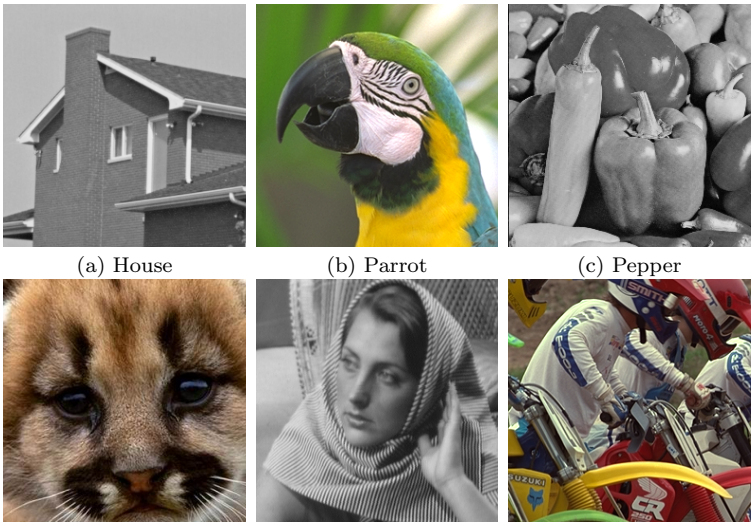
(d) L_2 + Tikhonov (e) L_2 + BTV (f) L_1 + BTV

(g) BEP (h) IRW (i) Proposed

Figure 5.10: Visual comparison of multi-frame super-resolved results from different methods for *Parrot* image with mixed noises.

For all of our simulated data, the mixed noises are added. Therefore, L_2 based methods such as $L_2+Tikhonov$ and L_2+BTV have poor performance to suppress the $Salt\&Pepper$ noise and registration errors. For L_1+BTV, it is robust to $Salt\&Pepper$ noise and estimation errors. Compared with $L_2+Tikhonov$, the results of L_1+BTV have better quality in most cases. However, since the additive noises contain Gaussian noise, the L_1+BTV introduces more observation errors than other methods. BEP algorithm uses an adaptive error norm in fidelity and regularization terms respectively. Therefore, it can adaptively choose error norm to suppress the $Salt\&Pepper$ noise and Gaussian noise. Our experiments show the results of BEP method are better than the traditional norm fixed methods, but some noise sill exists due to its simple initial estimation. Kohler *et al.* proposed a Iteratively Re-Weighted (IRW) minimization method in 2016 which is derived from a spatially adaptive Bayesian model to consider variant noise and outliers. It has good performance of suppressing mixed noises and preserving image texture. Compared with these state-of-the-art methods, our proposed algorithm has better quality in visual comparison and higher PSNR (dB) and SSIM values in most cases.

Table 5.2 shows the PSNR (dB) and SSIM comparisons of 7 different algorithms. Except for 4 commonly used images shown above, we also test on other six images to make our experiments more convincing. The other six images are shown in Figure 5.11 respectively.



(a) House (b) Parrot (c) Pepper

(c) Raccon (c) Barbara (c) Bike

Figure 5.11: Other images used to test the performance of our proposed method.

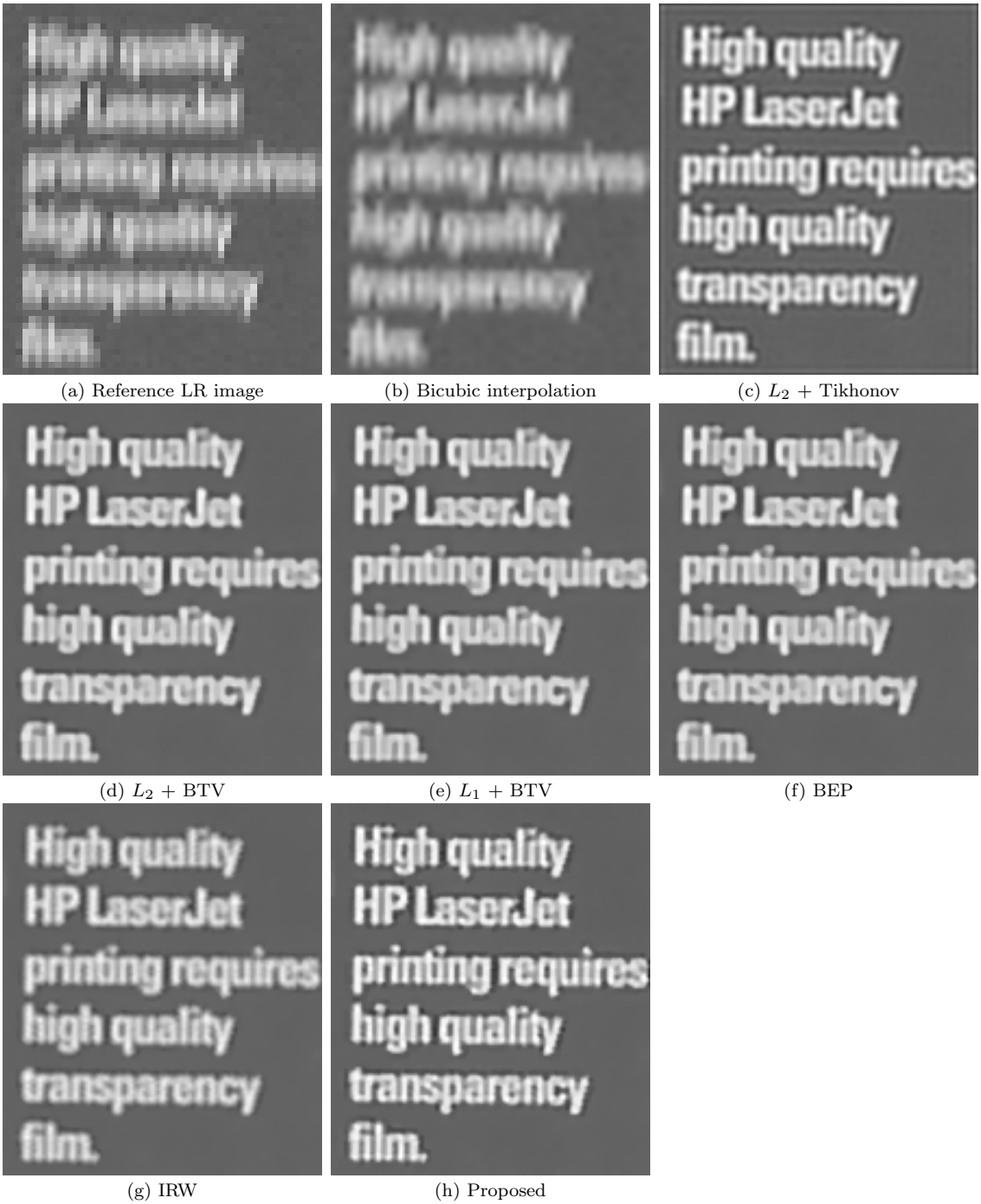
Table 5.2: PSNR(dB)/SSIM results of multi-frame super-resolved images from 7 different algorithms under ratio=2 and mixed noises

Images	Bicubic	L2+tikhonove	L2+BTV	L1+BTV	BEP	IRW	Proposed
Bike	20.6443	25.1011	25.1625	28.1603	28.7804	29.7308	**29.8507**
	0.7545	0.8714	0.8834	0.9247	0.9495	0.9582	**0.9593**
Butterfly	21.2965	25.9730	27.6616	28.9103	30.4342	31.7631	**32.1836**
	0.8938	0.9432	0.9653	0.9674	0.9820	0.9873	**0.9876**
Lena	22.5737	27.3270	29.4337	29.6870	30.7712	31.9757	**33.1956**
	0.8448	0.9094	0.9409	0.9426	0.9596	0.9670	**0.9748**
Hat	22.5973	27.1948	29.5654	30.0563	31.6839	32.0300	**33.7073**
	0.8702	0.9107	0.9632	0.9436	0.9750	0.9699	**0.9843**
Parrot	22.2565	27.2338	29.7220	29.9128	32.0377	34.3445	**34.4453**
	0.8318	0.8585	0.9363	0.9053	0.9570	**0.9757**	0.9735
Raccoon	22.5195	27.2443	27.7986	31.1457	30.3571	32.0768	**32.1816**
	0.7960	0.8771	0.8895	0.9467	0.9468	0.9619	**0.9626**
Cameraman	21.7009	25.6756	25.2985	27.3820	28.7688	29.9337	**30.1120**
	0.6222	0.6152	0.7566	0.8427	0.8722	0.8892	**0.8952**
House	23.6665	28.2798	31.6099	30.9898	32.8833	35.0856	**35.4820**
	0.6435	0.6215	0.8213	0.7246	0.8638	**0.9077**	0.9019
Pepper	22.2396	25.7537	29.1594	29.3307	32.1410	31.1859	**32.4112**
	0.6708	0.6586	0.8301	0.7825	0.8912	0.8955	**0.9017**
Barbara	21.8615	25.2559	25.2462	28.5042	27.1066	**28.5746**	28.4528
	0.6371	0.6920	0.7411	0.8573	0.8419	0.8796	**0.8888**

5.3.2 Experiments on real data

In this section, we use real data to test our proposed multi-frame super-resolution algorithm in practical application. The real data is provided by Multi-Dimensional Signal Processing Research Group (MDSP) [67], which is the most widely used dataset to test the performance of multi-frame SR methods. Since this data set is shot by real camera, there are no ground truth images. Therefore, the image assessment matrices such as PSNR and SSIM can not be used to evaluate the quality of real images. Nowadays, some blind image assessment methods are proposed in [70, 71]. But their assessment results are not very stable. Therefore, in our thesis, we only use visual comparison to assess image quality for real data. Moreover, for real data, the PSF kernel is unknown. To simplify this bind deblurring problem, we assume the unknown PSF kernel is a 4×4 Gaussian kernel with $\sigma = 0.4$, which is the same as the simulated data.

For the real data, large scale factors such as 3 and 4 are used to reconstruct our high-resolution images. Figure 5.12, Figure 5.13 and Figure 5.14 show the visual comparison of super-resolved results from different multi-frame SR methods with the scale factor of 3. In Figure 5.15 and Figure 5.16, the results of *disk* and *EIA* are presented with the scale factor of 4. Compared with other methods, our estimated HR images can effectively suppress the errors caused by noise, registration and bad estimation of unknown PSF kernels. Besides, the detail information in real images is preserved well, which makes our super-resolved images have good quality in visual comparison.

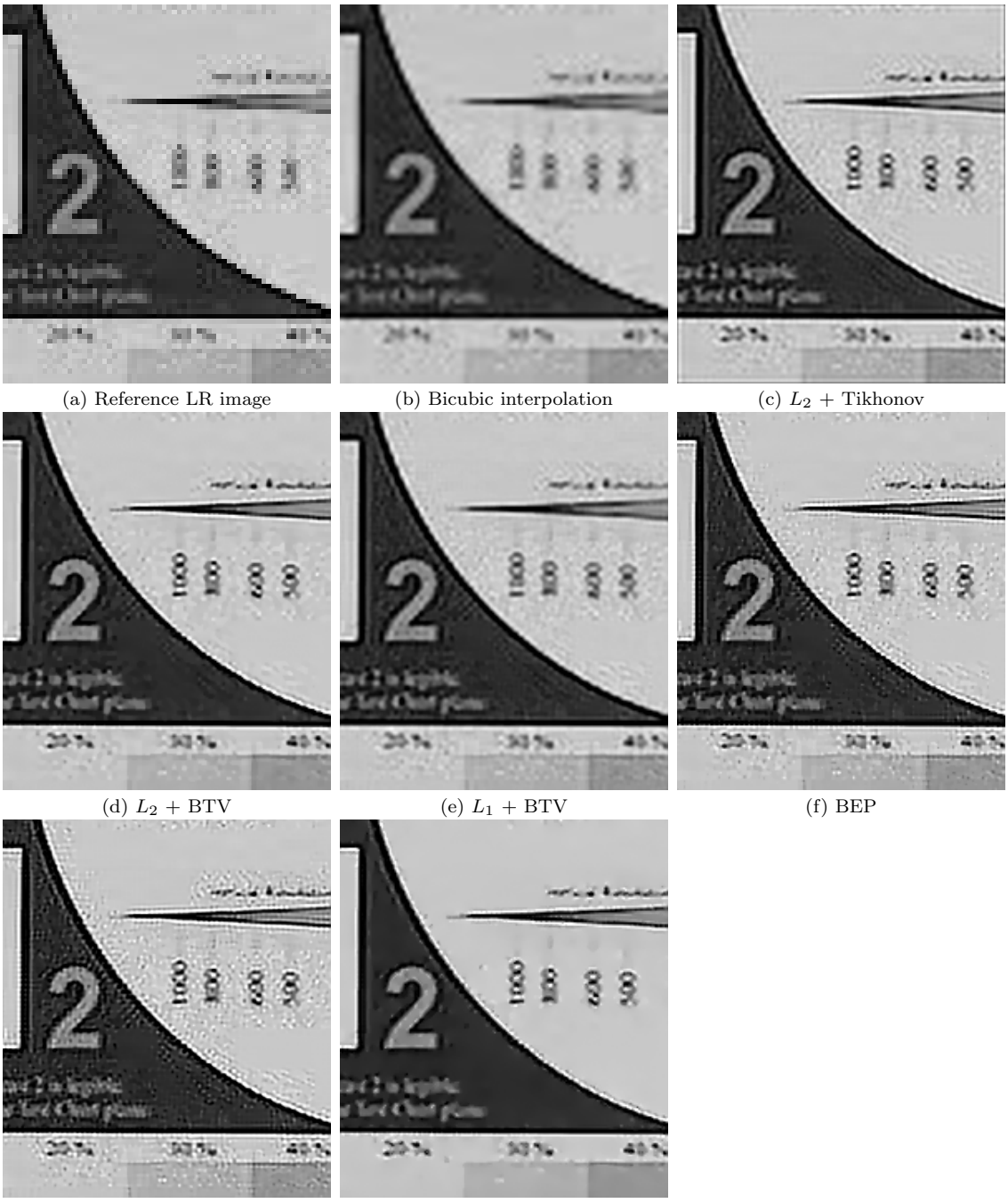


(a) Reference LR image (b) Bicubic interpolation (c) L_2 + Tikhonov

(d) L_2 + BTV (e) L_1 + BTV (f) BEP

(g) IRW (h) Proposed

Figure 5.12: Visual comparison of multi-frame super-resolved results from different algorithms for *text* frames (r=3).



(a) Reference LR image (b) Bicubic interpolation (c) L_2 + Tikhonov

(d) L_2 + BTV (e) L_1 + BTV (f) BEP

(g) IRW (h) Proposed

Figure 5.13: Visual comparison of multi-frame super-resolved results from different algorithms for *adyoron* frames (r=3).

(a) Reference LR image (b) Bicubic interpolation

(c) L_2 + Tikhonov (d) L_2 + BTV

(e) L_1 + BTV (f) BEP

(g) IRW (h) Proposed

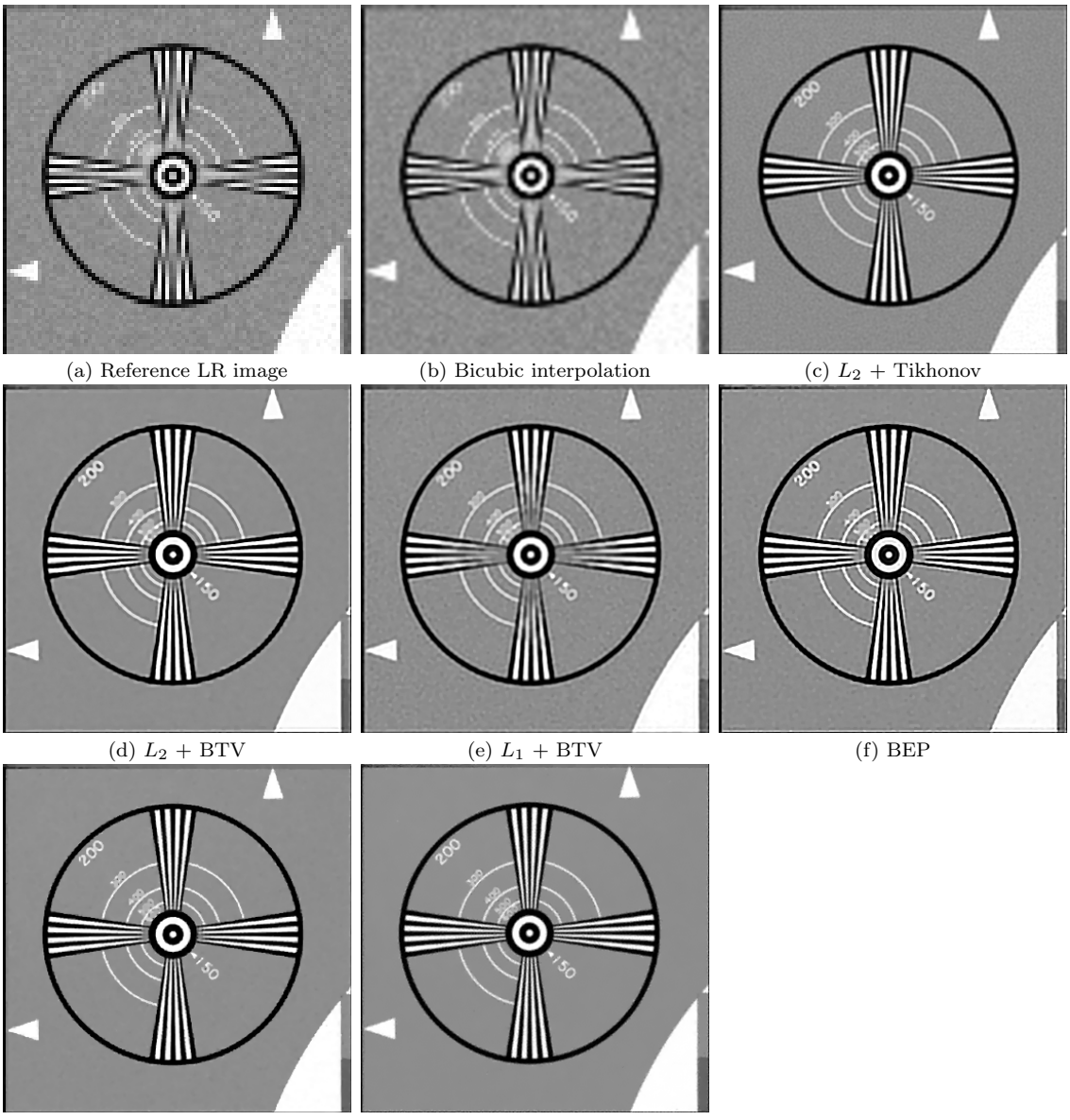
Figure 5.14: Visual comparison of multi-frame super-resolved results from different algorithms for *book* frames (r=3).

(a) Reference LR image (b) Bicubic interpolation (c) L_2 + Tikhonov

(d) L_2 + BTV (e) L_1 + BTV (f) BEP

(g) IRW (h) Proposed

Figure 5.15: Visual comparison of multi-frame super-resolved results from different algorithms for *disk* frames (r=4).



(a) Reference LR image (b) Bicubic interpolation (c) L_2 + Tikhonov

(d) L_2 + BTV (e) L_1 + BTV (f) BEP

(g) IRW (h) Proposed

Figure 5.16: Visual comparison of multi-frame super-resolved results from different algorithms for *EIA* frames (r=4).

5.4 Summary

In this chapter, we first presented the results of our initialization method in detail. Compared with other traditional initial settings, our novel initialization method can effectively remove most of the mixed noises in LR images, which directly improves the final results. Then we analyze the properties of difference curvature based indicator. In our experiments, this indicator has good performance of distinguishing image edges from mixed noises. Motivated by this property, we proposed the difference curvature based BTV regularization algorithm. Finally, the synthetic data and real data were both used to test our proposed framework. Compared with the other 6 methods, our super-resolved results have better visual and higher values in assessment matrices.

Chapter 6

Conclusions

Super-resolution problem has been studied for decades. In this thesis, we first introduced some basic concepts and methods in the field of multi-frame image super-resolution and then proposed our own algorithm. Different from single-image super-resolution, multi-frame super-resolution has more useful information hiding in low-resolution sequences. Therefore, the biggest problem in multi-frame super-resolution is how to obtain the useful information correctly. Our proposed method is based on Maximum a Posteriori (MAP) estimation, which is widely used to solve this problem.

Based on MAP estimation, we proposed our multi-frame super-resolution algorithm. In our framework, the initial HR image is estimated by a novel initialization method based on median operator. For our fidelity term, the half-quadratic estimation is used to make error norm adaptively chosen instead of using fixed L_1 or L_2 norm. And for the regularization term, we proposed a novel regularization method based on difference curvature to suppress mixed noises and preserve image edges automatically according to the local information. In our experimental results, both

simulated data and real data are tested to present the performance of our algorithm. Due to the improvements of initial value, fidelity term and regularization term, our super-resolved results have better quality in visual comparison and higher values of PSNR and SSIM when compared with other state-of-the-art methods in most cases.

References

[1] R. Tsai and T. S. Huang, "Multiframe image restoration and registration," *Advances in Computer Vision and Image Processing*, vol. 1, no. 2, pp. 317–339, 1984.

[2] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1187–1193, 2001.

[3] M.-C. Chiang and T. E. Boult, "Efficient super-resolution via image warping," *Image and Vision Computing*, vol. 18, no. 10, pp. 761–771, 2000.

[4] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[5] J. Tian and K.-K. Ma, "A survey on super-resolution imaging," *Signal, Image and Video Processing*, vol. 5, no. 3, pp. 329–342, 2011.

[6] H. Greenspan, "Super-resolution in medical imaging," *The Computer Journal*, vol. 52, no. 1, pp. 43–63, 2009.

[7] P. Milanfar, *Super-resolution imaging*. CRC Press, 2010.

[8] T. Köhler, X. Huang, F. Schebesch, A. Aichert, A. Maier, and J. Hornegger, "Robust multiframe super-resolution employing iteratively re-weighted minimization," *IEEE Transactions on Computational Imaging*, vol. 2, no. 1, pp. 42–58, 2016.

[9] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Overcoming registration uncertainty in image super-resolution: maximize or marginalize?" *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 1–14, 2007.

[10] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian image super-resolution, continued," in *Advances in Neural Information Processing Systems*, 2006, pp. 1089–1096.

[11] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[12] V. Patanavijit and S. Jitapunkul, "A robust iterative multiframe super-resolution reconstruction using a huber bayesian approach with huber-tikhonov regularization," in *IEEE Conference on International Symposium on Intelligent Signal Processing and Communications*, 2006, pp. 13–16.

[13] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47–57, 2004.

[14] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646–1658, 1997.

[15] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique," *Princeton University Bulletin*, vol. 13, no. 49-52, p. 28, 1902.

[16] G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov regularization and total least squares," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 1, pp. 185–194, 1999.

[17] S. Borman and R. L. Stevenson, "Super-resolution from image sequences-a review," in *Midwest Symposium on Circuits and Systems*, 1998, pp. 374–378.

[18] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

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

[20] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 1, pp. 13–27, 1984.

[21] P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 233–233, 2006.

[22] S. Kim, N. K. Bose, and H. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, 1990.

[23] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images," in *IEEE International Conference on Image Processing*, 1995, vol. 2, pp. 539–542.

[24] N. Nguyen, P. Milanfar, and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1299–1308, 2001.

[25] N. K. Bose, S. Lertrattanapanich, and M. B. Chappalli, "Superresolution with second generation wavelets," *Signal Processing: Image Communication*, vol. 19, no. 5, pp. 387–391, 2004.

[26] H. Song, X. He, W. Chen, and Y. Sun, "An improved iterative back-projection algorithm for video super-resolution reconstruction," in *2010 Symposium on Photonics and Optoelectronics*, 2010, pp. 1–4.

[27] S. Peleg, D. Keren, and L. Schweitzer, "Improving image resolution using subpixel motion," *Pattern Recognition Letters*, vol. 5, no. 3, pp. 223–226, 1987.

[28] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.

[29] S. Chaudhuri, *Super-resolution imaging*. Springer Science & Business Media, 2001, vol. 632.

[30] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *Journal of the Optical Society of America A*, vol. 6, no. 11, pp. 1715–1726, 1989.

[31] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-92.*, 1992, vol. 3, pp. 169–172.

[32] H. Trussell and M. Civanlar, "The feasible solution in signal restoration," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 201–212, 1984.

[33] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Robust methods for high-quality stills from interlaced video in the presence of dominant motion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 328–342, 1997.

[34] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1064–1076, 1997.

[35] J. Besag, J. York, and A. Mollié, "Bayesian image restoration, with two applications in spatial statistics," *Annals of the Institute of Statistical Mathematics*, vol. 43, no. 1, pp. 1–20, 1991.

[36] W. H. Richardson, "Bayesian-based iterative method of image restoration," *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55–59, 1972.

[37] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 721–741, 1984.

[38] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[39] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[40] H. Zhang, Y. Zhang, H. Li, and T. S. Huang, "Generative bayesian image super resolution with natural image prior," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4054–4067, 2012.

[41] R. R. Schultz and R. L. Stevenson, "A bayesian approach to image expansion for improved definition," *IEEE Transactions on Image Processing*, vol. 3, no. 3, pp. 233–242, 1994.

[42] D. J. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992.

[43] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to superresolution," in *SPIE's 48th Annual Meeting on Optical Science and Technology*, 2003, pp. 121–130.

[44] S. Borman and R. L. Stevenson, "Simultaneous multi-frame map super-resolution

video enhancement using spatio-temporal priors," in *IEEE International Conference on Image Processing, ICIP*, 1999, vol. 3, pp. 469–473.

[45] H. Shen, L. Zhang, B. Huang, and P. Li, "A map approach for joint motion estimation, segmentation, and super resolution," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 479–490, 2007.

[46] T. Kasetkasem, M. K. Arora, and P. K. Varshney, "Super-resolution land cover mapping using a markov random field based approach," *Remote Sensing of Environment*, vol. 96, no. 3, pp. 302–314, 2005.

[47] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[48] H. He and L. P. Kondi, "Resolution enhancement of video sequences with simultaneous estimation of the regularization parameter," *Journal of Electronic Imaging*, vol. 13, no. 3, pp. 586–596, 2004.

[49] 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 Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.

[50] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, 1996.

[51] J. Tian and K.-K. Ma, "A mcmc approach for bayesian super-resolution image

reconstruction," in *IEEE International Conference on Image Processing*, 2005, vol. 1, pp. I–45.

[52] J. Tian and K.-K. Ma, "Edge-adaptive super-resolution image reconstruction using a markov chain monte carlo approach," in *IEEE International Conference on Information, Communications & Signal Processing*, 2007, pp. 1–5.

[53] B. T. Oh and C.-C. J. Kuo, "Super-resolution texture synthesis using stochastic par/nl model," *Journal of Visual Communication and Image Representation*, vol. 23, no. 7, pp. 995–1007, 2012.

[54] R. Molina, M. Vega, J. Abad, and A. K. Katsaggelos, "Parameter estimation in bayesian high-resolution image reconstruction with multisensors," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1655–1667, 2003.

[55] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," Carnegie-Mellon University, Department of Computer Science, 1994.

[56] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993.

[57] L. Yue, H. Shen, Q. Yuan, and L. Zhang, "A locally adaptive L_1- L_2 norm for multi-frame super-resolution of images with mixed noise and outliers," *Signal Processing*, vol. 105, pp. 156–174, 2014.

[58] P. Rodriguez and B. Wohlberg, "An iteratively reweighted norm algorithm for total variation regularization," in *Fortieth Asilomar Conference on Signals, Systems and Computers*, 2006, pp. 892–896.

[59] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point Algorithms for Inverse Problems in Science and Engineering*. Springer, 2011, pp. 185–212.

[60] X. Zeng and L. Yang, "A robust multiframe super-resolution algorithm based on half-quadratic estimation with modified BTV regularization," *Digital Signal Processing*, vol. 23, no. 1, pp. 98–109, 2013.

[61] J. Nocedal and S. Wright, *Numerical Optimization*. Springer Science & Business Media, 2006.

[62] G. Wolberg, *Digital Image Warping*. IEEE Computer Society Press Los Alamitos, CA, 1990, vol. 10662.

[63] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Transactions on Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.

[64] Q. Chen, P. Montesinos, Q. S. Sun, P. A. Heng *et al.*, "Adaptive total variation denoising based on difference curvature," *Image and Vision Computing*, vol. 28, no. 3, pp. 298–306, 2010.

[65] R. Battiti, "First-and second-order methods for learning: between steepest descent and newton's method," *Neural Computation*, vol. 4, no. 2, pp. 141–166, 1992.

[66] T. Steihaug, "The conjugate gradient method and trust regions in large scale optimization," *SIAM Journal on Numerical Analysis*, vol. 20, no. 3, pp. 626–637, 1983.

[67] S. Farsiu, "MDSP super-resolution and demosaicing datasets," https://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html.

[68] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[69] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1858–1865, 2008.

[70] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[71] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.