

Research Article

A Total Variation Regularization Based Super-Resolution Reconstruction Algorithm for Digital Video

Michael K. Ng,¹ Huanfeng Shen,^{1,2} Edmund Y. Lam,³ and Liangpei Zhang²

¹Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

²The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, China

³Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong

Received 13 September 2006; Revised 12 March 2007; Accepted 21 April 2007

Recommended by Russell C. Hardie

Super-resolution (SR) reconstruction technique is capable of producing a high-resolution image from a sequence of low-resolution images. In this paper, we study an efficient SR algorithm for digital video. To effectively deal with the intractable problems in SR video reconstruction, such as inevitable motion estimation errors, noise, blurring, missing regions, and compression artifacts, the total variation (TV) regularization is employed in the reconstruction model. We use the fixed-point iteration method and preconditioning techniques to efficiently solve the associated nonlinear Euler-Lagrange equations of the corresponding variational problem in SR. The proposed algorithm has been tested in several cases of motion and degradation. It is also compared with the Laplacian regularization-based SR algorithm and other TV-based SR algorithms. Experimental results are presented to illustrate the effectiveness of the proposed algorithm.

Copyright © 2007 Michael K. Ng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Solid-state sensors such as CCD or CMOS are widely used nowadays in many image acquisition systems. Such sensors consist of rectangular arrays of photodetectors where their physical sizes limit the spatial resolution of acquired images. In order to increase the spatial resolution of images, one possibility is to reduce the size of rectangular array elements by using advanced sensor fabrication techniques. However, this method would lead to a small signal-to-noise ratio (SNR) because the amount of photons collected in each photodetector decreases correspondingly. On the other hand, the cost of manufacturing such sensors increases rapidly as the number of pixels in a sensor increases. Moreover, in some applications, we only obtain low-resolution (LR) images. In order to get a more desirable high-resolution (HR) images, super-resolution (SR) technique can be employed as an effective and efficient alternative.

Super-resolution image reconstruction refers to a process that produces an HR image from a sequence of LR images using the nonredundant information among them. It overcomes the inherent resolution limitation by bringing together the additional information from each LR image.

Generally, SR techniques can be divided into two classes of algorithms, namely, frequency domain algorithms and spatial domain algorithms. Most of the earlier SR work was developed in the frequency domain using discrete Fourier transform (DFT), such as the work of Tsai and Huang [1], Kim et al. [2, 3], and so on. More recently, discrete cosine transform- (DCT-) based [4] and wavelet transform-based [5–7] SR methods have also been proposed. In the spatial domain, typical reconstruction models include nonuniform interpolation [8], iterative back projection (IBP) [9], projection onto convex sets (POCS) [10–13], maximum likelihood (ML) [14], maximum a posteriori (MAP) [15, 16], hybrid ML/MAP/POCS [17], and adaptive filtering [18]. Based on these basic reconstruction models, researchers have developed algorithms with a joint formulation of reconstruction and registration [19–22], and other algorithms for multispectral and color images [23, 24], hyper-spectral images [25], and compressed sequence of images [26, 27].

In this paper, we study a total-variation- (TV-) based SR reconstruction algorithm for digital video. We remark that the TV-based regularization has been applied to SR image reconstruction in literature [24, 28–31]. The contributions of this paper are threefold. Firstly, we present an efficient

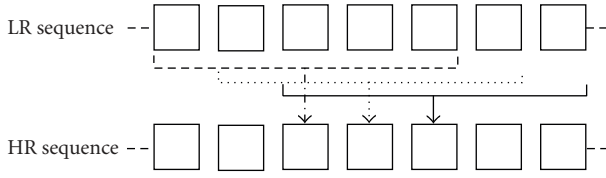


FIGURE 1: Illustration of the SR reconstruction of all frames in the video.

algorithm to solve the nonlinear TV-based SR reconstruction model using fixed-point and preconditioning methods. Preconditioned conjugate gradient methods with factorized banded inverse preconditioners are employed in the iterations. Experimental results show that our method is more efficient than the gradient descent method. Secondly, we combine image inpainting and SR reconstruction together to obtain an HR image from a sequence of LR images. We consider that there exist some missing and/or corrupted pixels in LR images. The filling-in of such missing and/or corrupted pixels in an image is called image inpainting [32]. By putting missing and/or corrupted pixels in the image observation model, the proposed algorithm can perform image inpainting and SR reconstruction simultaneously. Experimental results validate that it is more robust than the method of conducting image inpainting and SR reconstruction separately. Thirdly, while our algorithm is developed for the cases where raw uncompressed video data (such as a webcam directly linked to a host computer) is used, it can be applied to the MPEG compressed video. Simulation results show that the proposed algorithm is also capable of SR reconstruction with the compressed artifacts in the video.

It is noted that this paper aims to reconstruct an HR frame from several LR frames in the video. Using the proposed algorithm, all the frames in the video can be SR reconstructed in such a way [33]: for a given frame, a “sliding window” determines the set of LR frames to be processed to produce the output. The window is moved forward to produce successive SR frames in the output sequence. An illustration of this procedure is given in Figure 1.

The outline of this paper is as follows. In Section 2, we present the image observation model of the SR problem. The motion estimation methods used in this paper are described in Section 3. In Section 4, we present the TV regularization-based reconstruction algorithm. Experimental results are provided in Section 5. Finally, concluding remarks are given in Section 6.

2. IMAGE OBSERVATION MODEL

In SR image reconstruction, it is necessary to select a frame from the sequence as the referenced one. The image observation model is to relate the desired referenced HR image to all the observed LR images. Typically, the imaging process involves warping, followed by blurring and down-sampling to generate LR images from the HR image. Let the underlying HR image be denoted in the vector form by

$\mathbf{z} = [z_1, z_2, \dots, z_{L_1 N_1 \times L_2 N_2}]^T$, where $L_1 N_1 \times L_2 N_2$ is the HR image size. Letting L_1 and L_2 denote the down-sampling factors in the horizontal and vertical directions, respectively, each observed LR image has the size $N_1 \times N_2$. Thus, the LR image can be represented as $\mathbf{y}_k = [y_{k,1}, y_{k,2}, \dots, y_{k,N_1 \times N_2}]^T$, where $k = 1, 2, \dots, P$, with P being the number of LR images. Assuming that each observed image is contaminated by additive noise, the observation model can be represented as [17, 34, 35]

$$\mathbf{y}_k = \mathbf{D}\mathbf{B}_k\mathbf{M}_k\mathbf{z} + \mathbf{n}_k, \quad (1)$$

where \mathbf{M}_k is the motion (shift, rotation, zooming, etc.) matrix with the size of $L_1 N_1 L_2 N_2 \times L_1 N_1 L_2 N_2$, \mathbf{B}_k represents the blur (sensor blur, motion blur, atmosphere blur, etc.) matrix also of size $L_1 N_1 L_2 N_2 \times L_1 N_1 L_2 N_2$, \mathbf{D} is an $N_1 N_2 \times L_1 N_1 L_2 N_2$ down-sampling matrix, and \mathbf{n}_k represents the $N_1 N_2 \times 1$ noise vector.

In fact, in an unreferenced frame, there often exists occlusions that cannot be observed in the referenced frame. Obviously, these occlusions should be excluded in the SR reconstruction. Furthermore, there are also missing and/or corrupted pixels in the observed images in some cases. In order to deal with the occlusion problem and perform the image inpainting along with the SR, the observation model (1) should be expanded. We use the term *unobservable* to describe all the occluded, missing, and corrupted pixels, and *observable* to describe the other pixels. The unobservable pixels can be excluded by modifying the observation model as

$$\mathbf{y}_k^{\text{obs}} = \mathbf{O}_k(\mathbf{D}\mathbf{B}_k\mathbf{M}_k\mathbf{z} + \mathbf{n}_k), \quad (2)$$

where \mathbf{O}_k is an operator cropping the observable pixels from \mathbf{y}_k , and $\mathbf{y}_k^{\text{obs}}$ is the cropped result. This model provides the possibility to deal with the occlusion problem and to conduct simultaneous inpainting and SR. A block diagram corresponding to the degradation process of this model is illustrated in Figure 2.

3. MOTION ESTIMATION METHODS

Motion estimation/registration plays a critical role in SR reconstruction. In general, the subpixel motions between the referenced frame and the unreferenced frames can be modeled and estimated by a parameter model, or they may be scene dependent and have to be estimated for every point [36]. This section introduces the motion estimation methods employed in this paper. For a comparative analysis of the subpixel motion estimation methods in SR reconstruction, please refer to [37].

3.1. Parameter model-based motion estimation

Typically, if the objects in the scene remain stationary while the camera moves, the motions of all points often can be modeled by a parametric model. Generally, the relationship between the observed k th and l th frames can be expressed by

$$y_k(x_u, x_v) = y_l^{(l, \theta)}(x_u, x_v) + \varepsilon_{l,k}(x_u, x_v), \quad (3)$$



FIGURE 2: Block diagram illustration of the observation model (2), where the far left is the desired high-resolution image, and the far right is the observed image.

where (x_u, x_v) denotes the pixel site, $y_k(x_u, x_v)$ is a pixel in frame k , θ is the vector containing the corresponding motion parameters, $y_k^{(l, \theta)}(x_u, x_v)$ is the predicted pixel of $y_k(x_u, x_v)$ from frame l using parameter vector θ , and $\varepsilon_{l, k}(x_u, x_v)$ denotes the model error. In the literature, the six-parameter affine model and eight-parameter perspective model are widely used. Here we concentrate on the affine model, in which $y_k^{(l, \theta)}(x_u, x_v)$ can be expressed as

$$y_k^{(l, \theta)}(x_u, x_v) = y_l(a_0 + a_1 x_u + a_2 x_v, b_0 + b_1 x_u + b_2 x_v). \quad (4)$$

In this model, $\theta = (a_0, a_1, a_2, b_0, b_1, b_2)^T$ contains six geometric model parameters. To solve θ , we can employ the least square criterion, which has the following minimization cost function:

$$E(\theta) = \|\mathbf{y}_k - \mathbf{y}_k^{(l, \theta)}\|_2^2. \quad (5)$$

Using the Gaussian-Newton method, the six affine parameters can be iteratively solved by

$$\begin{aligned} \Delta\theta &= [(\mathbf{J}^n)^T \mathbf{J}^n]^{-1} [-(\mathbf{J}^n)^T \mathbf{r}^n], \\ \theta^{n+1} &= \theta^n + \Delta\theta. \end{aligned} \quad (6)$$

Here, n is the iteration number, $\Delta\theta$ denotes the corrections of the models parameters, \mathbf{r}^n is the residual vector that is equal to $\mathbf{y}_k - \mathbf{y}_k^{(l, \theta^n)}$, and $\mathbf{J}^n = \partial \mathbf{r}^n / \partial \theta^n$ denotes the gradient matrix of \mathbf{r}^n .

3.2. Optical flow-based motion estimation

In many videos, the scene may consist of independently moving objects. In this case, the motions cannot be modeled by a parametric model, but we can use optical flow-based methods to estimate the motions of all points. Here we introduce a simple MAP motion estimation method. Let us denote $\mathbf{m} = (\mathbf{m}_u, \mathbf{m}_v)$ as a 2D motion field which describes the motions of all points between the observed frames \mathbf{y}_k and \mathbf{y}_l with \mathbf{m}_u and \mathbf{m}_v being the horizontal and vertical fields, respectively, and $\mathbf{y}_k^{(l, \mathbf{m})}$ is the predicted version of \mathbf{y}_k from frame l using the motion field \mathbf{m} , the MAP motion estimation method has the following minimization function [38]:

$$E(\mathbf{m}) = \|\mathbf{y}_k - \mathbf{y}_k^{(l, \mathbf{m})}\|_2^2 + \lambda_1 U(\mathbf{m}), \quad (7)$$

where $U(\mathbf{m})$ describes prior information of the motion field \mathbf{m} , and λ_1 is the regularization parameter. In this paper, we

choose $U(\mathbf{m})$ as a Laplacian smoothness constraint consisting of the terms $\|\mathbf{Q}\mathbf{m}_u\|^2 + \|\mathbf{Q}\mathbf{m}_v\|^2$, where \mathbf{Q} is a 2D Laplacian operator. Using steepest descent method, we can iteratively solve the motion vector field by

$$\hat{\mathbf{m}}_u^{n+1} = \hat{\mathbf{m}}_u^n + \alpha \left[\frac{\partial \mathbf{y}_k^{(l, \mathbf{m})}}{\partial \mathbf{m}_u} (\mathbf{y}_k - \mathbf{y}_k^{(l, \mathbf{m})}) - \lambda_1 \mathbf{Q}^T \mathbf{Q} \mathbf{m}_u \right], \quad (8)$$

$$\hat{\mathbf{m}}_v^{n+1} = \hat{\mathbf{m}}_v^n + \alpha \left[\frac{\partial \mathbf{y}_k^{(l, \mathbf{m})}}{\partial \mathbf{m}_v} (\mathbf{y}_k - \mathbf{y}_k^{(l, \mathbf{m})}) - \lambda_1 \mathbf{Q}^T \mathbf{Q} \mathbf{m}_v \right],$$

where n again is the iteration number, and α is the step size. The derivative in the above equation is computed on a pixel-by-pixel basis, given by

$$\begin{aligned} \frac{\partial \mathbf{y}_k^{(l, \mathbf{m})}(x_u, x_v)}{\partial \mathbf{m}_u} &= \frac{\mathbf{y}_l(x_u + m_u + 1, x_v) - \mathbf{y}_l(x_u + m_u - 1, x_v)}{2}, \\ \frac{\partial \mathbf{y}_k^{(l, \mathbf{m})}(x_u, x_v)}{\partial \mathbf{m}_v} &= \frac{\mathbf{y}_l(x_u, x_v + m_v + 1) - \mathbf{y}_l(x_u, x_v + m_v - 1)}{2}. \end{aligned} \quad (9)$$

Whether using parameter-based model or using optical flow-based model, the unobservable pixels defined in Section 2 should be excluded in the SR reconstruction. Sometimes their positions are known, such as when some pixels (the corresponding sensor array elements) are not functional. However, in many cases when they are not known in advance, a simple way to determine them is to make a threshold judgment on the warping error of each pixel by

$$|y_k - y_k^{(l, \theta)}| < d \quad (10)$$

or

$$|y_k - y_k^{(l, \mathbf{m})}| < d \quad (11)$$

depending on which motion estimation model is used. Here, d is a scalar threshold.

4. TOTAL VARIATION-BASED RECONSTRUCTION ALGORITHM

4.1. TV-based SR model

In most situations, the problem of SR is an ill-posed inverse problem because the information contained in the observed LR images is not sufficient to solve the HR image. In order to obtain more desirable SR results, the ill-posed problem

should be stabilized to become well-posed. Traditionally, regularization has been described from both the algebraic and statistical perspectives [39]. Using regularization techniques, the desired HR image can be solved by

$$\hat{\mathbf{z}} = \arg \min \left[\sum_k \|\mathbf{y}_k^{\text{obs}} - \mathbf{O}_k \mathbf{DBM}_k \mathbf{z}\|^2 + \lambda_2 \Gamma(\mathbf{z}) \right], \quad (12)$$

where $\sum_k \|\mathbf{y}_k^{\text{obs}} - \mathbf{O}_k \mathbf{DBM}_k \mathbf{z}\|^2$ is the data fidelity term, $\Gamma(\mathbf{z})$ denotes the regularization term, and λ_2 is the regularization parameter. It is noted that we assume all the images have the same blurring function, so the matrix \mathbf{B}_k has been substituted by \mathbf{B} .

For the regularization term, Tikhonov and Gauss-Markov types are commonly employed. A common criticism to these regularization methods is that the sharp edges and detailed information in the estimates tend to be overly smoothed. When there is considerable motion error, noise, or blurring in the system, the problem is magnified. To effectively preserve the edge and detailed information in the image, some edge-preserving regularization should be employed in the SR reconstruction.

An effective total variation (TV) regularization was first proposed by Rudin et al. [40] in image processing field. The standard TV norm looks like

$$\Gamma(\mathbf{z}) = \int_{\Omega} |\nabla z| dx dy = \int_{\Omega} \sqrt{|\nabla z|^2} dx dy, \quad (13)$$

where Ω is the 2-dimensional image space. It is noted that the above expression is not differentiable when $\nabla z = 0$. Hence, a more general expression can be obtained by slightly revising (13), given as

$$\Gamma(\mathbf{z}) = \int_{\Omega} \sqrt{|\nabla z|^2 + \beta} dx dy. \quad (14)$$

Here, β is a small positive parameter which ensures differentiability. Thus the discrete expression is written as

$$\Gamma(\mathbf{z}) = \|\nabla z\|_{\text{TV}} = \sum_i \sum_j \sqrt{|\nabla z_{i,j}^1|^2 + |\nabla z_{i,j}^2|^2 + \beta}, \quad (15)$$

where $\nabla z_{i,j}^1 = z[i+1, j] - z[i, j]$ and $\nabla z_{i,j}^2 = z[i, j+1] - z[i, j]$. The TV regularization was first proposed for image denoising [40]. Because of its robustness, it has been applied to image deblurring [41], image interpolation [42], image inpainting [32], and SR image reconstruction [24, 28–31].

In [43], the authors used the l_1 regularization

$$\Gamma(\mathbf{z}) = \sum_i \sum_j (|\nabla z_{i,j}^1| + |\nabla z_{i,j}^2|) \quad (16)$$

to approximate the TV regularization. In [24, 31], Farsiu et al. proposed the so-called bilateral TV (BTV) regularization in SR image reconstruction. The BTV regularization looks like

$$\Gamma(\mathbf{z}) = \sum_{l=-P}^P \sum_{m=0}^P \alpha^{|m|+|l|} \|\mathbf{z} - \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{z}\|_1, \quad (17)$$

where operators \mathbf{S}_x^l and \mathbf{S}_y^m shift \mathbf{z} by l and m pixels in horizontal and vertical directions, respectively. The scalar weight α , $0 < \alpha < 1$, is applied to give a spatially decaying effect to the summation of the regularization terms [31]. The authors also pointed out that the l_1 regularization can be regarded as a special case the BTV regularization.

We call these two regularizations (l_1 and BTV) as TV-related regularizations in this paper. However, the distinction between these two regularizations and the standard TV regularization should be kept in mind. Bioucas-Dias et al. [44] have demonstrated that TV regularization can lead to better results than the l_1 regularization in image restoration. Therefore, we employ the standard TV regularization (15) in this paper. By substituting (15) in (12), the following minimization function can be obtained:

$$\hat{\mathbf{z}} = \arg \min \left[\sum_k \|\mathbf{y}_k^{\text{obs}} - \mathbf{O}_k \mathbf{DBM}_k \mathbf{z}\|^2 + \lambda_2 \|\nabla z\|_{\text{TV}} \right]. \quad (18)$$

4.2. Efficient optimization method

We should note that although the TV regularization has been applied to SR image reconstruction in [24, 28–31], most of these methods use the gradient descent method to solve the desired HR image. In this section, we introduce a more efficient and reliable algorithm for the optimization problem (18).

The Euler-Lagrange equation for the energy function in (18) is given by the following nonlinear system:

$$\nabla E(\mathbf{z}) = \sum_k \mathbf{M}_k^T \mathbf{B}^T \mathbf{D}^T \mathbf{O}_k^T (\mathbf{O}_k \mathbf{DBM}_k \mathbf{z} - \mathbf{y}_k^{\text{obs}}) - \lambda_2 \mathbf{L}_z \mathbf{z} = 0, \quad (19)$$

where \mathbf{L}_z is the matrix form of a central difference approximation of the differential operator $\nabla \cdot (\nabla / \sqrt{|\nabla z|^2 + \beta})$ with $\nabla \cdot$ being the divergence operator. Using the gradient descent method, the HR image \mathbf{z} is solved by

$$\mathbf{z}^{n+1} = \mathbf{z}^n - dt \nabla E(\mathbf{z}^n), \quad (20)$$

where n is the iteration number, and $dt > 0$ is the time step parameter restricted by stability conditions (i.e., dt has to be small enough so that the scheme is stable). The drawback of this gradient descent method is that it is difficult to choose time steps for both efficiency and reliability [43].

One of the most popular strategies to solve the nonlinear problem in (19) is the lagged diffusivity fixed point iteration introduced in [45, 46]. This method consists in linearizing the nonlinear differential term by lagging the diffusion coefficient $1/\sqrt{|\nabla z|^2 + \beta}$ one iteration behind. Thus \mathbf{z}^{n+1} is obtained as the solution to the linear equation

$$\begin{aligned} & \left[\sum_{k=1} \mathbf{M}_k^T \mathbf{B}^T \mathbf{D}^T \mathbf{O}_k^T \mathbf{O}_k \mathbf{DBM}_k - \lambda \mathbf{L}_z^n \right] \mathbf{z}^{n+1} \\ & = \sum_{k=1} \mathbf{M}_k^T \mathbf{B}^T \mathbf{D}^T \mathbf{O}_k^T \mathbf{y}_k^{\text{obs}}. \end{aligned} \quad (21)$$



FIGURE 3: The 24th frame in the “Foreman” sequence. (a) The original 352×288 image and (b) the extracted 320×256 image.

It has been showed in [45] that the method is monotonically convergent. To solve the above linear equation, any linear optimization solution can be employed. Generally, the preconditioned conjugate gradient (PCG) method is desirable. To suit the specific matrix structures in image restoration and reconstruction, several preconditioners have been proposed [47–51]. An efficient way to solving the matrix equations in high-resolution image reconstruction is to apply the factorized sparse inverse preconditioner (FSIP) [50]. Let \mathbf{A} be a symmetric positive definite matrix, and let its Cholesky factorization be $\mathbf{A} = \mathbf{G}\mathbf{G}^T$. The idea of FSIP is to find the lower triangular matrix \mathbf{L} with sparsity pattern \mathbf{S} such that

$$\|\mathbf{I} - \mathbf{G}\mathbf{L}\|_F \quad (22)$$

is minimized, where $\|\cdot\|_F$ denotes the Frobenius norm. Kolotilina and Yeremin [50] showed that \mathbf{L} can be obtained by the following algorithm.

Step 1. Compute $\hat{\mathbf{L}}$ with sparse pattern \mathbf{S} such that $(\hat{\mathbf{L}}\mathbf{A})_{x,y} = \delta_{x,y}, (x, y) \in \mathbf{S}$.

Step 2. Let $\hat{\mathbf{D}} = (\text{diag}(\hat{\mathbf{L}}))^{-1}$ and $\mathbf{L} = \hat{\mathbf{D}}^{1/2}\hat{\mathbf{L}}$.

According to this algorithm, m small linear systems need to be solved, where m is the number of rows in the matrix \mathbf{A} . These systems can be solved in parallel. Thus the above algorithm is also well suited for modern parallel computing.

Motivated by the FSIP preconditioner, we consider the factorized banded inverse preconditioner (FBIP) [47] which is a special type of FSIP. The main idea of FBIP is to approximate the Cholesky factor of the coefficient matrix by banded lower triangular matrices. The following theorem has been proved in [47].

Let \mathbf{T} be a Hermitian Toeplitz matrix, and let $\mathbf{B} = \mathbf{T}$ or $\mathbf{B} = \mathbf{I} + \mathbf{T}^T\mathbf{D}\mathbf{T}$ with \mathbf{D} be a positive diagonal matrix. Denote the k th diagonal of \mathbf{T} by t_k . Assume the diagonals of \mathbf{T} satisfy

$$|t_k| \leq ce^{-\gamma|k|} \quad (23)$$

for some $c > 0$ and $\gamma > 0$, or

$$|t_k| \leq c(|k| + 1)^{-s} \quad (24)$$

for some $c > 0$ and $s > 3/2$. Then for any given $\varepsilon > 0$, there exists a $p' > 0$ such that for all $p > p'$,

$$\|\mathbf{L}_p - \mathbf{C}^{-1}\| \leq \varepsilon, \quad (25)$$

where \mathbf{L}_p denotes the FBIP of \mathbf{B} with the lower bandwidth p , and \mathbf{C} is the Cholesky factor of \mathbf{B} . This theorem indicates that if the Toeplitz matrix \mathbf{T} has certain off-diagonal decay property, then the FBIPs of \mathbf{B} will be good approximation of \mathbf{B}^{-1} . Here we should note that even though the system matrix in (21) is not exactly in the Toeplitz form or in $\mathbf{I} + \mathbf{T}^T\mathbf{D}\mathbf{T}$ form, our experimental results indicate that the FBIP algorithm is still very efficient for this problem.

5. SIMULATION RESULTS

We tested the proposed TV-based SR reconstruction algorithm using a raw “Foreman” sequence and a realistic MPEG4 “Bulletin” sequence. The algorithm using Laplacian regularization (where the regularization term is $\|\mathbf{Q}\mathbf{z}\|^2$, with \mathbf{Q} being the 2-dimensional Laplacian operator) was also tested to make a comparative analysis. It is noted that the Laplacian regularization generally has stronger constraint on the image than the TV regularization because it is a square term and not extracted like the TV regularization, so it should require a smaller regularization parameter. In fact, we should respectively choose the optimal regularization parameters for the two different regularizations for a reasonable comparison. With this in mind, we tried a series of regularization parameters for the two regularizations in all the experiments. Furthermore, we also compared our proposed algorithm to other TV or TV-related algorithms in the “Foreman” experiments.

5.1. The “Foreman” sequence

We first tested the popular “Foreman” sequence with a 352×288 CIF format. One frame (the 24th) of this sequence is shown in Figure 3(a). It is seen that there are two dark regions, respectively, at the left and lower boundaries, and that there is also a labeled region around the top left corner. To make reliable quantitative analysis, most of the processing was restricted to the central 320×256 pixel region. The 320×256 extracted version of Figure 3(a) is shown in

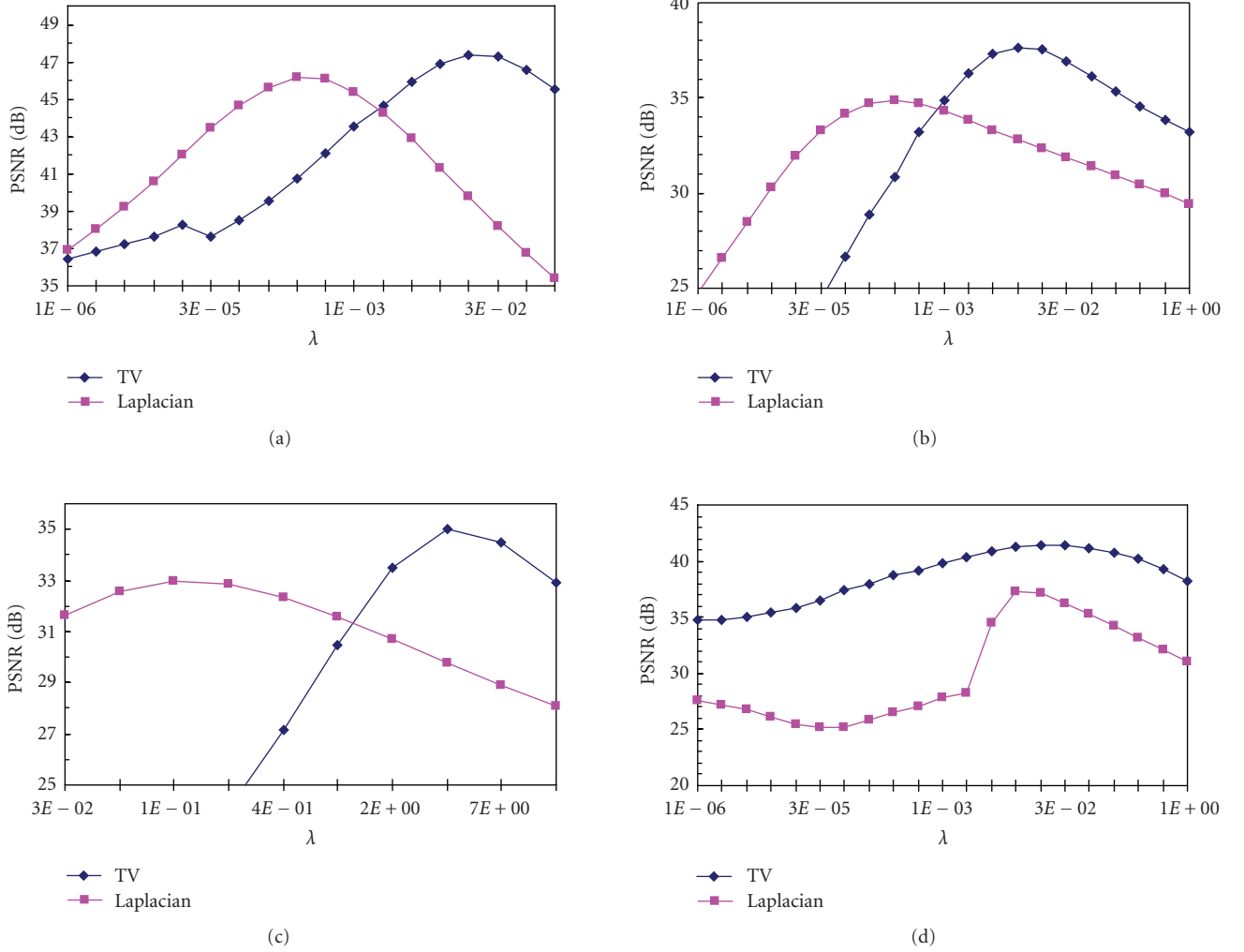


FIGURE 4: PSNR values versus the regularization parameter in the synthetic “Foreman” experiments: (a) the “motion only” case, (b) the “blurring” case, (c) the “noise” case, and (d) the “missing” case.

Figure 3(b). The following peak signal-to-noise ratio (PSNR) was employed as the quantitative measure:

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2 * L_1 N_1 L_2 N_2}{\|\hat{\mathbf{z}} - \mathbf{z}\|^2} \right), \quad (26)$$

where $L_1 N_1 L_2 N_2$ is the total number of pixels in the HR image, and $\hat{\mathbf{z}}$ and \mathbf{z} represent the reconstructed HR image and the original image, respectively.

5.1.1. Synthetic simulations

To show the feature and advantage of the TV-based reconstruction algorithm more sufficiently, we first implemented the synthetic experiments in which the LR images are simulated from a single frame of the “Foreman” sequence, frame 24 (the extracted 320×256 version). Using observation model (2), we simulated the LR frames in four different ways: (1) the “motion only” case, in which the original frame

was first warped and then the warped versions were down-sampled to obtain the LR frames; (2) the “blurring” case, in which the original frame was first blurred with a 5×5 Gaussian kernel before the warping; (3) the “noise” case, in which the LR frames obtained in the “motion only” case were then contaminated by Gaussian noise with 65.025 variance; and (4) the “missing” case, in which some missing regions were assumed to exist at the same positions of all the LR frames. For each case, the down-sampling factor was two, and four LR images were simulated using global translational motion model. PSNR values against the regularization parameter λ_2 in the four cases are demonstrated in Figures 4(a)–4(d), respectively. The SR reconstruction results are respectively shown in Figures 5–8.

In the “motion only” case, the best PSNR result using Laplacian regularization is 46.162 dB with $\lambda_2 = 0.000256$ and that of TV is 47.360 dB with $\lambda_2 = 0.016384$ (see Figure 4(a)). As expected, the use of TV regularization provided a higher PSNR value. However, since the motions were accurately



FIGURE 5: Experimental results in the synthetic “motion only” case. (a) LR frame, (b) Laplacian SR result with $\lambda_2 = 0.000256$ and (c) TV SR result with $\lambda_2 = 0.016384$.

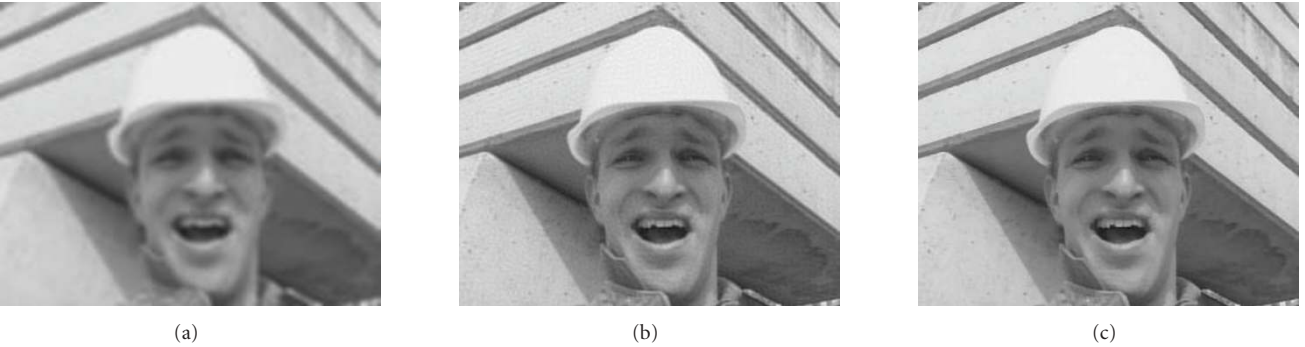


FIGURE 6: Experimental results in the synthetic “blurring” case. (a) LR frame, (b) Laplacian SR result with $\lambda_2 = 0.0001$ and (c) TV SR result with $\lambda_2 = 0.008192$.

known and there is no noise, blurring, or missing pixel in the image, the result using Laplacian regularization also has high quality. As a result, Figures 5(b) and 5(c) are almost indistinguishable visually.

From Figures 4(b) and 6, we can see the advantage of the TV-based reconstruction algorithm is much more obvious in the “blurring” case. Figure 6(b) is the Laplacian result with the best PSNR of 34.845 dB ($\lambda_2 = 0.00256$), and Figure 6(c) shows the TV result with the best PSNR of 37.663 dB ($\lambda_2 = 0.008192$). Visually, the use of Laplacian regularization leads to some artifacts in the reconstructed image. TV regularization, however, does well.

In the “noise” case, the best PSNR value for the Laplacian regularization is 32.968 dB with the regularization parameter being 0.1024. Using TV regularization, however, we obtained a best PSNR value of 34.987 dB when the regularization parameter is equal to 3.2768. The images corresponding to the best PSNR values are shown in Figures 7(b) and 7(c), respectively. Both images are still noisy to some extent although they have the highest PSNR values, and Figure 7(b) is more obvious. To further smooth the noise, larger regularization parameters should be chosen. Figure 7(d) is the Laplacian result with $\lambda_2 = 3.2768$, and Figure 7(e) is the TV result with $\lambda_2 = 6.5536$. The PSNRs of these two images are 29.797 dB (Laplacian) versus 34.459 dB (TV). The TV-based algorithm

is preferable again because it can provide simultaneous denoising and edge preservation.

Figures 4(d) and 8 show the “missing” case. This is a typical example of the simultaneous image inpainting and SR. The best PSNR values for Laplacian and TV are, respectively, 37.315 dB ($\lambda_2 = 0.008192$) and 41.400 dB ($\lambda_2 = 0.016384$). The corresponding results are shown in Figures 8(b) and 8(c), respectively. We also give the results using larger regularization parameters in Figure 8(d) (Laplacian, $\lambda_2 = 0.065536$, PSNR = 35.282 dB) and Figure 8(e) (TV, $\lambda_2 = 0.26214$, PSNR = 40.176 dB), respectively. These two images have better visual quality in the missing regions than their counterparts, Figures 8(b) and 8(c). We can clearly see that the missing regions can be desirably inpainted using the TV-based algorithm. However, the Laplacian regularization does not work well. Figure 8(f) shows the reconstruction result using TV regularization ($\lambda_2 = 0.26214$) by conducting image inpainting and SR separately. The missing regions cannot be inpainted as good as that in the simultaneous process case. The PSNR of Figure 8(f) is 35.003.

5.1.2. Nonsynthetic simulations

In the nonsynthetic experiments, the LR images used in the SR reconstruction are produced from the corresponding HR

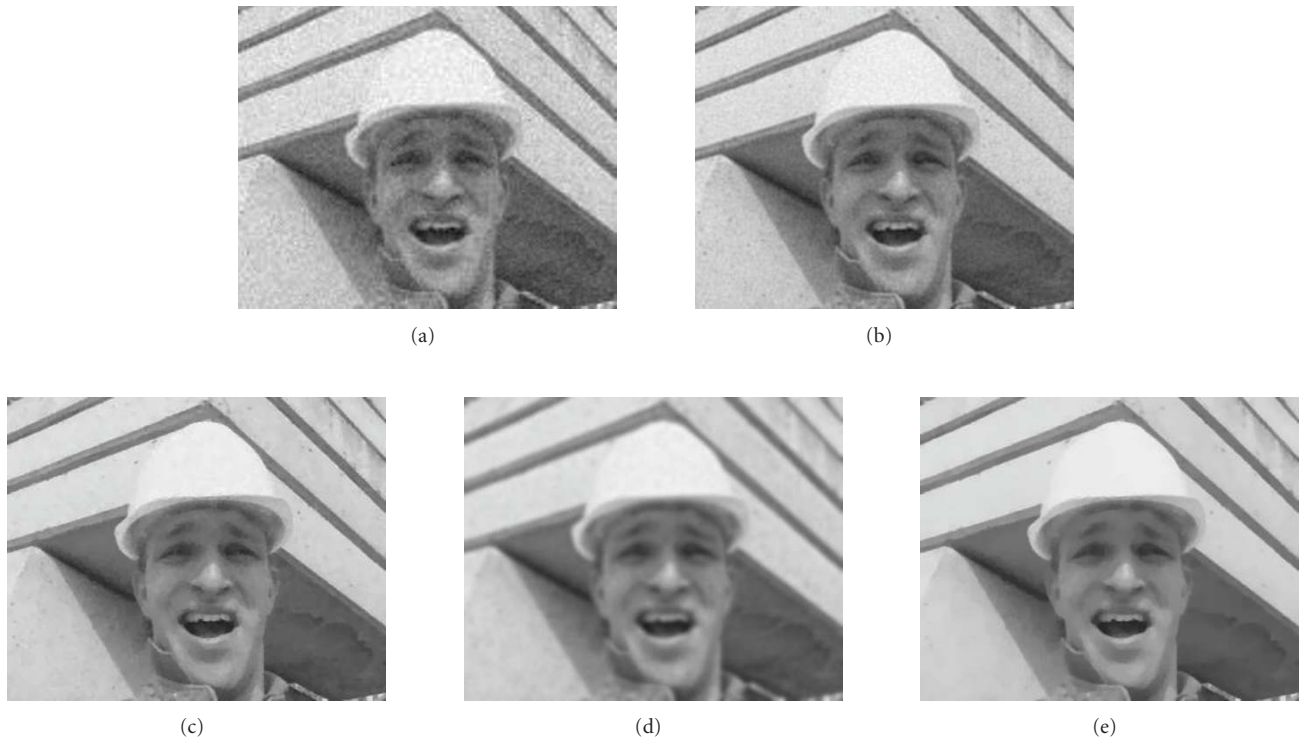


FIGURE 7: Experimental results in the synthetic “noise” case. (a) LR frame, (b) Laplacian SR result with $\lambda_2 = 0.1024$, (c) TV SR result with $\lambda_2 = 3.2768$, (d) Laplacian SR result with $\lambda_2 = 3.2768$ and (e) TV SR result with $\lambda_2 = 6.5536$.

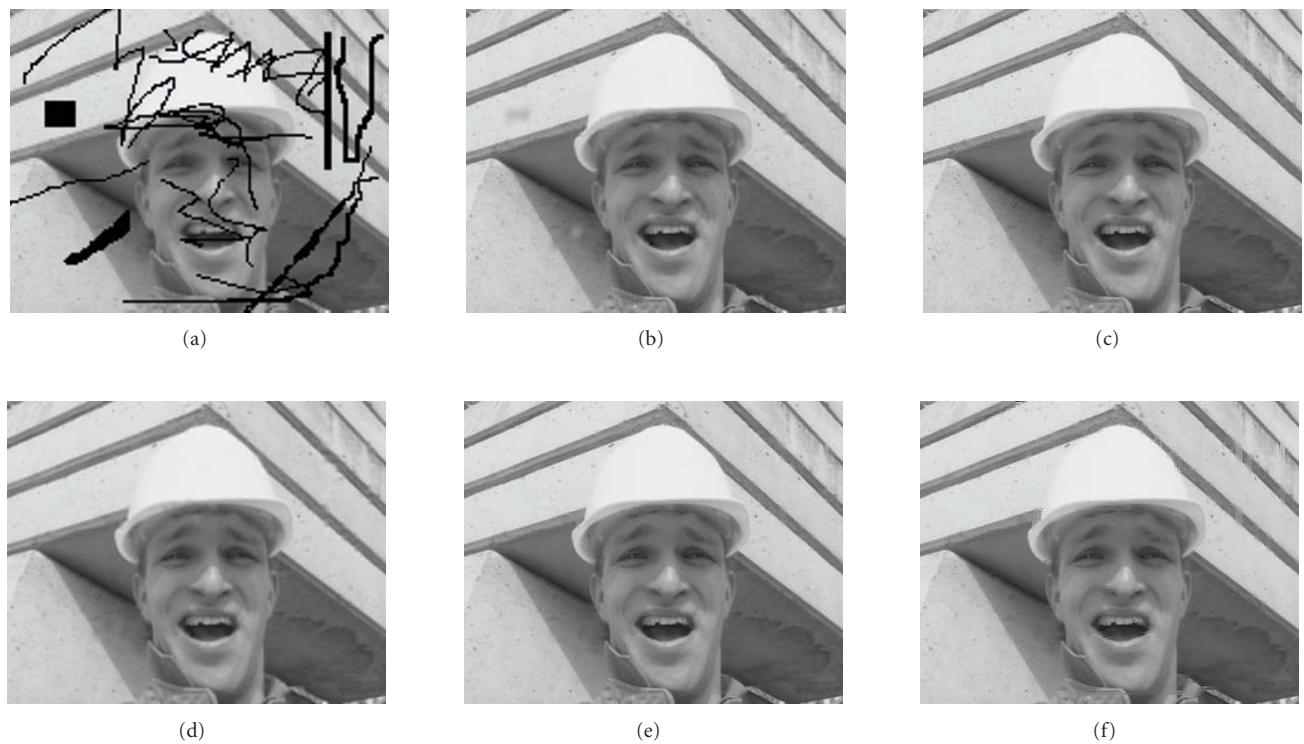


FIGURE 8: Experimental results in the synthetic “missing” case. (a) LR frame, (b) Laplacian simultaneous inpainting and SR result with $\lambda_2 = 0.008192$, (c) TV simultaneous inpainting and SR result with $\lambda_2 = 0.016384$, (d) Laplacian simultaneous inpainting and SR result with $\lambda_2 = 0.065536$, (e) TV simultaneous inpainting and SR result with $\lambda_2 = 0.26214$, and (f) TV result conducting inpainting and SR separately with $\lambda_2 = 0.26214$.



FIGURE 9: Motion estimates of frame 22 (a) and frame 25 (b) in the nonsynthetic “Foreman” experiment.



FIGURE 10: The unobservable pixels of frame 22 (a) and frame 25 (b) in the nonsynthetic “Foreman” experiment.

frames in the video with a downsampling factor of two. Here, we again demonstrate the reconstruction results of frame 24. Frames 22, 23, 25, and 26 were used as the unreferenced ones. We first tested the “motion only” case. It is noted that the motions are unknown and should be estimated in the nonsynthetic cases. We employed the motion estimation method introduced in Section 3.2, with $\lambda_1 = 10000$ and $\alpha = 10^{-6}$. The motion estimates of frames 22 and 25 are shown in Figure 9 as illustrations. After the motion estimation, (11) was used to determine the unobservable pixels, and the threshold d was chosen to be 6. Figures 10(a) and 10(b) illustrate the unobservable pixels of frame 22 and 25, respectively. Reconstruction methods using Laplacian regularization and TV regularization were respectively implemented. PSNR value against the regularization parameter λ_2 is demonstrated in Figure 11(a). The best PSNR result with Laplacian regularization is 36.185 dB with $\lambda_2 = 0.008$, and that of TV is 37.336 dB with $\lambda_2 = 0.512$. Again, the TV performs better than Laplacian quantitatively. Furthermore, unlike the synthetic “motion only” case, the advantage of the TV-based reconstruction is also visually obvious. The Laplacian result is shown in Figure 12(b), from which we can find that the sharp edges are obviously damaged due to the inevitable motion estimation errors. In the TV result shown in Figure 12(c), however, these edges are effectively preserved.

We also show the nonsynthetic “noise” case in which random Gaussian noise with 32.5125 variance was added to the down-sampled images. One of the noisy LR frames is shown in Figure 13(a). Figure 11(b) shows the curves of the PSNR value versus the regularization parameter. The best

PSNR values are, respectively, 32.040 dB and 33.851 dB for the Laplacian and TV. The corresponding reconstructed images are illustrated in Figures 13(b) and 13(c), and the results with larger regularization parameters which have better visual quality regarding the noise are shown in Figures 13(d) and 13(e), respectively. By comparisons, we see that the TV-based reconstruction algorithm outperforms the Laplacian-based algorithm in terms of both the visual evaluation and quantitative assessment again.

In order to demonstrate the efficacy of the proposed algorithm, we reconstructed the first 60 frames in the “Foreman” sequence and then combined them together to video format. The regularization parameters for all frames were the same, and the parameters used can provide almost the best visual equality in each case. The SR videos with WMV format can be found at the website <http://www.math.hkbu.edu.hk/mng/SR/VideoSR.htm>. It is noted that the original frames with size of 352×288 were used now. We also tried to deal with the missing and labeled regions in the original video frames in the “motion only” case. Actually, it is impossible to perfectly inpaint these regions because their areas are too large and they are located at the boundaries of the image. However, our experiment indicates that the TV-based reconstruction algorithm has the efficacy to provide a more desirable result as seen in Figure 14.

5.1.3. Comparison to other TV methods

In Sections 5.1.1 and 5.1.2, we compared the proposed TV regularization-based algorithm (FBIP TV algorithm) to the

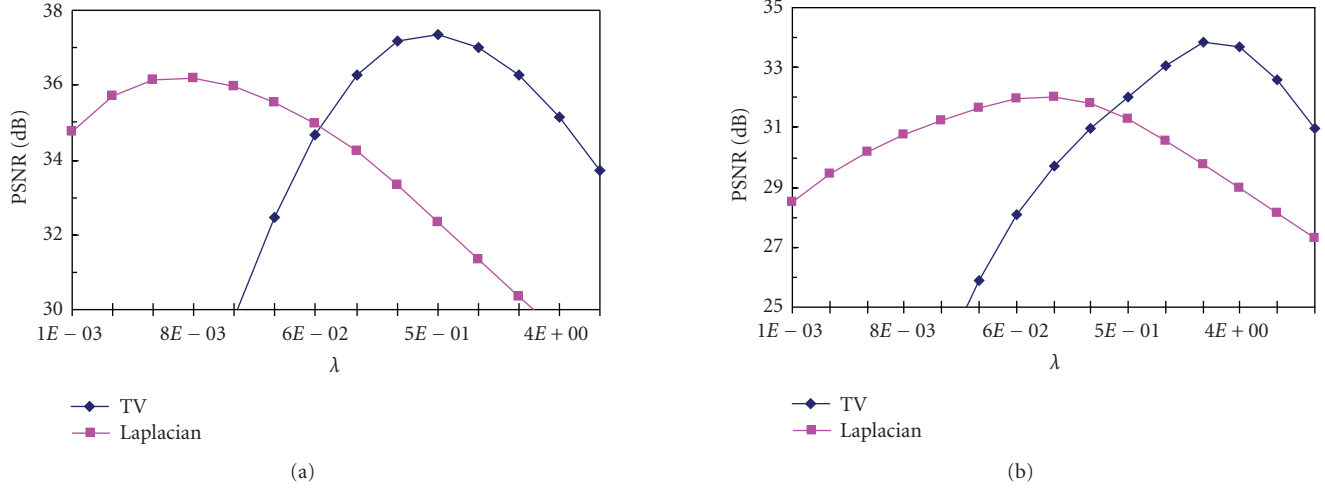


FIGURE 11: PSNR values versus the regularization parameter in the nonsynthetic “Foreman” experiments: (a) the “motion only” case, and (b) the “noise” case.



FIGURE 12: Experimental results in the nonsynthetic “motion only” case. (a) LR frame, (b) Laplacian SR result with $\lambda = 0.008$ and (c) TV SR result with $\lambda = 0.512$.

Laplacian regularization-based algorithm from the reliability perspective. In this subsection, we compare it to other TV-based algorithms which employ gradient descent (GD) method in terms of both efficiency and reliability. In the experiments, the iteration was terminated when the relative gradient norm $d = \|\nabla E(\mathbf{z}^n)\|/\|\nabla E(\mathbf{z}^0)\|$ was smaller or iteration number N was larger than some thresholds. We have mentioned that the drawback of the GD method is that it is difficult to choose time step dt for both efficiency and reliability. Therefore, we repeated several parameters in each case of the experiments. Here we show the reconstruction results using almost the optimal step parameters. We also tested the effect of parameter β in (14).

Table 1 shows the synthetic “noise-free” case with the full 4 frames being used. Since the problem is almost over-determined in this case, we believe most algorithms can be employed from the reliability perspective. From Table 1, we can see the PSNR value of the result using FBIP TV algorithm is even lower than that of the GD TV algorithm. But the GD TV algorithm is not stable when dt increases to 1.0. From the efficiency perspective, the FBIP TV algorithm is faster than

the GD TV and GD BTV algorithms. We also can see that a relatively larger parameter β leads to much faster convergence speed for the FBIP TV algorithm, but the efficiency effect of β to the GD TV algorithm is negligible. The reliability of both FBIP TV and GD TV algorithms is not sensitive to the choice of β .

Table 2 shows the synthetic “noise-free” case with only 2 frames being used. In this case, the problem is strongly under-determined. We can see that the efficiency advantage of the FBIP TV algorithm is very obvious. The FBIP TV algorithm also leads to higher PSNR values than the GD TV and BTV algorithms.

Table 3 shows the synthetic “missing” case. The FBIP TV algorithm is still very efficient when there are missing regions in the image. However, the convergence speed of the GD TV and GD BTV are extremely slow. Larger regularization or larger parameter P (in BTV) can speed up the processing, but cannot ensure the optimal solution.

Figure 15 shows the convergence performance in the nonsynthetic “noisy-free” case. Figure 15(a) illustrates the evolution of the gradient norm-based convergence condition



FIGURE 13: Experimental results in the nonsynthetic “noise” case. (a) LR frame, (b) Laplacian SR result with $\lambda = 0.128$, (c) TV SR result with $\lambda = 2.048$, (d) Laplacian SR result with $\lambda = 2.048$, and (e) TV SR result with $\lambda = 4.096$.

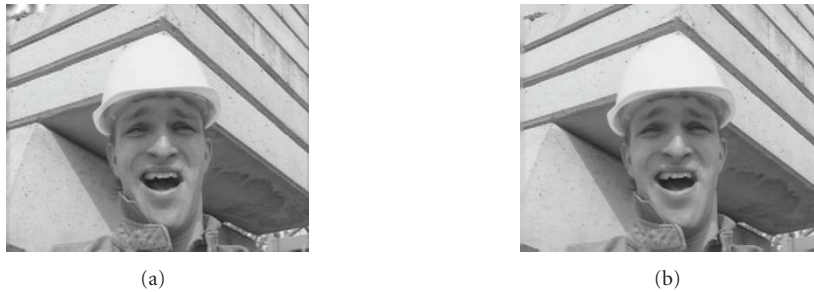


FIGURE 14: Reconstruction results of the “Foreman” with the original size. (a) Laplacian regularization and (b) TV regularization.

$\|\nabla E(\mathbf{z}^n)\|/\|\nabla E(\mathbf{z}^0)\|$ against the computational time, and Figure 15(b) is the demonstration of PSNR value versus the computational time. From both the gradient norm and PSNR convergence criteria, the FBIP TV algorithm greatly outperforms the GD TV algorithm and GD BTV algorithm. It is not only very efficient, but also very stable. In all the previous experiments, the differential operator in (19) was approximated by central difference for both GD TV and FBIP TV algorithms. In this case, we also tested the backward difference approximation for the GD TV algorithm (not that backward difference cannot be used for the FBIP TV algorithm because the corresponding system matrix is not symmetric and positive). The l_1 regularization was also tested. Table 4 shows the PSNR values of different algorithms. Here we note that the selected termination conditions can ensure

the optimal results for all the regularizations with the current parameter settings. It is seen that the proposed FBIP TV algorithm outperforms all other algorithms.

5.2. The “bulletin” sequence

In this experiment, we show the SR reconstruction of a “bulletin” sequence which was obtained from a consumer-level digital video camera. One frame of this sequence is shown in Figure 16. Although the original frame size is 640×480 pixels, our processing was restricted to a typical 450×80 pixel, as shown in (boxed in dashed) in Figure 16.

Here we used seven extracted 450×80 images to produce a 900×160 SR image. Since there is no independently

TABLE 1: Comparison with GD TV and BTV algorithms in the synthetic “noise-free” case (4 frames).

	λ	dt	β	Termination	Time (s)	PSNR (dB)
FBIP TV	0.016	—	10^{-5}	$d = 5 \times 10^{-4}$	16.063	47.360
		—	10^{-1}	$d = 5 \times 10^{-4}$	11.422	47.426
GD TV	0.016	0.2	10^{-5}	$d = 5 \times 10^{-4}$	73.031	47.707
			10^{-1}	$d = 5 \times 10^{-4}$	70.312	47.709
		0.6	10^{-5}	$d = 5 \times 10^{-4}$	24.531	47.722
			10^{-1}	$d = 5 \times 10^{-4}$	23.141	47.709
		1.0	10^{-5}	$d = 5 \times 10^{-4}$	15.063	47.733
			10^{-1}	$d = 5 \times 10^{-4}$	14.530	47.712
	10^{-5}	$N = 2000$	303.406	47.274		
GD BTV ($P = 1$)	0.007	0.6	—	$d = 5 \times 10^{-4}$	41.828	47.385
		1.0	—	$d = 5 \times 10^{-4}$	34.140	47.265

TABLE 2: Comparison with GD TV and BTV algorithms in the synthetic “noise-free” case (2 frames).

	λ	dt	β	Termination	Time (s)	PSNR (dB)
FBIP TV	0.016	—	10^{-5}	$d = 1 \times 10^{-3}$	18.140	40.201
		—	10^{-1}	$d = 1 \times 10^{-3}$	13.718	40.166
GD TV	0.016	0.5	10^{-5}	$d = 2 \times 10^{-3}$	93.734	39.038
GD BTV ($P = 1$)	0.007	0.5	—	$d = 2 \times 10^{-3}$	179.093	38.623

TABLE 3: Comparison with GD TV and BTV algorithms in the synthetic “missing” case (4 frames).

	λ	dt	β	Termination	Time (s)	PSNR (dB)
FBIP TV	0.016	—	10^{-5}	$d = 1 \times 10^{-5}$	33.875	41.373
		—	10^{-1}	$d = 1 \times 10^{-5}$	20.062	41.239
GD TV	0.016	1.0	10^{-5}	$N = 3000$	380.087	17.410
	0.1	1.0	10^{-5}	$N = 3000$	384.890	23.968
			10^{-5}	$N = 6000$	762.343	28.324
			10^{-5}	$N = 9000$	1139.56	32.577
BTV ($P = 1$)	0.1	1.0	—	$N = 3000$	335.187	26.883
				$N = 6000$	670.343	33.925
				$N = 9000$	1018.43	39.460
				$N = 10000$	1135.72	39.460
GD BTV ($P = 3$)	0.1	1.0	—	$N = 1000$	376.047	30.170
				$N = 2000$	749.141	38.839
				$N = 3000$	1191.70	38.861

TABLE 4: Comparison with GD TV and BTV algorithms in the nonsynthetic “noise-free” case (5 frames).

	λ	dt	β	Termination	PSNR (dB)
FBIP TV(Central)	0.512	—	10^{-5}	$d = 1 \times 10^{-3}$	37.336
GD TV (Central)	0.512	0.05	10^{-5}	$N = 1000$	37.206
GD TV (Backward)	0.512	0.05	10^{-5}	$N = 1000$	37.084
GD L1	0.2	0.05	—	$N = 1000$	36.576
GD BTV ($P = 1$)	0.2	0.05	—	$N = 1000$	36.854
GD BTV ($P = 2$)	0.2	0.05	—	$N = 1000$	36.875
GD BTV ($P = 3$)	0.1	0.05	—	$N = 1000$	36.873

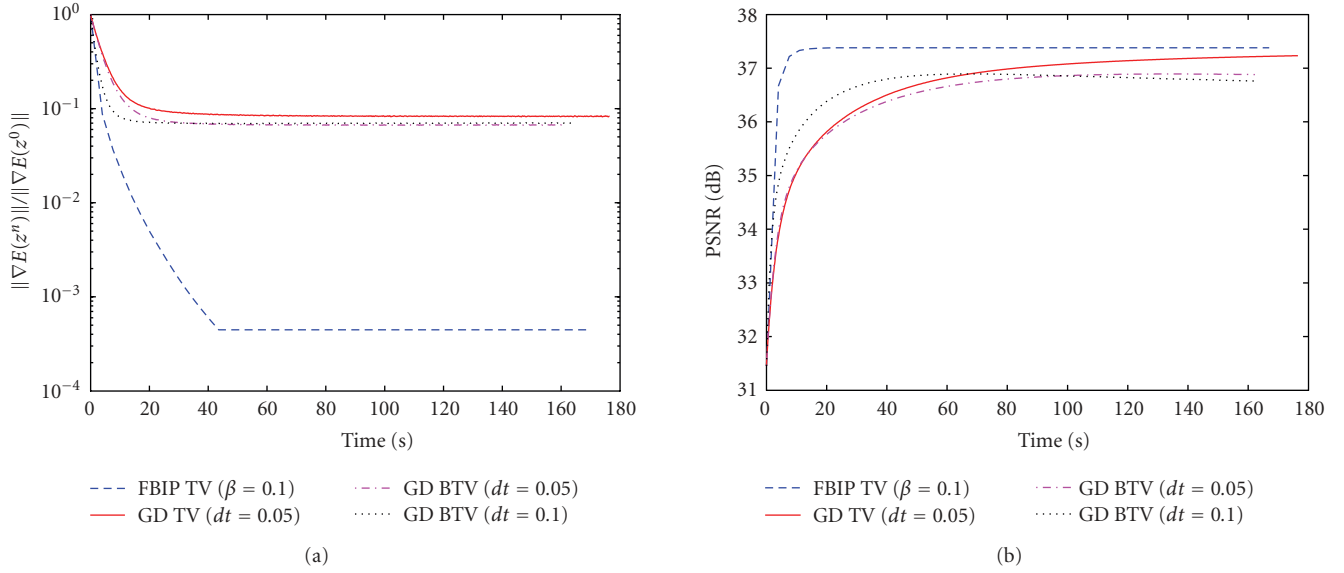


FIGURE 15: Convergence performance of different algorithms, (a) measured by gradient norm, (b) measured by PSNR value.

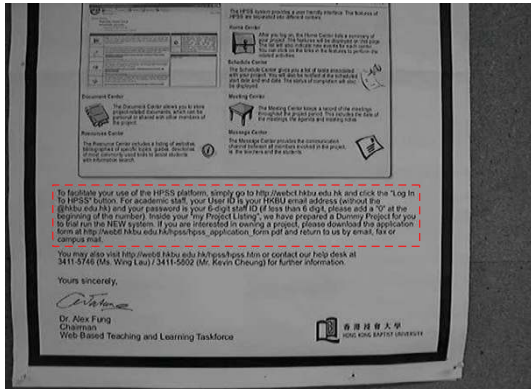


FIGURE 16: One 640×480 frame in the “bulletin” video. The region boxed in dashed is the interest.

moving object in the scene, the motions between the referenced image and the unreferenced images can be estimated by the affine parameter model introduced in Section 3.1. Figure 17(a) shows one of the extracted images. It is observed that there are obvious artifacts in most parts of this image. These artifacts mainly come from the compression process of the MPEG video. Figure 17(b) is the SR reconstruction result using Laplacian regularization with a relatively smaller regularization parameter 0.001. The characters in the image look clearer, but the compression artifact is aggravated. To solve this problem, higher regularization parameter should be chosen. In our experiment, we found the artifact problem could not be solved until we increase the regularization parameter to 0.5. The reconstructed image is shown in Figure 17(c). Although the artifacts are suppressed, the characters are too

smooth. However, the TV-based reconstruction algorithm can effectively solve this tradeoff. Figure 17(d) shows the TV reconstruction result with $\lambda_2 = 5.0$. Its overall visual quality is much better than that of Figures 17(a)–17(c). Detailed cropped regions from Figures 17(a)–17(d) are, respectively, demonstrated in Figures 18(a)–18(d), from which we can see the advantage of the TV-based SR reconstruction more easily.

It is noted that we did not consider the compression and decompression processes in the reconstruction model although the inputs of this experiment are the decoded frames from the MPEG video. But even in this case, we obtained desirable super-resolved result using the TV-based SR algorithm. If the compression and decompression processes also were included in the SR model such as the methods in [26, 27], the result should be better.

6. CONCLUSIONS

The SR reconstruction of digital video becomes difficult when there is blurring, noise, missing regions, compression artifacts, or inevitable motion estimation errors in the system. To address these challenges, we have proposed a robust and efficient TV-based SR algorithm in this paper. The proposed algorithm has been tested in different cases. Experimental results show that it performs quite well in terms of both robustness and efficiency. Nevertheless, further work can potentially improve upon the proposed algorithm, such as the inclusion of the compression process in the reconstruction model and the consideration of TV regularization in the motion estimation step. The adaptive determination of the number of the input frames in the SR reconstruction is also another possibility.

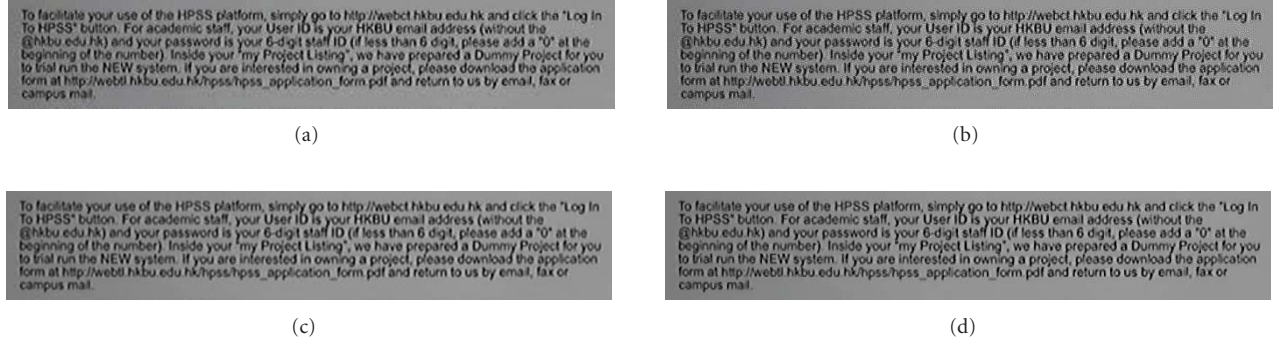


FIGURE 17: SR reconstruction results of the “bulletin” sequence. (a) LR frame, (b) Laplacian SR result with $\lambda_2 = 0.001$, (c) Laplacian SR result with $\lambda_2 = 0.5$, and (d) TV SR result with $\lambda_2 = 5.0$.

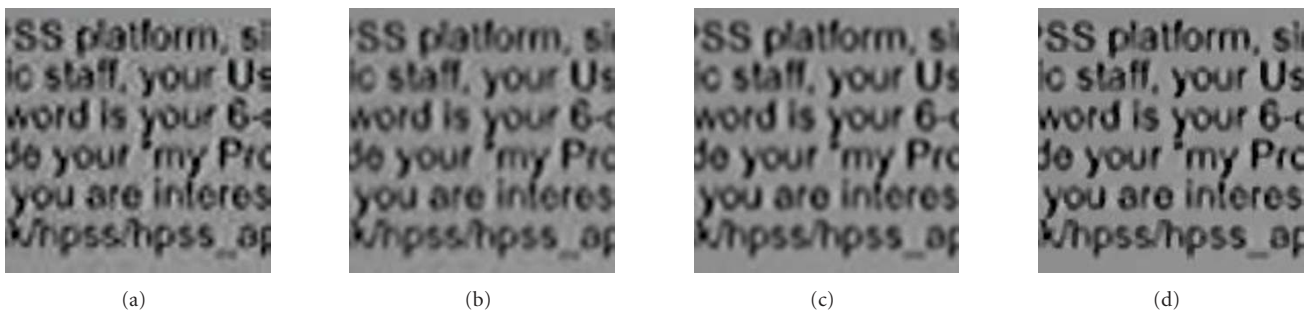


FIGURE 18: (a)–(d) Detail regions cropped from Figures 17(a)–17(d).

ACKNOWLEDGMENTS

Research supported in part by RGC 7035/04P and 7035/05P, and HKBU FRGs. H. Shen would like to thank M. K. Ng for his hospitality during his visit to Centre for Mathematical Imaging and Vision, Hong Kong Baptist University, from March 2006 to March 2007. This work was done during H. Shen visit to Hong Kong Baptist University. Research supported in part by RGC Grant HKU 7143/05E.

REFERENCES

- [1] R. Y. Tsai and T. S. Huang, “Multi-frame image restoration and registration,” *Advances in Computer Vision and Image Processing*, vol. 1, no. 2, pp. 317–339, 1984.
- [2] S. P. Kim, N. K. Bose, and H. M. Valenzuela, “Recursive reconstruction of high resolution image from noisy undersampled multiframe,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, 1990.
- [3] S. P. Kim and W.-Y. Su, “Recursive high-resolution reconstruction of blurred multiframe images,” *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 534–539, 1993.
- [4] S. Rhee and M. G. Kang, “Discrete cosine transform based regularized high-resolution image reconstruction algorithm,” *Optical Engineering*, vol. 38, no. 8, pp. 1348–1356, 1999.
- [5] R. H. Chan, T. F. Chan, L. Shen, and Z. Shen, “Wavelet algorithms for high-resolution image reconstruction,” *SIAM Journal of Scientific Computing*, vol. 24, no. 4, pp. 1408–1432, 2003.
- [6] M. K. Ng, C. K. Sze, and S. P. Yung, “Wavelet algorithms for deblurring models,” *International Journal of Imaging Systems and Technology*, vol. 14, no. 3, pp. 113–121, 2004.
- [7] N. Nguyen and P. Milanfar, “A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution),” *Circuits, Systems, and Signal Processing*, vol. 19, no. 4, pp. 321–338, 2000.
- [8] H. Ur and D. Gross, “Improved resolution from subpixel shifted pictures,” *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 2, pp. 181–186, 1992.
- [9] M. Irani and S. Peleg, “Improving resolution by image registration,” *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.
- [10] H. Stark and P. Oskoui, “High-resolution image recovery from image-plane arrays, using convex projections,” *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 6, no. 11, pp. 1715–1726, 1989.
- [11] 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 *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’92)*, vol. 3, pp. 169–172, San Francisco, Calif, USA, March 1992.
- [12] A. J. Patti, M. I. Sezan, and A. M. Tekalp, “High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur,” in *Proceedings of IEEE International Conference Image Processing (ICIP ’94)*, vol. 1, pp. 343–347, Austin, Tex, USA, November 1994.

- [13] 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.
- [14] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image from multiple-degraded misregistered low-resolution images," in *Visual Communications and Image Processing*, vol. 2308 of *Proceedings of SPIE*, pp. 971–981, Chicago, Ill, USA, September 1994.
- [15] 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.
- [16] R. C. Hardie, T. R. Tuinstra, J. Bogner, K. J. Barnard, and E. E. Armstrong, "High resolution image reconstruction from digital video with global and non-global scene motion," in *Proceedings of IEEE International Conference on Image Processing (ICIP '97)*, vol. 1, pp. 153–156, Santa Barbara, Calif, USA, October 1997.
- [17] 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.
- [18] M. Elad and A. Feuer, "Superresolution restoration of an image sequence: adaptive filtering approach," *IEEE Transactions on Image Processing*, vol. 8, no. 3, pp. 387–395, 1999.
- [19] J. Chung, E. Haber, and J. Nagy, "Numerical methods for coupled super-resolution," *Inverse Problems*, vol. 22, no. 4, pp. 1261–1272, 2006.
- [20] 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.
- [21] N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 201–213, 2006.
- [22] 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.
- [23] R. Sasahara, H. Hasegawa, I. Yamada, and K. Sakaniwa, "A color super-resolution with multiple nonsmooth constraints by hybrid steepest descent method," in *Proceedings of IEEE International Conference on Image Processing (ICIP '05)*, vol. 1, pp. 857–860, Genova, Italy, September 2005.
- [24] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 141–159, 2006.
- [25] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1860–1875, 2005.
- [26] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Transactions on Image Processing*, vol. 13, no. 7, pp. 898–910, 2004.
- [27] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 37–48, 2003.
- [28] D. Capel and A. Zisserman, "Super-resolution enhancement of text image sequences," in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR '00)*, vol. 1, pp. 600–605, Barcelona, Spain, September 2000.
- [29] Y. B. Han and L. N. Wu, "Super resolution reconstruction of video sequence based on total variation," in *Proceedings of International Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP '04)*, pp. 575–578, Hong Kong, October 2004.
- [30] C. Vazquez, H. Aly, E. Dubois, and A. Mitiche, "Motion compensated super-resolution of video by level sets evolution," in *Proceedings of IEEE International Conference on Image Processing (ICIP '04)*, vol. 3, pp. 1767–1770, Singapore, October 2004.
- [31] 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.
- [32] T. F. Chan and J. Shen, "Mathematical models for local non-texture inpaintings," *SIAM Journal on Applied Mathematics*, vol. 62, no. 3, pp. 1019–1043, 2002.
- [33] S. Borman and R. L. Stevenson, "Spatial resolution enhancement of low-resolution image sequences: a comprehensive review with directions for future research," Tech. Rep., Laboratory for Image and Signal Analysis (LISA), University of Notre Dame, Notre Dame, Ind, USA, July 1998.
- [34] 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.
- [35] 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.
- [36] D. Capel and A. Zisserman, "Computer vision applied to super resolution," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 75–86, 2003.
- [37] R. R. Schultz, L. Meng, and R. L. Stevenson, "Subpixel motion estimation for super-resolution image sequence enhancement," *Journal of Visual Communication and Image Representation*, vol. 9, no. 1, pp. 38–50, 1998.
- [38] A. M. Tekalp, *Digital Video Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.
- [39] 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.
- [40] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [41] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 813–824, 1998.
- [42] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1–2, pp. 89–97, 2004.
- [43] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image restoration," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 987–995, 1996.
- [44] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '06)*, vol. 2, pp. 861–864, Toulouse, France, May 2006.
- [45] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal of Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996.
- [46] C. R. Vogel, *Computational Methods for Inverse Problems*, Frontiers in Applied Mathematics, SIAM, Philadelphia, Pa, USA, 2002.

- [47] F.-R. Lin, M. K. Ng, and W.-K. Ching, "Factorized banded inverse preconditioners for matrices with Toeplitz structure," *SIAM Journal of Scientific Computing*, vol. 26, no. 6, pp. 1852–1870, 2005.
- [48] R. H. Chan, T. F. Chan, and C.-K. Wong, "Cosine transform based preconditioners for total variation deblurring," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1472–1478, 1999.
- [49] M. K. Ng, R. H. Chan, T. F. Chan, and A. M. Yip, "Cosine transform preconditioners for high resolution image reconstruction," *Linear Algebra and Its Applications*, vol. 316, no. 1–3, pp. 89–104, 2000.
- [50] L. Y. Kolotilina and A. Y. Yeremin, "Factorized sparse approximate inverse preconditionings I: theory," *SIAM Journal on Matrix Analysis and Applications*, vol. 14, no. 1, pp. 45–58, 1993.
- [51] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient superresolution image reconstruction algorithm," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 573–583, 2001.

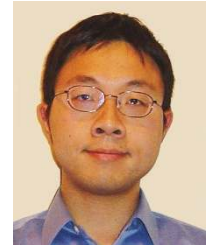
Michael K. Ng is a Professor in the Department of Mathematics at the Hong Kong Baptist University. He obtained his B.S. degree in 1990 and M.Phil. degree in 1992 at the University of Hong Kong, and Ph.D. degree in 1995 at Chinese University of Hong Kong. Michael won the Honourable Mention of Householder Award IX, in 1996, at Switzerland, an excellent young researcher's presentation at Nanjing International Conference on Optimization and Numerical Algebra, 1999, and the Outstanding Young Researcher Award of the University of Hong Kong. He supervised more than 20 graduate students. As an Applied Mathematician, Michael's main research areas include bioinformatics, data mining, operations research, and scientific computing. Michael has published and edited 5 books, published more than 140 journal papers. He has reviewed papers for more than 40 international journals. He currently serves on the editorial boards of *Journal of Computational and Applied Mathematics* (Principal Editor); *SIAM Journal on Scientific Computing*; *Numerical Linear Algebra with Applications*; *International Journal of Data Mining and Bioinformatics*; *Multidimensional Systems and Signal Processing*; *International Journal of Computational Science and Engineering*, and was guest editor of several special issues of the international journals (*Journal of Computational Mathematics*, *International Journal of Applied Mathematics*, *Applied Mathematics and Computation*, *EURASIP Journal on Applied Signal Processing*, *International Journal of Imaging Systems and Technology*).



Huanfeng Shen received the B.S. degree in surveying and mapping engineering from Wuhan University, Wuhan, China, in 2002. He is currently pursuing the Ph.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan. His current research interests focus on image reconstruction, remote sensing, image processing and application.



Edmund Y. Lam received the B.S. degree (with distinction) in 1995, the M.S. degree in 1996, and the Ph.D. degree in 2000, all in electrical engineering from Stanford University, Stanford, Calif. At Stanford, he developed image processing algorithms for the Programmable Digital Camera project. He also consulted for industry in the areas of digital camera systems design and algorithms development. Before returning to academia, he was affiliated with the Reticle and Photomask Inspection Division (RAPID) of KLA-Tencor Corporation in San Jose, Calif, as a Senior Engineer, working in the design of defect detection tools for the core die-to-die and die-to-database inspections. He is currently an Assistant Professor of electrical and electronic engineering at the University of Hong Kong, as well as the Director of its Imaging Systems Laboratory. His research interests include electronic and computational imaging, and image processing applications in semiconductor manufacturing, biomedical engineering, and sensor networks. He is a Senior Member of IEEE and a Member of SPIE.



Liangpei Zhang received the B.S. degree in physics from Hunan Normal University, ChangSha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998. From 1997 to 2000, he was a Professor of School of the Land Sciences in Wuhan University, Wuhan. In August 2000, he joined the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, as a Professor and Head of the Remote Sensing Section. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence. He has served as Cochair of the SPIE Series Conferences on Multispectral Image Processing and Pattern Recognition (MIPPR), the Conference on Asia Remote Sensing in 1999, Editor of the MIPPR01, MIPPR05 Symposiums, Associate Editor of *Geo-spatial Information Science Journal*, and Chinese National Committee for the International Geosphere-Biosphere Programme.

