Optical Engineering 46(12), 127003 (December 2007)

Zoom-based super-resolution reconstruction approach using prior total variation

Michael K. Ng
Hong Kong Baptist University
Department of Mathematics
Kowloon Tong, Hong Kong
E-mail: mng@math.hkbu.edu.hk

Huanfeng Shen
Wuhan University
School of Geodesy and Geomatics
Wuhan, Hubei, China

Subhasis Chaudhuri
Indian Institute of Technology, Bombay
Department of Electrical Engineering
Mumbai 400076, India

Andy C. Yau
The University of Hong Kong
Department of Mathematics
Pokfulam Road, Hong Kong

Abstract. We present a robust and efficient approach for zoom-based super-resolution (SR) reconstruction problems. We employ the total variation (TV) of the desired image *priori* in the maximum *a-posteriori* estimation. An efficient algorithm based on iterative methods and preconditioning techniques is employed to solve the resulting variational problem. To suit the proposed algorithm for realistic imaging situations, a registration method is presented to simultaneously solve the zooming factors, image center shifts, and photometric parameters. Experimental results show that the proposed TV-based algorithm performs quite well in terms of both quantitative measurements and visual evaluation. We also demonstrate that the proposed algorithm is robust for SR image inpainting, where some pixels are missed in the SR reconstruction model.
 [DOI: 10.1117/1.2818797]

Subject terms: zoom super-resolution reconstruction; total variation; preconditioning; registration; inpainting.

Paper 070164R received Feb. 28, 2007; revised manuscript received Jul. 6, 2007; accepted for publication Jul. 10, 2007; published online Dec. 7, 2007.

1 Introduction

High-resolution (HR) images are indispensable in applications such as health diagnosis and monitoring, military surveillance, terrain mapping by remote sensing, etc. However, due to the high cost and physical limitations of high precision optics and image sensors, it is not easy to obtain the desired HR images in many cases. Alternatively, super-resolution (SR) image reconstruction, which refers to a process that increases spatial resolution by fusing information from a sequence of images acquired in one or more of several possible ways, has been a very hot research field.

The multiframe SR idea was first proposed by Tsai and Huang.[1] They used a frequency domain approach to demonstrate the ability to reconstruct a single improved resolution image from several down-sampled, noise-free versions of it. Kim, Bose, and Valenzuela[2] extended the formulation to consider observation noise as well as the effects of spatial blurring. They solved the extended formulation by a weighted recursive least-squares method to improve computational efficiency. Rhee and Kang[3] proposed a discrete cosine transform (DCT)-based algorithm in which the computational costs were reduced by using DCT instead of the discrete Fourier transform (DFT). Furthermore, a couple of papers have appeared that concentrate on wavelet-transform-based SR methods.[4–6]

In the spatial domain, Ur and Gross[7] suggested a nonuniform interpolation method. Irani and Peleg[8] proposed an iterative back-projection method adapted from computer-aided topography (CAT). Stark and Oskoui[9] proposed a noteworthy projections onto convex sets (POCS)-based formulation to SR image reconstruction problems. Their approach was extended by Tekalp, Qzkan, and Sezan to include observation noise,[10] and by Patti, Sezan, and Tekalp to include motion blur.[11] Patti, Sezan, and Tekalp[12] extended the POCS approach to account for arbitrary sampling lattices and nonzero aperture time. Shultz and Stevenson[13] developed a popular maximum *a-posteriori* (MAP) formulation to the SR problem. This MAP estimator uses a motion-compensated subsampling matrix-based observation model and an edge preserving Huber–Markov random field for the image prior. Ng et al. used a preconditioned conjugate gradient method to speed up the convergence in the MAP reconstruction.[14–16] Hardie, Barnard, and Armstrong[17] developed an approach to simultaneously estimate the image registration parameters and the HR image. Woods, Galatsanos, and Katsaggelos[18] presented stochastic methods in which the parameters of registration, noise, and image statistics are estimated jointly based on the available observations. Also in the MAP framework, Shen et al.[19] proposed a joint estimation approach combining motion estimation, segmentation, and super-resolution together. Elad and Feuer[20] proposed a unified methodology for SR reconstruction from several geometrically wrapped, blurred, noisy, and down-sampled observations by combining the MAP and POCS approaches. Recently, some researchers paid much attention to color,[21,22] compressed,[23,24] and dynamic[25] SR methods.

Traditionally, the researchers use the motion (sensor motion or scene motion) cue to super-resolve the desired HR image. In some SR literatures,[8,26,27] although the employed motion models can cover the sensor zoom, the additional

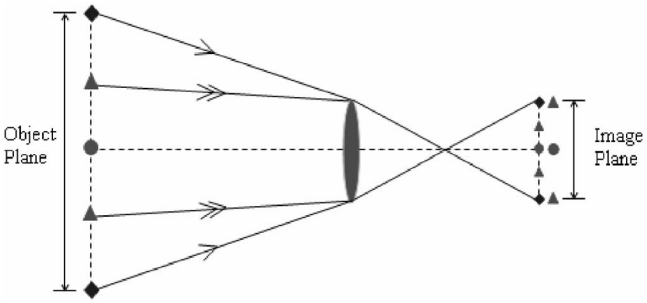

Fig. 1 Illustration of imaging with different focus. ≻ is the shorter focus, and ≫ is the longer focus.

information used for the SR reconstruction is still mainly rooted in the sensor motion and/or scene motion. However, it has been demonstrated that the sensor zoom itself also can be employed as an effective cue for SR problem in Refs. 28 to 31. We called this type of SR techniques zoom super-resolution (ZSR). Using an optical zoom camera, one can capture images with different zooming factors by adjusting the focus length of the camera lens. It is known that a long focus corresponds to higher spatial resolution, but to a narrower coverage of scene. For example, in Fig. 1, only the middle three objects can be observed under the longer focus condition. ZSR techniques can solve this tradeoff and obtain an image with high spatial resolution and wide coverage.[28] Li[29] suggests a line-geometric model for the video ZSR problem by assuming that the zooming speed is constant. In many cases, this assumption is somewhat restricted, especially for images captured by a still camera. In Ref. 30, Joshi, Chaudhuri, and Panuganti proposed a more elegant MAP method using a Markov prior to solve the ZSR problem. To preserve the discontinuities of an image, the authors include binary linear fields in the cost function and solve it using simulated annealing. The inclusion of a binary linear field can effectively preserve the edge information in an image. However, it is known that the simulated annealing often has a very slow convergence. Joshi, Chaudhuri, and Panuganti:[31] extended their method by using either a Markov random field (MRF) or a simultaneous autoregressive (SAR) model to learn the image prior from the most zoomed image. In theory, this method demands that the scene should be homogeneous to some extent.

The main aim of this work is to develop a robust but efficient ZSR algorithm, with which we can reconstruct an image with the entire scene as the least zoomed image and with the resolution as the most zoomed image from a sequence of images with different zooming factors. In addition, this algorithm is also expected to employ the image inpainting technique, which refers to the filling in of missing or corrupted regions in an image based on information available on the observed regions.[32–34] Thus, it can perform and conduct simultaneous image inpainting and ZSR for missing/corrupted pixels in the observed images. For this purpose, we use the total variation (TV) norm to model the image *priori* in the maximum *a-posterior* estimation to effectively preserve the edge information in the image. An efficient algorithm based on fixed-point iterations and preconditioning techniques is investigated to solve the associated Euler-Lagrange equations of the corresponding variational problem. To suit the proposed algorithm for realistic imaging situations, a registration method is presented to simultaneously solve the zooming factors, image center shifts, and photometric parameters. Experimental results show that the proposed method not only can effectively deal with the standard zoomed SR problem, but also has the efficacy to conduct simultaneous inpainting and SR. Moreover, the advantage of the proposed algorithm is further obvious when it is used to deal with noisy observations.

The rest of the work is organized as follows. In Sec. 2, we give an image observation model for the ZSR problem. We propose the TV-based ZSR reconstruction model in Sec. 3 and a preconditioned solver in Sec. 4. Section 5 presents the simultaneously geometric and photometric registration method. Experimental results are provided in Sec. 6 and Sec. 7 concludes the work.

2 Image Observation Model

The image observation model mainly relates the desired image with several observed images. For convenience, we name the desired image as the HR image and the observed images as LR images, although the desired image has the same resolution as the most zoomed observed one.

We assume the depth variation in the scene is not significant compared to the distance between the object and the camera. Thus, a pinhole imaging model can be assumed, and the depth-related perspective distortion in the image can be neglected.[30] We also assume the images are captured with relatively small aperture, so that we can neglect the effect of blurring due to the defocus. It has been mentioned that we want to obtain a HR image that has the same scene as the least zoomed image and the same resolution as the most zoomed image. Assuming the LR size is respectively N_1 and N_2 in the horizontal and vertical dimensions, the HR size denoted by HN_1 and HN_2 can be determined by

$$HN_1 = \text{round}\left(N_1 \times \frac{L_P}{L_1}\right)$$

$$HN_2 = \text{round}\left(N_2 \times \frac{L_P}{L_1}\right), \tag{1}$$

where L_1 and L_P are the zoom factors of the least zoomed image and the most zoomed image, respectively, and the mathematical function "*round*" rounds the value to the nearest integer.

To obtain the grid of the desired HR image, the simplest method is to divide the least zoomed image that has the entire scene into HN_1 by HN_2 units. This approach works well when the relative zoom factor L_P/L_1 is an integer value. However, when L_P/L_1 is fractional, it seems not to be perfect. In Fig. 2(a), the real line represents the LR grid of the least zoomed image, the dotted is the grid of the most zoomed image, and the dashed is the HR grid obtained by the method mentioned earlier. It can be seen that the most zoomed image needs to be resampled to fit the current HR grid. It is known that the sampling process will inevitably blur the edges and textures in the image. Since the most zoomed image has the same resolution as the desired HR image, it is better to directly move its pixels into the HR

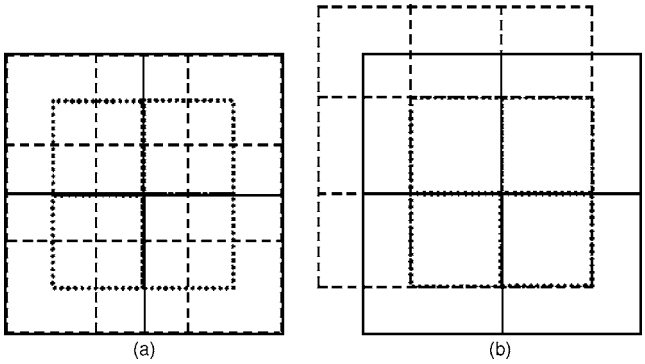

Fig. 2 Illustration of the adjustment of HR grid, (a) before and (b) after the adjustment.

grid without any geometric sampling. Therefore, we suggest slightly adjusting the HR grid to satisfy this requirement [see Fig. 2(b) for illustration].

In the grid system, each observed LR pixel can be regarded as a weighted sum of several HR pixels. Here, the weights relate with the relative zooming factor L_P/L_k (k is the image number) and the pixel position of the LR pixel in the HR grid. For example, in Fig. 3, the LR pixel g_1, which covers the HR pixels z_1, z_2, z_5, and z_6 partly or fully, can be expressed by

$$g_1 = (0.25 \times z_1 + 0.5 \times z_2 + 0.5 \times z_5 + z_6)/(1.5 \times 1.5), \quad (2)$$

where L_P/L_k is equal to 1.5. In this way, the relationship between the HR image and the observed LR image can be modeled in matrix notation as

$$\boldsymbol{g}_k = \boldsymbol{D}_k \boldsymbol{C}_k \boldsymbol{z} + \boldsymbol{n}_k. \quad (3)$$

In this expression, $\boldsymbol{g}_k$ is the lexicographically ordered LR image and $\boldsymbol{z}$ denotes the lexicographically ordered HR image. $\boldsymbol{C}_k$ is a cropping operator on $\boldsymbol{z}$ needed to handle the shrinkage of the view angle during the zooming process. It crops out the unobservable pixels from the HR image at an appropriate position.[28] $\boldsymbol{D}_k$ denotes the down-sampling matrix expressing the average relations in Eq. (2). Here $\boldsymbol{n}_k$ is the model noise.

In the prior image observation model, it is assumed that the image center does not change. Actually, the image center always move translationally as the lens parameters such as focus or zoom are varied, even if the camera is fixed.[30] With this in mind, we rewrite the observation model as

$$\boldsymbol{g}_k = \boldsymbol{D}_k \boldsymbol{C}_k \boldsymbol{z}_{\boldsymbol{a}_k} + \boldsymbol{n}_k, \quad (4)$$

where $z_{\boldsymbol{a}_k}(x,y) = z(x - a_{k,1}, y - a_{k,2})$ with x and y denoting the pixel locations, and $\boldsymbol{a}_k = (a_{k,1}, a_{k,2})$ represents the shift vector of the optical center due to zooming by the lens system for the k'th observation.

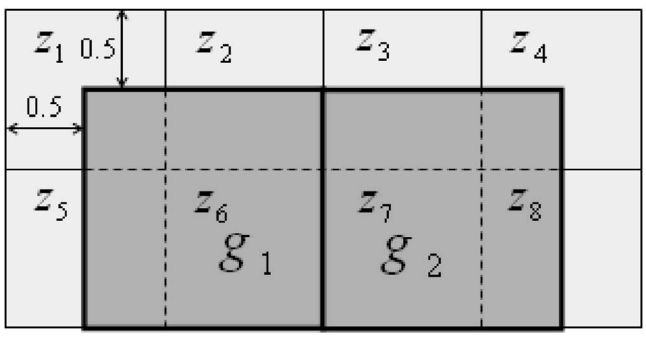

Fig. 3 Illustration of the relation between HR pixels and LR pixels.

3 Total Variation Zoom Super-Resolution Model

It is clear that ZSR reconstruction is an example of an inverse problem, where the source of information (desired HR image) is estimated from the observed data (zoomed LR images). In general, this inverse problem is ill-posed, because the information contained in the observed LR images is not sufficient to solve the desired HR image. Therefore, it is necessary to include regularization to stabilize the problem. Traditionally, regularization has been described from both algebraic and statistical perspectives.[35] Regularization can be regarded as some prior information of the desired image to constrain the image to some specific solution space.

We describe the ill-conditioned inverse problem from the statistical perspective using the maximum *a-posteriori* (MAP) estimator. Let the full set of P LR images be denoted as $\boldsymbol{g} = \{\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots \boldsymbol{g}_P\}$. The purpose is to realize the MAP estimate of HR image $\boldsymbol{z}$, given the observed LR images $\boldsymbol{g}$. The estimate can be computed by

$$\hat{\boldsymbol{z}} = \arg\max_{\boldsymbol{z}} p(\boldsymbol{z}|\boldsymbol{g}). \quad (5)$$

Applying Bayes' rule, Eq. (5) becomes:

$$\hat{\boldsymbol{z}} = \arg\max_{\boldsymbol{z}} \frac{p(\boldsymbol{g}|\boldsymbol{z})p(\boldsymbol{z})}{p(\boldsymbol{g})}. \quad (6)$$

Since $p(\boldsymbol{z}|\boldsymbol{g})$ is independent of $\boldsymbol{g}$, $p(\boldsymbol{g})$ can be considered a constant, and hence Eq. (6) can be rewritten as

$$\hat{\boldsymbol{z}} = \arg\max_{\boldsymbol{z}} p(\boldsymbol{g}|\boldsymbol{z})p(\boldsymbol{z}). \quad (7)$$

Assuming the LR images are independent, we obtain

$$\hat{\boldsymbol{z}} = \arg\max_{\boldsymbol{z}} \left[\prod_k p(\boldsymbol{g}_k|\boldsymbol{z}) \right] p(\boldsymbol{z}). \quad (8)$$

Using the monotonic log function, it can be expressed as

$$\hat{\boldsymbol{z}} = \arg\max_{\boldsymbol{z}} \left\{ \left[\sum_k \log p(\boldsymbol{g}_k|\boldsymbol{z}) \right] + \log p(\boldsymbol{z}) \right\}. \quad (9)$$

Usually, the noise $\boldsymbol{n}_k$ in Eq. (4) is assumed as a white zero mean Gaussian type, thus we give the likelihood distribution $p(\boldsymbol{g}_k|\boldsymbol{z})$ as

$$p(\boldsymbol{g}_k|\boldsymbol{z}) = \frac{1}{A_1} \exp\left\{ - \frac{\|\boldsymbol{O}_k(\boldsymbol{g}_k - \boldsymbol{D}_k \boldsymbol{C}_k \boldsymbol{z}_{\boldsymbol{\alpha}_k})\|^2}{2\sigma^2} \right\}, \quad (10)$$

where A_1 is a constant, and σ^2 is the error variance that is assumed to be the same in different observed images. Here, $\boldsymbol{O}_k$ is a diagonal matrix denoting which pixels are missing

in $\boldsymbol{g}_k$. Each pixel in $\boldsymbol{g}_k$ corresponds to a diagonal element in $\boldsymbol{O}_k$. If a pixel is missing, its corresponding diagonal element is 0, otherwise 1. The use of $\boldsymbol{O}_k$ removes the effects of the missing pixels from the distribution function, and provides the possibility to conduct the simultaneous inpainting and ZSR.

We assume the image prior $p(z)$ in Eq. (9) has the Gibbs form

$$p(z) = \frac{1}{A_2} \exp[-\beta U(z)], \tag{11}$$

where A_2 is also a constant, β is an adjusting factor, and $U(z)$ is the prior energy functional. Substituting Eqs. (10) and (11) in Eq. (9), after some manipulation, the maximization of this *posteriori* probability distribution is equivalent to the following minimization problem

$$\hat{z} = \arg\min\left[\sum_k \|\boldsymbol{O}_k(\boldsymbol{g}_k - \boldsymbol{D}_k\boldsymbol{C}_k\boldsymbol{z}_{\boldsymbol{\alpha}_k})\|^2 + \lambda U(z)\right], \tag{12}$$

where the first term $\Sigma_k\|\boldsymbol{O}_k(\boldsymbol{g}_k-\boldsymbol{D}_k\boldsymbol{C}_k\boldsymbol{z}_{\boldsymbol{\alpha}_k})\|^2$ provides a force of the conformance of the present HR image to the observed LR images according to the image observation model. The second term is the prior energy functional $U(z)$, which forces the prior on the image. $\lambda=2\beta\sigma^2$ is the confidence parameter or regularization parameter, which balances the two competing terms.

Traditionally, the Laplacian prior and Gauss-Markov image prior are commonly employed in the fields of image restoration and reconstruction. They regularize the corresponding ill-posed problem by forcing spatial smoothness on the image. However, some high frequency energy in the image tends to be removed in the use of these priors. This problem is more obvious when the images are contaminated by noise. Some edge-preserving priors have been used[13,21,26,36,37] in the literature. One of the successful methods is the TV prior, whose energy functional looks like[37]

$$U(z)| = \int_\Omega |\nabla z| \mathrm{d}x\mathrm{d}y = \int_\Omega \sqrt{|\nabla z|^2}\mathrm{d}x\mathrm{d}y, \tag{13}$$

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

$$U(z) = \int_\Omega \sqrt{|\nabla z|^2 + \beta}\mathrm{d}x\mathrm{d}y. \tag{14}$$

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

$$U(z) = \|\nabla z\|_{\mathrm{TV}} = \sum_x \sum_y \sqrt{|\nabla z_{x,y}|^2 + \beta}, \tag{15}$$

where $|\nabla z_{x,y}|^2$ can be approximated as

$$|\nabla z_{x,y}|^2 = (z[x+1,y] - z[x,y])^2 + (z[x,y+1] - z[x,y])^2. \tag{16}$$

By substituting Eq. (15) in Eq. (12), the following minimization function can be obtained

$$\hat{z} = \arg\min\left[\sum_k \|\boldsymbol{O}_k(\boldsymbol{g}_k - \boldsymbol{D}_k\boldsymbol{C}_k\boldsymbol{z}_{\boldsymbol{\alpha}_k})\|^2 + \lambda\|\nabla z\|_{\mathrm{TV}}\right]. \tag{17}$$

The distinctive feature of TV prior is that edges can be effectively preserved. The use of TV prior for image restoration has been proposed by Rudin, Osher, and Fatemi.[37] Chan and Shen[38] proposed a TV inpainting model for filling-in missing pixels and proved that the use of TV norm is desirable for the image inpainting problem. The TV prior has also been employed for the motion-based SR problem in some literatures,[36,39,40] in which the gradient descent algorithm, which often has slow convergence, is commonly used to solve the desired HR image. In the next section, we introduce a more efficient method for the optimization problem of Eq. (17).

4 Numerical Optimization

For expression convenience, we rewrite $\boldsymbol{z}_{\boldsymbol{\alpha}_k}$ in matrix form as $\boldsymbol{M}_k\boldsymbol{z}=\boldsymbol{z}_{\boldsymbol{\alpha}_k}$, with $\boldsymbol{M}_k$ being the shift operator corresponding to the shift vector a_k. Thus, the Euler-Lagrange equation for the energy function in Eq. (17) is given by the following nonlinear system

$$\nabla E(z) = \sum_k \boldsymbol{M}_k^T\boldsymbol{C}_k^T\boldsymbol{D}_k^T\boldsymbol{O}_k^T(\boldsymbol{O}_k\boldsymbol{D}_k\boldsymbol{C}_k\boldsymbol{M}_k\boldsymbol{z}-\boldsymbol{g}_k) - \lambda\boldsymbol{L}_z\boldsymbol{z} = 0, \tag{18}$$

where L_z is the matrix form of a central difference approximation of the differential operator

$$\nabla \cdot \left(\frac{\nabla}{\sqrt{|\nabla z|^2 + \beta}}\right).$$

here, $\nabla\cdot$ is the divergence operator.

For this optimization problem, Rudin, Osher, and Fatemi[37] proposed the artificial time marching scheme as the steady state of a parabolic partial differential equation. Indeed, it is equivalent to employing the following gradient descent (GD) method to solve the minimization problem

$$\boldsymbol{z}^{n+1} = \boldsymbol{z}^n - dt\,\nabla E_\lambda(\boldsymbol{z}^n), \tag{19}$$

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 main drawback of this method is the slow convergence.

In Ref. 41, the lagged diffusivity fixed point iteration is introduced. This method consists of linearizing the nonlinear differential term by lagging the diffusion coefficient

$$\frac{1}{\sqrt{|\nabla z|^2 + \beta}}$$

one iteration behind. It has been shown in Ref. 41 that the method is monotonically convergent. Using this strategy, $\boldsymbol{z}^{n+1}$ can obtained as the solution to the linear equation:

$$\left[\sum_k \boldsymbol{M}_k^T \boldsymbol{C}_k^T \boldsymbol{D}_k^T \boldsymbol{O}_k^T \boldsymbol{O}_k \boldsymbol{D}_k \boldsymbol{C}_k \boldsymbol{M}_k - \lambda \boldsymbol{L}_{z^n}\right] \boldsymbol{z}^{n+1} = \sum_k \boldsymbol{M}_k^T \boldsymbol{C}_k^T \boldsymbol{D}_k^T \boldsymbol{O}_k^T \boldsymbol{g}_k. \tag{20}$$

To solve Eq. (20), any linear optimization solution can be employed. Generally, the preconditioned conjugate gradient (PCG) method is desirable. Given a matrix equation $\boldsymbol{Ax}=\boldsymbol{b}$, there are two criteria for choosing a preconditioner for $\boldsymbol{A}$. First, the preconditioner should be a "good" approximation to $\boldsymbol{A}$. Second, it must be easily inverted.

To suit the specific matrix structures in image restoration and reconstruction, several preconditioners have been proposed.[42–45] Our experimental results indicate that an efficient way to solving the matrix equations arisen from high-resolution image reconstruction is to apply the factorized sparse inverse preconditioner (FSIP).[45] Let $\boldsymbol{A}$ be a symmetric positive definite matrix, and let its Cholesky factorization be $\boldsymbol{A}=\boldsymbol{GG}^T$. The idea of FSIP is to find the lower triangular matrix $\boldsymbol{L}$ with sparsity pattern S, such that

$$\|\boldsymbol{I}-\boldsymbol{GL}\|_F$$

is minimized, where $\|\cdot\|_F$ denotes the Frobenius norm. Kolotilina and Yeremin[45] showed that L can be obtained by the following algorithm.

1. Compute $\hat{\boldsymbol{L}}$ with sparse pattern S, such that $[\hat{\boldsymbol{L}}A]_{x,y} = \delta_{x,y}$, $(x,y) \in S$.
2. Let $\hat{\boldsymbol{D}}=[\mathrm{diag}(\hat{\boldsymbol{L}})]^{-1}$ and $\boldsymbol{L}=\hat{\boldsymbol{D}}^{1/2}\hat{\boldsymbol{L}}$.

According to this algorithm, n small linear systems need to be solved, where n is the number of rows in the matrix $\boldsymbol{A}$. These systems can be solved in parallel. Thus, the previous algorithm is also well suited for modern parallel computing. Motivated by the FSIP, we consider the factorized banded inverse preconditioner (FBIP), which is a special type of FSIP. It approximates the Cholesky factor $\boldsymbol{G}$ by banded lower triangular matrices.[42] In this work, the FBIP algorithm is employed to solve the ZSR problem by preconditioning the coefficient matrix in Eq. (20).

5 Registration

In most realistic situations, the relative zooming factor and the shift of image center are unknown. They should be estimated by some geometric registration method before implementing the SR scheme. Also, because of the possible illumination change across the scene and the automatic gain control (AGC) or automatic white balancing of the camera, the same objects often have different intensities in different acquired images. Therefore, the registration part should not only include geometric registration, but also include photometric registration, which refers to the estimation of the photometric transformations.

In Refs. 28, 30, and 31, the relative zooming and shift parameters between two observations are solved using a search strategy by minimizing the mean squared distance between an appropriate portion of the zoomed images. The photometric difference is separately registered and corrected. Here, we present a method to implement the geometric registration and photometric registration simultaneously. By using the global linear photometric transformation (suppose the illumination changes across the scene are global and can be expressed by a gain parameter and an offset parameter), which is commonly assumed,[26,46] the relation between the l'th and k'th images can be expressed as

$$g_k(x,y) = h_{k,1} \cdot g_l(q_k x + a_{k,1}, q_k y + a_{k,2}) + h_{k,0} + \varepsilon_k(x,y). \tag{21}$$

In this equation, $h_{k,1}$ and $h_{k,0}$ are the gain and offset values of the linear photometric transformation, respectively; q_k is the relative zooming factor; $a_{k,1}$ and $a_{k,2}$ are respectively the horizontal and vertical shifts (we remark that here the shift parameters are the LR versions, but we still use the same notations as their HR counterparts introduced in Sec. 2 for expression convenience); and $\varepsilon_k(x,y)$ denotes the model error. Let $\boldsymbol{\theta}$ denote the vector that contains both the two photometric parameters and four geometric parameters, and

$$f_k^{(l,\boldsymbol{\theta})}(x,y) = h_{k,1} \cdot g_l(q_k x + a_{k,1}, q_k y + a_{k,2}) + h_{k,0}$$

denote the predicted pixel of $g_k(x,y)$ from frame l using parameter vector $\boldsymbol{\theta}$. Thus, Eq. (21) can be rewritten by

$$g_k(x,y) = f_k^{(l,\boldsymbol{\theta})}(x,y) + \varepsilon_k(x,y). \tag{22}$$

The following quadratic cost function is employed as the minimization function

$$E(\boldsymbol{\theta}) = \|\boldsymbol{O}_k(\boldsymbol{g}_k - \boldsymbol{f}_k^{(l,\boldsymbol{\theta})})\|_2^2. \tag{23}$$

Here, $\boldsymbol{O}_k$ is used to remove the effects of the missing pixels as in the reconstruction part. Using the Gaussian-Newton method, the parameter vector $\boldsymbol{\theta}$ can be iteratively solved by

$$\Delta\boldsymbol{\theta}^n = [(\boldsymbol{J}^n)^T \boldsymbol{O}_k \boldsymbol{J}^n]^{-1}[-(\boldsymbol{J}^n)^T \boldsymbol{O}_k \boldsymbol{r}^n], \tag{24}$$

and

$$\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n + \Delta\boldsymbol{\theta}^n, \tag{25}$$

where n is the iteration number, $\Delta\boldsymbol{\theta}^n$ denotes the correction vector of $\boldsymbol{\theta}^n$, $\boldsymbol{r}^n$ is the image residual vector that is equal to $\boldsymbol{g}_k - \boldsymbol{f}_k^{(l,\boldsymbol{\theta}_n)}$, and $\boldsymbol{J}^n = (\partial\boldsymbol{r}/\partial\boldsymbol{\theta})^n$ denotes the gradient matrix of $\boldsymbol{r}^n$.

It is worthy to note that the registration parameters should be initiated before implementing this iteration process. In most cases, the center shift parameters $a_{k,1}$ and $a_{k,2}$ and the luminance offset $h_{k,0}$ can be initialized as 0, and the luminance gain $h_{k,1}$ can be initialized as 1. The relative zooming factor q_k, however, should be initiated more carefully. Otherwise, the optimization procedure could get trapped in a local minimum. Here, we employ the hierarchical searching strategy to obtain the initiated relative zooming factor from coarse to fine scales. Generally, the sum of squared difference (SSD) and the sum of absolute difference (SAD) are the two most commonly used matching criterions. However, the correlation coefficient criterion is more appropriate in our case, because it is not affected by linear transform. Thus, the effects of two photometric parameters can be neglected in the search process. The correlation coefficient is expressed as

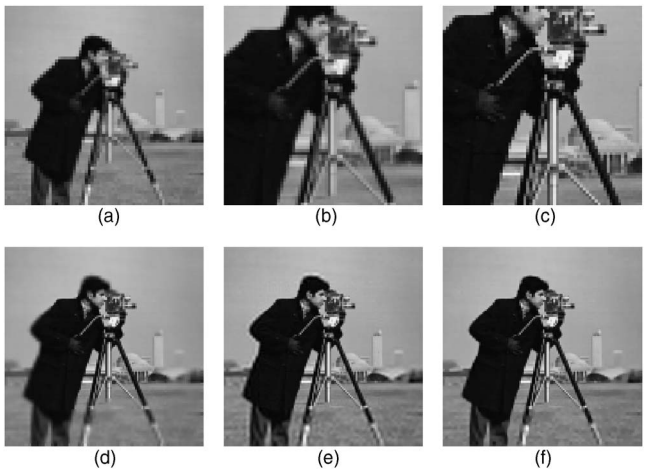

Fig. 4 HR estimates from noise-free observations in the synthetic experiment. (a) The least zoomed image, (b) the middle zoomed image, (c) the most zoomed image, (d) the bilinear interpolated image, (e) SR result using Laplacian prior, and (f) SR result using TV prior.

$$CC = \frac{\sigma_{gf}}{\sigma_g \sigma_f}. \tag{26}$$

Here, CC and σ_{gf} are respectively the correlation coefficient and covariance between $\boldsymbol{g}_k$ and $\boldsymbol{f}_k^{(l,\boldsymbol{\theta})}$, and σ_g and σ_f are their standard deviations.

Once the parameters are finally obtained after the initiation and iteration processes, the observed images first can be photometrically corrected by the photometric parameters, then the photometrically corrected images and the geometric parameters (the relative zooming factors and center shifts) are provided to the ZSR algorithm.

6 Experimental Results

We tested the efficacy of the proposed algorithm using a sequence of experiments. The Cameraman image was first used for synthetic simulations. Then, we used two real sequences to provide a more realistic demonstration of resolution improvement of the proposed algorithm.

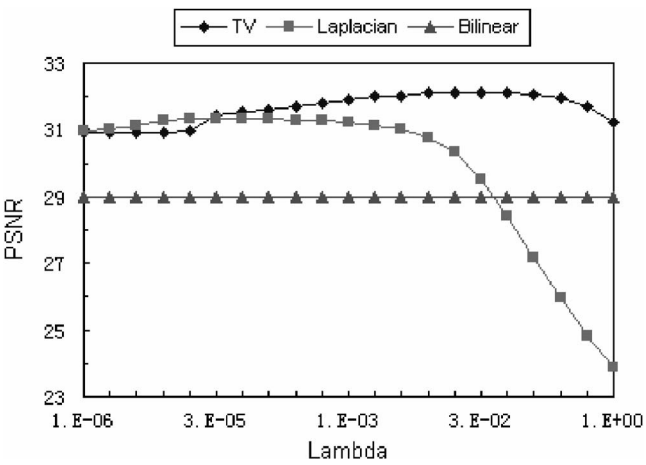

Fig. 5 PSNR values versus the confidence parameter in the Cameraman synthetic experiment on noise-free observations.

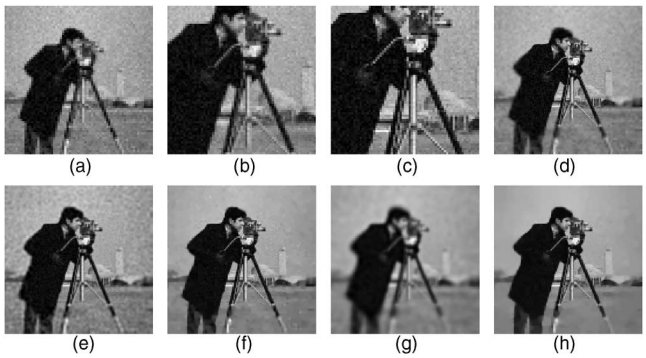

Fig. 6 HR estimates from noisy observations in the synthetic experiment. (a) The least zoomed image, (b) the middle zoomed image, (c) the most zoomed image, (d) the bilinear interpolated image, (e) SR result with the highest PSNR using Laplacian prior, (f) SR result with the highest PSNR using TV prior, (g) SR result with relative large confidence parameter using Laplacian prior, and (f) SR result with relative large confidence parameter using TV prior.

6.1 *Results Using Synthetic Images*

In the synthetic simulations, we experimented on the Cameraman image. The following peak signal-to-noise factor (PSNR) was employed as a quantitative measure.

$$\mathrm{PSNR} = 10 \log_{10}\left(\frac{255^2 * HN_1 * HN_2}{\|\hat{z} - z\|^2}\right), \tag{27}$$

where $HN_1 * HN_2$ is the total number of pixels in the HR image, and $\hat{z}$ and z represent the reconstructed HR image and the original image.

Following the image observation model, we created three LR images with the relative zooming factors being 1, 1.6, and 1.9. These images are respectively shown in Figs. 4(a)–4(c). Figure 4(d) is the bilinear interpolated image of the three LR images, which are obtained by interpolating the least zoomed image ($L=1$) first, then replacing the corresponding part by the interpolated version of the middle zoomed image ($L=1.6$), and finally replacing the corresponding pixels by the original pixels of the most zoomed

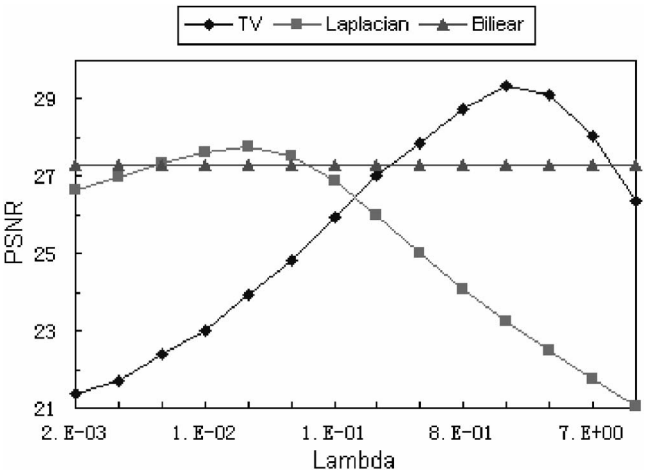

Fig. 7 PSNR values versus the confidence parameter in the Cameraman synthetic experiment on noisy observations.

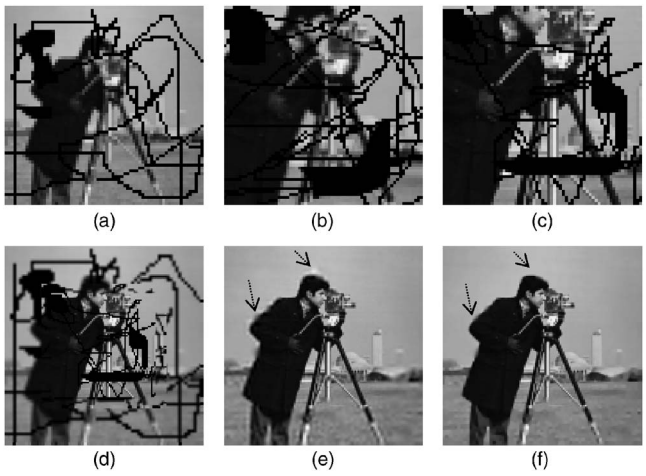

Fig. 8 HR estimates in the synthetic experiment with missing/corrupted pixels in the observations. (a) The least zoomed image, (b) the middle zoomed image, (c) the most zoomed image, (d) the bilinear interpolated image, (e) SR result using Laplacian prior, and (f) SR result using TV prior.

image ($L=1.9$). It is clear that the peripheral regions of this image are much smoother than those at the center. To obtain more desirable images, we implemented the proposed ZSR scheme regarding the bilinear interpolated image as an initial guess. The commonly employed Laplacian prior (the prior energy function is $\|\boldsymbol{Qz}\|^2$, where $\boldsymbol{Q}$ is the 2-D Laplacian operator) was also tested to make a comparative analysis with the TV prior. It is noted that the Laplacian prior generally has stronger constrain on the image than the TV prior, because its energy function is a square term and not extracted like that of the TV prior, so it should need a smaller confidence parameter than the TV. In fact, we should respectively choose the optimal confidence parameters λ for the two different priors to ensure fairness in the comparison. With this in mind, we tried a series of confidence parameters for both the two priors. Figure 5 demonstrates the PSNR values against the confidence parameter λ. The PSNR of bilinear interpolation is also plotted for comparison, although it has no relation with the confidence

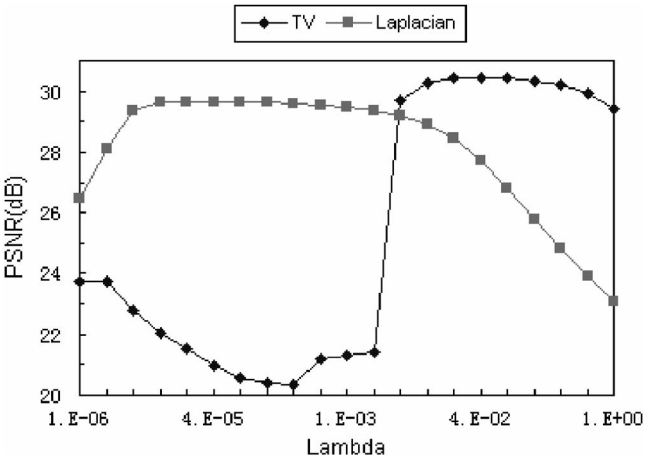

Fig. 9 PSNR values versus the confidence parameter in the Cameraman synthetic experiment on observations with missing pixels.

Table 1 PSNR values of TV reconstructions for various $\beta's$.

	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
Noise-free	32.107	32.116	32.130	32.130	32.139	32.138
Noisy	29.347	29.347	29.347	29.348	29.347	29.348
Missing	30.262	30.272	30.287	30.394	30.417	30.316

parameter. The PSNR value of the bilinear interpolation is fixed at 28.968. Using the Laplacian and TV priors, we can obtain the best PSNRs of 31.355 and 32.139 when λ is equal to 0.000064 and 0.032768, respectively. The SR images corresponding to the best PSNR values are shown in Figs. 4(e) and 4(f), respectively. It is found that both Figs. 4(e) and 4(f) are much clearer than Fig. 4(d). In addition, Fig. 4(f) has better visual quality than Fig. 4(e), especially around the edges. The visual evaluation agrees with the quantitative measurement.

We further tested our algorithm on the noisy observations. We added a zero mean Gaussian noise with variance 65.025 on Figs. 4(a)–4(c) in the experiment. The resulting noisy versions are demonstrated in Figs. 6(a)–6(c). Figure 7 demonstrates the PSNR values of the bilinear interpolation, Laplacian prior, and TV prior results versus the confidence parameter in the noisy case. The bilinear interpolation image has a constant PSNR value of 27.276 and is shown in Fig. 6(d). The best PSNR value for the Laplacian prior is 27.773, with confidence parameter λ being 0.0256. Note that this PSNR value is not much higher than that of bilinear interpolation. Using the TV prior, however, we obtained a best PSNR value of 29.347 when λ is equal to 1.6384. The images corresponding to the best PSNR values are shown in Figs. 6(e) and 6(f), respectively. It is observed that both the two images are still noisy to different extents. Therefore, we also give the results with larger confidence parameters. The parameter of the Laplacian is 1.6384, and the TV 3.2768. The SR results are shown in Figs. 6(g) and 6(h), respectively. It is clear the Laplacian-based algorithm fails to effectively remove the noise, even with a large λ, and smoothes much detailed information in the image. The TV-based algorithm can provide simultaneous denoisng and edge preservation. The PSNRs of these two images are 23.245 (Laplacian) versus 29.112 (TV).

We also tested our algorithm for simultaneous inpainting and ZSR. Figures 8(a)–8(c) are the simulated observations. In this case, the quality of the bilinear interpolated image is very poor [see Fig. 8(d)]. The PSNR is only 12.698. In the SR reconstructions, we again tried a series of confidence parameters to find the best PSNR results in the SR reconstructions. The Laplacian and TV priors obtained the best PSNR values of 29.648 and 30.417 with $\lambda=0.00004$ and $\lambda=0.02048$, respectively. PSNR values versus λ are shown in Fig. 9. Visually, the missing regions can be more desirably inpainted and SR reconstructed using the TV-based algorithm, especially around the image edges. Please see the arrow-indicated regions in Figs. 8(e) and 8(f).

In the previous three experiments (noise-free, noisy, and missing cases), the parameter β in Eq. (15) was fixed as 10^{-5}. The authors of Ref. 41 have shown that taking small

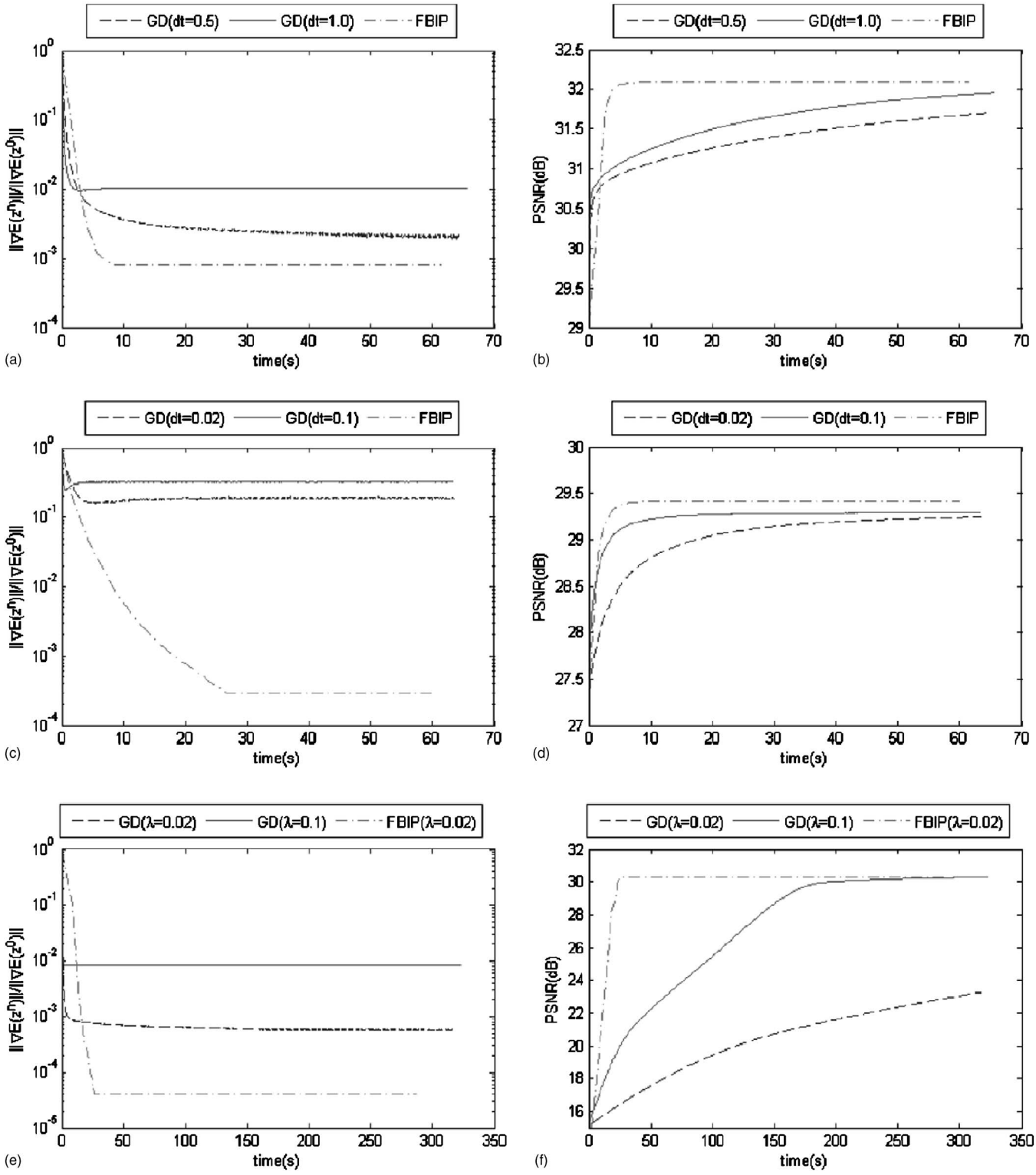

Fig. 10 Convergence performance of the FBIP algorithm compared to the GD algorithm. (a) Noise-free case measured by gradient norm, (b) noise-free case measured by PSNR value, (c) noisy case measured by gradient norm, (d) noisy case measured by PSNR value, (e) missing case measured by gradient norm, and (f) missing case measured by PSNR value.

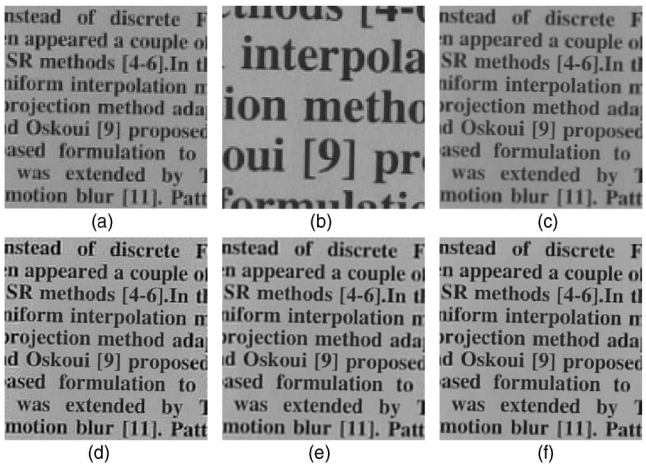

Fig. 11 HR estimates in the Text experiment. (a) The least zoomed image, (b) the most zoomed image, (c) the bilinear interpolated image, (d) Laplacian SR result with $\lambda=0.002$, (e) Laplacian SR result with $\lambda=0.064$, and (f) TV SR result with $\lambda=1$.

but positive β gives minimizers that are close to minimizers obtained with $\beta=0$ in the denoising problem. Here, we also tested the effect of β on the ZSR reconstruction results. The PSNR values for various $\beta's$ in the prior three cases are shown in Table 1 (for each case, the confidence parameter is fixed). We can see the robust performance of the TV algorithm is not sensitive to the choice of β.

The next issue is on the evaluation of the efficiency of the proposed algorithm. We compare the employed the FBIP preconditioning algorithm to the gradient descent (GD) algorithm. In Fig. 10, the left shows the evolution of the gradient-norm-based convergence condition $\|\nabla E(z^n)\|/\|\nabla E(z^0)\|$ against the computational time, and the right is the demonstration of PSNR value versus the computational time. It is noted that the selection of the step-size dt in the GD algorithm is not easy, because an overly small value results in a very slow convergence speed, and an overly large value often leads to divergence. Therefore, in the noise-free and noisy cases, we respectively give the results using two carefully selected step sizes (at least one of them is almost the best) to compare to the FBIP algorithm. Here, the confidence parameters were respectively fixed as 0.016384 (noise-free) and 1.6384 (noisy). In the missing case, since the convergence of the GD algorithm is extremely slow when the confidence parameter is equal to that of the FBIP algorithm ($\lambda=0.0248$), we also give the result with a larger confidence parameter $\lambda=0.1$. From both the gradient norm and PSNR convergence criterions, the FBIP algorithm greatly outperforms the GD algorithm.

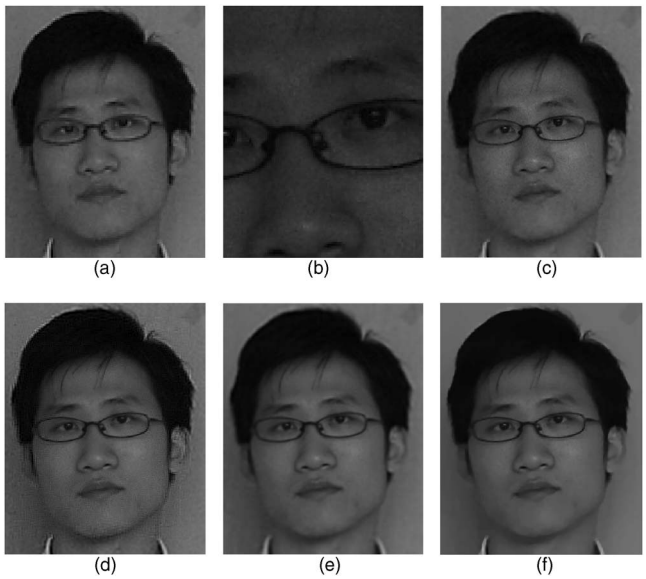

Fig. 12 HR estimates in the Face experiment. (a) The least zoomed image, (b) the most zoomed image, (c) the bilinear interpolated image, (d) Laplacian SR result with $\lambda=0.01$, (d) Laplacian SR result with $\lambda=0.64$, and (e) TV SR result with $\lambda=1.6$.

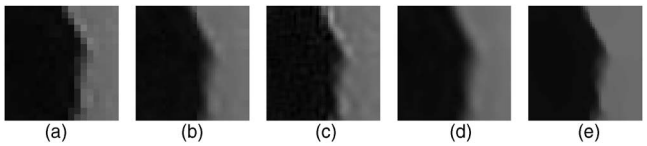

Fig. 13 (a) through (e) The first series of detained regions cropped from Fig. 12(a) and Figs. 12(c) through 12(f).

6.2 *Results Using Real Images*

Now we consider applying the proposed algorithm to real zoomed images obtained with a consumer-level digital camera. The registration method introduced in Sec. 5 is used to obtain the zooming, shifted, and photometric parameters. To reduce the computational expense, our processing was restricted to typical regions in the original images.

The first realistic imaging experiment was on the Text sequence, which contains six zoomed observations. The least and most zoomed images are respectively shown in Figs. 11(a) and 11(b). Figure 11(c) is the bilinear interpolated image. Figure 11(d) is the ZSR result using the Laplacian prior with relatively small confidence parameter 0.002. There are obvious artifacts in the image. Figure 11(e) is also the Laplacian result, but with relatively large confidence parameter 0.064. It can be seen that the artifacts are still perceptible. Moreover, the characters around the image boundary have been oversmoothed with the increase of the λ. We show the SR result of the proposed algorithm with $\lambda=1$ in Fig. 11(f). Obviously, this result is more visually attractive than the two former ones.

The second was on a Face sequence. Nine zoomed images were obtained under dark illumination conditions, which led to images contaminated by noise. The least and most zoomed images and the reconstructed results are shown in Fig. 12. In this experiment, we obtained similar

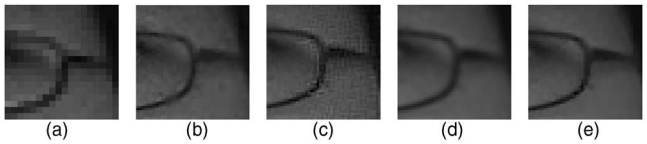

Fig. 14 (a) through (e) The second series of detained regions cropped from Fig. 12(a) and Figs. 12(c) through 12(f).

conclusions as in the synthetic experiment on noisy observations. From Figs. 12(d) and 12(e), which are respectively the results using the Laplacian prior with $\lambda=0.002$ and $\lambda=0.64$, we can see that the Laplacian prior does not remove the noises with a small confidence parameter, and it overly smooths the image if the confidence parameter is too large. The proposed TV-based algorithm, however, is more robust to noise [please see Fig. 12(f)], which is the result of using the TV prior with $\lambda=1.6$. For convenience of visual judgment, two series of detained regions cropped from Figs. 12(a) and Figs. 12(c)–12(f) are shown in Figs. 13(a)–13(e) and Figs. 14(a)–14(e), respectively.

7 Conclusions

We propose an effective and efficient algorithm for zoom-based SR image reconstruction. We employ the TV prior to set up the ZSR model and use the FBIP preconditioning algorithm to find the solution. For realistic imaging situations, a registration method is presented to simultaneously solve the zooming factors, image center shifts, and photometric parameters. Experiment results validate that the proposed reconstruction algorithm is very robust and efficient, and its advantage is further obvious when it is used to deal with noisy observations. We also demonstrate that the proposed algorithm also can effectively conduct simultaneous image inpainting and ZSR.

References

1. R. Y. Tsai and T. S. Huang, "Multi-frame image restoration and registration," *Comput. Graph. Image Process.* **1**(2), 317–339 (1984).
2. S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Process.* **38**(6), 1013–1027 (June 1990).
3. S. Rhee and M. G. Kang, "Discrete cosine transform based regularized high-resolution image reconstruction algorithm," *Opt. Eng.* **38**(8), 1348–1356 (1999).
4. R. Chan, T. Chan, L. Shen, and Z. Shen, "Wavelet algorithms for high-resolution image reconstruction," *SIAM J. Sci. Comput. (USA)* **24**(4), 1408–1432 (2003).
5. M. K. Ng, C. K. Sze, and S. P. Yung, "Wavelet algorithms for deblurring models," *Int. J. Imaging Syst. Technol.* **14**(3), 113–121 (2004).
6. N. Nguyen and P. Milanfar, "A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution)," *Circuits Syst. Signal Process.* **19**(4), 321–338 (2000).
7. H. Ur and D. Gross, "Improved resolution from sub-pixel shifted pictures," *CVGIP: Graph. Models Image Process.* **54**, 181–186 (Mar. 1992).
8. M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.* **53**(3), 231–239 (1991).
9. H. Stark and P. Oskoui, "High-resolution image recovery from image plane arrays, using convex projections," *J. Opt. Soc. Am. A* **6**, 1715–1726 (1989).
10. A. M. Tekalp and M. K. Ozkan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," *IEEE Proc. Intl. Conf. Acoustics, Speech Signal Process.*, San Francisco, CA (1992), Vol. **3**, pp. 169–172.
11. 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," *IEEE Proc. Intl. Conf. Image Process.*, Austin, TX (1994), pp. 343–347.
12. A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Process.* **6**(8), 1064–1076 (Aug. 1997).
13. R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.* **5**(6), 996–1011 (June 1996).
14. M. K. Ng and A. M. Yip, "A fast MAP algorithm for high-resolution image reconstruction with multisensors," *Multidimens. Syst. Signal Process.* **12**(2), 143–164 (Apr. 2001).
15. M. K. Ng, W. C. Kwan, and R. H. Chan, "A fast algorithm for high-resolution color image reconstruction with multisensors," in *Numerical Analysis and Its Applications*, Lecture Notes in Computer Science, vol. **1988**, pp. 615–627 (2001).
16. M. K. Ng and N. K. Bose, "Mathematical analysis of super-resolution methodology," *IEEE Signal Process. Mag.* **20**(3), 62–74 (2003).
17. R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.* **6**(12), 1621–1633 (Dec. 1997).
18. N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Trans. Image Process.* **15**(1), 201–213 (Jan. 2006).
19. H. Shen, L. Zhang, B. Huang, and P. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," *IEEE Trans. Image Process.* **16**(2), 479–490 (Feb. 2007).
20. M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Process.* **6**(12), 1646–1658 (Dec. 1997).
21. S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Trans. Image Process.* **15**(1), 141–159 (2006).
22. A. Zomet and S. Peleg, "Multi-sensor super-resolution," *IEEE Proc. Workshop Appl. Computer Vision* (2002), pp. 27–31.
23. B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of compressed video using transform-domain statistics," *IEEE Trans. Image Process.* **13**(1), 33–43 (2004).
24. C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video," *IEEE Signal Process. Mag.* **20**(3), 37–48 (May 2003).
25. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Dynamic demosaicing and color super-resolution of video sequences," *Proc. SPIE* **5562**, 169–178 (2004).
26. D. Capel and A. Zisserman, "Computer vision applied to super resolution," *IEEE Signal Process. Mag.* **20**(3), 75–86 (2003).
27. R. R. Schultz, L. Meng, and R. L. Stevenson, "Subpixel motion estimation for super-resolution image sequence enhancement," *J. Visual Commun. Image Represent* **9**(1), 38–50 (1998).
28. S. Chaudhuri and J. Manjunath, *Motion-Free Super-Resolution*, Springer-Verlag, New York (2005).
29. X. Li, "Super-resolution for synthetic zooming," *EURASIP J. Appl. Signal Process.* **2006**, 12 (2006).
30. M. V. Joshi, S. Chaudhuri, and R. Panuganti, "Super-resolution imaging: Use of zoom as a cue," *Image Vis. Comput.* **22**(14), 1185–1196 (2004).
31. M. V. Joshi, S. Chaudhuri, and R. Panuganti, "A learning-based method for image super-resolution from zoomed observations," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.* **35**(3), 527–537 (2005).
32. M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," *Proc. ACM SIGGRAPH 2000*, New Orleans, LA (2000), pp. 417–424.
33. T. F. Chan and J. H. Shen, "Variational image inpainting," *Commun. Pure Appl. Math.* **58**(5), 579–619 (May 2005).
34. M. Fornasier, "Nonlinear projection recovery in digital inpainting for color image restoration," *J. Math. Imaging Vision* **24**(3), 359–373 (2006).
35. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *Int. J. Imaging Syst. Technol.* **14**(2), 47–57 (2004).
36. S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.* **13**(10), 1327–1344 (Oct. 2004).
37. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, 259–268 (1992).
38. T. F. Chan and J. Shen, "Mathematical models for local nontexture inpaintings," *SIAM J. Appl. Math.* **62**(3), 1019–1043 (2002).
39. D. Capel and A. Zisserman, "Super-resolution enhancement of text image sequences," presented at *15th Intl. Conf. Patt. Recog.*, Barcelona, Spain (2000).
40. H. Yu Bing and L. Nan Wu, "Super resolution reconstruction of video sequence based on total variation," *Proc. Intl. Symp. Intelligent Multimedia, Video and Speech Process.*, Hong Kong, China (2004), pp. 575–578.
41. C. Vogel and M. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput. (USA)* **17**(1), 227–238 (1996).
42. F. R. Lin, M. K. Mg, and W. K. Ching, "Factorized banded inverse preconditioners for matrices with Toeplitz structure," *SIAM J. Sci. Comput. (USA)* **26**(6), 1852–1870 (2005).
43. R. H. Chan, T. F. Chan, and C. K. Wong, "Cosine transform based preconditioners for total variation deblurring," *IEEE Trans. Image Process.* **8**(10), 1472–1478 (Oct. 1999).
44. M. K. Ng, R. H. Chan, T. F. Chan, and A. M. Yip, "Cosine transform preconditioners for high resolution image reconstruction," *Linear Algebr. Appl.* **316**(1-3), 89–104 (Sep. 1 2000).

45. L. Kolotilina and A. Yeremin, "Factorized sparse approximate inverse preconditionings I: theory," *SIAM J. Matrix Anal. Appl.* **14**, 45–58 (1993).
46. D. Roberts, Y. Yamaguchi, and R. Lyon, "Calibration of airborne imaging spectrometer data to percent reflectance using field spectral measurements," presented at *19th Intl. Symp. Remote Sens. Environment*, MI (1985).

Huanfeng Shen received the BS degree in surveying and mapping engineering, and the PhD degree in photogrammetry and remote sensing from Wuhan University, China, in 2002 and 2007, respectively. His current, research interests focus on image reconstruction, and remote sensing image processing and applications.

Biographies and photographs of the other authors not available.