Signal Processing 90 (2010) 405–414

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A multi-frame image super-resolution method

Xuelong Li [a,*], Yanting Hu [b], Xinbo Gao [c], Dacheng Tao [d], Beijia Ning [c]

[a] State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China
[b] School of Medical Technology and Engineering, Xinjiang Medical University, Xinjiang 830001, PR China
[c] School of Electronic Engineering, Xidian University, Xi'an 710071, PR China
[d] School of Computer Engineering, Nanyang Technological University, 50 Nanyang Avenue, Blk N4, 639798, Singapore

ARTICLE INFO

Article history:
Received 30 March 2009
Received in revised form
27 May 2009
Accepted 28 May 2009
Available online 6 June 2009

Keywords:
Computer vision
Machine learning
Super-resolution
Regularization
Fuzzy entropy

ABSTRACT

Multi-frame image super-resolution (SR) aims to utilize information from a set of low-resolution (LR) images to compose a high-resolution (HR) one. As it is desirable or essential in many real applications, recent years have witnessed the growing interest in the problem of multi-frame SR reconstruction. This set of algorithms commonly utilizes a linear observation model to construct the relationship between the recorded LR images to the unknown reconstructed HR image estimates. Recently, regularization-based schemes have been demonstrated to be effective because SR reconstruction is actually an ill-posed problem. Working within this promising framework, this paper first proposes two new regularization items, termed as locally adaptive bilateral total variation and consistency of gradients, to keep edges and flat regions, which are implicitly described in LR images, sharp and smooth, respectively. Thereafter, the combination of the proposed regularization items is superior to existing regularization items because it considers both edges and flat regions while existing ones consider only edges. Thorough experimental results show the effectiveness of the new algorithm for SR reconstruction.

1. Introduction

In many military and civilian applications, high-resolution (HR) images are desirable and always required. HR means that the number of pixels within a given size of image is large. Therefore, an HR image usually offers important or even critical information for various practical applications. In recent decades, charge-coupled device (CCD) and CMOS image sensors have been widely used in imaging systems. Although these sensors work well for many imaging-based applications, the current resolution level in these sensors does not meet the increasing demands in the near future. Therefore, it is essential to find an effective way to expand the resolution of low-resolution (LR) images.

A straightforward solution to increase the spatial resolution of LR images is to reduce the pixel size by sensor manufacturing techniques, i.e., to increase the number of photo-detector for a given area of sensor chip. As the pixel size decreases, however, the power of light incident to each single photo-detector also decreases, and thus the image quality is degraded severely by the insufficient signal-to-noise ratio. Therefore, there exists a limitation of the pixel size reduction below which the suffering effect of shot noise could dominate. Unfortunately, the current image sensor technology has almost reached this limitation, i.e., it is impossible to obtain HR image through reducing the size of pixel [1].

Another approach to improve the spatial resolution of LR images is to increase the size of a sensor chip. This

* Corresponding author. +86 29 88887717; fax: +86 29 88887711.
E-mail address: xuelong_li@opt.ac.cn (X.L. Li).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.05.028

means more photo-detectors will be involved in an image-capturing device. However, a larger size of sensor leads to an increase in the capacitance on a chip, which in turn leads to lower charge transfer rate and a longer period of time to capture an image [1]. So this approach is inefficient. In addition, large size of sensor chips is inconvenient to many practical applications, e.g., satellite imagery. Finally, this approach will also increase the cost and it is not acceptable for cost-sensitive commercial applications.

As a consequence, there is an urgent need for developing post-acquisition signal processing techniques to enhance the resolution. These techniques offer flexibility as well as the cost benefit because there is no additional hardware involved. However, an increased computational cost may be the burden that a user has to suffer. Such a resolution enhancement is called super-resolution (SR) image reconstruction. SR reconstruction restores an HR image by using several LR images or a video sequence, while eliminating noises and blurs introduced by optical devices and the limited size of embedded sensor chips. It is an effective way to increase the resolution of a sequence of degraded images and has attracted extensive attention of researchers in signal processing, computer vision, and machine learning. Also, it has been widely applied to many applications, e.g., remote sensing, medical imaging, data mining, petroleum exploration, military information gathering and high definition television (HDTV).

Popular SR reconstruction algorithms can be roughly divided into two categories: frequency domain algorithms and spatial domain algorithms.

For frequency domain algorithms, Tsai and Huang [1] proposed the first work for the SR reconstruction by estimating the relative shifts between observations. Their approach is based on the following three aspects: the property of shifting of Fourier transform, the spectral aliasing principle, and the limited bandwidth of the original HR image. Based on this algorithm, a series of improved SR reconstruction algorithms had been proposed [2–4].

For spatial domain algorithms, representative works are given as follows. The non-uniform interpolation-based approaches [5,6]: their common advantage is that their computational cost is relatively low so they are ready for real-time applications. However, degradation models are not applicable in these approaches if the blur and the noise characteristics are different for LR images. Projections on a convex set (POCS)-based methods [7,8]: their common advantage is simplicity, i.e., the utilization of the spatial domain observation model and inclusion of a priori information. However, their disadvantages are non-uniqueness of solutions, slow convergence rate and heavy computational load. Iterative back projection (IBP)-based approaches [13]: they conduct SR reconstruction in a straightforward way. However, they have no unique solution due to the ill-posed nature of the inverse problem and some parameters are difficult to choose. Additionally, it is difficult to combine priori constraints with these approaches. Bayesian *Maximum A-posteriori* (MAP) estimation-based methods [12,20,21]: compared with IBP-based methods, they explicitly use the *priori* information in the form of a prior probability density on an HR image and provide a rigorous theoretical framework. In [21], an MAP-based joint formulation is proposed and it judiciously combines motion estimation, segmentation, and super resolution together. This formulation is used for a complex super-resolution problem in which the scenes contain multiple independently moving objects. Regularization-based approaches [9–11], learning-based SR method [14,18] , space-time SR method [15] and color image SR method [22,23] have been proposed. Spatial domain approaches are better in adaptability and lead to better SR reconstruction results than frequency domain approaches, and thus become popular in recent years.

Among all spatial domain approaches, what is worthy of mentioning is the regularization-based methods, which are effective to solve the multi-frame SR reconstruction problem, the focus of this paper. Because SR reconstruction is an ill-posed problem, their common point is to integrate a priori knowledge (represented by a regularization item) into the process of SR reconstruction to obtain a stable solution. Tikhonov regularization reconstruction method [19] is one of the most representative regularization-based algorithms for SR reconstruction. It introduces smoothness constraints to suppress the noise in reconstructed images, but it loses some details (e.g., edges) in LR images. Another representative work is proposed by Farsiu et al. [16], who introduced the bilateral total variation (BTV) operator as a regularization term measured by L_1 norm. This approach is more robust and can preserve more details (e.g., edges) than Tikhonov regularization method. However, this approach fails to consider the partial smoothness of an image, i.e., it is not locally adaptive, and thus it has limited adaptive capability in the process of SR reconstruction and cannot balance the suppression of noise against the preservation of image details.

To reduce shortcomings of the aforementioned Tikhonov regularization method and Farsiu's SR reconstruction algorithm, in this paper, we propose a new SR reconstruction algorithm, which can, respectively, keep edges and flat regions implicitly described in LR images sharp and smooth in the restored HR image. In the proposed approach, locally adaptive bilateral total variation (LABTV) operator, which is measured by the fuzzy-entropy-based neighborhood homogeneous measurement, is used as a regularization item to constrain the smoothness of the reconstructed images. At the same time, gradient error term is introduced as gradient homogeneity constraint term to further improve the reconstructed images. To improve the robustness to estimation error of this method, LABTV regularization term is measured with adaptive L_p norm, while data error term and gradient error term are measured with L_1 norm.

The rest of the paper is organized as follows. Section 2 introduces the imaging degradation model. In Section 3, we first describe the data error term in SR reconstruction, then propose two new regularization items, i.e., LABTV and consistency of the gradient (CG), and finally show how to implement the proposed SR reconstruction algorithm. Section 4 evaluates the new algorithm in

comparison with representative SR reconstruction methods and Section 5 concludes.

2. Imaging degradation model

It is necessary to set up a suitable imaging model, which can mimic the physical process of imaging degradation, to implement the super-resolution reconstruction. In the practical process of image sampling, there exist some degrading factors, i.e., atmospheric turbulence, object motion, optical blurring, sampling devices, and sampling process. Considering these main factors, the block diagram of the imaging degradation system is shown in Fig. 1.

Since optical blurring from camera has greater effect on images than atmospheric blurring in conventional imaging system, only optical blurring is considered in this paper. We apply the imaging degradation model developed by Farsiu et al. [16] to construct a connection between high-resolution images and corresponding low-resolution images.

Suppose that the expected high-resolution image vector is $\underline{X}=[x_1,x_2,\ldots,x_{Q^2}]^T$ and $Q=rM$, where r is the down-sampling scalar. The size of the kth observed image Y_k is $M\times M$, wherein $\underline{Y}_k=[y_1^k,y_2^k,\ldots y_{M^2}^k]$, $k=1,2,\ldots,N$, and N is the number of corresponding low-resolution images. According to [16], by imposing the additive noises on, the observation model $\underline{Y}_k$ is given by

$$\underline{Y}_k = D_k H_k^{cam} F_k \underline{X} + \underline{V}_k, \quad k=1,2,\ldots,N, \tag{1}$$

where F_k is the motion matrix for modeling the motion degradation process of the kth LR image, H_k^{cam} the blur matrix for representing the point spread function (PSF) of the camera sensor, D_k the down-sampling matrix, and $\underline{V}_k$ the system additive noise.

3. Adaptive edge-preserving regularized SR

The observation model defined in Eq. (1) describes the direct LR image acquisition process by an imaging degradation system. Based on Eq. (1), we can estimate the corresponding HR image from observed LR images, and this process is terms as the super-resolution. However, the operators H_k, related to the point spread functions, are derived from the discretization of compact operator, so SR process is ill-posed. Thus, even small changes in LR images can result in large perturbations in the final solution and there exist an infinite number of solutions when (1) is solved directly. To obtain a stable solution, a specific regularization $\Upsilon(\underline{X})$ is always imposed on the observation model. The regularization $\Upsilon(\underline{X})$ can incorporate prior knowledge of the desirable HR solution, e.g., degree of smoothness. So, additional constraints that favor well-behaved solutions can be enforced by specific regularization to remove artifacts from final result. Accordingly, SR process can be converted to a generalized minimization cost function [16] i.e.,

$$\min J(\underline{X}), \quad J(\underline{X}) = \sum_{k=1}^{N} \rho(\underline{Y}_k, D_k H_k F_k \underline{X}) + \lambda \Upsilon(\underline{X}), \tag{2}$$

where $\underline{X}$ is the unknown high-resolution image to be estimated, $\underline{Y}_k$ the kth observed LR image, λ the Lagrangian constant coefficient, and ρ the distance between the observation and an estimation. F_k, H_k^{cam}, and D_k in (2) are motion matrix, blur matrix, and down-sampling matrix, respectively.

To better preserve edges in the estimated HR image, a gradient-based constraint is introduced to Eq. (2),

$$J(\underline{X}) = \sum_{k=1}^{N} \rho(\underline{Y}_k D_k H_k F_k \underline{X}) + \lambda_1 \Upsilon(\underline{X}) + \lambda_2 \sum_{k=1}^{N} \rho'(\nabla \underline{Y}_k, D_k F_k \nabla \underline{X}), \tag{3}$$

where ρ' measures the distance between the gradient of LR images and the estimated HR image. The first item in the right-hand side of Eq. (3), namely with the data error term, measures the reconstruction error to ensure that pixels in the reconstructed HR image are approaching to real values. The second item in the right-hand side of Eq. (3), namely with the regularization term, controls the smoothness of the reconstructed HR image. The third item in the right-hand side of Eq. (3), namely with the gradient error term, guarantees the consistency between the gradient of estimated HR image and that of original LR images. Coefficients λ_1 and λ_2 balance mentioned terms.

3.1. Data error term

Usually, the data error term used in Eqs. (2) and (3) is defined by the $L_p(1\le p\le 2)$ norm of the residual, i.e.,

$$\sum_{k=1}^{N} \rho(\underline{Y}_k, D_k H_k F_k \underline{X}) = \sum_{k=1}^{N} \|D_k H_k F_k \underline{X} - \underline{Y}_k\|_p^p. \tag{4}$$

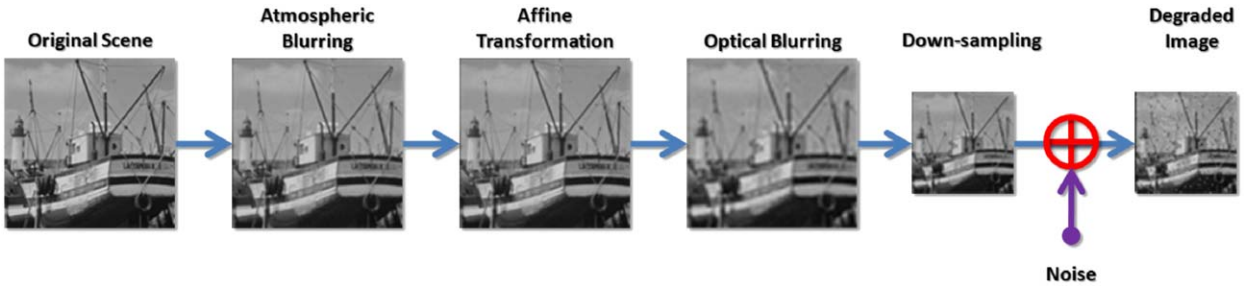

Fig. 1. The block diagram of an imaging system.

According to [16], L_1 norm leads to a more robust result in error estimation than L_2 norm, so we adopt this result in this paper.

3.2. Locally adaptive bilateral total variation

The regularization term controls the perturbation of the solution, solves the ill-posed problem for SR reconstruction, and guarantees a stable HR estimation. Farsiu et al. [16] proposed the bilateral total variation model for regularization by combining the total variation and the bilateral filter,

$$\Upsilon_{BTV}(\underline{X}) = \sum_{l=-w}^{w}\sum_{m=0}^{w} \alpha^{|m|+|l|} \| \underline{X} - S_x^l S_y^m \underline{X} \|, \tag{5}$$

where $l+m \geq 0$, S_x^l,and S_y^m are shift matrices to present l and m pixels shift in horizontal and vertical directions, respectively, and α $(0<\alpha<1)$ is the weighting coefficient.

The weighting coefficient α affects the estimated HR image significantly. Small α sharpens edges while amplifying noises in the estimated HR image, as shown in Fig. 2b. Large α helps to suppress noise while smoothing the estimated HR image, i.e., edges in the estimated HR image will be blurred, as shown in Fig. 2c.

Based on the above fact, it is desirable to impose small α on edges and large α to smooth regions for the HR image estimation, so that both the edge preservation and noise suppression can be achieved simultaneously. To achieve this objective, we propose a novel regularization item, termed locally adaptive bilateral total variation.

3.2.1. Definition of LABTV

Suppose that the vector form of a $Q \times Q$ image $\underline{X} = [x_{i,j}]_{Q\times Q}$ is given by $\underline{X} = [x_1, x_2, \ldots, x_{Q^2}]^T$ and the LABTV regularization is defined as

$$\Upsilon_{ABTV}(\underline{X}) = \sum_{l=-w}^{w}\sum_{m=-w}^{w} \frac{1}{p_{\underline{X}}(m,l)} \Phi(m,l)^{|m|+|l|} \| \underline{X} - S_x^l S_y^m \underline{X} \|_{p_{\underline{X}}(m,l)}^{p_{\underline{X}}(m,l)}, \tag{6}$$

where

$$P_{\underline{X}}(m,l) = [p_{x_k}(m,l)]_{k=1,\ldots,Q^2} = [p_{i,j}(m,l)]_{\substack{i=1,2,\ldots,Q \\ j=1,2,\ldots,Q}} \quad (k=(j-1)\times Q + i), \tag{7}$$

where the range of the kth entry of $p_{\underline{X}}(m,l)$ is $1 \leq p_{x_k}(m,l) \leq 2$ and its value varies adaptively with pixels

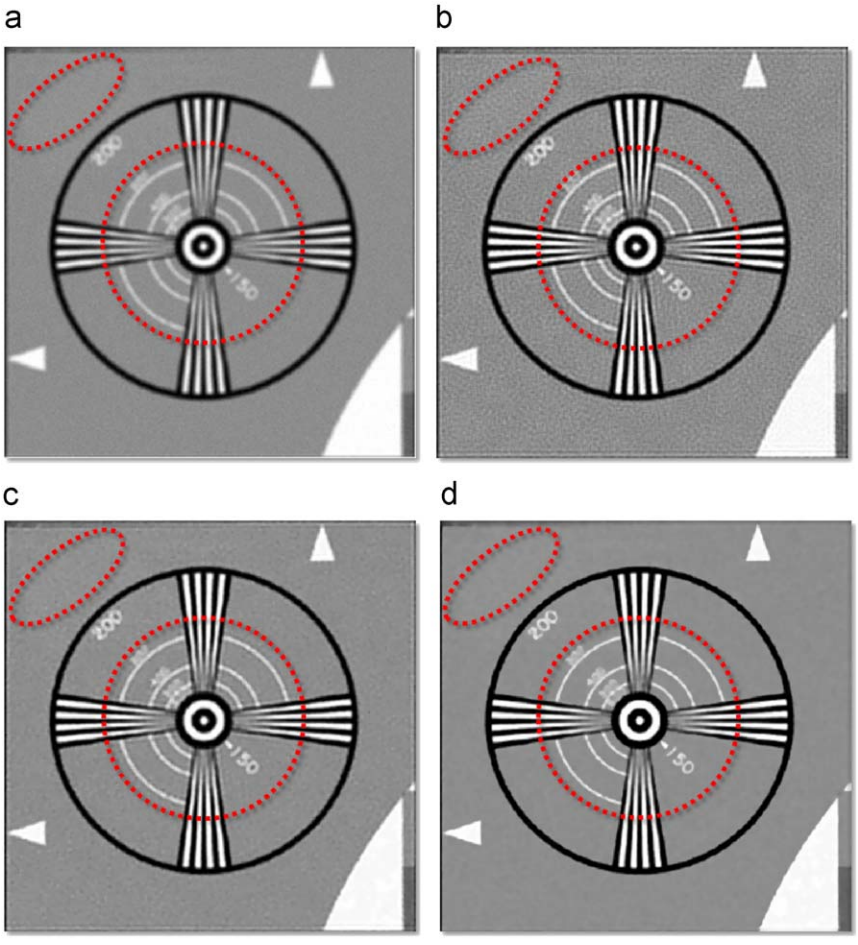

Fig. 2. The proposed LABTV for restoration in comparison with the original BTV with two different settings for SR reconstruction: LABTV outperforms BTV because it preserves edges and suppresses noises: (a) degraded image, (b) L_1+BTV ($\alpha = 0.1$), (c) L_1+BTV ($\alpha = 0.9$), and (d) L_1+LBTV.

and their shifts and the diagonal $Q^2 \times Q^2$ adaptive weighting coefficient matrix $\Phi(m,l)$ contains adaptive weighing coefficients.

3.2.2. Determination of $\Phi(m, l)$

Diagonal entries of $\Phi(m, l)$ are obtained from image local features, which responses to local smoothness. In this paper, we apply the fuzzy entropy-based neighborhood homogeneity $R_{i,j}$ to measure the local smoothness for the $(2n+1)\times(2n+1)$ patch centered at (i, j), i.e.,

$$R_{i,j} = \frac{1}{(2n+1)^2}\sum_{k=-n}^{n}\sum_{l=-n}^{n} H_t(u_t(x_{i+k,j+l})), \tag{8}$$

where $H_t(u_t)$ is the fuzzy entropy of u_t and $u_t = 1/(1 + |x_{i,j} - t|/D)$ is the membership function of pixel $x_{i,j}$. The range of $R_{i,j}$ is [0, 1]. The smaller the value of $R_{i,j}$ is, the smoother the patch centered at (i, j) will be. Moreover, $R_{i,j}$ is robust to noise (i.e., it is noise resisting) and sensitive to weak edges. Therefore, $R_{i,j}$ is a suitable choice to determine $\Phi(m, l)$.

With $R_{i,j}$, we define the kth diagonal entry of $\Phi(m, l)$ as

$$\phi_{k,k}(m,l) = \begin{cases} \alpha_{i,j}, & \alpha_{i,j} \le \alpha_{i+m,j+l} \\ \alpha_{i+m,j+l}, & \alpha_{i,j} > \alpha_{i+m,j+l} \end{cases} \quad (k = (j-1)\times Q + i), \tag{9}$$

where $\alpha_{i,j} = 1/(1+\eta R_{i,j})$ and η is a non-zero constant to control the value and scope of $\alpha_{i,j}$. Based on the definition of $\phi_{k,k}(m, l)$ in Eq. (9), we have the following observations: (1) the range of $\phi_{k,k}(m, l)$ is $(0,1)$; (2) $\phi_{k,k}(m, l)\to 1$ if the pixel (i, j) and its neighbor $(i+m, j+l)$ are in smooth regions because both $\alpha_{i,j}$ and $\alpha_{i+m,j+l}$ are approaching to 1; and (3) on the contrary to (2), $\phi_{k,k}(m, l)\to 0$ if the pixel (i, j) and its neighbor $(i+m, j+l)$ are in non-smooth regions. In summary, $\phi_{k,k}(m, l)$ is adaptive to the smoothness of a patch and achieves our objective for designing weighing coefficient.

3.2.3. The adaptive selection of L_p norm

In order to overcome shortcomings of L_2 and L_1 norms while inheriting their advantages, L_1 norm in Eq. (5) is replaced by L_p $(1<p<2)$ norm over pixels, and p is determined by the difference between a specific pixel and its neighbors. To reduce noise for smooth regions, p should be set with large values, i.e., $p\to 2$, while to preserve edges for non-smooth regions, p should be set with small values, i.e., $p\to 1$. Therefore, $p_{i,j}(m, l)$, the p norm for the pixel (i, j) with its neighbors (m, l), is given by

$$p_{i,j}(m,l) = 1 + \frac{1}{1+\delta|x_{i,j} - x_{i+m,j+l}|}, \tag{10}$$

where δ is a positive constant. According to Eq. (10), we know $P_{i,j}(m, l)\to 1$, i.e., $p_{\underline{X}}(m, l)\to 1$, in an edge area, so the L_1 norm is used to preserve edges. On the contrary, in a smooth area, $P_{i,j}(m, l)\to 2$, i.e., $p_{\underline{X}}(m, l)\to 2$ in Eq. (6), so the L_2 norm is used to smooth the area. Therefore, L_p norm is adaptively determined.

3.2.4. Analysis of the proposed adaptive regularization

In the proposed adaptive regularization, both the weighting coefficient matrix $\Phi(m, l)$ and the norm parameter $p_{\underline{X}}(m, l)$ (i.e., $p_{i,j}(m, l)$) can both be estimated by the initially reconstructed HR image. And the adaptive regularization term can be calculated from Eq. (6). This regularization term has different effects on pixels with different degrees of smoothness. Its smooth effect on a non-border pixel $x_{i,j}$ in an image $\underline{X}$ can be presented by the variation of the LABTV operator, i.e., $\partial\Upsilon_{ABTV}(\underline{X})/\partial x_{i,j}$.

For simplicity, supposing $w = 1$ in Eq. (6), let

$$\varphi(x_{i,j} - x_{i+m,j+l}) = \frac{1}{p_{i,j}(m,l)}(x_{i,j} - x_{i+m,j+l})^{p_{i,j}(m,l)}, \tag{11}$$

then we have

$$\begin{aligned}\frac{\partial\Upsilon_{ABTV}(\underline{X})}{\partial x_{i,j}} &= \frac{\partial}{\partial x_{i,j}}\sum_{l=-1}^{1}\sum_{m=-1}^{1}\phi_{k,k}^{|m|+|l|}(m,l)\varphi(x_{i,j} - x_{i+l,j+m}) \\ &= \sum_{l=-1}^{1}\sum_{m=-1}^{1}\phi_{k,k}^{|m|+|l|}(m,l)\varphi'(x_{i,j} - x_{i+l,j+m}).\end{aligned} \tag{12}$$

When $(x_{i,ja} - x_{i+m,j+l})\to 0$, we have $P_{i,j}(m, l)\to 2$, so

$$\lim_{(x_{i,j}-x_{i+m,j+l})\to 0}\frac{\varphi'(x_{i,j} - x_{i+m,j+l})}{(x_{i,j} - x_{i+m,j+l})} = 1.$$

Therefore, each item in Eq. (12) can be written as

$$\phi_{k,k}^{|m|+|l|}(m,l)\frac{\varphi'(x_{i,j} - x_{i+m,j+l})(x_{i,j} - x_{i+m,j+l})}{(x_{i,j} - x_{i+m,j+l})}$$

and Eq. (12) can be rewritten as

$$\frac{\partial\Upsilon_{ABTV}(\underline{X})}{\partial x_{i,j}} = \xi_{\Sigma}x_{i,j} - \sum_{l=-1}^{1}\sum_{m=-1}^{1}\xi_{m,l}x_{i+m,j+l}, \tag{13}$$

where

$$\xi_{m,l} = \phi_{k,k}^{|m|+|l|}(m,l)\frac{\varphi'(x_{i,j} - x_{i+m,j+l})}{(x_{i,j} - x_{i+m,j+l})},$$

$$\xi_{\Sigma} = \sum_{l=-1}^{1}\sum_{m=-1}^{1}\xi_{m,l},$$

and indices l and m cannot be zero at the same time. The LABTV regularization term processes the pixel $x_{i,j}$ by weighted summation of all its 8-neighbors according to the following weighting matrix:

$$W_{ABTV} = -\begin{bmatrix} \xi_{-1,-1} & \xi_{-1,0} & \xi_{-1,1} \\ \xi_{0,-1} & -\xi_{\Sigma} & \xi_{0,1} \\ \xi_{1,-1} & \xi_{1,0} & \xi_{1,1} \end{bmatrix}. \tag{14}$$

For a smooth area in an image, $p_{i,j}(m,l)\to 2$ in Eq. (10) because $\phi_{k,k}^{|m|+|l|}(m, l)\to 1$ in Eq. (9). Then we have,

$$\xi_{m,l} = \lim_{(x_{i,j}-x_{i-1,j+1})\to 0}\phi_{k,k}^{|m|+|l|}(m,l)\frac{\varphi'(x_{i,j} - x_{i+m,j+l})}{(x_{i,j} - x_{i+m,j+l})} = 1$$

and

$$\xi_{\Sigma} = \sum_{l=-1}^{1}\sum_{m=-1}^{1}\xi_{m,l}$$

is around 8. Therefore, W_{ABTV} in Eq. (14) is reduced to a smoothing filter.

For a non-smooth area in an image, we can assume $(x_{i,j}-x_{i-1,j+1})\to\infty$ (a large number), and then we have $p_{i,j}(-1,1)\to 1$ and $\phi_{k,k}(-1,1)\to 0$. Therefore, $\xi_{-1,1}=0$, that is pixel $x_{i-1,j+1}$ almost contributes nothing to smooth pixel $x_{i,j}$.

In summary, the proposed LABTV regularization term is adaptive to different types of areas, i.e., LABTV can preserve edges for non-smooth areas and reduce noises for smooth areas. To further justify the effectiveness of LABTV at empirical level, we conduct a simple experiment as shown in Fig. 2, which visually evaluates the restorations in comparison with BTV [16] with different settings.

Fig. 2a is the blurred and Gaussian noised image. Fig. 2b and c are restorations by using the robust SR method (L_1+BTV) in [16] with weight α being 0.1 and 0.9, respectively. Fig. 3d shows the restoration by replacing BTV with the proposed LABTV in the place of that in the robust SR method [16] (L_1+LABTV), while the iterative step is kept and other parameters in the method are the same as those in obtaining (b) and (c). The peak signal-to-noise ratio (PSNR) in (b)–(d) are 23.40, 24.23 and 26.39, respectively. In Fig. 2b, edges are well preserved while noises are amplified because α is small. In Fig. 2c, noises are suppressed while edges are smoothed because α is large. From Fig. 2d, we can find edges are well preserved and noises are suppressed because LABTV is an adaptive regularization term.

3.3. The consistency of gradients

To further preserve the edge information, we can consider the consistency of the gradient between the reconstructed HR image and the observed LR image. The CG is defined as

$$\rho'(\nabla\underline{Y}_k, D_kF_k\nabla\underline{X}) = \sum_{k=1}^{N}\sum_{i=1}^{4}|D_kF_k(\nabla_i\underline{X}) - \nabla_i\underline{Y}_k|, \tag{15}$$

where the gradient is calculated in four direction without considering the blurring effect, the L_1 norm is used to measure the difference between the gradient maps, F_k is the motion matrix, D_k is the down-sampling matrix, and $\nabla_1\underline{X}$, $\nabla_2\underline{X}$, $\nabla_3\underline{X}$, and $\nabla_4\underline{X}$, are gradient vectors in horizontal, vertical, and two diagonal directions, respectively.

In this paper, $\nabla_1\underline{X}$, $\nabla_2\underline{X}$, $\nabla_3\underline{X}$, and $\nabla_4\underline{X}$ are calculated according to

$$\begin{aligned}
\nabla_1\underline{X} &= S_x^1\underline{X} + S_x^{-1}\underline{X} - 2\underline{X};\\
\nabla_2\underline{X} &= S_y^1\underline{X} + S_y^{-1}\underline{X} - 2\underline{X};\\
\nabla_3\underline{X} &= \tfrac{1}{2}(S_x^{-1}S_y^1\underline{X}) + \tfrac{1}{2}(S_x^1S_y^{-1}\underline{X}) - \underline{X};\\
\nabla_4\underline{X} &= \tfrac{1}{2}(S_x^1S_y^1\underline{X}) + \tfrac{1}{2}(S_x^{-1}S_y^{-1}\underline{X}) - \underline{X},
\end{aligned} \tag{16}$$

where $S_x^{\pm1}$ and $S_y^{\pm1}$ are shift matrices and $\nabla_i\underline{Y}_k$ the gradient of the observed LR image. The calculation of $\nabla_i\underline{Y}_k$ is the same as that of $\nabla_i\underline{X}$.

In Fig. 3, we show the effectiveness of CG for edge preservation in SR reconstruction. Fig. 3a is an observed LR image, Fig. 3b is the reconstructed HR image by using the robust SR method in [16], and Fig. 3c is the reconstructed HR image by using the combination of Eq. (16) and BTV in [16]. As shown in Fig. 3, the proposed CG can further enhance the PSNR from 20.866 to 21.203.

3.4. Implementation

Based on the proposed LABTV regularization term in Eq. (6) and CG defined in Eq. (15), the Lagrangian objective function for HR image reconstruction is defined as

$$\begin{aligned}
\hat{\underline{X}} = \underset{\underline{X}}{Arg\,Min}\Bigg[&\sum_{k=1}^{N}\|D_kH_kF_k\underline{X} - \underline{Y}_k\|_1 + \lambda_1\sum_{l=-w}^{w}\sum_{m=-w}^{w}\frac{1}{p_{\underline{X}}(m,l)}\\
&\times\Phi(m,l)^{|m|+|l|}\|\underline{X} - S_x^lS_y^m\underline{X}\|_{p_{\underline{X}}(m,l)}^{p_{\underline{X}}(m,l)}\\
&+\lambda_2\sum_{k=1}^{N}\sum_{i=1}^{4}\|D_kF_k(\nabla_i\underline{X}) - \nabla_i\underline{Y}_k\|_1\Bigg].
\end{aligned} \tag{17}$$

There is no closed form solution for Eq. (17) and the steepest descent is adopted in this paper to find a solution of (18). The calculation iteration is given by

$$\begin{aligned}
\hat{\underline{X}}_{n+1} = \hat{\underline{X}}_n - \beta\Bigg\{&\sum_{k=1}^{N}F_k^TH_k^TD_k^T\,\mathrm{sign}(D_kH_kF_k\hat{\underline{X}}_n - \underline{Y}_k)\\
&+\lambda_1\sum_{l=-w}^{w}\sum_{m=-w}^{w}\Phi(m,l)^{|m|+|l|}[I - S_x^{-l}S_y^{-m}]\\
&\times\mathrm{sign}(\hat{\underline{X}}_n - S_x^lS_y^m\hat{\underline{X}}_n)|\hat{\underline{X}}_n - S_x^lS_y^m\hat{\underline{X}}_n|^{p_{\underline{X}}(m,l)-1}\\
&+\lambda_2\sum_{k=1}^{N}\sum_{i=1}^{4}F_k^TD_k^T\,\mathrm{sign}(D_kF_k(\nabla_i\hat{\underline{X}}_n) - \nabla_i\underline{Y}_k)\frac{\partial\nabla_i\hat{\underline{X}}_n}{\partial\hat{\underline{X}}_n}\Bigg\},
\end{aligned} \tag{18}$$

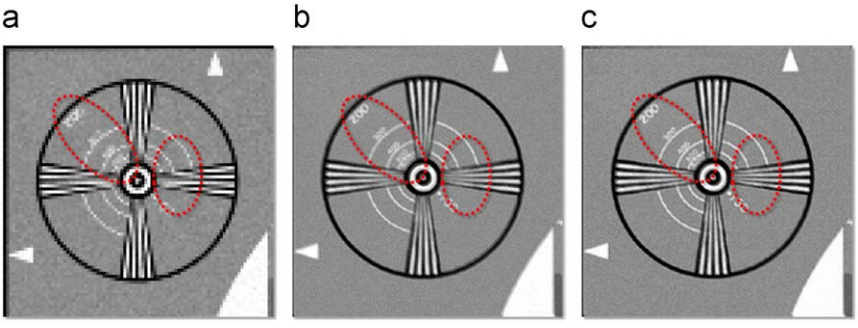

Fig. 3. CG is helpful to further enhance edges in the reconstructed HR image: (a) degraded image, (b) L_1+BTV, and (c) L_1+BTV+CG.

where β is the learning rate; S_x^{-l} and S_y^{-m} are transposes of matrices S_x^l and S_y^m to represent opposite shifting directions of S_x^l and S_y^m, respectively; and $\partial\nabla_i \underline{X}/\partial \underline{X}$ is the partial derivative of gradients in one of the four directions with respect to $\underline{X}$. The multiplication between matrices F, H, D, S with their transposes and the image vector can be implemented by shift, shift-back, convolution, down-sampling and up-sampling operations on the image directly to speed up the SR reconstruction process and to save both the time and space costs for computation [16,17].

4. Experimental results and analysis

In order to show the SR reconstruction improvement achievable with the new algorithm based on the proposed LABTV and gradient consistency, three typical 256-gray-level HR images, i.e., Eia, Boat and Shopping, in image processing were selected for performance evaluation. To validate the effectiveness and robustness of the proposed algorithm (denoted as "L_1+LABTV+CG") for SR reconstruction, it was evaluated subjectively and objectively in comparison with bi-cubic interpolation (denoted as "bi-cubic") and the robust SR method [16] (denoted as "L_1+BTV"). We also test the effectiveness of the proposed LABTV and denote it as "L_1+LABTV".

In the following experiments, HR images were degraded by using Eq. (1) and several LR images were obtained. The parameter selection criterion in each experiment for each algorithm was to choose parameters to produce visually most appealing results. For fair evaluation, we applied each algorithm for SR reconstruction several times based on different parameters and the best result of each time was chosen as the output of the algorithm for this experiment.

Fig. 4 evaluates the SR reconstructions based on the proposed algorithm in comparison with various existing algorithms. Fig. 4a shows the original HR image Eia with size of 360×360. The HR image was degraded with 16 predefined shift vectors and blurred by a Gaussian low-pass filter with size of 4×4 and standard deviation of 1. These 16 degraded images were then down-sampled horizontally and vertically with factor of 4 and Gaussian noise was further introduced on these LR images. Fig. 4b shows one representative LR example. Fig. 4c shows the SR reconstruction obtained by using the bi-cubic interpolation. Fig. 4d shows the result obtained by using the robust SR method [16]. Fig. 4e shows the effectiveness of the ABVT and Fig. 4f shows the SR reconstruction of the proposed "L_1+LABTV+CG". The PSNRs of different approaches are given in Table 1.

Fig. 4 shows that bi-cubic interpolation method performs worst because it conducts SR reconstruction based on a single LR image. In comparison with the SR reconstruction result obtained by L_1+BTV [16], the result shown in Fig. 4e obtained by the proposed L_1+LABTV looks better because LABTV has better capability to restore

Table 1
Comparison of PSNR from SR reconstruction methods.

Image	Figure ID	Bi-cubic	L_1+BTV	L_1+LABTV	L_1+LABTV+CG
Eia	4	17.5218	20.8661	21.6246	**21.953**
Boat	5	20.4988	28.1957	32.1679	**32.3295**
	6	23.2732	28.9194	32.1229	**32.328**
Shopping	7	20.9798	29.0454	32.2163	**32.3349**

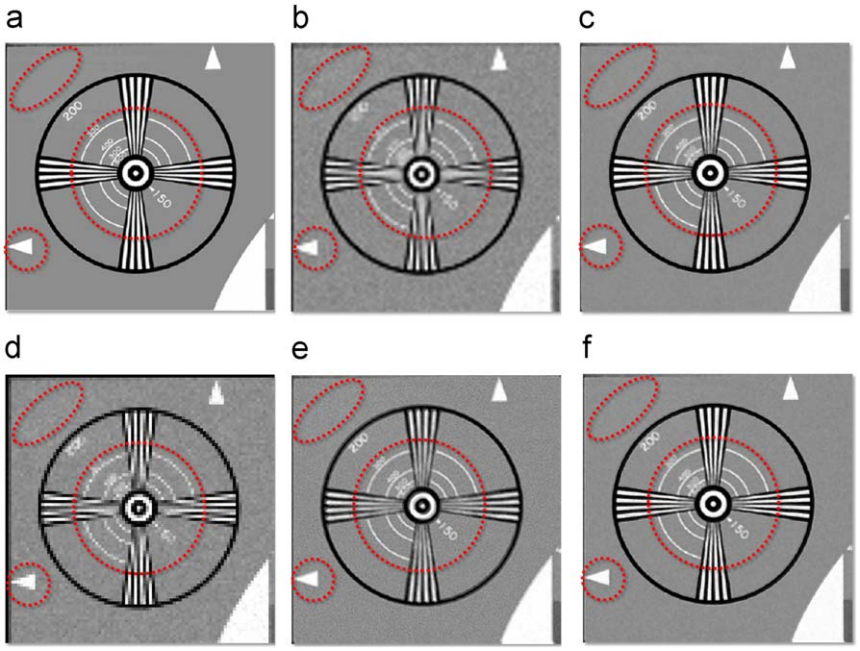

Fig. 4. Comparison of reconstruction results of image Eia: (a) original image, (b) degraded image, (c) bi-cubic, (d) L_1+BTV, (e) L_1+ABTV, and (f) L_1+ABTV+CG.

image details, to preserve edges, to reduce noises in smooth areas, and to increase the brightness of the whole image. The result shown in Fig. 4f obtained by the proposed L_1+LABTV+CG is better than Fig. 4e because the CG constraint is helpful to preserve more edge information than the algorithm without this constraint.

Experiments 2 and 3 were designed to evaluate the robustness of the proposed algorithm. In experiment 2, we evaluate SR reconstruction results obtained by different algorithms by using an incorrect noise model. In experiment 3, we evaluate SR reconstruction results obtained by different algorithms by using an incorrect motion estimation model. Both experiments were conducted on *Boat* with size of 360×360 and 9 LR images were obtained in the same way as used in the experiment 1, i.e., down-sampling factor was 3, the size of the blurring operator was 5×5, and the standard deviation in the Gaussian low-pass filter was 1. The PSNRs of different approaches are given in Table 1.

Fig. 5b shows that one of 9 LR images noised by the salt-and-pepper noise in experiment 2. For SR reconstruction, the Gaussian white noise with zero mean was used as the noise model. The SR reconstruction result obtained by L_1+BTV method [16] is shown in Fig. 5d. Fig. 5e and f illustrate the SR reconstruction results obtained by using the proposed L_1+LABTV and L_1+LABTV+CG.

In the experiment 3, 2 out of 9 generated images were selected arbitrarily. One was processed by an affine transformation, and the other was magnified and clipped, as shown in Fig. 6a and b, respectively. For all SR reconstruction algorithms, we assumed there was only shift between LR images. The original image used in this experiment was the same as that in the experiment 2, as shown in Fig. 5a. Fig. 6e is the result obtained by using L_1+BTV [16]. Results obtained by using the proposed L_1+LABTV and L_1+LABTV+CG are shown in Fig. 6e and Fig. 6f. The PSNRs of different approaches are given in Table 1.

By comparing Fig. 5e and 6e and f with Fig. 5d and 6d, we find the reconstruction results obtained by using the proposed L_1+LABTV and L_1+LABTV+CG are still better than those obtained by using L_1+BTV [16]. In addition, L_1+LABTV+CG performs better than L_1+LABTV because CG can further enhance the reconstructed edges. Therefore, the combination of the proposed constraints LABTV and CG is much more robust than BTV when incorrect motion and noise models are chosen for reconstruction, i.e., LABVT+CG smoothes image background while preserving sharp edges, e.g., like letters, scratches and poles on the boat hull.

Experiment 4 was conducted to examine the effectiveness of the proposed constraints when the number of LR image frames was insufficient for SR reconstruction [16]. The PSNRs of different approaches are given in Table 1. The testing image was *Shopping* with size 300×300. Nine LR images were generated in the same way as that used in experiment 1 while the down-sampling factor in this experiment was 4 and Gaussian low-pass filter was with size 3×3 standard deviation 1.2. Experimental results in Fig. 7e and f show the proposed L_1+LABTV and L_1+LABTV+CG balance well between noise suppression and detail preservation even under the circumstance of insufficient information, e.g., noise was reduced and letters were kept sharply on the box surface. Therefore, the proposed algorithms have good visual effects.

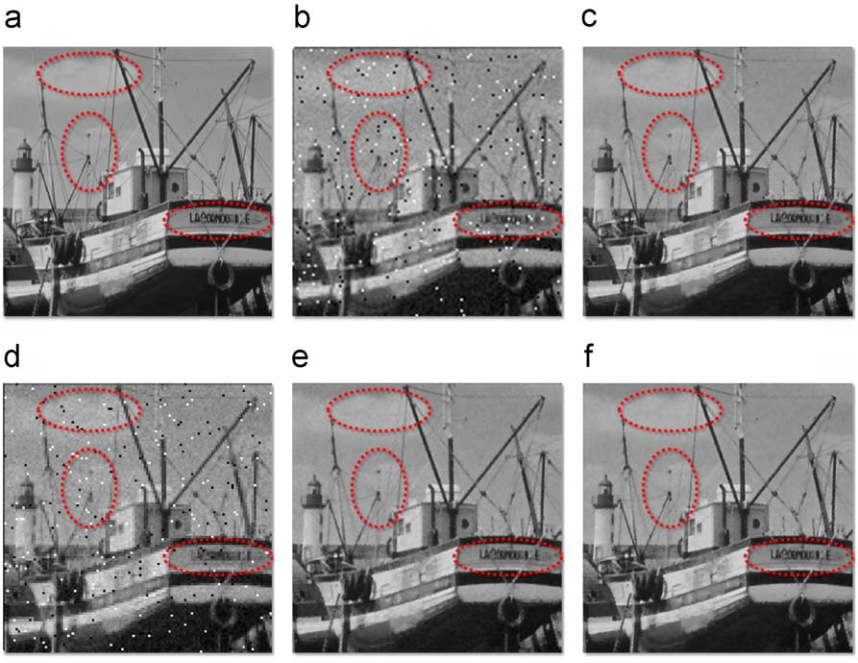

Fig. 5. Comparison of reconstruction results of image Boat, when noise model is used wrongly: (a) original image, (b) degraded image, (c) bi-cubic, (d) L_1+BTV, (e) L_1+ABTV, and (f) L_1+ABTV+CG.

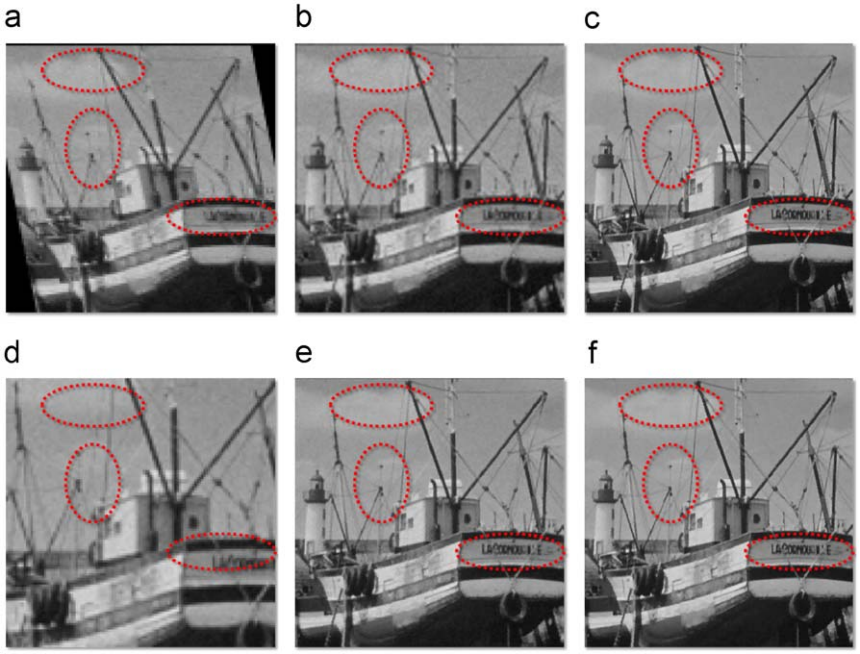

Fig. 6. Comparison of reconstruction results of image Boat, when motion model is used wrongly: (a) LR image after affine transformation, (b) LR image after magnification and clip, (c) bi-cubic, (d) L_1+BTV, (e) L_1+ABTV, and (f) L_1+ABTV+CG.

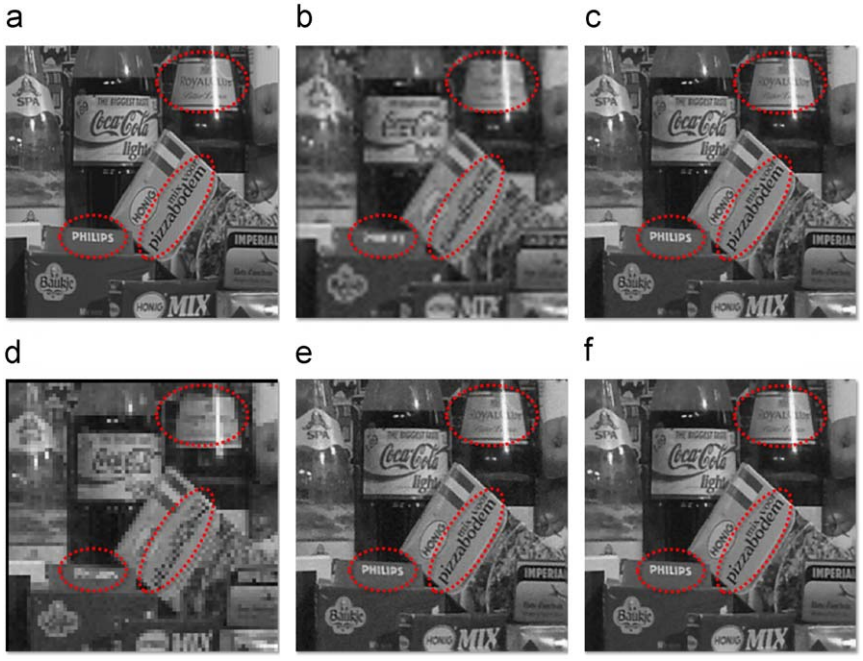

Fig. 7. Comparison of reconstruction results of image Shopping, when pixels are insufficient: (a) original image, (b) LR image, (c) bi-cubic, (d) L_1+BTV, (e) L_1+ABTV, and (f) L_1+ABTV+CG.

5. Conclusion

In this paper, a new multi-frame SR reconstruction algorithm is proposed based on image local characteristics, which can balance noise suppression against detail preservation and improve adaptively in the SR reconstruction process. In this approach, locally adaptive bilateral total variation model is used as a regularization item and measured with adaptive L_p norm. Also, L_1 norm is used to measure data error term and newly introduced gradient error term which is used as gradient consistent constraint. Experimental results show that the proposed method

leads to better results with noise suppression and detail preservation. Moreover, they are robust to fluctuations in reconstruction results due to misuse of some models.

The proposed methods can be further extended to improve the performance of multi-frame SR reconstruction methods. For example, the number of iterative steps can be reduced by adjusting step size adaptively according to the number of iterative steps. Estimation of blurring operator can be introduced into the reconstruction process for HR image estimation. And LR images can be weighted according to their differences in SR reconstruction process.

Acknowledgments

We want to thank the helpful comments and suggestions from the anonymous reviewers. This research was supported by National Science Foundation of China (60771068, 60702061, 60832005), the Open Project Program of the National Laboratory of Pattern Recognition in China, the Program for Changjiang Scholars and innovative Research Team in University of China (IRT0645), and the 100 Talents Program of The Chinese Academy of Sciences.

References

[1] R. Tsai, T. Huang, Multi-frame image restoration and registration, in: Advances in Computer Vision and Image Processing, vol. 1, no. 2, JAI Press Inc., Greenwich, CT, 1984, pp. 317–339.
[2] A. Tekalp, M. Ozkan, M. 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, vol. 3, San Francisco, CA, USA, 1992, pp. 169–172.
[3] S.H. Rhee, M.G. Kang, Discrete cosine transform based regularized high-resolution image reconstruction algorithm, Optical Engineering 38 (8) (1999) 1348–1356.
[4] S.C. Park, M.K. Park, M.G. Kang, Super-resolution image reconstruction: a technical overview, IEEE Signal Processing Magazine 20 (3) (2003) 21–36.
[5] S. Lertrattanapanich, N.K. Bose, High resolution image formation from low resolution frames using Delaunay triangulation, IEEE Transactions on Image Processing 11 (12) (2002) 1427–1441.
[6] N. Nguyen, P. Milanfar, An efficient wavelet-based algorithm for image super resolution, in: Proceedings of International Conference on Image Processing, vol. 2, Vancouver, BC, Canada, 2000, pp. 351–354.
[7] H. Stark, P. Oskoui, High resolution image recovery from image plane arrays, using convex projections, Journal of the Optical Society of America A 6 (11) (1989) 1715–1726.
[8] A.J. Patti, Y. Altunbasak, Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants, IEEE Transactions on Image Processing 10 (1) (2001) 179–186.
[9] T.F. Chan, C.K. Wong, Multichannel image deconvolution by total variation regularization, in: Proceedings of SPIE, vol. 3162, San Diego, CA, USA, 1997, pp. 358–366.
[10] R.R. Schulz, R.L. Stevenson, Extraction of high-resolution frames from video sequences, IEEE Transactions on Image Processing 5 (6) (1996) 996–1011.
[11] A.K. Katsaggelos, Digital Image Restoration, in: Springer Series in Information Science, Springer, Berlin, 1991.
[12] M. Elad, A. Feuer, Restoration of a single super-resolution image from several blurred, noisy, and under-sampled measured images, IEEE Transactions on Image Processing 6 (12) (1997) 1646–1658.
[13] M. Irani, S. Peleg, Improving resolution by image registration, CVGIP: Graphical Models and Image Processing 53 (3) (1991) 231–239.
[14] W.T. Freeman, T.R. Jones, et al., Example-based super resolution, IEEE Computer Graphics and Applications 22 (2) (2002) 56–65.
[15] E. Shechtman, Y. Caspi, M. Irani, Space–time super-resolution, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (4) (2005) 531–545.
[16] S. Farsiu, M.D. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super resolution, IEEE Transactions on Image Processing 13 (10) (2004) 1327–1344.
[17] A. Zomet, S. Peleg, Efficient super-resolution and application to mosaics, in: Proceedings of IEEE 15th International Conference on Pattern Recognition, vol. 1, Barcelona, Spain, 2000, pp. 579–583.
[18] A. Chakrabarti, A.N. Rajagopalan, R. Chellappa, Super-resolution of face images using kernel PCA-based prior, IEEE Transactions on Multimedia 9 (4) (2007) 888–892.
[19] X. Zhang, E.Y. Lam, E.X. Wu, K.K. Wong, Application of Tikhonov regularization to super-resolution reconstruction of brain MRI image, in: Lecture Notes in Computer Science, vol. 4987, 2008, pp. 51–56.
[20] L. Xiao, Z.H. Wei, A super-resolution reconstruction via local and contextual information driven partial differential equations, in: Proceedings of Fourth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 4, Haikou, China, 2007, pp. 726–730.
[21] H. Shen, L. Zhang, B. Huang, P. Li, A MAP approach for joint motion estimation segmentation and super resolution, IEEE Transactions on Image Processing 16 (2) (2007) 479–490.
[22] S. Farsiu, M.D. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super-resolution, IEEE Transactions on Image Processing 13 (10) (2004) 1327–1344.
[23] N.A. El-Yamany, P.E. Papamichalis, Robust color image super-resolution: an adaptive M-estimation framework, EURASIP Journal on Image and Video Processing 2008 (2008) 12 Article ID 763254.