Joint Learning for Single-Image Super-Resolution via a Coupled Constraint

Xinbo Gao, *Senior Member, IEEE*, Kaibing Zhang, Dacheng Tao, *Member, IEEE*, and Xuelong Li, *Senior Member, IEEE*

***Abstract*—The neighbor-embedding (NE) algorithm for single-image super-resolution (SR) reconstruction assumes that the feature spaces of low-resolution (LR) and high-resolution (HR) patches are locally isometric. However, this is not true for SR because of one-to-many mappings between LR and HR patches. To overcome or at least to reduce the problem for NE-based SR reconstruction, we apply a joint learning technique to train two projection matrices simultaneously and to map the original LR and HR feature spaces onto a unified feature subspace. Subsequently, the k-nearest neighbor selection of the input LR image patches is conducted in the unified feature subspace to estimate the reconstruction weights. To handle a large number of samples, joint learning locally exploits a coupled constraint by linking the LR–HR counterparts together with the K-nearest grouping patch pairs. In order to refine further the initial SR estimate, we impose a global reconstruction constraint on the SR outcome based on the maximum *a posteriori* framework. Preliminary experiments suggest that the proposed algorithm outperforms NE-related baselines.**

***Index Terms*—Grouping patch pairs (GPPs), joint learning, neighbor embedding (NE), super-resolution (SR).**

I. Introduction

DUE to the limitations of digital imaging systems and the imaging environment, it is not always easy to capture an image at a desired high-resolution (HR) level. However, in many practical applications such as computer vision, medical imaging, video surveillance, and entertainment, HR images are required usually to obtain a robust performance. Therefore, it becomes popular to synthesize a new HR image by using one or more low-resolution (LR) images [1].

Although simple yet fast interpolation methods such as bilinear, bicubic, and other resampling methods [2]–[4] can increase the size of LR input, they are prone to blur high-frequency details and end up with unsatisfactory performance for larger magnification ratios. Thus, a large number of super-resolution (SR) reconstruction techniques have been proposed in recent years, which can be divided into two categories: multiimage-based SR methods and example learning-based SR methods [33], [35].

There are two types of multiimage-based SR methods: the frequency-domain methods [6]–[8] and spatial-domain methods [9]–[18]. The SR capability of the frequency-domain methods is inherently limited due to the assumed strict imaging model and the preclusion of any prior knowledge from the reconstruction process. By contrast, the spatial-domain methods have received a lot of attention because certain prior knowledge can facilitate the reconstruction process to enhance the SR performance. Representative spatial-domain methods include the regularized SR reconstruction approaches [9]–[11], the projection-onto-convex-sets approach [12], iterative back projection (IBP) [13], adaptive filtering [14], nonlocal means [15]–[17], and nonlocal kernel regression [18]. However, these methods cannot function well without an adequate number of LR images. Example learning-based SR techniques do not face this problem [19]. This group of methods assumes that the high-frequency details lost in an LR image can be learned from a training set of LR and HR image pairs, i.e., the relationship between LR image patches and the corresponding HR patches can be used to estimate the missing HR frequency details in the given LR input. Fundamentally, they are nonparametric and do not depend on any presumed model. A dozen of methods have been developed for example learning-based SR. For example, Freeman *et al.* [20] utilized the Markov network [36] to model the relationship between local regions of observations and underlying scenes. This approach is computationally intensive and sensitive to training examples. Chang *et al.* [21] introduced locally linear embedding (LLE) [22] to estimate HR patches by linearly combining k-candidate HR patches in the training data set. They assumed that the counterparts of LR–HR image patch pairs (or their feature representations) are locally isometric. In contrast to [20], this neighbor-embedding (NE)-based method does not require a large number of samples and achieves top level performance. Since then, there are a lot of consequent results such as in [23]–[27], [37], and [42].

With the increase in magnification (e.g., more than $3\times$ upscaling), the SR problem becomes severely undetermined.

Manuscript received July 25, 2010; revised January 29, 2011, April 22, 2011 and June 05, 2011; accepted June 10, 2011. Date of publication July 12, 2011; date of current version January 18, 2012. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB316404; by the National Natural Science Foundation of China under Grant 61125204, Grant 61125106, Grant 60832005, Grant 41031064, and Grant 61072093; by the Ph.D. Programs Foundation of the Ministry of Education of China under Grant 20090203110002; by the Open Project Foundation of the State Key Laboratory of Industrial Control Technology of Zhejiang University under Grant ICT1105; by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2009JM8004; and by the Key Science and Technology Program of Shaanxi Province of China under Grant 2010K06-12. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bulent Sankur.

X. Gao and K. Zhang are with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: xbgao@mail.xidian.edu.cn; kbzhang0505@gmail.com).

D. Tao is with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney NSW 2007, Australia (e-mail: dacheng.tao@uts.edu.au).

X. Li is with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2161482

Thus, the correspondence between the LR image patches and their counterparts (HR image patches) in NE-based methods becomes ambiguous, i.e., the neighborhood relationship cannot be preserved perfectly due to the "one-to-many" mapping existing between one LR image and many HR images. For NE-based methods, a crucial problem is preserving the neighborhood relationship of each LR and its HR counterpart as consistently as possible. In addition, an inappropriate number of neighbors may end up with underfitting or overfitting of the result [32]. To mitigate the adverse effect of the aforementioned problems, many variant improvements to NE for SR have been proposed. For example, Chan *et al.* [23] used histogram matching to choose proper training images. However, this method does not radically attack the problem. Considering that edges and the neighborhood size are important in reconstruction, Chan *et al.* [24] put much emphasis on edge detection and feature selection by linearly combining the first-order gradient features and the weighted luminance features to depict each LR image patch. Note that the choice of the neighborhood size affects linear embedding of edge patches and of nonedge patches. Despite the face that their approach produces somewhat sharper edge details compared with that in [21], their algorithm produces artifacts in the case of incorrect edge detection. Based on [21] and [24], Zhang *et al.* [27] introduced the idea of multiple manifolds and proposed a partially supervised NE (PSNE) algorithm for image SR reconstruction. This work achieved a better overall recovery compared with [21] and [24]. Li *et al.* [26] enforced locality preserving constraints (LPCs) to avoid ambiguities through explicitly enhancing the consistency of localities on LR and HR manifolds. Compared with [21], the new method reduces blurs and artifacts. Recently, Yang *et al.* [31], [32] proposed a sparse-coding method for adaptively choosing the relevant reconstruction neighbors to represent the relationship between the training data and the input patches. Glasner *et al.* [33] exploited the redundancy of local patches in natural images and combined both classical and example learning-based SR techniques for SR reconstruction.

NE is promising, except for its limitation of a locally isometric assumption in the LR and HR feature spaces. To address this problem, the existing variations of the NE algorithm for image SR mainly concentrate on two aspects: one is to select more suitable features to characterize LR image patches such that the neighborhood relationship between LR–HR patch pairs can be preserved as consistently as possible [23]–[25], and the other is to build a better reconstruction function by imposing some consistency constraints on HR–LR pairs [25], [26].

Inspired by [34], we propose a novel NE-based method for SR reconstruction by joint learning via a coupled constraint. In contrast to existing NE-based methods, the proposed method finds the k-nearest neighbors (k-NNs) for linear embedding in a unified feature subspace spanned by LR–HR image patches rather than in the original LR feature space alone. In particular, the new SR algorithm mainly contains three parts. First, preprocessing is used to construct K-nearest grouping patch pairs (GPPs) by linking the LR and its corresponding HR features together. In the second step, joint learning is performed to learn two projection matrices of the nearest GPPs associated with each LR input patch such that the difference between LR and HR is reduced as much as possible. Thereafter, the measurements of LR and HR features in the corresponding GPPs are projected onto a unified feature subspace. Then, we use the NE algorithm to estimate the optimal weights in the learned unified feature subspace and combine linearly the corresponding high-frequency patches with the estimated weights to synthesize the HR image patches. Finally, we incorporate back projection [41] into the maximum *a posteriori* (MAP) framework, where the global construction constraint and the prior knowledge (that the final output of the HR image should be as close as possible to the initial SR estimate) are combined to enhance the initial SR image. The contributions are summarized below.

1) The coupled constraint is applied to group K-nearest LR–HR patch pairs into GPPs, which simultaneously considers the similarity between LR patches as well as that between the corresponding HR patches.
2) Joint learning on GPPs is proposed to project LR and HR feature spaces onto a unified feature subspace such that the difference between LR–HR counterparts is reduced as much as possible, which can alleviate the ambiguity between LR and HR image patches.
3) A MAP SR framework is approximated to boost the quality of the SR result, in which the global reconstruction constraint and the consistency between the initial estimate and final output are taken into consideration.

The remainder of this paper is organized as follows. Section II reviews briefly the NE-based image SR reconstruction. Section III presents the proposed method. Experimental results are presented in Section IV. Section V concludes this paper.

II. Review of the NE-Based SR Reconstruction

NE for SR reconstruction [21] uses LLE [22] to estimate the HR patches corresponding to the LR inputs by assuming that the two manifolds in the HR and the LR image patch spaces are locally similar. Let $X_s = \{x_s^i\}_{i=1}^N$ be the training data set of the LR image patches, $Y_s = \{y_s^i\}_{i=1}^N$ be the training data set of the corresponding HR image patches, $X_t = \{x_t^j\}_{j=1}^M$ be the test LR data set, and $Y_t = \{y_t^j\}_{j=1}^M$ be the estimated HR image patches. Here, N and M are the number of training image patches in the training data set and that of the test image patches in the test data set, respectively. The NE algorithm for SR reconstruction is summarized in algorithm 1.

From algorithm 1, we find that the original NE algorithm for SR reconstruction assumes that the two manifolds spanned by the feature spaces of the LR and HR image patches are locally similar. Based on this assumption, this algorithm maps the relationship between each input LR image patch and its fixed-number k-neighbors to estimate the HR image patch as a linear combination of the corresponding HR patches of neighbors.

However, according to [26] and [37], the above assumption does not hold inherently due to that "one-to-many" mappings that commonly exist between one LR image and many HR images. Therefore, the NE algorithm for SR reconstruction may not perform well. In general, there are two problems. One is to select suitable features to represent the LR and HR image patches such that the two manifolds between them can be preserved as consistently as possible. The other is to build a suitable reconstruction function by incorporating the prior knowledge into the SR process to enhance the performance. To ad-

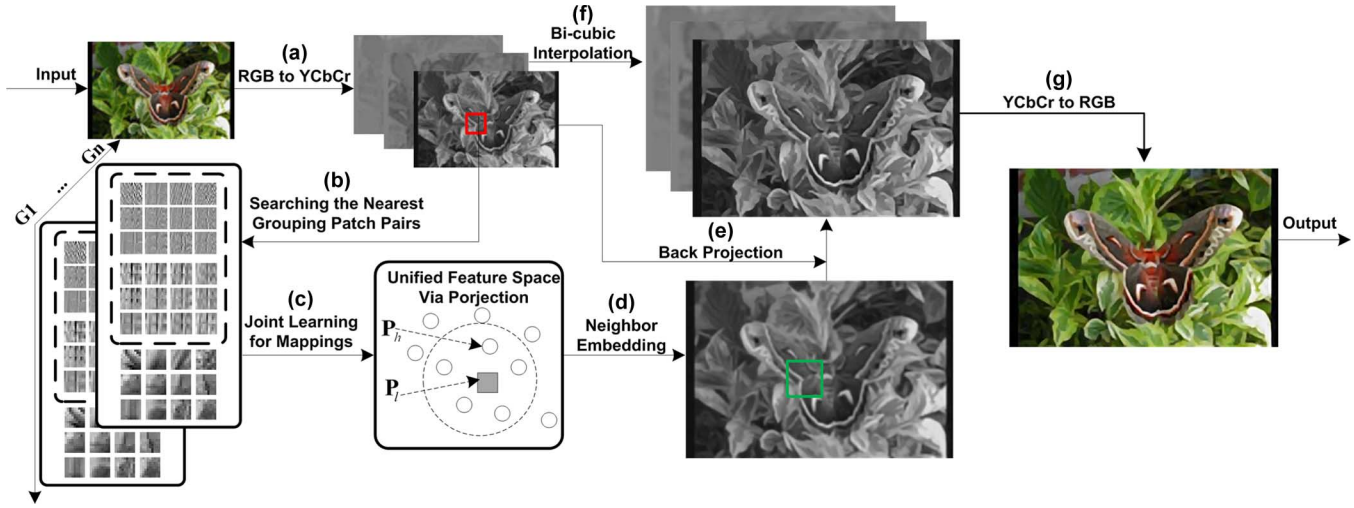

Fig. 1. Reconstruction framework of the proposed method.

dress the first problem, we apply joint learning to find a unified feature subspace spanned by the LR and HR features. For the second problem, we propose a MAP reconstruction framework for better SR recovery.

Algorithm 1 (NE for SR Reconstruction)

Input:

- Training data sets $X_s = \{x_s^i\}_{i=1}^N$ and $Y_s = \{y_s^i\}_{i=1}^N$;
- Test data set $X_t = \{x_t^j\}_{j=1}^M$;
- The number of nearest neighbor k.

Output: $Y_t = \{y_t^j\}_{j=1}^M$.

1) For each test patch x_t^j in X_t, find its k-nearest neighbor set $Ng(j)$ in the training data set $X_s = \{x_s^i\}_{i=1}^N$.

2) Compute the reconstruction weights of the neighbors that minimize the error of reconstructing x_t^j, i.e.,

$$\varepsilon^j = \min \left\| x_t^j - \sum_{i \in Ng(j)} w_{ij} x_s^i \right\|_2^2$$
$$\text{s.t.} \quad \sum_{i \in Ng(j)} w_{ij} = 1 \quad \text{and} \quad w_{ij} = 0 \quad \text{if} \quad i \notin Ng(j).$$

3)Reconstruct each HR patch y_t^j corresponding to x_t^j with the estimated weights w_{ij} as

$$y_t^j = \sum_{i \in Ng(j)} w_{ij} y_s^i.$$

III. Joint-Learning-Based SR

In this section, we first present a MAP reconstruction framework for SR. A coupled constraint on K-NNs of LR–HR counterparts, i.e., GPPs, is then established for the joint learning process. Thereafter, for each LR image patch to be super-resolved, its nearest GPPs (including K-NNs of LR–HR image patch pairs) is searched to perform joint learning for the unified feature subspace. Subsequently, the selection of k-NNs and the optimal weights for reconstruction are performed in the unified feature subspace for the initial HR estimate.

In the postprocessing phase, an IBP algorithm combining the global reconstruction constraint and the consistency prior is applied to improve the resultant image quality. Fig. 1 illustrates the reconstruction framework of the proposed algorithm. As demonstrated, the difference between the proposed method and other NE-related baselines is twofold (see steps (b) and (c) shown in Fig. 1).

1) The coupled constraint on GPPs consisting of the LR–HR image patches is applied to the learning process, which ensures a better consistency between the local manifolds of the LR and HR image patches.
2) Joint learning is performed to learn the mapping matrices $\mathbf{P}_l$ and $\mathbf{P}_h$ such that the difference between LR–HR counterparts is reduced as much as possible in the unified feature subspace. The optimal reconstruction weights for SR reconstruction are then estimated in the unified feature subspace rather than solely in the LR feature space.

A. MAP-Based Reconstruction Framework

HR images can be divided into three subbands [20]: the high-frequency band I_h, the middle-frequency band I_m, and the low-frequency band I_l. The middle-frequency components can be recovered based upon a simple interpolation algorithm (e.g., bilinear or bicubic). Thus, the goal of SR reconstruction is to estimate the missing high-frequency component with a single LR input. If prior knowledge, which is denoted as manifold MF, can be incorporated into the SR process, the problem of SR reconstruction can be written as

$$I_h^* = \arg\max_{I_h} P(I_h | I_l, \text{MF}). \tag{1}$$

By using the Bayes rule, (1) can be rewritten as

$$\begin{aligned} I_h^* &= \arg\max_{I_h} P(I_h | I_l, \text{MF}) \\ &= \arg\max_{I_h} P(\text{MF} | I_h, I_l) P(I_l | I_h) P(I_h). \end{aligned} \tag{2}$$

Taking the negative log likelihood $E = -\log(P(\cdot))$ in (2), we can get

$$I_h^* = \arg\min_{I_h} E(\text{MF}|I_h, I_l) + E(I_l|I_h) + E(I_h) \quad (3)$$

where $E(\text{MF}|I_h, I_l)$ denotes the data cost from the LR image patches and HR image patches on manifold $\text{MF} = \{M_j\}_{j=1}^M$. In particular, the first term in (3) can be formulated as

$$E(\text{MF}|I_h, I_l) = \sum_j \sum_{i \in M_j} \left\| f_l\left(x_s^i\right) - f_h\left(y_s^i\right) \right\|_2^2 \quad (4)$$

where f_l and f_h denote two mapping functions used to convert two different feature vectors x_s^i and y_s^i into a unified feature space. In this way, we can use a common distance metric to measure the similarity between them.

The term $E(I_l|I_h)$ stands for the data cost of the reconstruction error of all input LR image patches with respect to their neighborhood $Ng(\cdot)$, i.e.,

$$E(I_l|I_h) = \sum_j \left\| f_l\left(x_t^j\right) - \sum_{i \in Ng(j)} w_{ij} f_h\left(y_s^i\right) \right\|_2^2. \quad (5)$$

The term $E(I_h)$ can be seen as the smoothness prior of the HR image [26], which can be approximated by overlapping one or several pixels within the adjacent image patches. Therefore, the key problem to solve is finding the two mapping functions f_l and f_h to transform the LR and HR feature spaces into a unified feature subspace. Once achieved, we can utilize the traditional NE algorithm to estimate the optimal weights and generate the desired HR image patch as a linear combination of the corresponding HR image patches from its neighbors. In addition, the problem of finding M_j associated with each x_t^j remains unresolved. This problem is addressed by grouping the K-nearest group patch pairs of the LR and HR features in the training data set.

B. Coupled Constraint

In Section II, we have defined the training data set of the LR image patches $X_s = \{x_s^i\}_{i=1}^N$ and that of their corresponding HR image patches $Y_s = \{y_s^i\}_{i=1}^N$, where i represents the index of each image patch and N is the number of image patches collected. In the traditional NE for SR [21], x_s^i represents a d-dimensional feature vector of the ith LR image patch by concatenating the first- and second-order gradient features in horizontal and vertical directions, respectively. The y_s^i is the high-frequency details of the ith HR image patch. In order to apply joint learning to two different feature spaces of LR and HR patches, we augment another training data set $Z_s = \{z_s^i\}_{i=1}^N$ that consists of the first- and second-order gradient features of HR image patches in horizontal and vertical directions, which is similar to the representation of each LR image patch. In such way, z_s^i denotes an m-dimensional feature vector of the ith HR image patch.

As is known, most existing NE algorithms [21], [23], [27] perform SR reconstruction without considering the correlation between the LR and HR image patches. That means that they solely apply an isometric assumption to synthesize the expected HR estimates. However, this assumption does not hold perfectly for the SR problem. To solve or at least to reduce this problem, we consider the similarity between the LR image patch and the HR image patch with a coupled constraint as follows.

Let $C = \{c^i\}_{i=1}^N$ be a coupled set by concatenating each feature vector x_s^i and z_s^i. Thus, each column measurement c^i in set C is a $(d+m)$-dimensional feature vector, i.e.,

$$c^i = \begin{bmatrix} x_s^i/\sqrt{d} \\ z_s^i/\sqrt{m} \end{bmatrix}. \quad (6)$$

To adjust the dynamic changes of the concatenated feature vector, we normalize it to a unit two-norm. For each vector c^i in set C, we select the K-NNs associated with it and group them together by

$$G^i \triangleq \{c_j\}_{j \in N_K(i)} \quad (7)$$

where $G^i (1 \le i \le N)$ stands for the ith GPPs related to x_s^i and $N_K(i)$ represents the index set of K-NNs of c_i. Clearly, (7) requires a lot of memory to store GPPs because it has to replicate all feature vectors in set C redundantly. For efficient processing, we define G^i as the corresponding index set specified in $N_K(i)$. As a result of this simplification, G^i can be rewritten as

$$G^i \triangleq \{j\}_{j \in N_K(i)}. \quad (8)$$

We then can perform the joint learning to build the mapping functions f_l and f_h for each GPP.

C. Joint Learning on Example Patch Pairs

Under the MAP reconstruction framework described in Section III-A, it is necessary to find the two mapping functions f_l and f_h to project the LR and HR feature spaces onto a unified feature subspace. To this end, we apply joint learning to transform two feature spaces spanned by different dimensional feature vectors into a unified feature subspace.

Let $L^i \in \mathbb{R}^{d \times K} (i = 1, \ldots, N)$ denote the constraint patch matrix by stacking K d-dimensional column vectors whose indices are specified in G_i, i.e., $L^i = \{x_s^r\}_{r \in G^i}$. Similarly, we have $H^i \in \mathbb{R}^{m \times K} (i = 1, \ldots, N)$ and $H^i = \{z_s^r\}_{r \in G^i}$. The existing NE-based SR (e.g., [21], [23]–[27]) measure the similarity in the LR feature space. Generally, these algorithms make an assumption that the relationship between LR and HR feature spaces is locally isometric, i.e., the neighbor relationship between the LR and HR feature spaces is locally preserved. However, this assumption is inconsistent with the SR problem due to the fact that the LR patches are neighbors, whereas their corresponding HR patches may not be neighbors and *vice versa* [37]. Therefore, how to preserve the neighborhood relationship between LR patches and their corresponding HR patches is of crucial importance.

Motivated by [34], LR and HR features can share a unified feature subspace in which they are more closely associated with each other. Based on this consideration, the feature representations of LR and HR counterparts are projected onto a unified feature subspace by joint learning via a coupled constraint. Following this, the selection of k-NNs is conducted in this unified feature subspace instead of the original LR space. In contrast to [34], we perform the joint learning on the GPP of each LR input locally, rather than on all training samples, which is efficient and tractable for a training data set containing an enormous number of samples.

We cannot measure directly two different dimensional feature vectors in a common metric. To unify them into a common distance metric, an effective approach is to search for two mappings to convert them into a unified feature subspace of the same dimension. Concretely, one is for the LR feature space $f_l : \mathbb{R}^d \to \mathbb{R}^p$, and the other is for the HR space $f_h : \mathbb{R}^m \to \mathbb{R}^p$. In this way, we can project simultaneously the LR and HR feature spaces onto a unified feature subspace and measure their similarity by

$$d_{ij} = \text{Dist}\left(f_l\left(x_s^i\right), f_h\left(z_s^j\right)\right) \tag{9}$$

where function Dist($\bullet$) represents a certain distance metric such as the Euclidean distance. For (9), the problem is converted to construct the mapping functions f_l and f_h such that distance d_{ij} should be as close as possible in the unified feature subspace. If the similarity is measured by the Euclidean distance, we have

$$\arg \min_{\{f_l, f_h\}} \sum_{i \in G^i} \left\| f_l\left(x_s^i\right) - f_h\left(z_s^i\right) \right\|_2^2. \tag{10}$$

Suppose that the two mapping functions f_l and f_h are represented by projection matrices $\mathbf{P}_l \in \mathbb{R}^{d \times p}$ and $\mathbf{P}_h \in \mathbb{R}^{m \times p} (p \leq d)$. We can reformulate (10) as

$$\arg \min_{\{\mathbf{P}_l, \mathbf{P}_h\}} \sum_{i \in G^i} \left\| \mathbf{P}_l^T x_s^i - \mathbf{P}_h^T z_s^i \right\|_2^2. \tag{11}$$

Similar to [34], we expand (11) to obtain the two projection matrices as follows:

$$\begin{aligned} &\arg \min_{\{\mathbf{P}_l, \mathbf{P}_h\}} \sum_i^K \left\| \mathbf{P}_l^T x_s^i - \mathbf{P}_h^T z_s^i \right\|^2 \\ &= \arg \min_{\{\mathbf{P}_l, \mathbf{P}_h\}} tr\left(\begin{bmatrix} \mathbf{P}_l \\ \mathbf{P}_h \end{bmatrix}^T \begin{bmatrix} \mathbf{L}^i & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^i \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \right. \\ &\qquad \left. \times \begin{bmatrix} \mathbf{L}^i & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^i \end{bmatrix}^T \begin{bmatrix} \mathbf{P}_l \\ \mathbf{P}_h \end{bmatrix} \right) \end{aligned} \tag{12}$$

where $\mathbf{I}$ denotes a $K \times K$ identity matrix. Let $\mathbf{P} = \begin{bmatrix} \mathbf{P}_l \\ \mathbf{P}_h \end{bmatrix}$, $\mathbf{S} = \begin{bmatrix} \mathbf{L}^i & \mathbf{0} \\ \mathbf{0} & \mathrm{H}^i \end{bmatrix}$, and $\mathbf{A} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix}$; then, we have

$$\begin{aligned} &\arg \min_{\{\mathbf{P}\}} tr(\mathbf{P}^T \mathbf{S} \mathbf{A} \mathbf{S}^T \mathbf{P}) \\ &\quad \text{s.t.} \quad \mathbf{P}^T \mathbf{S} \mathbf{S}^T \mathbf{P} = \mathbf{I} \quad \text{and} \quad \mathbf{P}^T \mathbf{S} \mathbf{1} = 0, \end{aligned} \tag{13}$$

where $\mathbf{1} = [1, \ldots, 1]^T$ is the vector of ones with $2K$ entries. Suppose that $\mathbf{E} = \mathbf{S}\mathbf{A}\mathbf{S}^T$ and $\mathbf{F} = \mathbf{S}\mathbf{S}^T$, the optimization problem with respect to $\mathbf{P}$ can be obtained by the eigenvectors p of $\mathbf{E}p = \lambda \mathbf{F}p$ associated with the second to the pth ($p \leq d$) smallest eigenvalues. Here, matrices $\mathbf{E}$ and $\mathbf{F}$ are of size $(d + m) \times (d + m)$. For the solutions to p_l and p_h, we can expand $\mathbf{E}p = \lambda \mathbf{F}p$ to two linear equations, i.e.,

$$\mathbf{L}^i (\mathbf{L}^i)^T p_l - \mathbf{L}^i (\mathbf{H}^i)^T p_h = \lambda \mathbf{L}^i (\mathbf{L}^i)^T p_l \tag{14}$$

$$\mathbf{H}^i (\mathbf{H}^i)^T p_h - \mathbf{H}^i (\mathbf{L}^i)^T p_l = \lambda \mathbf{H}^i (\mathbf{H}^i)^T p_h. \tag{15}$$

From (15), the solution to p_h can be obtained from

$$p_h = \frac{\left(\mathbf{H}^i (\mathbf{H}^i)^T\right)^{-1} \mathbf{H}^i (\mathbf{L}^i)^T p_l}{(1 - \lambda)}. \tag{16}$$

According to (16), we can eliminate p_h in (14) and obtain

$$\mathbf{L}^i (\mathbf{H}^i)^T \left(\mathbf{H}^i (\mathbf{H}^i)^T\right)^{-1} \mathbf{H}^i (\mathbf{L}^i)^T p_l = (1 - \lambda)^2 \mathbf{L}^i (\mathbf{L}^i)^T p_l. \tag{17}$$

Intrinsically, (17) is a generalized eigendecomposition. Let $\mathbf{U} = \mathbf{L}^i (\mathbf{H}^i)^T (\mathbf{H}^i (\mathbf{H}^i)^T)^{-1} \mathbf{H}^i (\mathbf{L}^i)^T$ and $\mathbf{V} = \mathbf{L}^i (\mathbf{L}^i)^T$; then, (17) is reformulated as

$$\mathbf{U} p_l = (1 - \lambda)^2 \mathbf{V} p_l. \tag{18}$$

Once p_l is obtained, we can substitute it into (16) for p_h. By selecting the appropriate dimension p of the subspace (or the unified feature subspace), we can use the two derived projection matrices $\mathbf{P}_l$ and $\mathbf{P}_h$ to transform the original LR and HR feature spaces into a common subspace. In this way, the selection of k-NNs can be conducted within the unified feature subspace. Joint learning for the projection matrices is summarized in algorithm 2.

Algorithm 2 (Joint Learning for Projection Matrices)

Input:

- Training data set $X_s = \{x_s^i\}_{i=1}^N$ and $Z_s = \{z_s^i\}_{i=1}^N$;
- GPPs set $\{G^i\}_{i=1}^N$;
- LR patch x_t^j in the test data set X_t.

Output:

- Projection matrices $\mathbf{P}_l$ and $\mathbf{P}_h$;
- Constraint patch matrices $\mathbf{L}^i$ and $\mathbf{H}^i$.

1) Find the nearest neighbor x_s^i related to x_t^j in the training data set X_s.

2) Obtain the GPPs G^i corresponding to x_s^i.

3) According to the index set specified in G^i, build the constraint patch matrices $\mathbf{L}^i$ and $\mathbf{H}^i$, respectively.

4) Use (16) and (18) to construct the projection matrices $\mathbf{P}_l$ and $\mathbf{P}_h$.

D. Summary of the Proposed Algorithm

Usually, the initial SR estimate produced by the NE algorithm does not meet perfectly with the global reconstruction constraint. Thus, in order to enhance further the quality of the SR result, the global reconstruction constraint as well as the consistency between the initial SR estimation and the final outcome can be enforced to make an improvement on the initial result. In this paper, we utilize the iterative back-projection algorithm [41] for this purpose. Suppose that X represents the HR image and Y is the corresponding LR observation degraded by the operators of blurring B and downsampling D, i.e.,

$$Y = DBX. \tag{19}$$

By using a smoothness prior (i.e., the initial estimate and the final output should be as close as possible), we can optimize the initial estimate X_0 with the reconstruction constraint in (19) according to

$$X^* = \frac{1}{2}\arg\min_X \|Y - DBX\|_2^2 + \frac{\mu}{2}\|X_0 - X\|_2^2 \quad (20)$$

where parameter μ balances the prior and the back-projection constraint. To obtain solution X^*, we use the gradient-descent-based minimization to update the solution as follows:

$$X_{n+1} = X_n + \alpha\left[B^T D^T (Y - DBX_n) + \mu(X_0 - X_n)\right] \quad (21)$$

where X_n denotes the estimate of the HR image after the nth iteration, and α is the step size of the gradient descent. Through predetermining the maximum number of iterations or the error changes between two sequent iterations, we can obtain solution X^* that satisfies the reconstruction constraint and is close to the initial estimate X_0.

Based on algorithms 1 and 2, we summarize the proposed SR procedure in algorithm 3.

Algorithm 3 (Joint Learning for Single-Image SR via Coupled Constraint)

Input:

- Training data sets $X_s = \{x_s^i\}_{i=1}^N$, $Y_s = \{y_s^i\}_{i=1}^N$, and $Z_s = \{z_s^i\}_{i=1}^N$;
- GPPs set $\{G^i\}_{i=1}^N$; LR image Y;
- The size of LR image patch $q \times q$;
- Neighborhood size k;
- Dimension p of the unified feature subspace to be projected.

Output:

- HR image X.

1) Partition Y into $q \times q$ image patches with one or two pixels overlapped in raster-scan order to construct the test data set $X_t = \{x_t^j\}_{j=1}^M$.

2) For each patch x_t^j in X_t, execute the following steps repeatedly:

- Compute the mean values $\bar{x}$ of patch x_t^j.
- Construct the constraint patch matrices L^i and H^i related to the nearest neighbor x_s^i of x_t^j, and compute the projection matrices $\mathbf{P}_l$ and $\mathbf{P}_h$ associated with x_t^j with algorithm 2.
- Compute the transformed feature of x_t^j via $\mathbf{P}_l^T x_t^j$, and project the coupled constraint patches $\mathbf{L}^i$ and $\mathbf{H}^i$ into the unified feature subspace via $\mathbf{P}_l^T\mathbf{L}^i$ and $\mathbf{P}_h^T\mathbf{H}^i$, respectively.
- Compute the optimal weights by minimizing reconstruction error as follows:

$$\arg\min_{\{w_{ij}\}} \left\| \mathbf{P}_l^T x_t^j - \sum_{i \in Ng(j)} w_{ij}\mathbf{P}_h^T\mathbf{H}^i \right\|_2^2.$$

- Reconstruct each y_t^j corresponding to x_t^j with the optimal weights w_{ij} as follows:

$$y_t^j = \sum_{i \in Ng(j)} w_{ij} y_s^i.$$

- Sum up mean values $\bar{x}$ and y_t^j together to generate the HR patch y_t^j and append it to Y_t.

3) Produce the initial HR image X_0 by merging all the HR patches in the set $Y_t = \{y_t^j\}_{j=1}^M$. For the overlapping regions between those adjacent patches, averaging fusion is applied to obtain the estimated pixels.

4) Build the final HR image X^* with the iterative back-projection procedure as follows:

$$X^* = \arg\min \|DBX - Y\|_2^2 + \mu\|X - X_0\|_2^2.$$

E. Computational Complexity

Suppose that there are N samples in the training set and M samples in the test set. Let the feature representation of the LR and HR image patches be of d_1- and d_2-dimensional vectors, respectively. In the training stage, the original NE algorithm [21] takes the $O(N)$ time complexity to generate the training set and the $O(N(d_1 + d_2))$ space complexity to store the training samples.

In the synthesis stage of the original NE algorithm, it takes $O(d_1 M)$ to partition the LR image into patches, $O(kd_1 MN)$ to find k-NNs for embedding, and $O(d_1 k^3 M)$ to compute the optimal weights. In addition, the time complexity for synthesizing all HR image patches is $O(d_2 kM)$. For space complexity, it is $O((d_1 + d_2)(N + M))$.

The proposed algorithm requires additional computing cost in the training stage on the construction of the GPPs of each coupled patch in the training set, where the time complexity and the space complexity are $O((d_1 + m)KN^2)$ and $O(KN)$, respectively. In the synthesis stage, it first takes $O(d_1 MN)$ to find the nearest neighbor GPPs based on the sequential search algorithm. To build the two projection matrices in (16) and (18), the main computational cost is on finding the inverse of the real symmetric matrix $\mathbf{H}^i(\mathbf{H}^i)^T$ and the projection matrix $\mathbf{P}_l$ in (18). The generalized eigendecomposition problem in (18) requires the $O(d_1^3 M)$ time complexity. The calculation of the optimal weights in the unified feature subspace needs $O(pk^3 M)$. In addition to the extra space $O(KN)$ to store GPPs, the proposed algorithm shares similar calculation procedures to NE-based SR [21], including the partition of the LR image patches and the merging of the estimated HR image patches into a whole image. Table I summarizes the computational complexities of NE-based SR and that of the proposed method.

IV. Experimental Results and Discussion

To validate the effectiveness of the proposed algorithm for image SR reconstruction, 3× magnification experiments are conducted to compare with several NE-related SR approaches (including NESR [21], NeedFS [24], and PSNE [27]) and the bicubic interpolation.

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITIES BETWEEN NE-BASED SR [21] AND THE PROPOSED ALGORITHM

Complexity	Methods	Stages	
		Training	Synthesis
Time	NESR	$O(N)$	$O\begin{pmatrix} d_2M + kd_1MN \\ +d_1k^3M + d_2kM \end{pmatrix}$
	Proposed	$O\left(N + (d_1 + m)KN^2\right)$	$O\begin{pmatrix} d_2M + d_1MN \\ +d_1^3M + pk^3M \\ +d_2kM \end{pmatrix}$
Space	NESR	$O(d_1N + d_2N)$	$O(d_1(N+M) + d_2N)$
	Proposed	$O((d_1 + d_2 + K)N)$	$O\begin{pmatrix} d_1(N+M) \\ +d_2N + KN \end{pmatrix}$

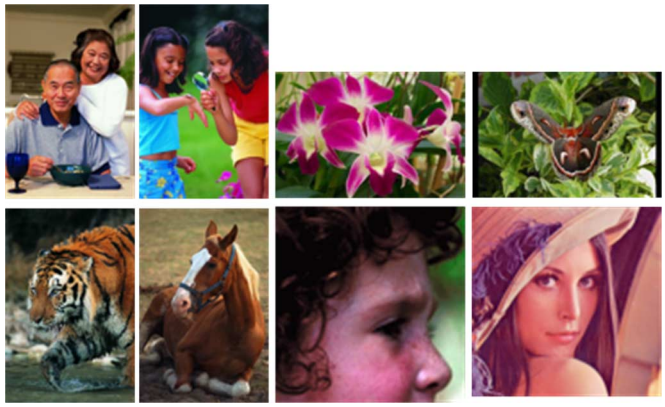

Fig. 2. Gallery of test images. We refer to them from left to right and from top to bottom as the first image to the eighth image.

A. Experimental Setting

In example learning-based image SR reconstruction, the training data set affects greatly the quality of the super-resolved HR image. In our experiments, we downloaded the software package for [32] from the author's homepage,[1] and 61 HR training images were selected to prepare our training database. In order to mimic the formation process of the LR image from its HR version, all the training HR images are downsampled by using the bicubic interpolation by a factor of 3 to obtain the corresponding LR images.

Since the human visual system (HVS) is more sensitive to the luminance component than the chrominance components, we use the YCbCr color space for color images and only perform SR reconstruction in the luminance component. Considering that the middle-frequency information of LR images has greater correlation with high frequency than low frequency, we first magnify the original LR input by a factor of 2 with the bicubic interpolation and then perform SR on it.

We refer to the interpolated images as LR images. Four convolution operations are conducted on the original HR and the interpolated version of each LR image to produce four gradient images. Next, we extract randomly four image patches of size 6 $\times$ 6 from the four gradient representations and expand them into a column vector. Thus, each LR image patch is represented by a column vector of size 144. Correspondingly, each HR image patch is a feature vector of size 324. Similar to most example learning-based SR algorithms, high-frequency components of each HR patch are generated by a high-pass filter. This can be done simply by subtracting its mean value from each pixel in the HR image patch. To avoid uninformative patches affecting the efficiency of learning, the HR image patches whose norms are close to zero are excluded from the training data set. By this manipulation, 19 906 image patch pairs are collected in the training data set.

A given LR image in the test phase is first transformed into the YCbCr color space. The chrominance components (Cb and Cr) are upscaled directly to the desired size with the bicubic interpolation, but the luminance component is upscaled using the bicubic interpolation by a factor of 2 to recover part of the middle-frequency information. Meanwhile, the four gradient images are partitioned into the image patches of size 6 $\times$ 6 in raster-scan order. For consistency in the HR image to be super-resolved, two pixels are overlapped between the adjacent patches in the interpolation image. Correspondingly, three pixels are overlapped between the adjacent HR image patches. Fig. 2 shows eight test LR images, including humans, animals, and flowers.

[1]http://www.ifp.illinois.edu/~jyang29/

TABLE II
RMSE, PSNR, AND SSIM FOR 3$\times$ SCALE FACTOR

Img.	Methods				
	bi-cubic	NESR	NeedFS	PSNE	Proposed
1	13.545	11.292	10.926	11.108	9.817
	25.495	27.076	27.362	27.239	28.291
	0.7758	0.7963	0.8036	0.7973	0.8404
2	10.338	9.166	8.922	9.027	8.206
	27.842	28.887	29.122	29.02	29.848
	0.8116	0.8222	0.8317	0.8286	0.8557
3	12.659	11.475	9.812	9.153	8.929
	28.788	26.936	28.296	28.899	29.115
	0.7996	0.7845	0.8358	0.8381	0.8630
4	16.947	15.037	13.957	14.273	13.534
	23.549	24.587	25.235	25.041	25.502
	0.6562	0.6582	0.7028	0.6965	0.7402
5	17.323	16.565	15.907	15.85	14.908
	23.359	23.747	24.099	24.13	24.662
	0.6352	0.6478	0.6670	0.6742	0.7051
6	9.272	8.088	8.244	8.139	6.998
	26.082	29.974	29.808	29.919	31.231
	0.8134	0.8165	0.8023	0.819	0.8623
7	6.024	7.057	7.863	6.588	6.457
	32.534	31.158	30.219	31.755	31.93
	0.7933	0.7876	0.7731	0.7885	0.7912
8	8.027	8.755	12.849	8.71	8.682
	30.039	29.286	25.953	29.33	29.358
	0.8494	0.8558	0.8337	0.8601	0.8677
Avg.	11.767	10.929	11.06	10.356	9.692
	27.211	27.706	27.512	28.167	28.742
	0.7668	0.7711	0.7813	0.7878	0.8157

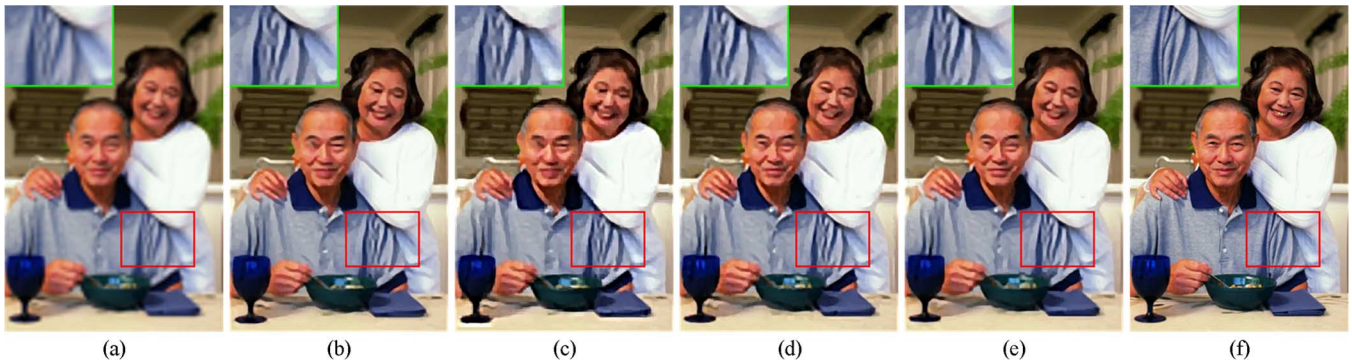

Fig. 3. Comparison with [21], [24], and [27] of the SR result on the first test image (magnified by a factor of 3). Local magnification in red rectangle is shown in the upper left corner in each example. (a) Bicubic interpolation. (b) NESR. (c) NeedFS. (d) PSNE. (e) Proposed method. (f) Original.

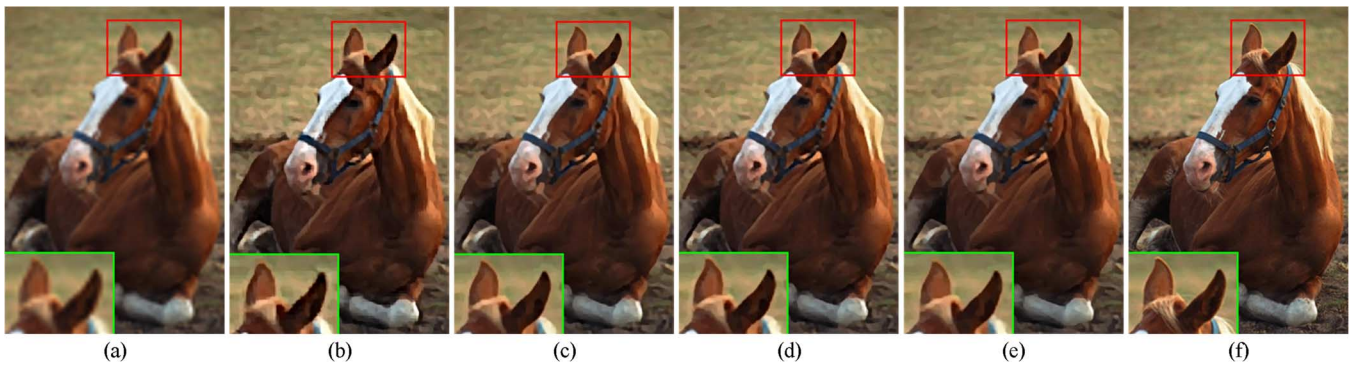

Fig. 4. Comparison with [21], [24], and [27] of the SR result on the sixth test image (magnified by a factor of 3). Local magnification in red rectangle is shown in the lower left corner in each example. (a) Bicubic interpolation. (b) NESR. (c) NeedFS. (d) PSNE. (e) Proposed method. (f) Original.

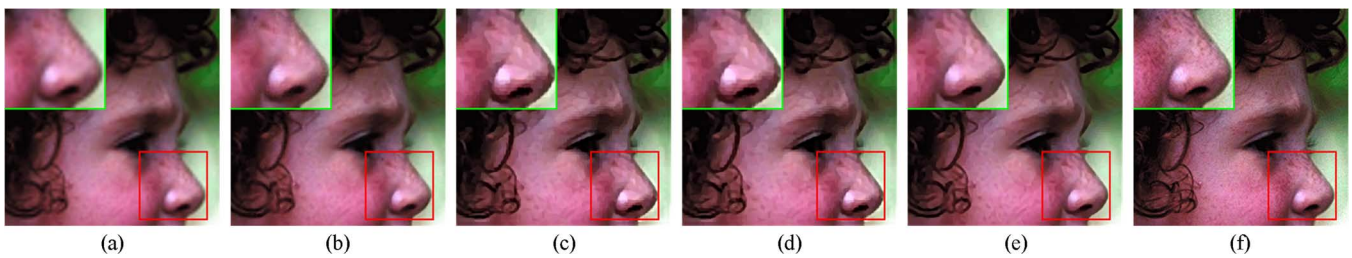

Fig. 5. Comparison with [21], [24], and [27] of the SR result on the seventh test image (magnified by a factor of 3). Local magnification in red rectangle is shown in the upper left corner in each example. (a) Bicubic interpolation. (b) NESR. (c) NeedFS. (d) PSNE. (e) Proposed method. (f) Original.

B. Experimental Results

We evaluate the performance of the proposed algorithm with other NE-related baselines [21], [24], [27].[2] The numbers of neighbors are different in these approaches. For NESR, five neighbors are suggested in [21] for embedding, but for NeedFS, the number of neighbors is determined by the type of each input LR image patch. In particular, two nearest neighbors are selected for edge patches, whereas four nearest neighbors are selected for nonedge patches. Because the class information of each LR patch is exploited in PSNE [27], slightly more neighbors are chosen for embedding. Experimentally, the neighborhood size for all the test images is set to ten. Meanwhile, the supervised parameter α for all test images is set to 0.1. For the proposed method, there are two important parameters to determine: the neighborhood size and the dimension of the unified feature subspace. In all the experiments, we retain 85 dimensions for the unified feature subspace while selecting five neighbors for linear embedding. The validation experiments are shown in Sections IV-C and IV-D. In addition, the number of K-NNs imposed on the coupled constraint is set to 128 in all the experiments.

To assess objectively the quality of the SR reconstruction, root-mean-square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) [43], [44] are used to evaluate the quality of reconstruction. Since the SR process is conducted only on the luminance component of the color image, we only compare the quantitative difference of this part between the original and the SR output. In addition to the

[2]The three NE-related algorithms use different feature representations for the LR patch. For NESR [21] and the proposed method, the first- and second-order gradient features are extracted to represent LR image patches. NeedFS [24] uses the weighted combination of the first-order gradient features and the normalized luminance features.

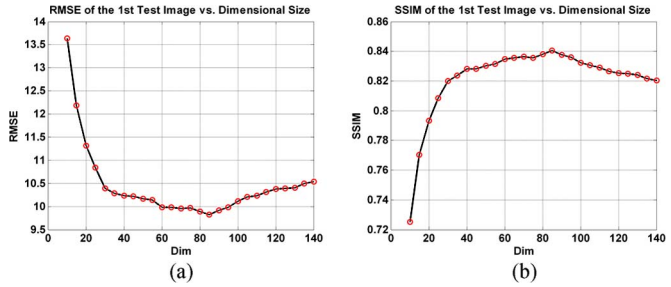

Fig. 6. Performance comparison of the first test image with different dimensions of a unified feature subspace. (a) RMSE of the first test image versus different dimensions. (b) SSIM of the first test image versus different dimensions.

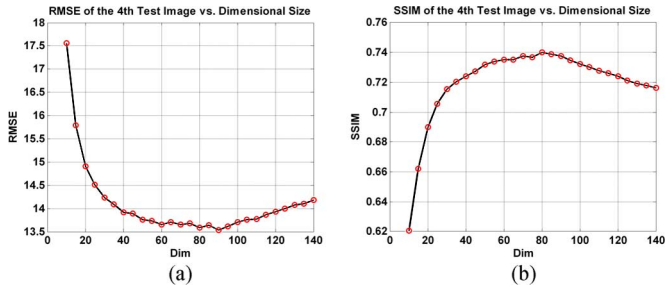

Fig. 7. Performance comparison of the fourth test image with different dimensions of a unified feature subspace. (a) RMSE of the fourth test image versus different dimensions. (b) SSIM of the fourth test image versus different dimensions.

bicubic interpolation, Table II shows the RMSE, PSNR, and SSIM indexes of eight test images.

Table II shows that the RMSE values of the proposed method on eight test images are smaller than those of other NE-related algorithms. Correspondingly, the PSNR values show that the proposed method is superior to the others. This is partly due to the coupled constraint on the LR and HR image patches as well as the fact that the joint learning is able to probe better matching for the desired HR patches. In other words, the NE for SR reconstruction in the unified feature subspace enforces a much stronger constraint, which reduces the ambiguity between the LR image patches and the HR patches. Moreover, to evaluate better the advantages of the proposed method, we also report the SSIM scores of the experimental results. In contrast to RMSE (or PSNR), SSIM is based on the HVS. The SSIM scores also suggest the effectiveness of the proposed method.

Figs. 3–5 compare different methods on the first, sixth, and seventh test images with a common magnification factor of 3. The figures show that the bicubic interpolation constantly produces the lowest visual quality with blurring and jagging artifacts along the edge regions. NESR [21] partially recovers high-frequency details and produces artifacts. NeedFS [23] provides slightly better results than NESR, particularly on the edge regions. This is because NeedFS emphasizes edge patches. However, the performance of NeedFS depends highly on the edge detection. When it fails to detect edges in blurred regions, aliasing and artifacts will be introduced. The SR results by PSNE [27] are also better than those of NESR but not as good as those produced by NeedFS, particularly on edge regions. In contrast to the above methods, the proposed method learns a unified feature subspace spanned by LR–HR patch pairs for linear embedding and achieves more reliable recovery. Therefore, the resultant SR images have less aliasing and artifacts than those produced by the others.

To substantiate the above assessment, it is observed that some visible artifacts on the eyes in the "parents" image are introduced in Fig. 3(b) and (c). In contrast, the proposed approach does not lead to the presence of these undesired artifacts. We also demonstrate the local magnification results in the red rectangle region segmented from each example. It is clear that the SR results of the proposed approach are much more competitive than the others. In other salient regions such as the brows, the hands, and the bowl, the proposed method produces much sharper details. For further comparison, the recovered results of the sixth test example magnified by a factor of 3 are shown in Fig. 4. Compared with the outcomes of the others, our approach affords much more details in the synthesized SR output. Furthermore, the local magnified regions in the red rectangle prove that the proposed method can produce sharper edges along the horse's ears than the others. In addition, Fig. 5 compares different methods on the seventh test image. As shown, the images in Fig. 5(b)–(d) appear smooth. Fortunately, in Fig. 5(e), our result shows better competitive reconstruction than the others because the freckles on the girl's face are more visible.

C. *Effects of the Dimension of the Unified Feature Subspace*

Experimentally, we find that the SR results generated by our method are highly correlated with the dimension of the unified feature subspace spanned by LR and HR image patches. To obtain the optimal dimension of a unified feature subspace, we conduct the experiments on each test image by changing the dimension from 10 to 140 at an interval of five while fixing the neighborhood size to five. RMSE and SSIM are evaluated to determine the dimension of the unified feature subspace for linear embedding. We choose the first and the fourth test images for demonstration. The curved lines of the values of RMSE and SSIM varying with the dimension of the unified feature subspace for the first and the fourth test images are plotted in Figs. 6 and 7, respectively.

As shown in Fig. 6(a), the values of RMSE decrease stably with the increase in the dimension of the unified feature space, whereas those of SSIM keep steadily increasing. Nonetheless, when the dimension reaches 85, the values of RMSE begin to increase steadily again. The inverse tendency of SSIM is presented in Fig. 6(b). A similar conclusion can be drawn for the fourth test image from the changes in Fig. 7(a) and (b). From the above experiments, good performance can be achieved at a dimension of about 85 with the neighborhood size fixed as five.

In addition, Fig. 8 demonstrates visually the reconstructed results of the first test image using different dimensions, and the corresponding values of RMSE and SSIM. Among the six resulted images, quantitatively, the least reconstruction error is yielded at the dimension of 85.

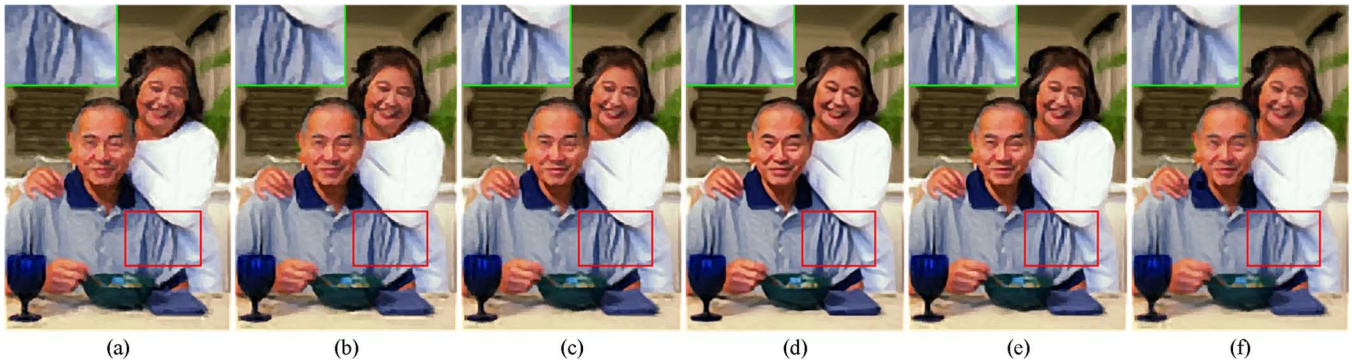

Fig. 8. Visual quality of the first test image with different dimensions of unified feature subspace (magnified $\times 3$). (a) $p = 10$, RMSE: 13.64, SSIM: 0.725. (b) $p = 20$, RMSE: 11.318, SSIM: 0.793. (c) $p = 40$, RMSE: 10.23, SSIM: 0.828. (d) $p = 85$, RMSE: 9.82, SSIM: 0.840. (e) $p = 120$, RMSE: 10.37, SSIM: 0.825. (f) $p = 140$, RMSE: 10.54, SSIM: 0.82.

D. Effects of the Neighborhood Size

Similar to other NE-related methods for SR reconstruction, the neighborhood size is important for the proposed method. For edge regions, the NE algorithm can blur image details if the neighborhood size is too large. For smooth regions, aliasing artifacts (e.g., introducing noisy results) can be introduced by very small neighborhood size. Thus, choosing a proper neighborhood size for each test patch is an open problem given that the training samples are generally nonuniformly distributed [5].

As a tradeoff, we assume that the distribution of training samples is scattered equally within the feature space. Therefore, we adopt a fixed neighborhood size for all the test images. Experimental results show that, when the neighborhood size is selected properly, we can also achieve a relatively satisfactory SR reconstruction. Similar to [21], we adjust the number of neighbors from 1 to 20 while fixing the dimension of the unified feature subspace to 85, to examine how performance changes with the neighborhood size [cf. Fig. 9(a) and (b)]. The figures show that the best performance of RMSE is achieved at a neighborhood size of six, whereas that of SSIM is achieved at a neighborhood size of five.

In addition, we conduct the same experiment on the fourth test image. Fig. 10 illustrates the changes on RMSE and SSIM with different neighborhood sizes, where the best values of RMSE and SSIM are achieved simultaneously at a neighborhood size of five when the dimension of the unified feature subspace is fixed to 85. Since SSIM score is consistent with HVS, we choose five in our experiments.

The corresponding SR results of the fourth and the eighth test images with different neighborhood sizes are shown in Figs. 11 and 12.

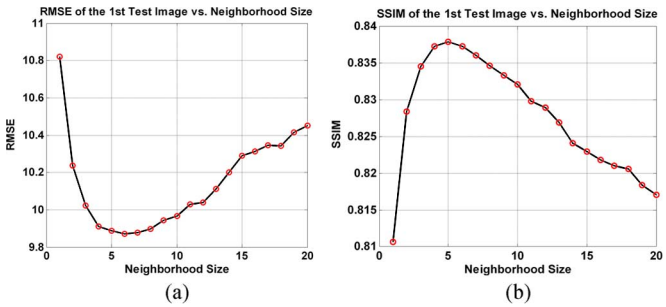

Fig. 9. Performance comparison of the first test image with different nearest neighbors used in unified feature subspace. (a) RMSE versus different number of neighbors. (b) SSIM versus different number of neighbors.

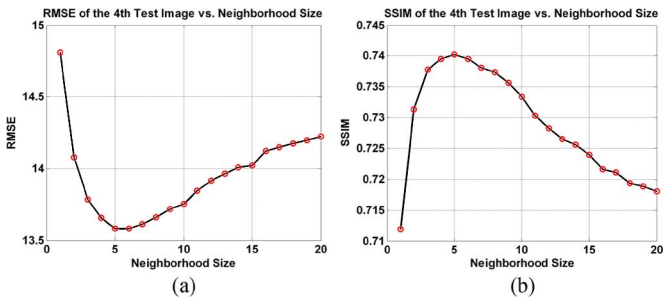

Fig. 10. Performance comparison of the fourth test image with different nearest neighbors used in unified feature subspace. (a) RMSE versus different number of neighbors. (b) SSIM versus different number of neighbors.

V. Conclusion

This paper have introduced joint learning via a couple constraint to the NE-based single-image SR reconstruction. Thorough experimental results demonstrate that the proposed method outperforms the representative baselines. The distinction of the proposed method lies in the selection of k-NNs and the linear embedding being performed in the unified feature subspace rather than in the original LR space. In our implementation, we apply the joint learning technique on GPPs that are most relevant to each LR input patch to obtain the desired unified feature subspace. In principle, the proposed method has the potential to be extended to other SR applications such as face image hallucination. In addition, the construction of optimal GPPs rather than a fixed neighborhood size is challenging for the proposed method. In the future, we will further investigate a more reliable method to construct GPPs, such as clustering and subspace learning techniques [28]–[30], [38]–[40], to improve the performance of the proposed SR method.

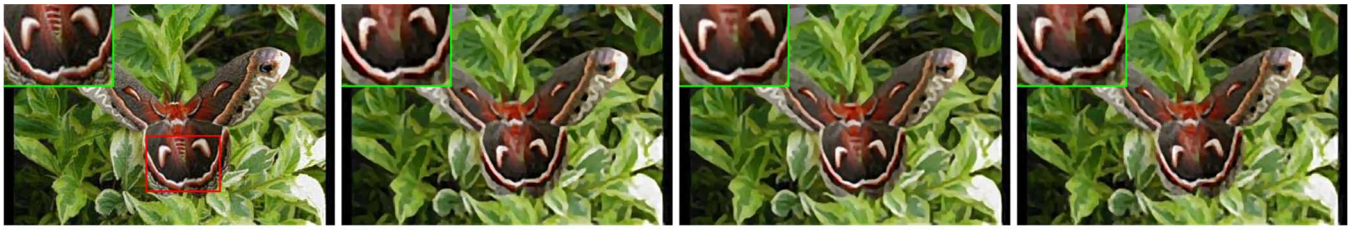

Fig. 11. Super-resolved results of the fourth test image with different neighborhood sizes (magnified $\times 3$). (a) Original image. (b) $k = 1$, RMSE: 14.57, SSIM: 0.712. (c) $k = 5$, RMSE: 13.534, SSIM: 0.740. (d) $k = 20$, RMSE: 14.23, SSIM: 0.718.

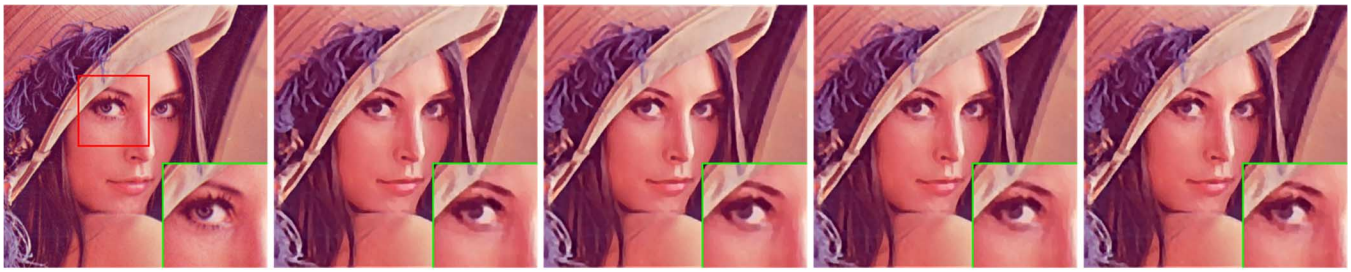

Fig. 12. Super-resolved results of the eighth test image with different neighborhood sizes (magnified $\times 3$). (a) Original image. (b) $k = 1$, RMSE: 8.982, SSIM: 0.865. (c) $k = 5$, RMSE: 8.682, SSIM: 0.868. (d) $k = 10$, RMSE: 8.759, SSIM: 0.854. (e) $k = 20$, RMSE: 8.821, SSIM: 0.841.

ACKNOWLEDGMENT

The authors would like to thank Prof. B. Sankur and the anonymous reviewers for their constructive comments and suggestions.

REFERENCES

[1] S. Park, M. Park, and M. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[2] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.

[3] N. A. Dodgson, "Quadratic interpolation for image resampling," *IEEE Trans. Image Process.*, vol. 6, no. 9, pp. 1322–1326, Sep. 1997.

[4] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.

[5] K. S. Ni and T. Q. Nguyen, "An adaptable k-nearest neighbors algorithm for MMSE image interpolation," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1976–1987, Sep. 2009.

[6] T. S. Huang and R. Y. Tsai, "Multi-frame image restoration and registration," *Adv. Comput. Vis. Image Process.*, vol. 1, no. 2, pp. 317–339, 1984.

[7] 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.*, vol. 38, no. 6, pp. 1013–1027, Jun. 1990.

[8] S. P. Kim and W. Y. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Trans. Image Process.*, vol. 2, no. 4, pp. 534–539, Oct. 1993.

[9] M. C. Hong, M. G. Kang, and A. K. Katsaggelos, "A regularized multichannel restoration approach for globally optimal high resolution video sequence," in *Proc. SPIE VCIP*, San Jose, CA, Feb. 1997, vol. 3024, pp. 1306–1317.

[10] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[11] X. Li, X. Gao, Y. Hu, D. Tao, and B. Ning, "A multi-frame image super-resolution method," *Signal Process.*, vol. 90, no. 2, pp. 405–414, Feb. 2010.

[12] H. Stark and P. Oskoui, "High-resolution image recovery from image plane arrays, using convex projections," *J. Opt. Soc. Amer. A, Opt. Image Sci., Vis.*, vol. 6, no. 11, pp. 1715–1726, Nov. 1989.

[13] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP, Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, May 1991.

[14] M. Elad and A. Feuer, "Superresolution restoration of an image sequence: Adaptive filtering approach," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 387–395, Mar. 1999.

[15] M. Protter and M. Elad, "Super resolution with probabilistic motion estimation," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1899–1904, Aug. 2009.

[16] H. Takeda, P. Milanfar, M. Protter, and M. Elad, "Super-resolution without explicit subpixel motion estimation," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1958–1975, Sep. 2009.

[17] X. Gao, Q. Wang, X. Li, D. Tao, and K. Zhang, "Zernike-moment-based image super resolution," *IEEE Trans. Image Process.*, vol. 20, no. 10, Oct. 2011.

[18] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang, "Non-local kernel regression for image and video restoration," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 566–579.

[19] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[20] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar./Apr. 2002.

[21] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jul. 2004, pp. 275–282.

[22] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[23] T.-M. Chan and J. Zhang, "An improved super-resolution with manifold learning and histogram matching," in *Proc. IAPR Int. Conf. Biometric*, 2006, pp. 756–762.

[24] T.-M. Chan, J. Zhang, J. Pu, and H. Huang, "Neighbor embedding based super-resolution algorithm through edge detection and feature selection," *Pattern Recognit. Lett.*, vol. 30, no. 5, pp. 494–502, Apr. 2009.

[25] W. Fan and D. Y. Yeung, "Image hallucination using neighbor embedding over visual primitive manifolds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2007, pp. 1–7.

[26] B. Li, H. Chang, S. Shan, and X. Chen, "Locality preserving constraints for super resolution with neighbor embedding," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 1189–1192.
[27] K. Zhang, X. Gao, X. Li, and D. Tao, "Partially supervised neighbor embedding for example-based image super-resolution," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 2, pp. 230–239, Apr. 2011.
[28] T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: A unified framework for sparse dimension reduction," *Data Mining Knowl. Discov.*, vol. 22, no. 3, pp. 340–371, May 2011.
[29] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.
[30] W. Bian and D. Tao, "Biased discriminant Euclidean embedding for content based image retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 545–554, Feb. 2010.
[31] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2008, pp. 1–8.
[32] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
[33] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 349–356.
[34] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled local preserving mappings," *IEEE Signal Process. Lett.*, vol. 17, no. 1, pp. 20–23, Jan. 2010.
[35] J. Sun, Q. Chen, S. Yan, and L.-F. Cheong, "Selective image super-resolution," Computer Vision and Pattern Recognition 2010, arXiv: 1010.5610v1.
[36] M. Song, D. Tao, C. Chen, X. Li, and C. W. Chen, "Color to gray: Visual cue preservation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1537–1552, Sep. 2010.
[37] K. Su, Q. Tian, Q. Que, N. Sebe, and J. Ma, "Neighborhood issue in single-frame image super-resolution," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2005, pp. 1122–1125.
[38] W. Bian and D. Tao, "Max-min distance analysis by using sequential SDP relaxation for dimension reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 1037–1050, May 2011.
[39] S. Si, D. Tao, and B. Geng, "Bregman divergence based regularization for transfer subspace learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 7, pp. 929–942, Jul. 2010.
[40] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[41] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," *J. Vis. Commun. Image Represent.*, vol. 4, no. 4, pp. 324–335, Dec. 1993.
[42] Y. Tang, P. Yan, Y. Yuan, and X. Li, "Single-image super resolution via local learning," *Int. J. Mach. Learn. Cybern.*, vol. 2, no. 1, pp. 15–23, Mar. 2011, DOI: 10.1007/s13042-011-0011-6.
[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[44] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1409–1423, May 2009.

Xinbo Gao (M'02–SM'07) was born in Shandong, China, in 1972. He received the B.Sc., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Postdoctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University, where he is currently a Professor of pattern recognition and intelligent system and the Director with the VIPS Laboratory. He is the author of four books and around 100 scientific papers in refereed journals and proceedings, including IEEE TIP, TCSVT, TNN, TSMC, etc. His research interests include computational intelligence, machine learning, computer vision, pattern recognition, and artificial intelligence.

Dr. Gao is a member of IEEE Xi'an Section Executive Committee and the Chair of the Membership Development Committee, a Fellow of IET, and the Vice Chair of IET Xi'an Network. He is a member of the Editorial Boards of many journals, including *EURASIP Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier). He served as the General Chair/Co-Chair or Program Committee Chair/Co-Chair or PC member for around 30 major international conferences.

Kaibing Zhang was born in Xiaogan, China, in 1975. He received the B.S.degree in computer science and technology from Wuhan University of Technology, Wuhan, China, in 1997 and the M.Eng. degree in computer software and theory from Xihua University, Chengdu, China, in 2005. He is currently working toward the Ph.D. degree with the Pattern Recognition & Intelligent Systems, Xidian University, Xi'an, China.

His research interests include pattern recognition and computer vision.

Dacheng Tao (M'07) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China; the M.Phil. degree from the Chinese University of Hong Kong, Shatin, Hong Kong; and the Ph.D. degree from the University of London, London, U.K.

He is currently a Professor of computer science with the Centre for Quantum Computation and Information Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. He mainly applies statistics and mathematics for data analysis problems in data mining, computer vision, machine learning, multimedia, and video surveillance. He is the author and coauthor of more than 100 scientific articles at top venues, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *Knowledge and Data Engineering*, *Image Processing*, NIPS, ICDM, AISTATS, AAAI, CVPR, ECCV, *ACM Transactions on Knowledge Discovery from Data*, and KDD.

Dr. Tao is the recipient of the Best Theory/Algorithm Paper Runner-Up Award in the IEEE ICDM'07.

Xuelong Li (M'02–SM'07) is a Full Professor with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.