

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221696851>

# Image Super-Resolution With Sparse Neighbor Embedding

Article in IEEE Transactions on Image Processing · March 2012

DOI: 10.1109/TIP.2012.2190080 · Source: PubMed

CITATIONS

222

READS

1,319

4 authors, including:



Xinbo Gao

Xidian University

684 PUBLICATIONS 12,192 CITATIONS

SEE PROFILE



Kaibing Zhang

Hubei Engineering University

31 PUBLICATIONS 1,464 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Image Quality Assessment [View project](#)



Heterogeneous Image Synthesis [View project](#)

# Image Super-Resolution With Sparse Neighbor Embedding

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

**Abstract**—Until now, neighbor-embedding-based (NE) algorithms for super-resolution (SR) have carried out two independent processes to synthesize high-resolution (HR) image patches. In the first process, neighbor search is performed using the Euclidean distance metric, and in the second process, the optimal weights are determined by solving a constrained least squares problem. However, the separate processes are not optimal. In this paper, we propose a sparse neighbor selection scheme for SR reconstruction. We first predetermine a larger number of neighbors as potential candidates and develop an extended Robust-SL0 algorithm to simultaneously find the neighbors and to solve the reconstruction weights. Recognizing that the  $k$ -nearest neighbor ( $k$ -NN) for reconstruction should have similar local geometric structures based on clustering, we employ a local statistical feature, namely histograms of oriented gradients (HoG) of low-resolution (LR) image patches, to perform such clustering. By conveying local structural information of HoG in the synthesis stage, the  $k$ -NN of each LR input patch is adaptively chosen from their associated subset, which significantly improves the speed of synthesizing the HR image while preserving the quality of reconstruction. Experimental results suggest that the proposed method can achieve competitive SR quality compared with other state-of-the-art baselines.

**Index Terms**—Histograms of oriented gradients (HoG), neighbor embedding (NE), sparse representation, super-resolution (SR).

## I. INTRODUCTION

**I**MAGE super-resolution (SR) reconstruction is the process of generating an image at a higher spatial resolution by using one or more low-resolution (LR) inputs from a scene. By

super-resolving an LR image, more robust performance can be achieved in many applications such as computer vision, medical imaging, video surveillance, and entertainment [1].

Over more than two decades, many methods for SR reconstruction have been proposed, which can broadly be grouped into three categories: interpolation-based methods [2]–[4], multi-image-based methods [5]–[8], and example-learning-based methods [9]–[28]. The focus of this paper is the example-learning-based method because methods of this kind exhibit stronger SR capability when a larger magnification ratio (e.g., a factor of more than double) is performed.

Example-learning-based SR approaches, also called “image hallucination”, assume that high-frequency details lost in an LR image can be predicted from a training data set. There are two main types: coding-based methods [9]–[15] and regression-based [21]–[26] methods. The coding-based methods first prepare the counterparts or cooccurrence dictionaries of LR–HR image patch pairs for learning. Next, the relation between the LR image patches in the test data set and the training data set is estimated to synthesize the desired HR image patches. The representative works include  $k$ -nearest neighbors ( $k$ -NN) learning methods [9]–[14] and sparse coding methods [15]–[20]. Freeman *et al.* [9] first proposed modeling the relation between the local regions of images and underlying scenes with a Markov network [29], and applied a belief propagation algorithm to train the Markov network for image SR reconstruction. However, their algorithm depends heavily on a large training data set and is weak in generality. Chang *et al.* [10] introduced locally linear embedding (LLE) [30] from manifold learning, assuming that two manifolds of the LR image patches and the corresponding HR patches are locally in similar geometries. Taking this assumption, the neighbor-embedding-based (NE) algorithm was proposed to estimate HR image patches by linearly combining the HR counterparts of neighbors. Yang *et al.* [15] proposed an example-learning-based SR method by employing an occurrence relation that the LR–HR image patch pairs share with the same sparse representation with respect to their own dictionaries. Recently, Dong *et al.* [19] presented a unified image restoration framework by integrating adaptive sparse domain selection and adaptive regularization, which performs well on image denoising, deblurring, and SR reconstruction. By considering both the local sparsity and the nonlocal sparsity constraints, they presented a sparse representation model for image restoration [20]. Another class of methods is the regression-based methods, which directly estimate the HR pixels via a set of learned mapping functions.

Manuscript received January 05, 2011; revised October 27, 2011 and February 16, 2012; accepted February 22, 2012. Date of publication March 09, 2012; date of current version June 13, 2012. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB316400; by the National Natural Science Foundation of China under Grant 61125106, Grant 91120302, Grant 61125204, Grant 61172146, and Grant 61072093; by the Ph.D. Programs Foundation of the Ministry of Education of China under Grant 20090203110002; by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2009JM8004; by the State Administration of STIND under Grant B1320110042; and by the Australian ARC discovery project under Grant ARC DP-120103730. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiaolin Wu.

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 & Intelligent Systems and the Faculty of Engineering & 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, 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.2012.2190080

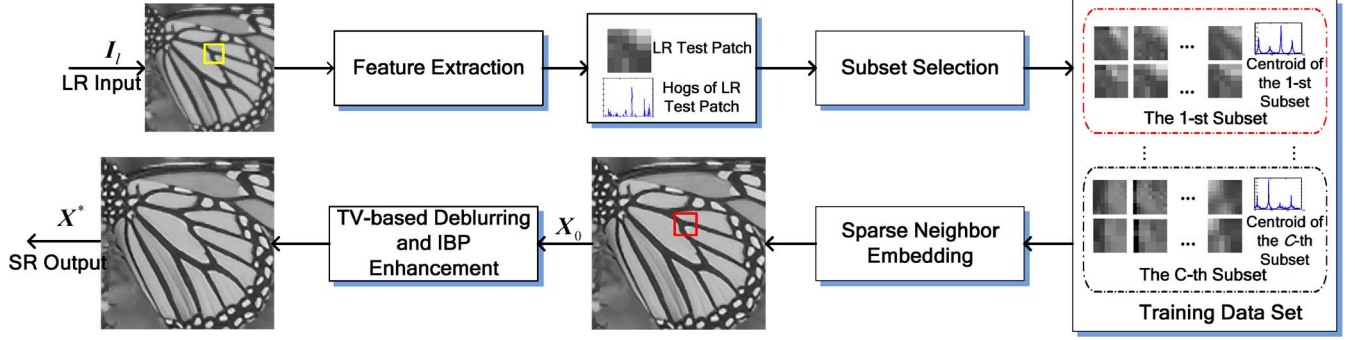


Fig. 1. Reconstruction framework of the proposed method.

In this paper, we focus on the NE-based methods [10]–[14]. In contrast to [9], the NE-based SR algorithms can represent more patterns even if a relatively smaller training data set is available and thus show a much stronger generalization ability for a variety of images. Nonetheless, the limitation of the NE-based method is twofold: 1) The LR–HR feature mappings cannot be established well because features in a high-dimensional space cannot be represented perfectly in a low-dimensional space, leading to some ambiguity between LR–HR patch pairs; and 2) the strictly fixed neighborhood size usually ends up with blurring effects, due to over- or under-fitting [15].

To date, many variant improvements over the NE-based SR method have been proposed [11]–[14]. For example, Chan *et al.* [11] utilized histogram matching to choose more reasonable training images and achieved better results than with the NE-based SR method. Fan *et al.* [12] introduced a primitive prior and established a refined training data set to hallucinate generic images. In addition, residual errors related to the reconstructed image patches were estimated to compensate for the information lost in the local averaging process. Considering the importance of both edges and neighborhood size for reconstruction, Chan *et al.* [13] elaborated on edge detection and feature selection to mitigate the adverse effect of the representation of LR patches. Compared with that in [10], this approach generates relatively sharper edges. Taking a multimanifold assumption, Zhang *et al.* [14] proposed a partially supervised NE SR method by employing an unsupervised Gaussian mixture model to predict class label information for neighbor search.

Note that the aforementioned NE-based SR methods unexceptionally use the Euclidean distance metric to search a fixed number of neighbors for linear embedding. Due to blurring, downsampling, and noisy data, this simple tactic for neighbor selection cannot perform well for image SR problems because the neighbor search and minimum reconstruction error are separated into two independent processes. Moreover, neighbor search is often manipulated within the whole training data set, leading to a computationally intensive process. To reduce these problems and to further improve the NE-based algorithm for SR reconstruction, we propose a sparse-NE-based (SpNE) method that takes the following factors into account.

- 1) How can the required  $k$ -candidates be searched to minimize reconstruction error?

- 2) How can the computational cost suffered by the NE-based algorithms be reduced effectively?

To target the first problem, we develop a sparse neighbor selection (SpNS) algorithm to simultaneously search neighbors and estimate weights by incorporating the Robust-SL0 algorithm [33], [34] and the  $k/K$ -NN criterion [35]. For the second problem, we suggest that the  $k$ -NN chosen for reconstruction should have similar local geometric structures. To this end, we partition the whole training data set into a set of subsets by clustering histograms of oriented gradients (HoG) [36], i.e., a powerful descriptor of local geometry representation. Through clustering, a group of medium-scale subsets can be constructed, which can effectively reduce computational time while preserving SR quality.

The reconstruction framework of the proposed method is illustrated in Fig. 1. As shown, the reconstruction process takes place in the following stages: 1) Both the HoG and the first- and second-order gradient features are extracted in a raster-scan order from the upscaled version of the LR input by using the bicubic (BI) interpolation with a factor of 2; 2) subset selection that the HoG feature of each LR input matches the centroids of clusters is performed to find a medium-scale subset close to the LR input for synthesis process; 3) SpNE is applied to synthesize the HR image patch of the LR input, in which searching neighbors and estimating weights are simultaneously conducted; and 4) after constructing all the HR patches and obtaining the initial HR image, the total-variation-based (TV) deblurring [42] and the iterative back-projection (IBP) algorithm [37] are sequentially performed to obtain the final HR outcome.

The remainder of this paper is organized as follows. The previous studies that are associated with this paper are briefly reviewed in Section II. The proposed SpNE algorithm is presented in Section III. The experimental results are demonstrated in Section IV. The conclusion is presented in Section V.

## II. RELATED WORKS

In this section, we will briefly review the robust-SL0 algorithm for sparse representation [33] and the  $k/K$ -NN selection [35], which are both important to our work.

### A. Robust-SL0 for Sparse Representation

In recent years, deriving relevant sparse solutions of under-determined inverse problems has become one of the most spot-

lighted research topics in the signal processing community and has successfully been applied in a wide variety of signal processing tasks, e.g., image recovery [15], [44], pattern recognition [31], and compressed sensing [32]. The objective of signal sparse representation is to represent a signal as a sparse linear combination with respect to an overcomplete dictionary. Suppose that the matrix  $\mathbf{X} \in \mathbb{R}^{d \times N}$  is an overcomplete dictionary, in which each column vector is a  $d$ -dimensional atom. The number of columns in the matrix  $\mathbf{X}$  is far greater than the number of rows, i.e.,  $N > d$ , which ensures that the dictionary is overcomplete. Given a signal  $\mathbf{x} \in \mathbb{R}^d$ , its sparse representation can be seen as finding a sparse vector  $\mathbf{w} = [w_1, \dots, w_N]^T \in \mathbb{R}^N$  by solving the following optimization problem:

$$\min_{\mathbf{w}} \|\mathbf{w}\|_0 \quad \text{s.t.} \quad \|\mathbf{x} - \mathbf{X}\mathbf{w}\|_2^2 \leq \varepsilon^2 \quad (1)$$

where  $\|\cdot\|_0$  denotes the  $\ell_0$ -norm, which counts the number of nonzero elements in a vector. Theoretically, the solution of (1) cannot directly be achieved due to the combinatorial search required. Moreover, since any small amount of noise completely changes the  $\ell_0$ -norm of the solution, this method is prone to errors in noisy settings. Therefore, alternative approaches have been proposed to pursue sparse solutions. Under a mild constraint, solving

$$\min_{\mathbf{w}} \|\mathbf{w}\|_1$$

subject to  $\|\mathbf{x} - \mathbf{X}\mathbf{w}\|_2^2 \leq \varepsilon^2$  can approximate the solution of (1) by LASSO  $\ell_1$ -norm regularization [38], the feature sign search algorithm [39], BP algorithm, [40] or FOCUSS [41]. These algorithms are typically associated with prohibitive computational complexity. Considering that  $\|\mathbf{w}\|_0$  is a discontinuous function of  $\mathbf{w}$ , Mohimani *et al.* [33] proposed to solve the minimum solution of  $\ell_0$ -norm by maximizing a family of continuous functions as

$$\lim_{\delta \rightarrow 0} \arg \max_{\mathbf{w}} F_{\delta}(\mathbf{w}) = \lim_{\delta \rightarrow 0} \arg \max_{\mathbf{w}} \sum_{i=1}^N f_{\delta}(w_i) \quad \text{s.t.} \quad \|\mathbf{x} - \mathbf{X}\mathbf{w}\|_2^2 \leq \varepsilon^2 \quad (2)$$

where  $f_{\delta}(w_i)$  can use the Gaussian family of functions to represent, i.e.,  $f_{\delta}(w_i) = e^{-w_i^2/\delta^2}$  satisfying

$$\lim_{\delta \rightarrow 0} f_{\delta}(w_i) = \begin{cases} 1, & \text{if } w_i = 0 \\ 0, & \text{if } w_i \neq 0. \end{cases} \quad (3)$$

Consequently, the minimum of  $\|\mathbf{w}\|_0$  can approximate the maximum of  $F_{\delta}(\mathbf{w})$  by iteratively decaying  $\delta$  with gradient descent. The convergence evidence can be referred to [34] for details.

### B. $k/K$ -NN Selection

In pattern recognition and statistical analysis, the  $k$ -NN algorithm is popular because it is among the simplest of all machine learning algorithms. With the  $k$ -NN criterion, an observation can be best represented with a weighted average of its neighbors. Taking this assumption, the LLE problem is modeled as an approximately equivalent optimization problem [35], i.e.,

$$\arg \min_{\mathbf{w}_i} \lim_{\lambda \rightarrow \infty} \|\mathbf{x}_i - \mathbf{X}\mathbf{w}_i\|_2^2 + \lambda \sum_{j \neq i} u(w_{ij}) \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \quad \text{s.t.} \quad \|\mathbf{w}_i\|_0 = k, \mathbf{w}_i^T \mathbf{1} = 1, w_{ii} = 0 \quad (4)$$

where  $\mathbf{x}_i$  represents the  $i$ th sample,  $\mathbf{X} \in \mathbb{R}^{d \times N}$  is a data matrix whose columns are formed by  $N$   $d$ -dimensional feature vectors,  $\mathbf{w}_i \in \mathbb{R}^N$  denotes the linear weight vector for representing  $\mathbf{x}_i$ , and  $\text{supp}(\mathbf{w}_i)$  is the supporting set of  $\mathbf{w}_i$  whose elements are nonzero. In addition,  $u(\cdot)$  stands for a step function (or staircase function) and  $\mathbf{1} = [1 \dots 1]^T \in \mathbb{R}^N$ . As stated in [35], the optimization problem in (4) cannot identify outliers and excludes them. To address this problem, Eftekhari *et al.* [35] suggested that minimizing the objective function in (4) should contribute to the choice of neighbors. Moreover, the minimum number of samples is required to best represent each observation. If an error upper limit on  $\sum_{j \neq i} u(w_{ij}) \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$  is considered, the supporting set  $\text{supp}(\mathbf{w}_i)$  of  $\mathbf{w}_i$  can be extended to a relatively larger number of neighbors, i.e.,  $N_K(i)$  ( $K > k$ ), representing the  $K$ -NNs of  $\mathbf{x}_i$ . On the basis of the given consideration, an improved  $k/K$ -NN criterion is reformulated by

$$\min_{\mathbf{w}_i} \|\mathbf{w}_i\|_0 \quad \text{s.t.} \quad \|\mathbf{x}_i - \mathbf{X}\mathbf{w}_i\|_2^2 \leq \varepsilon^2, \text{supp}(\mathbf{w}_i) \subset N_K(i), \mathbf{w}_i^T \mathbf{1} = 1. \quad (5)$$

With (5), the modified  $k/K$ -NN scheme has two advantages over (4). First, the choice of  $k$ -NN is extended to a relatively larger number of neighbors (i.e., with a size of  $K > k$ ) rather than an absolutely fixed number  $k$ , which theoretically is able to probe some potential samples beyond the scope of  $k$ -NN. Second, minimizing the reconstruction error can contribute to the selection of neighbors; therefore, the required  $k$ -NN are adaptively selected to better represent the data point  $\mathbf{x}_i$ .

## III. PROPOSED METHOD

To target the aforementioned problems in the NE-based SR algorithm, we first exploit the HoG feature to characterize the local geometric structure of LR image patches and divide the whole training data set into a set of medium-scale subsets to accelerate the SR process. To overcome the limitation of previous NE methods that use the Euclidean distance metric to search neighbors and to improve the quality of reconstruction, we develop a SpNS scheme by using a variation of sparse representation with  $\ell_0$ -norm [33].

### A. Representation of LR Image Patch With HoG

In order to represent a variety of image patterns, example-learning-based SR methods often collect a large number of samples for learning. However, a very large training data set would lead to a computationally intensive load. To mitigate the problem, an alternative approach is to search the  $k$ -NN of a given sample within a subset close to the input. To achieve the objective, we can use low-level but efficient features to characterize the local structure of image patches and perform clustering on them. In this regard, we use HoG [36], a rather good geometric descriptor that uses the distribution of local intensity gradients or edge directions. We choose HoG rather than other low-level features, such as pixel intensities, gradient information or a combination of both, because pixel intensities exhibit their variance to intensity difference between image patches, whereas gradient features are sensitive to noise. By contrast, the HoG feature does not have either problem.

To extract the HoG, a gradient detection operator is first conducted on the input image in the horizontal and vertical directions. Once achieved, the gradients of each pixel can be represented by a vector  $\vec{p}_i = \{dx_i, dy_i\}$ , where  $dx_i$  and  $dy_i$  denote the horizontal and vertical derivatives of the  $i$ th pixel point, respectively. For the gradient direction that falls into the range of  $-\pi/2 - \pi/2$  in the radian form, we can transform it into  $0^\circ - 180^\circ$  via  $((\arctan(dy_i/dx_i) + \pi/2) * 180)/\pi$ .

Next, the discrete directions should be determined. We specify the orientation bins evenly spaced over  $0^\circ - 180^\circ$  at intervals of  $5^\circ$  and round the continuous angle of each pixel to a discrete value, i.e., downward or upward. We then take a weighted vote on the discrete direction and accumulate these votes into a set of orientation bins within the  $3 \times 3$  local spatial regions called “cells”. Finally, the number of pixels falling into the same bin is calculated for edge orientation histograms. By linking the edge orientation histograms of each cell in an LR image patch of size  $6 \times 6$  and by normalizing the patch to the unit  $\ell_2$ -norm, a 144-dimensional HoG feature is constructed.

### B. Clustering on HoG

In the previous subsection, it is suggested that HoG can characterize the local geometry structure of LR image patches well. With HoG, we can segment the image patch pairs in the training data set into a set of subsets. Each subset shares a similar geometric structure. Specifically, all training samples comprise a union of such clusters as follows:

$$\begin{cases} \mathbf{X}_s = \bigcup_{c=1}^C \mathbf{X}_s^{(c)} = \bigcup_{c=1}^C \{\mathbf{x}_s^i | i \in \Omega_c\} \\ \mathbf{Y}_s = \bigcup_{c=1}^C \mathbf{Y}_s^{(c)} = \bigcup_{c=1}^C \{\mathbf{y}_s^i | i \in \Omega_c\} \end{cases} \quad (6)$$

where  $\mathbf{X}_s$  and  $\mathbf{Y}_s$  denote the matrices of the training data set  $\{\mathbf{x}_s^i\}_{i=1}^N$  and  $\{\mathbf{y}_s^i\}_{i=1}^N$  by stacking the features of LR and HR image patches in column form and  $C$  represents the number of clusters. Correspondingly,  $\mathbf{X}_s^{(c)}$  and  $\mathbf{Y}_s^{(c)}$  denote the data matrices consisting of the features of LR and HR image patches in the  $c$ th cluster, respectively. In addition,  $\Omega_c$  stands for the index set of  $\mathbf{X}_s^{(c)}$  or  $\mathbf{Y}_s^{(c)}$ . To accomplish clustering, we make use of a version of the standard K-means clustering algorithm [43], which is relatively fast and can successfully partition the training samples into satisfying subsets. When clustering, the simple  $\ell_2$ -norm is used for the distance metric. With the use of K-means clustering, we can divide the training data set into the prefix number (e.g.,  $C$ ) of clusters by minimizing the inter-cluster variance such that

$$J = \sum_{c=1}^C \sum_{i \in \Omega_c} \left\| \mathbf{g}_i - \bar{\mathbf{g}}^{(c)} \right\|_2^2 \quad (7)$$

where  $\bar{\mathbf{g}}^{(c)}$  is the mean vector, i.e., the centroid of the  $c$ th cluster. In this paper, we collect 99 918 training samples and segment them into 40 clusters by 30 iterations. In Fig. 2, three typical clustering results, which are learned from the training data set, in which the top row shows the distributions of the centroids in the second, 20th, and 40th clusters and in which the bottom

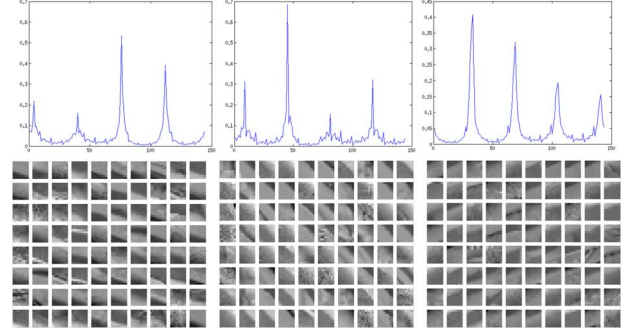


Fig. 2. Three paradigms of 40 clusters based upon HoG. The number of clusters from left to right: the second, the 20th and the 40th clusters. Top row: Curved lines of distribution of centroids. Bottom row: Top 80 high-frequency patches (with  $9 \times 9$  size) corresponding to each cluster.

row shows the top 80 high-frequency image patches in the corresponding subset are shown. It is clearly shown that the high-frequency image patches having similar local structures are well grouped into the same cluster via clustering on HoG features.

### C. Sparse Neighborhood Selection

In Section II-B, we give a brief review of the Robust-SL0 algorithm, which can be used for general sparse representation over an overcomplete dictionary. Note that the  $k/K$  neighborhood selection differs from the original NE-based algorithm in that there exists an extra constraint term, i.e.,  $\text{supp}(\mathbf{w}_i) \subset N_K(i)$  in (5). To solve this, we can divide the whole index set of  $\mathbf{w}_i$  into two index sets  $S$  and  $\bar{S}$ , i.e.,

$$\begin{cases} S = \{j \in N_K(i)\} \\ \bar{S} = \{j \notin N_K(i)\} \end{cases} \quad (8)$$

Similar to [35], we set  $w_{ij} = 0$  when  $j \notin N_K(i)$  and add an extra term that admits small functional values for  $w_{ij}$  when  $j \notin N_K(i)$ . By modifying the Robust-SL0 algorithm, we solve the problem in (5) by

$$\min \|\mathbf{w}_i\|_0 \text{ s.t. } \left\| \mathbf{x}_i - \mathbf{X} (\delta_S(\mathbf{w}_i) + \delta_{\bar{S}}(\mathbf{w}_i)) \right\|_2^2 \leq \varepsilon^2 \quad (9)$$

where  $\delta_S : \mathbb{R}^N \rightarrow \mathbb{R}^N$  is a characteristic function that selects the weights associated with subset  $S$ . To be more specific,  $\delta_S(\mathbf{w}_i) \in \mathbb{R}^N$  produces a new vector in which the indexes corresponding to  $S$  remain unchanged and sets rest to zero. Correspondingly,  $\delta_{\bar{S}}(\mathbf{w}_i) \in \mathbb{R}^N$  generates a new vector in which the indexes corresponding to  $\bar{S}$  remain unchanged and sets rest to zero. Similar to (2), we can divide the optimization problem in (9) into two parts as

$$\begin{aligned} \lim_{\delta \rightarrow 0} \max_{\mathbf{w}_i} F_\delta(\mathbf{w}_i) &= \lim_{\delta \rightarrow 0} \max_{\mathbf{w}_i} \sum_{j \in S} (1-b) e^{-w_{ij}^2/2\delta^2} \\ &\quad + \sum_{j \in \bar{S}} (1+b) e^{-w_{ij}^2/2\delta^2} \\ \text{s.t. } \left\| \mathbf{x}_i - \mathbf{X} (\delta_S(\mathbf{w}_i) + \delta_{\bar{S}}(\mathbf{w}_i)) \right\|_2^2 &\leq \varepsilon^2 \end{aligned} \quad (10)$$

where  $b$  is a constant taking  $0 \leq b < 1$ . In contrast to (2), the additional term, namely, the second term in (10), provides a smoother approximation of the minimum of  $\|\mathbf{w}_i\|_0$ . We can use a gradient descent algorithm to obtain the solution vector

$\mathbf{w}_i^*$  by decaying  $\delta$  with a factor of  $\xi$  at each iteration. Let  $t$  ( $t = 0, \dots, T$ ) be the iterative times, then the minimum  $\ell_2$ -norm solution  $\mathbf{w}_i^{(0)}$  can be obtained from the pseudoinverse  $\mathbf{X}^+$  of  $\mathbf{X}$ , i.e.,

$$\mathbf{w}_i^{(0)} = \mathbf{X}^+ \mathbf{x}_i. \quad (11)$$

With the constraint condition  $\text{supp}(\mathbf{w}_i) \subset N_K(i)$ , we reset the  $j$ th element in the solution  $\mathbf{w}_i^{(t)}$  ( $t = 0, \dots, T$ ) to zero if  $j \notin N_K(i)$  and keep the rest unchanged. In terms of (10), the derivation of  $F_\delta(\mathbf{w}_i^{(t)})$  with respect to  $\mathbf{w}_i^{(t)}$  is formulated as

$$\begin{aligned} \nabla F_\sigma(\mathbf{w}_i^{(t)}) = & - \sum_{j \in S} (1-b) w_{ij} e^{-w_{ij}^2/2\sigma^2} \\ & - \sum_{j \in \bar{S}} (1+b) w_{ij} e^{-w_{ij}^2/2\sigma^2}. \end{aligned} \quad (12)$$

Subsequently,  $\mathbf{w}_i^{(t+1)}$  is updated via

$$\mathbf{w}_i^{(t+1)} = \mathbf{w}_i^{(t)} + \eta \nabla F_\delta(\mathbf{w}_i^{(t)}) \quad (13)$$

where  $\eta$  is a positive constant that specifies the iteration step. Sequentially,  $\mathbf{w}_i^{(t+1)}$  is projected onto the feasible set by

$$\mathbf{w}_i^{(t+1)} = \mathbf{w}_i^{(t+1)} - \mathbf{X}^+ (\mathbf{X} \mathbf{w}_i^{(t+1)} - \mathbf{x}_i). \quad (14)$$

After  $T$  times iterations have finished, we choose the indexes of the top  $k$  nonzero elements of  $\mathbf{w}_i^{(T)}$  as the desired neighbors  $N_k(i)$  for linear embedding.

#### D. Summary of Proposed Algorithm

For example-learning-based SR approaches, patchwise synthesis is adopted to estimate all HR patches, and averaging fusion is applied to sequentially merge these estimated HR patches into a whole HR image. However, the local averaging process would result in blurred details and unwanted artifacts. To obtain a local optimal solution of SR, we employ the TV-based regularization for image deblurring [43] and use the IBP algorithm [37] to further enhance the deblurring result by imposing the global reconstruction constraint that the HR image should meet the LR input via the degradation process. Based upon Sections III-A–III-C, the proposed SpNE algorithm for image SR reconstruction is summarized in algorithm 1.

---

#### Algorithm 1: Image SR with Sparse NE.

---

##### Input:

- The LR image  $I_l$  and the size of LR patch  $s \times s$ ;
- The neighborhood size  $k$  and an integer  $K$ ;
- The HoG centroid set  $\mathbf{G} = \{\bar{\mathbf{g}}^{(c)}\}_{c=1}^C$  obtained by K-means clustering;
- The training data set  $\mathbf{X}_s = \cup_{c=1}^C \mathbf{X}_s^{(c)}$  and  $\mathbf{Y}_s = \cup_{c=1}^C \mathbf{Y}_s^{(c)}$ .

##### Output:

- HR image  $X^*$ .

**(1)Feature extraction:** Partition  $I_l$  into a sequence of  $s \times s$  image patches with two pixels overlapped in raster-scan

order and generate the test data set  $\mathbf{X}_t = \{\mathbf{x}_t^j\}_{j=1}^M$  and the corresponding HoG feature set  $\mathbf{G}_t = \{\mathbf{g}_t^j\}_{j=1}^M$ ;

**(2)Iteration:** For each  $\mathbf{x}_t^j$  in  $\mathbf{X}_t$ , perform the following steps iteratively:

- Compute the mean values  $\bar{\mathbf{x}}$  of the  $j$ th test image patch.
- Choose the closest subsets  $\mathbf{X}_s^{(c)}$  and  $\mathbf{Y}_s^{(c)}$  according to  $c = \arg \min_{c=1, \dots, C} \{\|\mathbf{g}_t^j - \bar{\mathbf{g}}^{(c)}\|_2^2\}$ .
- Find the  $K$ -NN of  $\mathbf{x}_t^j$  by using the Euclidean distance metric within  $\mathbf{X}_s^{(c)}$  and construct its neighbor set  $N_K(j)$ .
- Given a test sample  $\mathbf{x}_t^j$ , calculate  $\mathbf{w}_j$  according to (11)–(14) with  $T$  times iterations.
- Select the top  $k$  nonzero elements of  $\mathbf{w}_j$  as the neighbor set  $N_k(j)$ .
- Normalize  $\mathbf{w}_j$  whose indexes belong to  $N_k(j)$  according to  $w_{pj} = w_{pj} / \sum_{p \in N_k(j)} w_{pj}$ .
- Synthesize  $\mathbf{y}_t^j$  corresponding to  $\mathbf{x}_t^j$  according to  $\mathbf{y}_t^j = \sum_{p \in N_k(j)} w_{pj} \mathbf{y}_s^p$ .
- Sum up the mean values  $\bar{\mathbf{x}}$  and  $\mathbf{y}_t^j$  to generate the HR image patch and attach it to  $\mathbf{Y}_t$ .

**(3)Fusion:** Produce the initial HR image  $X_0$  by sequentially merging all HR patches in  $\mathbf{Y}_t = \{\mathbf{y}_t^j\}_{j=1}^M$ . For the overlapped regions between the adjacent patches, averaging fusion is used to obtain the final pixel value.

**(4)Deblurring:** Perform the TV-based deblurring on the initial estimation  $X_0$  and obtain the deblurred HR image  $X_0$ .

**(5)IBP enhancement:** Reconstruct the final HR image  $X^*$  by carrying out the IBP procedure on the deblurred HR image  $X_0$ .

---

#### E. Computational Cost Evaluation

Suppose that there are  $N$  samples in the training data set and  $M$  samples in the test data set. Let the feature representation of LR image patches be a  $d$ -dimensional vector. Compared with the traditional NE-based algorithm for SR reconstruction, algorithm 1 incurs additional costs  $O(M)$  in extracting the HoG feature in Step (1) and  $O(MCd)$  for selecting the matched subset in Step 2(b). In Step 2(c), the proposed algorithm takes about  $O(KdMN/C)$  for the neighbor search if all training samples are equally divided into  $C$  clusters. The traditional NE-based algorithm takes  $O(kdMN)$  for  $k$ -NN search and  $O(dk^3M)$  for obtaining the reconstruction weights within the whole training data set. In the new algorithm, the neighbor search and reconstruction weights calculation are simultaneously conducted in Step 2(d), which is affected by the iteration number  $T$  (usually less than 10), the initial deviation value  $\delta_{\max}$ , the minimal deviation value  $\delta_{\min}$ , and the decaying factor  $\xi$ . It takes  $O(MNT\delta_{\max}\xi/(\delta_{\min}C))$  for test samples. Except for the given steps, the new algorithm shares similar calculation procedures with the traditional NE-based algorithm. Thus, the new algorithm is more efficient than the NE-based algorithm.

#### IV. EXPERIMENTAL RESULTS

In this section, we perform  $3 \times$  magnification experiments on eight test images from [19] to validate the effectiveness of the proposed SR method.

TABLE I  
PARAMETER CONFIGURATIONS

Methods	Parameters
NE	$k = 12$
NeedFS	$k = 5, 15$
Proposed	$k = 15, K = k + 40, T = 20, \xi = 0.8,$ $\delta_{\min} = 0.0001, \eta = 4, b = 0.9$

### A. Experimental Configuration

In our experiments, we download the software for [15] from the author’s homepage, and 61 HR images are used to prepare our training data set. To mimic the real imaging system, all the training images are blurred by a  $9 \times 9$  Gaussian filter with standard deviation 1.1 and downsampled by a decimation factor of 3 to produce the corresponding LR training images. Since the human visual system is more sensitive to the luminance channel than the chrominance channels, we transform RGB values into YCbCr color space and only carry out the SR process on this part. We directly magnify the chrominance (Cb and Cr) channels to the desired size with the BI interpolation.

For a fair comparison, we use the same feature representation of the LR and HR image patches as [10], [13], and [15]. One-hundred-thousand image patches are randomly extracted from the training image pairs to build the training data set. To avoid uninformative image patches affecting the learning efficiency, we exclude the patches whose norms are close to zero from the training data set, which results in 99 918 training pairs being available for learning. To segment the whole training data set into subsets, we perform K-means clustering [43] on the HoG feature. As a tradeoff between the processing speed and reconstruction quality, 40 clusters are constructed in all the experiments. In the synthesis phase, all the LR test images are degraded in the same manner as in the training phase. For the methods in [10] and [13], only the neighborhood size  $k$  is determined. In addition to this parameter, the proposed method requires another major parameter  $K$  to probe the potential neighbors beyond the neighborhood size  $k$ . We experimentally tune this with  $K = k + 40$ . The initial deviation value  $\delta_{\max}$  is set to the double maximum absolute values of the pseudoinverse solution from (11). The remaining parameters are listed in Table I.

### B. Experimental Results

We compare the proposed method with five representative algorithms: NE-based SR [10], NeedFS [13], SC-based SR [15], TV-based SR [18], and ASDS [19]. Quantitatively, we employ peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [46], [47], and feature similarity (FSIM) [45] to assess the performance of different algorithms. Note that we introduce a blurring process in our experiments, which is not employed in [10], [13], and [15]. For fair comparison, we use TV-based regularization for image deblurring [42], and 30 times iterations are carried out on the initial estimate to obtain the deblurring results because the TV-based deblurring can remove unwanted artifacts while preserving important high-frequency details such as edges. Finally, we perform IBP algorithm [37] to enhance the deblurring result so that the final HR estimate is perfectly consistent with the LR input via the degradation process.

TABLE II  
PSNR (IN DECIBELS), SSIM, AND FSIM RESULTS OF RECONSTRUCTED IMAGES BY DIFFERENT METHODS. FOR EACH IMAGE, WE HAVE THREE ROWS. THE FIRST ROW IS PSNR (IN DECIBELS), THE SECOND ROW IS SSIM, AND THE THIRD ROW IS FSIM

Images	BI	[10]	[13]	[15]	[18]	[19]	Proposed
Butterfly	22.09	23.28	21.55	24.13	25.63	27.09	27.08
	0.773	0.83	0.811	0.866	0.884	0.898	0.903
	0.784	0.831	0.823	0.844	0.846	0.872	0.884
Parrots	26.48	27.38	25.89	27.41	27.73	29.97	29.48
	0.846	0.872	0.846	0.887	0.885	0.909	0.906
	0.905	0.906	0.884	0.918	0.921	0.941	0.944
Bike	21.82	22.70	21.84	23.18	23.13	24.48	24.34
	0.631	0.728	0.715	0.754	0.762	0.794	0.784
	0.774	0.802	0.797	0.829	0.825	0.851	0.849
Flower	26.22	26.47	25.73	27.11	27.37	29.19	28.76
	0.73	0.791	0.781	0.812	0.825	0.848	0.838
	0.822	0.827	0.824	0.867	0.848	0.890	0.881
Girl	31.67	31.55	30.08	31.72	31.57	33.54	33.07
	0.767	0.788	0.775	0.801	0.807	0.824	0.817
	0.871	0.878	0.869	0.895	0.882	0.918	0.906
Hat	28.27	29.05	27.95	29.20	29.14	30.92	30.97
	0.809	0.85	0.834	0.854	0.856	0.871	0.873
	0.839	0.863	0.85	0.881	0.88	0.900	0.895
Leaves	21.85	22.31	20.94	23.26	23.81	26.72	25.99
	0.734	0.832	0.802	0.855	0.877	0.905	0.892
	0.775	0.825	0.836	0.843	0.858	0.877	0.883
Plants	29.72	30.51	28.64	31.41	31.34	33.47	32.69
	0.829	0.865	0.851	0.883	0.885	0.909	0.902
	0.878	0.89	0.874	0.901	0.91	0.928	0.919
Average	26.01	26.66	25.33	27.18	27.46	29.42	29.05
	0.765	0.820	0.802	0.839	0.848	0.870	0.864
	0.831	0.853	0.845	0.872	0.871	0.897	0.895

Different algorithms in terms of PSNR, SSIM, and FSIM are compared in Table II. The table indicates that the proposed method outperforms BI interpolation, NE-based SR [10], NeedFS [13], and SC-based SR [15]. Furthermore, the quantitative evaluations of the proposed method are superior to those of the TV-based method [18]. Although some of the quantitative evaluations of the proposed method are slightly inferior to those of ASDS [19], the average quantitative evaluations of the proposed method are comparable to those of ASDS. In particular, the main merit of the proposed method is that the model is simple and easy to implement. By contrast, the model of the ASDS method is much more complex because the sparse prior regularization [15], the nonlocal similarity regularization [49], and the local smooth regularization [50] are simultaneously incorporated for the SR problem. In terms of the perceptual quality assessment immediately following, we see that the proposed method can achieve a competitive SR result.

To further illustrate the effectiveness of the proposed method, The  $3\times$  magnification results of the “Butterfly” and “Leaves” images for visual quality comparison are shown in Figs. 3 and 4, respectively. As shown in Fig. 3, the BI interpolation consistently produces the least pleasing results with blurred edges. Although the NE-based SR can reproduce some high-frequency details, it also produces ringing effects along the edge regions due to the participation of noisy neighbors during reconstruction. NeedFS provides sharper edges because edge regions are emphasized by this method, but the resultant image is not photorealistic. The SC-based SR produces oversmooth results; thus,



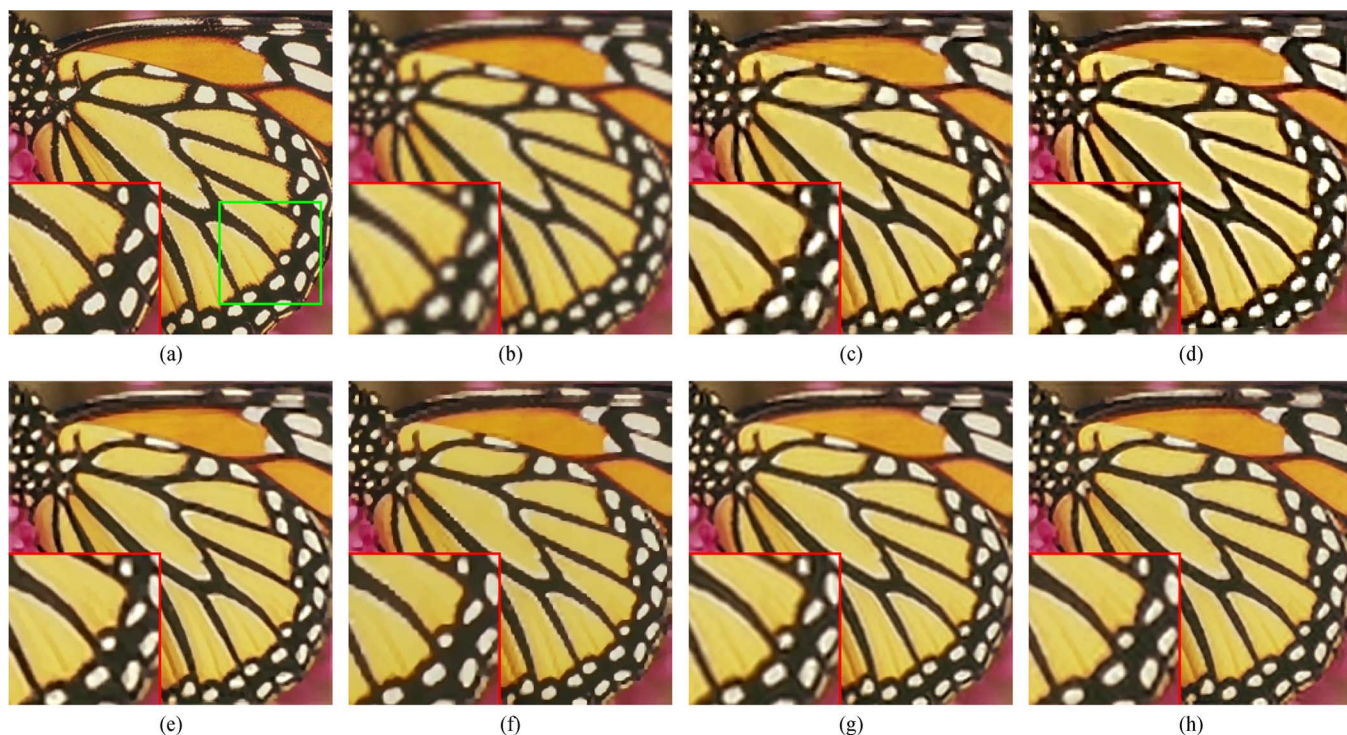


Fig. 3. Comparison ( $\times 3$ ) of SR results on “Butterfly” image by different methods. (a) Original. (b) BI. (c) NE-based SR [10]. (d) NeedFS SR [13]. (e) SC-based SR [15]. (f) TV-based SR [18]. (g) ASDS SR [19]. (h) Proposed.

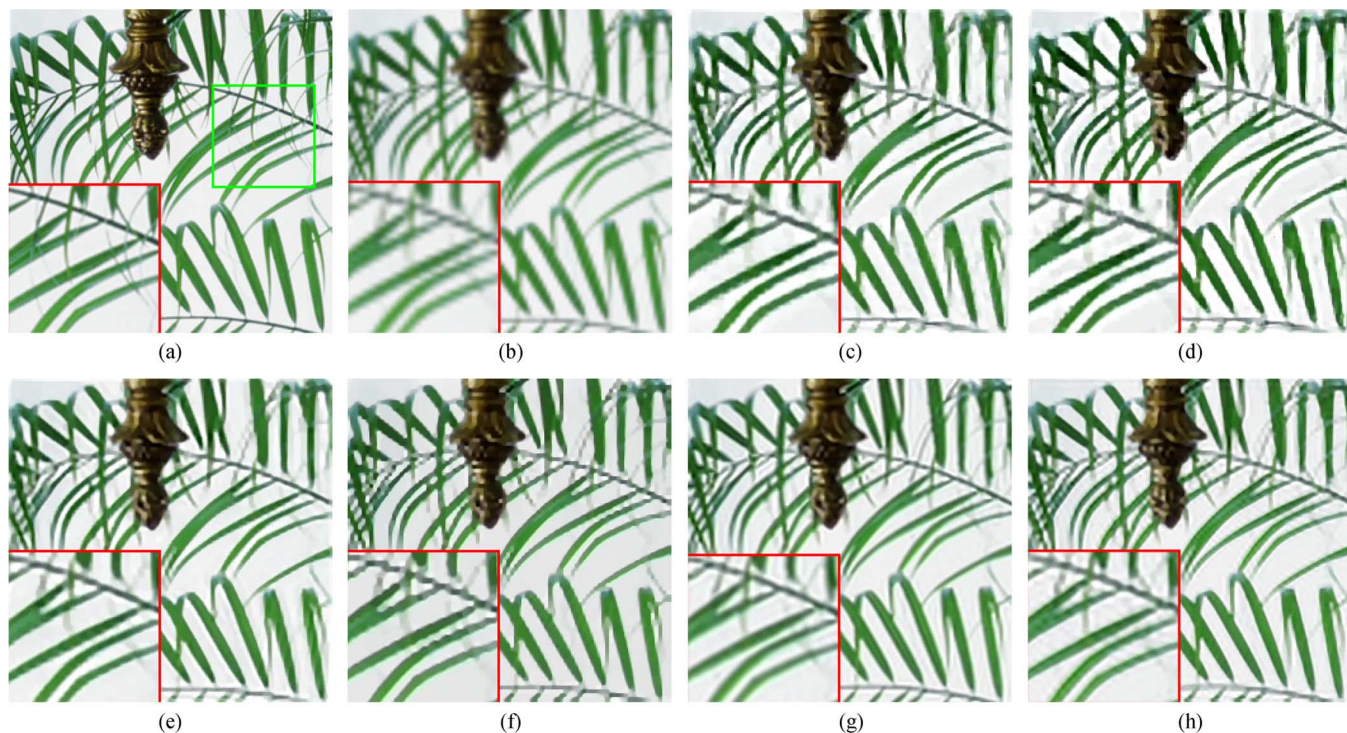


Fig. 4. Comparison ( $\times 3$ ) of SR results on “Leaves” image by different methods. (a) Original. (b) BI. (c) NE-based SR [10]. (d) NeedFS SR [13]. (e) SC-based SR [15]. (f) TV-based SR [18]. (g) ASDS SR [19]. (h) Proposed.

fine details can be recovered well. The TV-based method is capable of suppressing the ringing effects, but it generates obvious jaggy artifacts along edges and removes rich details in textural regions, leading to an oversmoothed result. The ASDS method can recover many fine high-frequency details due to its combined regularization priors. However, the reconstructed image

tends toward blurred edges in terms of subjective visual evaluation. The proposed method produces top-level visual quality, which is most faithful to the ground truth. In Fig. 3(h), we can see that our result demonstrates comparable quality in terms of the sharpness of edges and fineness of textures. As shown in Fig. 4, we obtain similar observations to those obtained in Fig. 3.



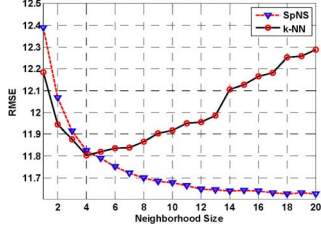


Fig. 5. RMSE comparison of results on “Butterfly” image with SpNS and  $k$ -NN methods by varying different neighborhood sizes from 1 to 20.

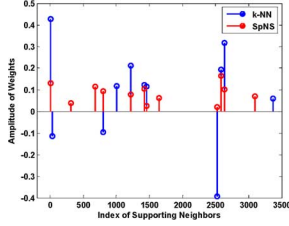


Fig. 6. Weight coefficients versus supporting neighbors with SpNS and  $k$ -NN methods for the 2000th raw image patch in “Parrots” image.

### C. Effectiveness of SpNS

Apart from the comparisons in the previous subsection, we further validate the effectiveness of the SpNS algorithm for reconstruction quality. All the experiments are performed based upon the SubS method. First, we compare the performance of different numbers of neighbors by ranging the neighborhood size  $k$  from 1 to 20 for the test images. Correspondingly, we perform the SpNS method by using a larger neighborhood size  $K$  to yield different reconstructions in the same experimental context. We take the test image “Butterfly” as an example. The curved lines of the RMSE varying with the neighborhood size are shown in Fig. 5. As we can see, the RMSE attained by the  $k$ -NN method is first decreasing and then increasing with the changes of neighborhood size. In contrast, the SpNS method shows a steadily downward trend within the range. This is in part due to the SpNS method being able to probe the required samples between  $k$ -NN and  $K$ -NN for better reconstruction. It also substantiates the view that the simple  $k$ -NN strategy is insufficient for the SR problem. To further illustrate the different level of efficiency between the  $k$ -NN and SpNS methods, we conduct another experiment on the 2000th raw image patch in the “Parrots” image to reflect the influence on the supporting neighbors from two methods. The contrasted results are shown in Fig. 6, and we observe that different supporting neighbors with different weighting coefficients exist between the two methods. By contrast to the results from the  $k$ -NN method, the amplitudes of the weights computed by the SpNS are more uniform and positive, which means the neighbors found by the SpNS have a positive correlation with the test sample. However, the results from  $k$ -NN show a remarkable difference, in which obvious negative correlation samples are found when linearly embedding. Finally, we provide the RMSE values by synthesizing all the raw HR image patches extracted from eight test images. There are 6724 raw patches segmented from each test image of size  $255 \times 255$ . We synthesize the corresponding HR image patches with the  $k$ -NN and SpNS methods, and the values of

TABLE III  
RMSES OF RECONSTRUCTED RAW IMAGE PATCHES BY  
 $k$ -NN AND SPNS METHODS

Images	$k$ -NN	SpNS	Images	$k$ -NN	SpNS
Butterfly	23.38	15.13	Girl	9.08	6.51
Parrots	11.25	7.54	Hat	10.28	6.78
Bike	25.26	17.03	Leaves	27.03	17.18
Flower	16.38	10.31	Plants	10.46	6.41

RMSE between the synthesized and the raw values are then calculated to assess their reconstruction capability. The averaged RMSE values are presented in Table III. As reported, the results achieved by the SpNS steadily outperform those by  $k$ -NN. It follows that the neighbors found by the SpNS are more reliable than the  $k$ -NN, leading to less reconstruction error.

### D. Experiments on Noisy Image

In this subsection, we conduct the noisy experiments on the cropped “Lena” image of size  $300 \times 300$  with two different levels of Gaussian noise to justify the belief that SpNS has stronger robustness against noise. The different neighbor selection criteria are based upon the SubS method for a fair comparison. The standard deviations are set to 2 and 4.

To analyze the effect of neighborhood size, we model the NE-based SR reconstruction as a MAP framework, similar to [15], i.e.,

$$\mathbf{w}_j^* = \arg \max P(\mathbf{w}_j) \cdot P(\mathbf{x}_t^j | \mathbf{w}_j, \mathbf{X}_s^{(c)}) \quad (15)$$

in which

$$P(\mathbf{w}_j) = \frac{1}{2a} \exp\left(-\frac{\|\mathbf{w}_j\|_0}{a}\right)$$

$$P(\mathbf{x}_t^j | \mathbf{w}_j) = \frac{1}{2\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \left\| \mathbf{X}_s^{(c)} \mathbf{w}_j - \mathbf{x}_t^j \right\|_2^2\right). \quad (16)$$

Here,  $a$  is the variance of the Laplacian prior on  $\mathbf{w}_j$ , whereas  $\sigma^2$  is the variance of the noise assumed on  $(\mathbf{x}_t^j \in \Omega_c)$ . Taking the negative log-likelihood in (15), we can get

$$\arg \min_{\mathbf{w}} \left\| \mathbf{X}_s^{(c)} \mathbf{w}_j - \mathbf{x}_t^j \right\|_2^2 + \frac{\sigma^2}{a} \|\mathbf{w}_j\|_0. \quad (17)$$

From (17), we can see that if parameter  $a$  is fixed, the larger the noise of the data, the smaller the value of  $\|\mathbf{w}_j\|_0$  should be. In other word, the lesser number of neighborhood size  $k$  should be chosen for linear embedding. Similarly, the proposed SpNE method should also comply with this law. That is to say, we should tune a relatively smaller  $K$  to suppress noise. To test the performance on noisy data, we compare the results of the proposed method against NE-based [10], SC-based [15], TV-based [18], and ASDS [19] methods. For the NE-based method, the smaller neighborhood size achieves better results as the noise becomes heavier. For the SC-based method, the robustness to noise is closely related to the regularization parameter  $\lambda$ . Concretely, the larger the value of  $\lambda$  we set, the stronger its capability to suppress noise the model can achieve. Therefore, we set  $\lambda$  to 0.2 and 0.4 for the noisy images at two different levels. For the proposed SpNS method, we use the same number of neighbors as the NE-based method with  $K = k + 40$ . The visual quality evaluation is shown in Fig. 7. As shown, the NE-based method is good at preserving edges but fails to distinguish the signal from



Fig. 7. Comparison ( $\times 3$ ) of SR results on the noisy “Lena” image of size  $300 \times 300$ . Noise level (standard deviation of Gaussian noise) from top to bottom: 2 and 4. (a) Result of the NE-based method ( $k = 12$ ). (b) Result of the SC-based method ( $\lambda = 0.2$ ). (c) Result of the TV-based method. (d) Result of the ASDS method. (e) Result of the SpNE method ( $k = 12$ ). (f) Result of the NE-based method ( $k = 8$ ). (g) Result of the SC-based method ( $\lambda = 0.4$ ). (h) Result of the TV-based method. (i) Result of the ASDS method. (j) Result of the SpNE method ( $k = 8$ ).

TABLE IV  
RMSEs OF RECONSTRUCTION OF THE LENA IMAGE CONTAMINATED BY DIFFERENT NOISE LEVELS

Noise Levels	NE-based	SC-based	TV-based	ASDS	SpNS
2	8.79	8.51	8.49	6.57	6.98
4	8.87	8.76	9.04	6.99	7.49

noise, leading to noisy results. The SC-based method is prone to smoothing details because it purchases the denoised result at the expense of smoothing away fine details. The TV-based method generates the noisiest outputs. Due to the combined priors used, the ASDS method shows an outstanding capability to recover high frequency, as well as denoising, leading to a less noisy output. Nonetheless, visually, the resultant images trend toward smooth results. By contrast, the proposed SpNS algorithm is capable of suppressing noise and preserving more details; thus, the resultant images look more natural than others. Additionally, we report the RMSE values of the reconstructed images from different levels of noisy data in Table IV. As presented, the results of the proposed method are superior to those of NE-based, SC-based, and TV-based methods in all situations and slightly inferior to those of ASDS. This indicates that the proposed method has stronger robustness against noise.

#### E. Effects of Patch Size

For example-learning-based SR approaches, an important issue is the size of the image patch. If the image patch size is too big, some fine details cannot be properly recovered, and the SR result will be smooth. Alternatively, unwanted artifacts will be introduced if too small an image patch size is used, e.g., noises appear in the smooth regions and jagged effects appear along edges in the SR result. In order to validate the effects of different patch sizes for SR, we conduct a set of

TABLE V  
PSNR (IN DECIBELS), SSIM, AND FSIM RESULTS OF RECONSTRUCTED IMAGES BY DIFFERENT PATCH SIZES. FOR EACH IMAGE, WE HAVE THREE ROWS. THE FIRST ROW IS PSNR (IN DECIBELS), THE SECOND ROW IS SSIM, AND THE THIRD ROW IS FSIM

Images	$5 \times 5$	$7 \times 7$	$9 \times 9$	$11 \times 11$	$13 \times 13$
Butterfly	26.66	26.77	27.08	26.36	26.25
	0.896	0.897	0.903	0.883	0.881
	0.881	0.878	0.884	0.859	0.854
Parrots	29.27	29.25	29.48	29.17	29.12
	0.899	0.900	0.906	0.895	0.896
	0.929	0.929	0.944	0.926	0.926
Bike	24.13	24.20	24.34	24.08	24.06
	0.777	0.779	0.784	0.771	0.770
	0.845	0.847	0.849	0.843	0.843
Flower	28.54	28.60	28.76	28.44	28.42
	0.832	0.833	0.838	0.826	0.826
	0.877	0.880	0.881	0.876	0.876
Girl	32.90	32.95	33.07	32.82	32.87
	0.813	0.814	0.817	0.812	0.813
	0.902	0.903	0.906	0.904	0.904
Hat	30.78	30.86	30.97	30.60	30.56
	0.870	0.871	0.873	0.865	0.864
	0.894	0.895	0.895	0.890	0.891
Leaves	25.72	25.77	25.99	25.41	25.33
	0.887	0.887	0.892	0.873	0.871
	0.882	0.878	0.883	0.860	0.857
Plants	32.48	32.56	32.69	32.33	32.24
	0.899	0.901	0.902	0.896	0.896
	0.918	0.918	0.919	0.914	0.912
Average	28.81	28.87	29.05	28.65	28.61
	0.859	0.860	0.864	0.853	0.852
	0.891	0.891	0.895	0.884	0.883

experiments on different patch sizes of  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ,  $11 \times 11$ , and  $13 \times 13$ . We first collect five training data sets consisting of 100 000 patch pairs. These different training data sets are then applied to synthesize the HR images of eight test images. The experimental results with different patch sizes are presented in Table V, from which we can see that different



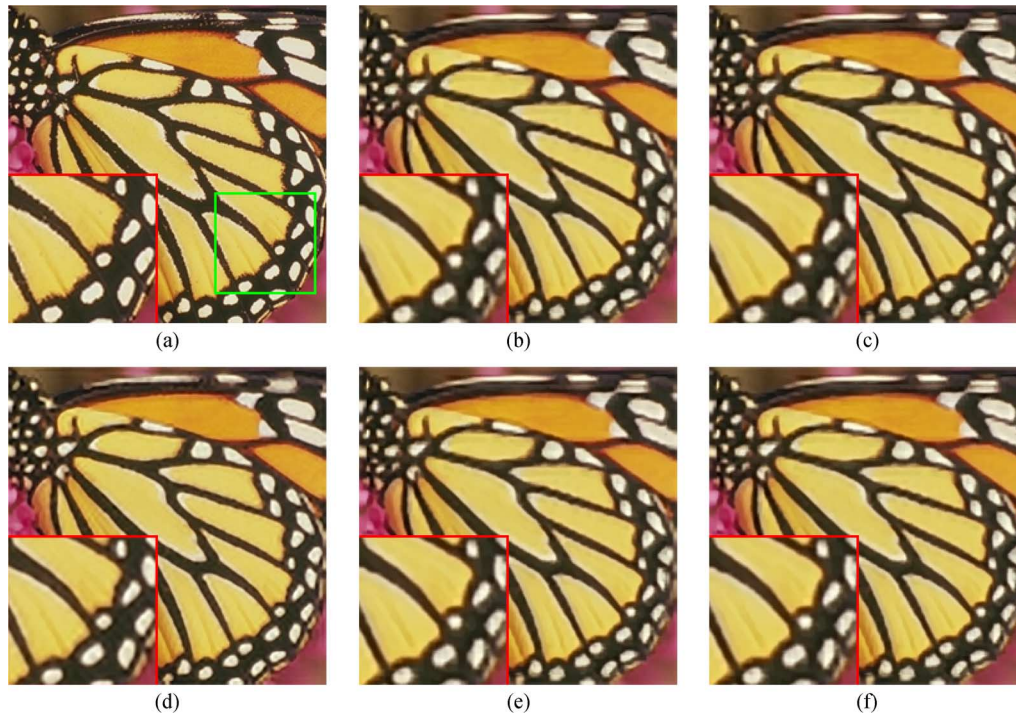


Fig. 8. Comparison ( $\times 3$ ) of SR results on “Butterfly” image with different patch sizes. (a) Original. (b)  $5 \times 5$  result. (c)  $7 \times 7$  result. (d)  $9 \times 9$  result. (e)  $11 \times 11$  result. (f)  $13 \times 13$  result.

patch sizes result in different PSNR, SSIM, and FSIM values. It is observed that the results obtained from the  $9 \times 9$  patch size are the best. Visually, we present the SR results of the “Butterfly” image from different patch sizes, from which we can see that a smaller patch size (e.g.,  $5 \times 5$ ) tends to produce some unwanted artifacts in the resultant images, whereas a larger patch size (e.g.,  $13 \times 13$ ) produces smooth results, as shown in Fig. 8. Therefore, we adopt an image patch size of  $9 \times 9$  when carrying out SR reconstruction.

## V. CONCLUSION

In this paper, we have presented an improved NE-based algorithm for image SR reconstruction by combining the sparse neighbor search and subset selection based on HoG clustering. To accelerate the speed of SR reconstruction, we first employ clustering on HoG features to partition the training data set into a set of subsets. Second, to surmount the drawback of the  $k$ -NN criterion with Euclidean distance metric, we develop a novel neighbor selection scheme by introducing a variation of the Robust-SL0 algorithm. Competitive experimental results validate the efficiency of this scheme. However, one challenging problem that remains is how to establish optimal subsets. This problem can be mitigated by cluster validity analysis [48] to determine a more rational number of subsets. Furthermore, for complicated image structures, such as textural regions, the proposed method does not perform well. In future work, we will further investigate texture similarity. In addition, constructing a more reliable sparse representation model for the SR problem, such as group sparse representation via structural clustering and self-adaptive dictionary learning, is expected to improve the performance of example-learning-based image SR.

## 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] 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.
- [6] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning, “A multi-frame image super-resolution method,” *Signal Process.*, vol. 90, no. 2, pp. 405–414, Feb. 2010.
- [7] M. Protter, M. Elad, H. Takeda, and P. Milanfar, “Generalizing the non-local-means to super-resolution reconstruction,” *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 36–51, Jan. 2009.
- [8] 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, pp. 2738–2747, Oct. 2011.
- [9] 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.
- [10] 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.
- [11] T.-M. Chan and J. Zhang, “An improved super-resolution with manifold learning and histogram matching,” in *Proc. IAPR Int. Conf. Biometrics*, 2006, pp. 756–762.
- [12] 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.
- [13] T.-M. Chan, J. Zhang, J. Pu, and H. Huang, “Neighbor embedding based super-resolution algorithm through edge detection and feature selection,” *Pattern Recog. Lett.*, vol. 30, no. 5, pp. 494–502, Apr. 2009.

- [14] 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.
- [15] 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.
- [16] Y. Tang, X. Pan, Y. Yuan, P. Yan, L. Li, and X. Li, "Single-image super-resolution based on semi-supervised learning," in *Proc. Asian Conf. Pattern Recog.*, 2011, pp. 52–56.
- [17] G. Mu, X. Gao, K. Zhang, X. Li, and D. Tao, "Single image super resolution with high resolution dictionary," in *Proc. IEEE Int. Conf. Image Process.*, 2011, pp. 1165–1168.
- [18] A. Marquina and S. J. Osher, "Image super-resolution by TV-regularization and Bregman iteration," *J. Sci. Comput.*, vol. 37, no. 3, pp. 367–382, Dec. 2008.
- [19] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.
- [20] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1259–1266.
- [21] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010.
- [22] Y. Tang, Y. Yuan, P. Yan, and X. Li, "Single-image super-resolution via sparse coding regression," in *Proc. IEEE Int. Conf. Image Graph.*, 2011, pp. 267–272.
- [23] Y. Tang, X. Pan, Y. Yuan, P. Yan, L. Li, and X. Li, "Local semi-supervised regression for single-image super-resolution," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, 2011, pp. 52–56.
- [24] X. Lu, H. Yuan, Y. Yuan, P. Yan, L. Li, and X. Li, "Local learning-based image super-resolution," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, 2011, pp. 1–5.
- [25] 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.
- [26] Y. Tang, Y. Yuan, P. Yan, and X. Li, "Greedy regression in sparse coding space for single-image super-resolution," *J. Vis. Commun. Image Represent.*, 2012, to be published.
- [27] X. Lu, H. Yuan, P. Yan, Y. Yuan, and X. Li, "Utilizing homotopy for single image superresolution," in *Proc. Asian Conf. Pattern Recog.*, 2011, pp. 316–320.
- [28] X. Gao, K. Zhang, X. Li, and D. Tao, "Joint learning for single-image super-resolution via a coupled constraint," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 469–480, Feb. 2012.
- [29] M. Song, D. Tao, C. Chen, X. Li, and C.-W. Chen, "Colour to gray: Visual cue preservation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1537–1552, Sep. 2010.
- [30] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [31] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [32] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–124, Jul. 2007.
- [33] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed L0 norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, Jan. 2009.
- [34] A. Eftekhar, M. Babaie-Zadeh, C. Jutten, and H. Abrishami Moghaddam, "Robust-SL0 for stable sparse representation in noisy settings," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2009, pp. 3433–3436.
- [35] A. Eftekhar, H. A. Moghaddam, and M. Babaie-Zadeh, " $k/K$ -nearest neighborhood criterion for improving locally linear embedding," in *Proc. CGIV*, 2009, pp. 392–397.
- [36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.
- [37] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, May 1991.
- [38] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [39] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," *Adv. Neural Inf. Process. Syst.*, vol. 19, pp. 801–808, 2007.
- [40] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [41] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted norm minimization algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
- [42] P. Getreuer, "Variational imaging methods for denoising, deconvolution, inpainting, and segmentation," 2009. [Online]. Available: <http://www.math.ucla.edu/~getreuer/tvreg.html>
- [43] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
- [44] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 457–464.
- [45] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [46] 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.
- [47] 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, Jul. 2009.
- [48] X. Gao, J. Li, D. Tao, and X. Li, "Fuzziness measurement of fuzzy sets and its application in cluster validity analysis," *Int. J. Fuzzy Syst. (IJFS)*, vol. 9, no. 4, pp. 188–197, 2007.
- [49] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 60–65.
- [50] X. Wu, X. Zhang, and J. Wang, "Model-guided adaptive recovery of compressive sensing," in *Proc. Data Compress. Conf.*, 2009, pp. 123–132.



**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. His research interests include computational intelligence, machine learning, computer vision, pattern recognition, and artificial intelligence.

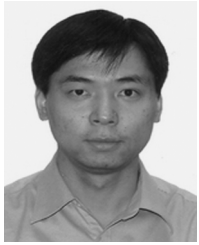
Dr. Gao is a member of IEEE Xi'an Section Executive Committee and the Chair of the Membership Development Committee, a Fellow of IET, and the Vice Chair of IET Xi'an Network. He served as the General Chair/Co-Chair or Program Committee Chair/Co-Chair or PC member for around 30 major international conferences.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



**Kaibing Zhang** was born in Xiaogan, China, 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. Since March 2009, he has been working towards the Ph. D. degree in pattern recognition and intelligent systems at Xidian University, Xi'an, China.

His research interests include pattern recognition and computer vision.





**Dacheng Tao** (M'07–SM'12) 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 was a recipient of the Best Theory/Algorithm Paper Runner-Up Award in the *IEEE ICDM'07*.

**Xuelong Li** (M'02–SM'07–F'12) is currently a Full Professor 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, China.