Computing and Informatics, Vol. 38, 2019, 1301–1319, doi: 10.31577/cai_2019_6_1301

IMAGE SUPER-RESOLUTION BASED ON SPARSE CODING WITH MULTI-CLASS DICTIONARIES

Xiuxiu Liao, Kejia Bai, Qian Zhang
Xiping Jia, Shaopeng Liu, Jin Zhan

School of Computer Science
Guangdong Polytechnic Normal University
No. 293, Zhongshan Avenue West, Tianhe District
510665 Guangzhou, China
e-mail: xxliao2013@aliyun.com, baikejia@gmail.com,
{234482377, 260628072, 149265005, 82952836}@qq.com

Abstract. Sparse coding-based single image super-resolution has attracted much interest. In this paper, a super-resolution reconstruction algorithm based on sparse coding with multi-class dictionaries is put forward. We propose a novel method for image patch classification, using the phase congruency information. A sub-dictionary is learned from patches in each category. For a given image patch, the sub-dictionary that belongs to the same category is selected adaptively. Since the given patch has similar pattern with the selected sub-dictionary, it can be better represented. Finally, iterative back-projection is used to enforce global reconstruction constraint. Experiments demonstrate that our approach can produce comparable or even better super-resolution reconstruction results with some existing algorithms, in both subjective visual quality and numerical measures.

Keywords: Image patch classification, multi-class dictionaries, phase congruency, sparse coding, super-resolution

1 INTRODUCTION

Image super-resolution (SR) refers to the problem of using signal processing techniques to estimate a high-resolution (HR) image $\boldsymbol{X}$ with better quality from an observed low-resolution (LR) image $\boldsymbol{Y}$. The image observation model is usually de-

scribed as [1, 2]:

$$\boldsymbol{Y} = SH\boldsymbol{X} + \boldsymbol{V} \tag{1}$$

where H is a blurring operator, S is a down-sampling operator and $\boldsymbol{V}$ is additive noise.

In recent years, learning-based SR methods [3] have been extensively studied, which use a learned co-occurrence to predict the correspondence between LR and HR patches. The learning algorithms including Markov network [4, 5, 6], neighbor embedding [7, 8, 9, 10], dictionary learning [11, 12, 13, 14], anchored neighborhood regression [16, 15], random forests [17], and deep learning [18, 19, 20, 21].

Freeman et al. [4] propose an approach named VISTA (Vision by Image/Scene TrAining). They generate a synthetic world of scenes and their corresponding rendered images, modeling their relationships with a Markov network. Bayesian belief propagation is used to efficiently find a local maximum of the posterior probability for the scene if an image is given. They apply VISTA to the super resolution problem, such as estimating high frequency details from a low-resolution image. Stephenson and Chen [5] propose to use even stronger prior information by extending Markov random filed (MRF)-based super-resolution to use adaptive observation and transition functions, that is, to make these functions region-dependent. Ma et al. [6] learn the parameters of the network from training set, which computes probability distribution by K-means algorithm. Given a low-resolution image as input, Chang et al. [7] recover its high-resolution counterpart using a set of training examples. Specifically, small image patches in the low and high-resolution images form manifolds with similar local geometry in two distinct feature spaces. As in locally linear embedding (LLE), local geometry is characterized by how a feature vector corresponding to a patch can be reconstructed by its neighbors in the feature space. Besides using the training image pairs to estimate the high-resolution embedding, they also enforce local compatibility and smoothness constraints between patches in the target high-resolution image through overlapping. Zhang et al. [8] propose a partial least squares (PLS) method, called locality preserving PLS (LPPLS), to find a unified feature space where the correlation between LR and HR image patches on that space is maximized. Applying the proposed LPPLS, they learn the joint mapping of LR and HR image patches simultaneously and then map these image patches onto the unified feature space. The k-nearest neighbor (k-NN) searching and the optimal reconstruction weights computing are performed in this unified feature space as well. Rahiman and George [9] propose learning-based approaches for single image super-resolution using sparse representation and neighbor embedding. Separate prediction models are trained for each cluster, and the model parameters are updated with each input image to adapt to input test image. Gao et al. [10] propose a sparse neighbor selection scheme for SR reconstruction. They first predetermine a larger number of neighbors as potential candidates and develop an algorithm to simultaneously find the neighbors and to solve the reconstruction weights. Recognizing that the k-nearest neighbor for reconstruction should have similar local geometric structures based on clustering, they employ a local statisti-

cal feature, namely histograms of oriented gradients (HoG) of 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. Wang et al. [11] propose a semi-coupled dictionary learning (SCDL) model to solve cross-style image synthesis problems. Under SCDL, a pair of dictionaries and a mapping function will be simultaneously learned. The dictionary pair can well characterize the structural domains of the two styles of images, while the mapping function can reveal the intrinsic relationship between the two styles' domains. The two dictionaries will not be fully coupled so that much flexibility can be given to the mapping function for an accurate conversion across styles. Moreover, clustering and image nonlocal redundancy are introduced to enhance the robustness of SCDL. He et al. [12] apply a Bayesian method using a beta process prior to learn the over-complete dictionaries. They not only provide dictionaries that customized to each feature space, but also add more consistent and accurate mapping between the two feature spaces. The proposed algorithm is able to learn sparse representations that correspond to the same dictionary atoms with the same sparsity but different values in coupled feature spaces, thus bringing consistent and accurate mapping between coupled feature spaces.

Timofte et al. [15] propose fast super-resolution methods while making no compromise on quality. First, they support the use of sparse learned dictionaries in combination with neighbor embedding methods. In this case, the nearest neighbors are computed using the correlation with the dictionary atoms rather than the Euclidean distance. Second, they show that using global collaborative coding has considerable speed advantages, reducing the super-resolution mapping to a precomputed projective matrix. Third, they propose the anchored neighborhood regression (ANR) algorithm to anchor the neighborhood embedding of a low resolution patch to the nearest atom in the dictionary and to precompute the corresponding embedding matrix. In their later work [16], an improved variant of ANR (A+) is proposed which combines the best qualities of ANR and simple functions (SF) [22]. A+ builds on the features and anchored regressors from ANR. Instead of learning the regressors on the dictionary, it uses the full training material, similar to SF. Schulter et al. [17] propose to directly map from low to high-resolution patches using random forests. They demonstrate how random forests nicely fit into this framework. During the process of trees training, they optimize a novel and effective regularized objective that not only operates on the output space but also on the input space, which especially suits the regression task. During inference, they comprise the same well-known computational efficiency that has made random forests popular for many computer vision problems.

Dong et al. [18] propose a deep learning method for single image super-resolution which directly learns an end-to-end mapping between the low/high-resolution images. The mapping is represented as a deep convolutional neural network (CNN) that takes the low-resolution image as the input and outputs the high-resolution one. Their deep CNN has a lightweight structure, which demonstrates state-of-the-

art restoration quality, and achieves fast speed for practical on-line usage. They extend the network to cope with three color channels simultaneously, and show better overall reconstruction quality. Kim et al. [19] use a very deep convolutional network inspired by Visual Geometry Group (VGG)-net for ImageNet classification. Lai et al. [20] propose the Laplacian Pyramid Super-Resolution Network (LapSRN) to progressively reconstruct the sub-band residuals of high-resolution images. At each pyramid level, the model takes coarse-resolution feature maps as input, predicts the high-frequency residuals, and uses transposed convolutions for upsampling to the finer level. They train the proposed LapSRN with deep supervision using a robust Charbonnier loss function and achieve high-quality reconstruction. Furthermore, the network generates multi-scale predictions in one feed-forward pass through the progressive reconstruction, thereby facilitates resource-aware applications. Liu et al. [21] argue that domain expertise from the conventional sparse coding model can be combined with the key ingredients of deep learning to achieve further improved results.

Yang et al. [13] propose sparse coding-based SR framework. Image patches are assumed to have a sparse representation with respect to an over-complete dictionary, and the most relevant reconstruction neighbors are adaptively selected based on sparse coding, avoiding under- or over-fitting. Zeyde et al. [14] embark from the work of [13], similarly assume a local Sparse-Land model on image patches serving as regularization, but use a different training approach for the dictionary pair. Both methods aim at learning a universal dictionary. However, the contents may vary significantly across different image patches, and sparse decomposition over a universal dictionary is potentially unstable [23]. Adaptive sparse coding via multiple dictionaries has been proposed. Yang et al. [24] employ multiple dictionaries learned from K-means clustered patches. Dong et al. [25] use principal component analysis (PCA) technique to learn the sub-dictionaries, and autoregressive and nonlocal self-similarity are introduced as regularization terms. These methods do not use the geometric information as a supervised prior to guide the image patch clustering.

In this paper, a novel method for image patch classification is proposed, and it is integrated into the multiple dictionaries learning SR framework, called SR reconstruction based on Sparse Coding with Multi-Class Dictionary (SC-MCD). Employing the Phase Congruency (PC) measurement [26], image patches are divided into non-smooth patches with different orientations and smooth patches. PC provides an absolute measure of the significance of a local structure, and it is invariant to changes in illumination and magnification. The example patches are classified into several categories, and each category consists of patches with similar patterns, where a sub-dictionary can be learned. For an image patch to be coded, the sub-dictionary that belongs to the same category is selected adaptively. Since the given patch has similar pattern with the selected sub-dictionary, it can be better represented, and the whole image can be more accurately reconstructed. Besides, we use iterative back-projection (IBP) [27] to enforce global reconstruction constraint, which is simple but effective.

The remainder of this paper is organized as follows. In Section 2, we present the proposed SC-MCD algorithm in detail. Experimental results are then presented in Section 3. Finally, Section 4 gives some concluding remarks and discussions on future works.

2 PROPOSED APPROACH

2.1 Sparse Representation-Based SR

Following the image observation model (1), the task of SR is to estimate an HR image $\boldsymbol{X}$ satisfying reconstruction constraint, which requires that $\boldsymbol{X}$ should be consistent with the observed LR image $\boldsymbol{Y}$ with respect to (1). It is ill-posed as many HR images satisfy the reconstruction constraint. Sparse representation-based SR methods use sparse prior on local patches to regularize the estimate of HR image.

Let x, y denote the HR and LR image patches, respectively, two dictionaries $\boldsymbol{D}_h$ and $\boldsymbol{D}_l$ are trained to have the same sparse representations for each HR and LR image patch pair. The recovery of x from y under the sparse prior can be described as:

$$\min_{\alpha,x} \|\alpha\|_0 \text{ s.t. } \|F\boldsymbol{D}_l\alpha - Fy\|_2^2 \leq \epsilon_1,\ \|F\boldsymbol{D}_h\alpha - Fx\|_2^2 \leq \epsilon_2,\ \|SH\boldsymbol{X} - \boldsymbol{Y}\|_2^2 \leq \epsilon_3 \tag{2}$$

where F is a (linear) feature extraction operator, α is the sparse representation coefficients, $\|\alpha\|_0$ represents the number of non-zero coefficients in α, and ϵ_i, $i = 1, 2, 3$ are the admissible errors.

2.2 Classification of Image Patches

Learning a universal dictionary able to optimally represent image patches with various patterns is very difficult. So it is meaningful to learn multiple dictionaries with different patterns. In this section, a novel method for image patch classification is proposed, employing the PC information.

Rather than defining features directly at points with sharp changes in intensity, the PC model postulates that features are perceived at points where the Fourier components are maximal in phase. Kovesi [26] proposes to calculate PC with logarithmic Gabor wavelets. PC at location i is expressed as the summation over orientation o and scale n:

$$PC(i) = \frac{\sum_0 \sum_n W_0(i) \lfloor A_{no}(i)\Delta\Phi_{no}(i) - T_0 \rfloor}{\sum_0 \sum_n A_{no}(i) + \epsilon} \tag{3}$$

where $\lfloor x \rfloor$ equals to x when $x > 0$ and it equals to 0 when $x < 0$. A represents the amplitude of the Fourier component, and $\Delta\Phi$ is phase deviation. T compensates for the influence of noise, W is the weighting function for frequency spread, and ϵ is a small constant to avoid division by zero.

Once the PC map of the image is obtained, a threshold P_0 is used to get a binary image, where '1' indicates feature points. At the same time, an orientation image is computed, recording the direction angle in which local energy is a maximum for each pixel. The direction angles are uniform sampled in the range $[0°, 180°]$, and the sample interval is determined by the number of orientation.

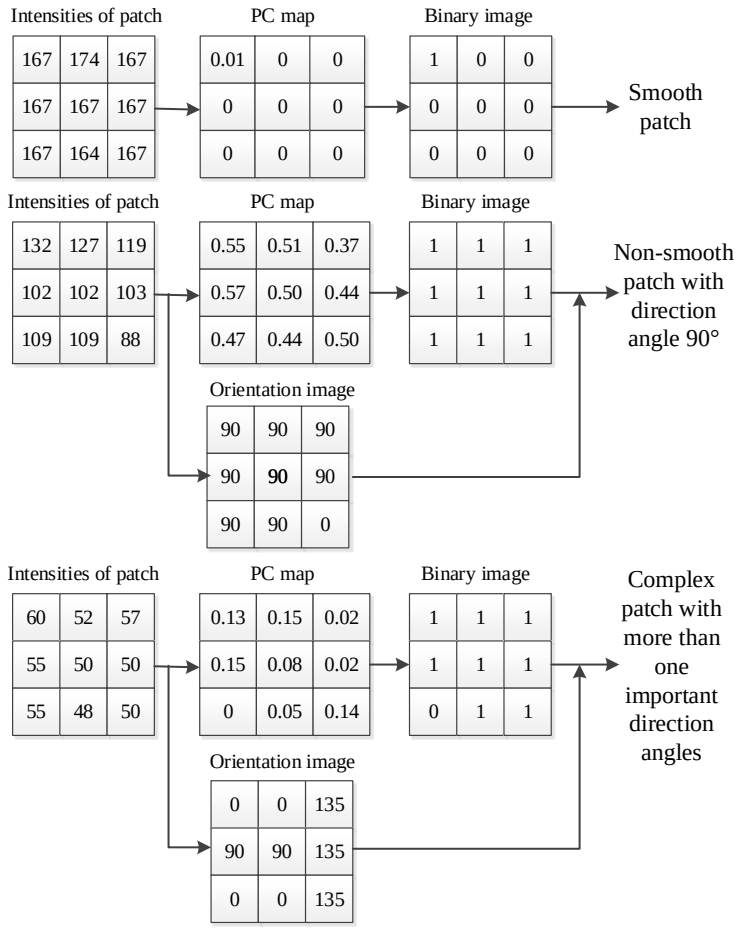


Figure 1. Examples of image patch classification

Extract patches from the binary image and orientation image, denote as $\{b_i\}$ and $\{o_i\}$. Count the number of '1' for each b_i, and if it is smaller than 1/3 of the total number of pixels in the patch, then patch i is classified as a smooth patch. For each non-smooth patch, find the direction angle d_i that repeats most times

in o_i. Classify non-smooth patches that have identical d_i into the same category. However, if the number of occurrences of d_i is less than half of the total number of image patches, the pixel block may have more than one main orientation, and it is judged as a complex patch. So if we compute PC map for $J-1$ orientations, we will get J categories of patches. Figure 1 shows examples of image patch classification.

Using the above classification method, example patches are classified into different categories with similar patterns, and for each category a sub-dictionary can be learned. For an image patch to be processed, the sub-dictionary that belongs to the same category is selected adaptively.

2.3 SR Reconstruction Based on Sparse Coding with Multi-Class Dictionary (SC-MCD)

The algorithm consists of two parts: multi-class dictionaries training and SR reconstruction.

Part I: Multi-Class Dictionary Pairs Training (Can Be Done Offline)

The first step is to extract feature vector pairs from the HR and LR training images. Firstly, LR patches of size $n \times n$ pixels are extracted from the LR training images, and the classification algorithm described in Section 2.2 is applied to the patches, so each patch gets a category label $j, j = 1, 2, \cdots, J$. Secondly, high-pass filters are used to extract features from the LR training images. The four filters used are:

$$f_1 = [-1, 1], f_2 = f_1^T, f_3 = [1, -2, 1]/2, f_4 = f_3^T \tag{4}$$

where the superscript "T" means transpose. Applying these filters yields four vectors for each LR patch, which are concatenated into one vector q_l^j as the feature vector of the LR patch, where the superscript j is the category label of the patch. The feature vectors with the same category labels are grouped together, so the LR feature vector set $\left\{ \boldsymbol{Q}_l^{(1)}, \boldsymbol{Q}_l^{(2)}, \ldots, \boldsymbol{Q}_l^{(J)} \right\}$ is obtained. Finally, the HR training images are subtracted by the interpolated images scaled up from the corresponding LR training ones and the high frequency parts are kept. HR patches of size $R_n \times R_n$ pixels are extracted from the high frequency images, where R is the SR ratio. Each HR patch is arranged in vector form q_h^j as the feature vector, where the superscript j is the category label of the corresponding LR patch. The feature vectors with the same category labels are grouped together, so the HR feature vector set $\left\{ \boldsymbol{Q}_h^{(1)}, \boldsymbol{Q}_h^{(2)}, \ldots, \boldsymbol{Q}_h^{(J)} \right\}$ is obtained.

Suppose that there are totally J categories of feature vector pairs set for HR and LR image patches respectively, $\left\{ \left\langle \boldsymbol{Q}_l^{(1)}, \boldsymbol{Q}_h^{(1)} \right\rangle, \left\langle \boldsymbol{Q}_l^{(2)}, \boldsymbol{Q}_h^{(2)} \right\rangle, \ldots, \left\langle \boldsymbol{Q}_l^{(J)}, \boldsymbol{Q}_h^{(J)} \right\rangle \right\}$ for HR and LR image patches, respectively, with the same number of columns for each $\left\langle \boldsymbol{Q}_l^{(j)}, \boldsymbol{Q}_h^{(j)} \right\rangle$ pairs. We want to learn J LR sub-dictionaries $\left\{ \boldsymbol{D}_l^{(1)}, \boldsymbol{D}_l^{(2)}, \ldots, \boldsymbol{D}_l^{(J)} \right\}$

under which the patches in $\left\{\boldsymbol{Q}_l^{(1)}, \boldsymbol{Q}_l^{(2)}, \ldots, \boldsymbol{Q}_l^{(J)}\right\}$ can be sparsely represented, respectively. The problem can be written as:

$$\left\{\boldsymbol{D}_l^{(j)}, \boldsymbol{A}^{(j)}\right\} = \arg\min \left\|\boldsymbol{D}_l^{(j)}\boldsymbol{A}^{(j)} - \boldsymbol{Q}_l^{(j)}\right\|_2^2 \ s.t. \ \left\|\alpha_i^{(j)}\right\|_0 \leq C, j = 1, 2, \ldots, J \quad (5)$$

where $\boldsymbol{A}^{(j)}$ is the coding coefficients of the patch vectors set $\boldsymbol{Q}_l^{(j)}$ under the sub-dictionary $\boldsymbol{D}_l^{(j)}$, $\alpha_i^{(j)}$ is the i^{th} column of $\boldsymbol{A}^{(j)}$, $\left\|\alpha_i^{(j)}\right\|_0$ represents the number of non-zero coefficients in $\alpha_i^{(j)}$, C is the sparsity threshold. K-SVD dictionary learning algorithm [28] is used to simultaneously determine the sub-dictionaries and the coding coefficients.

The HR sub-dictionaries can be computed by the following Pseudo-Inverse expression:

$$\boldsymbol{D}_h^{(j)} = \boldsymbol{Q}_h^{(j)}\left(\boldsymbol{A}^{(j)}\right)^+ = \boldsymbol{Q}_h^{(j)}\boldsymbol{A}_h^{(j)T}\left(\boldsymbol{A}^{(j)}\boldsymbol{A}^{(j)T}\right)^{-1} \quad (6)$$

where Q_h and Q_l are matrices of feature vectors of HR and LR image patches, D_h and D_l are the HR and LR dictionaries, $Q_h = A_h D_h$, $Q_l = A_l D_l$, where A_h and A_l are sparse matrices. We suppose the two dictionaries D_h and D_l are trained to have the same sparse representations for each HR and LR image patch pair, so $A_h = A_l = A$. The coefficient matrix A is obtained during the training process of D_l, and the dictionary D_h is obtained from D_h and A from the formula

$$\boldsymbol{D}_h^{(j)} = \arg\min \left\|\boldsymbol{Q}_h^{(j)} - \boldsymbol{A}^{(j)}\boldsymbol{D}_h^{(j)}\right\|_F^2. \quad (7)$$

Multi-class dictionary pairs are trained, instead of only one pair of dictionary $\langle D_l, D_h\rangle$:

$$\left\{\left\langle \boldsymbol{D}_l^{(1)}, \boldsymbol{D}_h^{(1)}\right\rangle, \left\langle \boldsymbol{D}_l^{(2)}, \boldsymbol{D}_h^{(2)}\right\rangle, \ldots, \left\langle \boldsymbol{D}_l^{(J)}, \boldsymbol{D}_h^{(J)}\right\rangle\right\}. \quad (8)$$

Part II: SR Reconstruction

For a given LR test image $\boldsymbol{Y}$, patches of size $n \times n$ pixels are extracted, with $n-1$ pixels overlap in each direction. For each patch, extract feature vector y, and calculate which category it belongs to. For each y that belongs to j-th category, sub-dictionary pair $\left\langle \boldsymbol{D}_l^{(j)}, \boldsymbol{D}_h^{(j)}\right\rangle$ is selected, and the sparse coding can be written as

$$\mu^{(j)*} = \arg\min \left\|\boldsymbol{D}_l^{(j)}\mu^{(j)} - y\right\|_2^2 \ s.t. \ \left\|\mu^{(j)}\right\|_0 \leq C. \quad (9)$$

OMP algorithm [29] is used to calculate the optimal solution $\mu^{(j)*}$, and the corresponding HR feature vector is obtained by $x = \boldsymbol{D}_h^{(j)}\mu^{(j)*}$. When all the HR features are obtained, the high frequency HR image $\boldsymbol{X}_k$ can be constructed by enforcing local compatibility and smoothness constraints between adjacent patches. The target HR image $\boldsymbol{X}_0$ is the summation of the high frequency image $\boldsymbol{X}_k$ and the interpolated image $\boldsymbol{X}_i$, which is scaled up from the LR image $\boldsymbol{Y}$ by bicubic interpolation.

The HR image produced by the above sparse representation method may not satisfy the reconstruction constraint exactly. IBP [27] is employed to alleviate this problem:

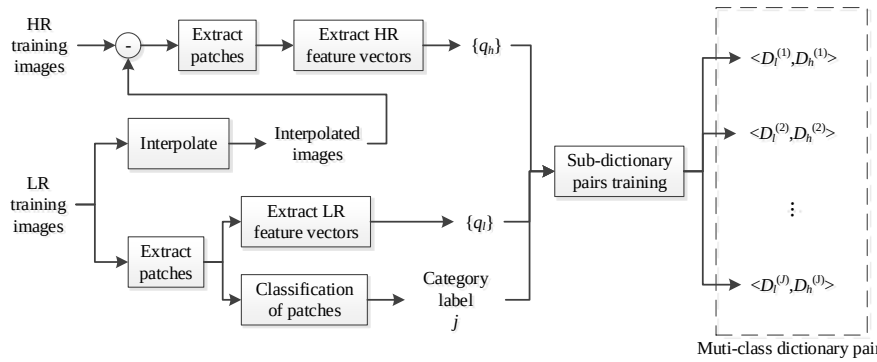
$$\boldsymbol{X}^* = \arg\min \|SH\boldsymbol{X} - \boldsymbol{Y}\|_2^2 . \tag{10}$$

The solution can be efficiently computed using gradient descent:

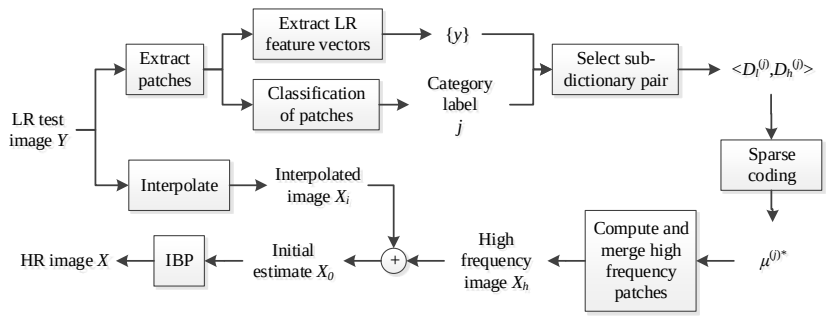
$$\boldsymbol{X}_{t+1} = \boldsymbol{X}_t + \tau \left[H^T S^T (\boldsymbol{Y} - SH\boldsymbol{X}_t) \right] . \tag{11}$$

where τ is the step size of the gradient descent, $\boldsymbol{X}_t$ is the estimate of HR image after the t^{th} iteration, and $\boldsymbol{X}_0$ is used as the initial estimate of $\boldsymbol{X}$.

The block diagram of the proposed SC-MCD algorithm is shown in Figure 2.



a) Multi-class dictionary pairs training



b) SR reconstruction

Figure 2. The block diagram of the proposed SC-MCD algorithm

3 EXPERIMENTAL RESULTS

We use several benchmarks, including Set14 [14], Set5 [30] and B100 [16] as our testing set. The magnification factor is 4. Those 3×3 patches are extracted in LR images, and the corresponding HR patches are 12×12. About 80 000 training patch pairs are collected from the training image set used in [13] and 1 024 atoms are trained for each sub-dictionary. Sparsity threshold C is set to be 5. All the simulations are conducted in MATLAB R2016a on PC with Intel® Core™ i7/3.6 GHz/4 GB.

If the number of categories is too small, the differences between the geometric features of the image patches corresponding to each sub-dictionary will be rather large. On the other hand, if there are a lot of categories, the discriminations between sub-dictionaries are low. Therefore, both the factors of accuracy and discrimination are taken into account in the expression of sub-dictionaries to choose the appropriate number of categories. In the experiments, the number of categories is increased from 5 to 9, and the average peak signal-noise ratio (PSNR) of each data set is calculated. The results are shown in Table 1. Experiments show that the reconstruction results are better when the number of categories (J) is set to 6 or 7. In the later experiments, J is set to be 6.

	J = 5	J = 6	J = 7	J = 8	J = 9
Set5	27.787	27.787	**27.814**	27.785	27.762
Set14	25.095	**25.117**	25.101	25.095	25.097
B100	25.331	**25.335**	25.332	25.327	25.324

Table 1. Mean PSNR of SR reconstructed images using different number of categories

3.1 MCD vs. SCD

In this part, we evaluate the influence of multi-class dictionaries to the quality of reconstructed HR images. We test two methods: SR reconstruction with single class dictionary (denote as SCD) [14], and SR reconstruction with multi-class dictionaries (denote as MCD). In order to find out the contribution of the multi-class dictionaries independently, we do not apply IBP to enforce global reconstruction constraint in the experiments for this part.

We select 7 images from Set14 as the testing set. For each test image, the mean squared error (MSE) between the reconstructed image and the original one is calculated, and the result is shown in Table 2. We can see that, compared to SCD, MCD averagely reduces the squared error by 7.541 per pixel.

3.2 MD_PC vs. MD_KM

In this part, we compare the SR performance by different multi-class dictionaries. The proposed image patches classification method based on phase congruency is

	SCD	MCD
Barbara	271.231	253.994
Coastguard	237.873	233.966
Face	77.549	76.640
Foreman	109.968	99.542
Man	200.533	197.136
Pepper	108.982	105.698
Zebra	252.086	238.458
Average	179.746	172.205

Table 2. MSE of SR reconstructed images using SCD and MCD

denoted as MD_PC. In [25] Dong et al. use k-means for image patches clutering, denoted as MD_KM. In order to highlight the impact of the image patches classification algorithms on SR performance, the SR reconstruction process here does not introduce any global reconstruction constraints.

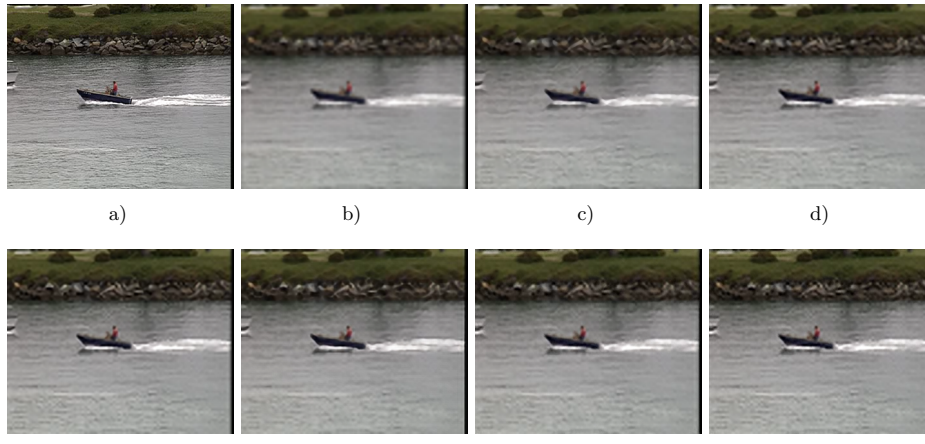
Measures	MD_KM		MD_PC	
	PSNR (dB)	FSIM	PSNR (dB)	FSIM
Barbara	24.038	0.933	24.083	0.938
Coastgrard	24.337	0.725	24.439	0.729
Face	29.322	0.871	29.286	0.870
Foreman	27.060	0.879	28.151	0.896
Man	25.098	0.932	25.183	0.936
Pepper	27.713	0.961	27.890	0.965
Zebra	24.075	0.909	24.357	0.920
Average	25.949	0.887	26.198	0.894

Table 3. Numerical measurements of the reconstructed images by multi-class dictionaries

The results are compared by PSNR and Feature Similarity (FSIM) [31]. The higher of PSNR and FSIM means much similar of the reconstructed image to the original image. As shown in Table 3, the numerical measurements of PSNR and FSIM obtained by proposed MD_PC are higher than that of MD_KM used in [25].

3.3 SC_MCD vs. Other Algorithms

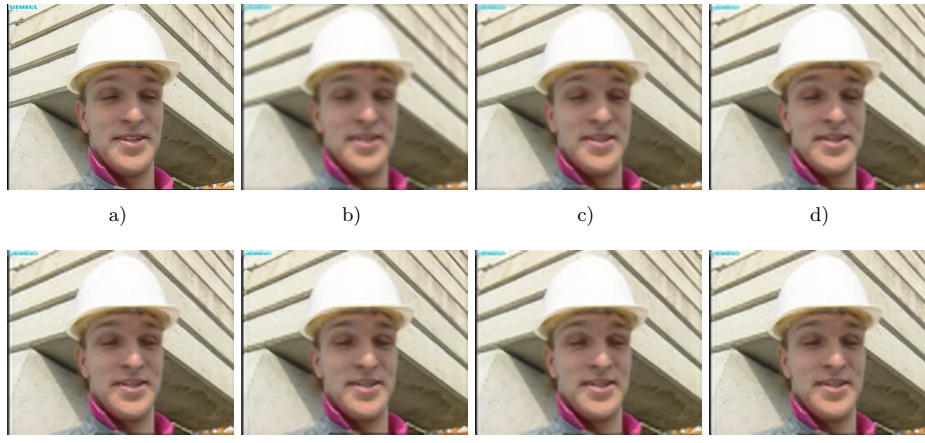
We compare the proposed MCD (without IBP) and MCD_IBP methods with Bicubic interpolation method, Kim's method using sparse regression and natural image prior denoted as SR_NIP [32], Zeyde's sparse coding-based SR using a universal dictionary denoted as SC_SR [14], Dong's method by k-means clustering, adaptive sparse domain selection and adaptive regularization denoted as ASDS_AR_NL [25], and Dong's method by deep convolutional neural network denoted as SRCNN [18]. The comparison experiments are based on the matlab versions of the source code



a) b) c) d)

e) f) g) h)

Figure 3. Original "Coastguard" image and images reconstructed by different methods. a) Original image, b) Bicubic, c) SR_NIP [32], d) SC_SR [14] e) ASDS_AR_NL [25], f) SRCNN [18], g) MCD, h) MCD_IBP.



a) b) c) d)

e) f) g) h)

Figure 4. Original "Foreman" image and images reconstructed by different methods. a) Original image, b) Bicubic, c) SR_NIP [32], d) SC_SR [14], e) ASDS_AR_NL [25], f) SRCNN [18], g) MCD, h) MCD_IBP.

or demos provided by the above papers, so it is very fair. The results are compared by visual quality subjectively and by numerical measurements of PSNR and FSIM.

Image	Measures	Bicubic	SR_NIP [32]	SC_SR [14]	ASDS_AR_NL [25]	SRCNN [18]	MCD	MCD_IBP
Baboon	PSNR (dB)	20.199	20.436	20.407	20.476	20.423	20.476	**20.499**
	FSIM	0.849	0.890	0.892	**0.904**	0.886	**0.904**	**0.904**
Barbara	PSNR (dB)	23.440	24.077	23.797	**24.120**	23.982	24.083	24.118
	FSIM	0.896	0.934	0.931	0.936	0.932	**0.938**	**0.938**
Bridge	PSNR (dB)	22.873	23.440	23.519	23.561	23.421	23.552	**23.610**
	FSIM	0.870	0.904	0.909	0.913	0.901	**0.916**	**0.916**
Coastguard	PSNR (dB)	23.821	24.333	24.367	24.374	24.339	24.439	**24.479**
	FSIM	0.639	0.685	0.696	0.724	0.692	**0.729**	**0.729**
Comic	PSNR (dB)	22.022	20.767	20.730	**20.909**	20.820	20.782	20.841
	FSIM	0.702	0.750	0.749	**0.765**	0.760	0.761	0.761
Face	PSNR (dB)	28.620	29.156	29.235	**29.437**	29.015	29.286	29.333
	FSIM	0.826	0.845	0.856	**0.872**	0.850	0.870	0.870
Flowers	PSNR (dB)	23.560	24.552	24.403	**24.662**	24.601	24.491	24.548
	FSIM	0.770	0.809	0.806	**0.822**	0.815	0.819	0.819
Foreman	PSNR (dB)	25.756	27.640	27.718	27.553	26.596	28.151	**28.222**
	FSIM	0.850	0.888	0.888	0.890	0.882	**0.896**	**0.896**
Lena	PSNR (dB)	27.973	29.186	29.040	**29.333**	29.029	29.151	29.201
	FSIM	0.935	0.959	0.960	**0.967**	0.956	0.965	0.965
Man	PSNR (dB)	24.187	25.234	25.109	**25.315**	25.212	25.183	25.251
	FSIM	0.892	0.930	0.930	**0.939**	0.929	0.936	0.936
Monarch	PSNR (dB)	25.867	27.780	27.300	27.764	**28.125**	27.419	27.517
	FSIM	0.922	0.954	0.945	**0.956**	**0.956**	0.949	0.949
Pepper	PSNR (dB)	26.974	27.864	27.757	**27.923**	27.660	27.890	27.918
	FSIM	0.937	0.963	0.962	**0.967**	0.960	0.965	0.965
PPT3	PSNR (dB)	20.203	21.465	21.341	21.435	**21.920**	21.569	21.635
	FSIM	0.824	0.883	0.886	0.892	**0.896**	0.889	0.889
Zebra	PSNR (dB)	22.438	24.380	24.115	24.329	24.457	24.357	**24.466**
	FSIM	0.851	0.914	0.910	0.919	0.916	**0.920**	**0.920**
Average	PSNR (dB)	23.995	25.022	24.917	25.085	24.971	25.059	**25.117**
	FSIM	0.840	0.879	0.880	**0.890**	0.881	**0.890**	**0.890**

Table 4. Numerical measurements of the reconstructed images of Set14 by different methods

Figure 3 shows the original "Coastguard" image and images reconstructed by different methods. Figure 4 gives the results of "Foreman" image. Table 4 shows numerical measurements of the reconstructed images of Set14. Table 5 shows the average PSNR and FSIM measurements on several benchmarks, including Set5, Set14 and B100. We can see ASDS_AR_NL [25], the proposed MCD and MCD_IBP methods have best results, which outperform SR_NIP [32], SC_SR [14] and SRCNN [18].

Method	Measures	Set5	Set14	B100
Bicubic	PSNR (dB)	26.226	23.995	24.463
	FSIM	0.837	0.840	0.791
SR_NIP [32]	PSNR (dB)	27.638	25.022	25.220
	FSIM	0.879	0.879	0.784
SC_SR [14]	PSNR (dB)	27.569	24.917	25.204
	FSIM	0.872	0.880	0.793
ASDS_AR_NL [25]	PSNR (dB)	**27.951**	25.085	**25.355**
	FSIM	0.883	**0.890**	**0.812**
SRCNN [18]	PSNR (dB)	27.677	24.971	25.200
	FSIM	**0.885**	0.881	0.791
MCD_IBP	PSNR (dB)	27.787	**25.117**	25.332
	FSIM	0.878	**0.890**	**0.812**

Table 5. Average numerical measurements of the reconstructed images of Set 5, Set14 and B100

Image	SR_NIP [32]	SC_SR [14]	ASDS_AR_NL [25]	SRCNN [18]	MCD	MCD_IBP
Baboon	31.596	1.583	103.039	10.481	1.897	1.954
Barbara	44.494	2.679	185.560	19.267	3.306	3.391
Bridge	41.816	1.684	117.051	11.446	2.061	2.127
Coastguard	11.041	0.635	36.238	2.574	0.775	0.817
Comic	21.218	0.557	34.803	2.410	0.683	0.706
Face	4.294	0.471	25.832	2.047	0.575	0.603
Flowers	25.238	1.163	75.181	7.510	1.398	1.452
Foreman	13.245	0.635	36.788	2.671	0.776	0.803
Lenna	23.240	1.684	110.869	11.246	2.048	2.138
Man	39.288	1.682	112.917	11.226	2.070	2.135
Monarch	39.786	2.521	183.439	17.681	3.099	3.207
Pepper	24.084	1.669	110.913	11.209	2.067	2.127
PPT3	46.084	2.102	163.759	16.059	2.303	2.392
Zebra	39.680	1.436	99.953	9.544	1.764	1.841
Average	28.936	**1.464**	99.739	9.669	**1.773**	**1.835**

Table 6. Reconstruction time(s) of different methods

Table 6 shows the reconstruction time to investigate the time complexity. Among the four algorithms with the best reconstruction effect, the proposed MCD and MCD_IBP are much faster than SRCNN and ASDS_AR_NL. The average time for reconstructing an image is less than 2 seconds using the proposed MCD and MCD_IBP algorithms, whereas the SRCNN needs about 10 seconds and ASDS_AR_NL needs nearly 100 seconds.

Considering all the factors including subjective visual quality, objective assessment and time complexity, the proposed method obtains good performance for image

SR reconstruction. The proposed methods not only significantly improve the image reconstruction speed, but also significantly improve the image reconstruction quality.

4 CONCLUSIONS

In this paper, we propose a SR reconstruction algorithm based on sparse coding with multi-class dictionaries. A novel method for image patch classification is put forward, and a sub-dictionary is selected adaptively for each given image patch. IBP is used to enforce global reconstruction constraint. Our approach produces comparable or even better SR reconstruction results with some existing algorithms. The robustness of the proposed algorithm under different imaging conditions will be our future work.

Acknowledgment

This work is supported by Young Creative Talents Project of Department of Education of Guangdong Province (Natural Science) (2016KQNCX092, 2017KQNCX117, 2015KQNCX084, 2016KQNCX089), National Natural Science Foundation of China (61772144, 61702119), Innovation Research Project of Education Department of Guangdong Province (Natural Science) (2016KTSCX077), Foreign Science and Technology Cooperation Plan Project of Guangzhou Science Technology and Innovation Commission (201807010059), National Natural Science Foundation of Guangdong (2016A030313472, 2018A030313994, 2018A0303130187), Science and Technology Program of Guangzhou (201607010152).

REFERENCES

[1] Park, S. C.—Park, M. K.—Kang, M. G.: Super-Resolution Image Reconstruction: A Technical Overview. IEEE Signal Processing Magazine, Vol. 20, 2003, No. 3, pp. 21–36, doi: 10.1109/MSP.2003.1203207.

[2] Zhang, K.—Zuo, W.—Zhang, L.: Learning a Single Convolutional Super-Resolution Network for Multiple Degradations. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3262–3271, doi: 10.1109/CVPR.2018.00344.

[3] Pandey, G.—Ghanekar, U.: A Compendious Study of Super-Resolution Techniques by Single Image. Optik, Vol. 166, 2018, pp. 147–160, doi: 10.1016/j.ijleo.2018.03.103.

[4] Freeman, W. T.—Pasztor, E. C.—Carmichael, O. T.: Learning Low-Level Vision. International Journal of Computer Vision (IJCV), Vol. 40, 2000, No. 1, pp. 25–47, doi: 10.1023/A:1026501619075.

[5] Stephenson, T. A.—Chen, T.: Adaptive Markov Random Fields for Example-Based Super-Resolution of Faces. EURASIP Journal on Advances in Signal Processing, 2006, Art. No. 031062, 11 pp., doi: 10.1155/ASP/2006/31062.

[6] Ma, Y. J.—Zhang, H.—Xue, Y.—Zhang, S.: Super-Resolution Image Reconstruction Based on K-Means-Markov Network. Proceedings of the 2009 Fifth International Conference on Natural Computation, Tianjin, China, IEEE, 2009, pp. 316–318, doi: 10.1109/ICNC.2009.608.

[7] Chang, H.—Yeung, D.-Y.—Xiong, Y.: Super-Resolution Through Neighbor Embedding. Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2004), 2004, pp. 275–282, doi: 10.1109/CVPR.2004.1315043.

[8] Zhang, Z.—Qi, C.—Hao, Y.: Locality Preserving Partial Least Squares for Neighbor Embedding-Based Face Hallucination. Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 409–413, doi: 10.1109/ICIP.2016.7532389.

[9] Rahiman, V. A.—George, S. N.: Single Image Super Resolution Using Neighbor Embedding and Statistical Prediction Model. Computer and Electrical Engineering, Vol. 62, 2017, pp. 281–292, doi: 10.1016/j.compeleceng.2016.12.018.

[10] Gao, X.—Zhang, K.—Tao, D.—Li, X.: Image Super-Resolution with Sparse Neighbor Embedding. IEEE Transactions on Image Processing, Vol. 21, 2012, No. 7, pp. 3194–3205, doi: 10.1109/TIP.2012.2190080.

[11] Wang, S.—Zhang, L.—Liang, Y.—Pan, Q.: Semi-Coupled Dictionary Learning with Applications to Image Super-Resolution and Photo-Sketch Synthesis. Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2216–2223, doi: 10.1109/CVPR.2012.6247930.

[12] He, L.—Qi, H.—Zaretzki, R.: Beta Process Joint Dictionary Learning for Coupled Feature Spaces with Application to Single Image Super-Resolution. Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 345–352, doi: 10.1109/CVPR.2013.51.

[13] Yang, J.—Wright, J.—Huang, T. S.—Ma, Y.: Image Super-Resolution via Sparse Representation. IEEE Transactions on Image Processing, Vol. 19, 2010, No. 11, pp. 2861–2873, doi: 10.1109/TIP.2010.2050625.

[14] Zeyde, R.—Elad, M.—Protter, M.: On Single Image Scale-Up Using Sparse-Representations. Proceedings of the 7th International Conference on Curves and Surfaces, 2010. Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 6920, 2012, pp. 711–730, doi: 10.1007/978-3-642-27413-8_47.

[15] Timofte, R.—De, V.—Van Gool, L.: Anchored Neighborhood Regression for Fast Example-Based Super-Resolution. Proceedings of the 2013 IEEE International Conference Computer Vision, 2013, pp. 1920–1927, doi: 10.1109/ICCV.2013.241.

[16] Timofte, R.—De Smet, V.—Van Gool, L.: A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution. In: Cremers, D., Reid, I., Saito, H., Yang, M. H. (Eds.): Computer Vision – ACCV 2014. Springer, Cham, Lecture Notes in Computer Science, Vol. 9006, 2014, pp. 111–126, doi: 10.1007/978-3-319-16817-3_8.

[17] Schulter, S.—Leistner, C.—Bischof, H.: Fast and Accurate Image Upscaling with Super-Resolution Forests. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3791–3799, doi: 10.1109/CVPR.2015.7299003.

[18] Dong, C.—Loy, C. C.—He, K.—Tang, X.: Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 38, 2016, No. 2, pp. 295–307, doi: 10.1109/TPAMI.2015.2439281.

[19] Kim, J.—Lee, J. K.—Lee, K. M.: Accurate Image Super-Resolution Using Very Deep Convolutional Networks. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1646–1654, doi: 10.1109/CVPR.2016.182.

[20] Lai, W.-S.—Huang, J.-B.—Ahuja, N.—Yang, M.-H.: Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5835–5843, doi: 10.1109/CVPR.2017.618.

[21] Liu, D.—Wang, Z.—Wen, B.—Yang, J.—Han, W.—Huang, T. S.: Robust Single Image Super-Resolution via Deep Networks with Sparse Prior. IEEE Transactions on Image Processing, Vol. 25, 2016, No. 7, pp. 3194–3207, doi: 10.1109/TIP.2016.2564643.

[22] Yang, C. -Y.—Yang, M.-H.: Fast Direct Super-Resolution by Simple Functions. Proceedings of the 2013 IEEE International Conference on Computer Vision, 2013, pp. 561–568, doi: 10.1109/ICCV.2013.75.

[23] Elad, M.—Yavneh, I.: A Plurality of Sparse Representation Is Better Than the Sparsest One Alone. IEEE Transactions on Information Theory, Vol. 55, 2009, No. 10, pp. 4701–4714, doi: 10.1109/TIT.2009.2027565.

[24] Yang, S.—Liu, Z.—Wang, M.—Sun, F.—Jiao, L.: Multitask Dictionary Learning and Sparse Representation Based Single-Image Super-Resolution Reconstruction. Neurocomputing, Vol. 74, 2011, No. 17, pp. 3193–3203, doi: 10.1016/j.neucom.2011.04.014.

[25] Dong, W.—Zhang, L.—Shi, G.—Wu, X.: Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization. IEEE Transactions on Image Processing, Vol. 20, 2011, No. 7, pp. 1838–1857, doi: 10.1109/TIP.2011.2108306.

[26] Kovesi, P.: Phase Congruency: A Low-Level Image Invariant. Psychological Research, Vol. 64, 2000, No. 2, pp. 136–148, doi: 10.1007/s004260000024.

[27] Irani, M.—Peleg, S.: Improving Resolution by Image Registration. CVGIP: Graphical Models and Image Processing, Vol. 53, 1991, No. 3, pp. 231–239, doi: 10.1016/1049-9652(91)90045-L.

[28] Aharon, M.—Elad, M.—Bruckstein, A.: K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. IEEE Transactions on Signal Processing, Vol. 54, 2006, No. 11, pp. 4311–4322, doi: 10.1109/TSP.2006.881199.

[29] Pati, Y. C.—Rezaiifar, R.—Krishnaprasad, P. S.: Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition. Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers, 1993, pp. 40–44, doi: 10.1109/ACSSC.1993.342465.

[30] Bevilacqua, M.—Roumy, A.—Guillemot, C.—Alberi Morel, M.-L.: Low-Complexity Single-Image Super-Resolution Based on Nonnegative Neighbor Embedding. Proceedings of the British Machine Vision Conference, 2012, pp. 135.1–135.10, doi: 10.5244/C.26.135.

[31] Zhang, L.—Zhang, L.—Mou, X.—Zhang, D.: FSIM: A Feature Similarity Index for Image Quality Assessment. IEEE Transactions on Image Processing, Vol. 20, 2011, No. 8, pp. 2378–2386, doi: 10.1109/TIP.2011.2109730.

[32] Kim, K. I.—Kwon, Y.: Single-Image Super-Resolution Using Sparse Regression and Natural Image Prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, 2010, No. 6, pp. 1127–1133, doi: 10.1109/TPAMI.2010.25.

Xiuxiu Liao received her Ph.D. degree in computer application technology from the South China University of Technology in 2013. Her research interests include image processing, compressive sensing and machine learning. She is currently working as Lecturer in the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China.

Kejia Bai received his Ph.D. in system engineering from the South China University of Technology, Guangzhou, China, in 2009. He is currently Associate Professor in the Guangdong Polytechnic Normal University, Guangzhou, China. His research interests include image processing, pattern recognition, artificial intelligence, super-resolution, and target tracking.

Qian Zhang received her Ph.D. degree in computer application technology from the South China University of Technology in 2013. Her research interests include image processing and deep learning. She is currently working as Lecturer in the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China.

Xiping Jia received his Ph.D. degree in computer application technology from the South China University of Technology in 2008. His research interests include image processing, data mining and machine learning. He is currently working as Associate Professor in the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China.

Shaopeng Liu received his Ph.D. degree from the Sun Yat-Sen University, Guangdong, China, in 2013. Currently, he is Lecturer in the School of Computer Science, Guangdong Polytechnic Normal University, China. His research interests include machine learning, computational intelligence and medical image analysis.

Jin Zhan received her B.Sc. and Ph.D. degrees in computer application technology from the Sun Yat-Sen University in 2004 and 2015, respectively, and she is currently Associate Professor in the School of Computer Sciences, Guangdong Polytechnic Normal University. Her research interests include image processing, video analysis, machine learning and applications.