

Image Super-Resolution via Hierarchical and Collaborative Sparse Representation

Xianming Liu¹, Deming Zhai^{1*}, Debin Zhao¹, Wen Gao^{1,2}

¹School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

²School of Electronic Engineering & Computer Science, Peking University, Beijing, China

Abstract: In this paper, we propose an efficient image super-resolution algorithm based on hierarchical and collaborative sparse representation (HCSR). Motivated by the observation that natural images typically exhibit multi-modal statistics, we propose a hierarchical sparse coding model which includes two layers: the first layer encodes individual patches, and the second layer jointly encodes the set of patches that belong to the same homogeneous subset of image space. We further present a simple alternative to achieve such target by identifying optimal sparse representation that is adaptive to specific statistics of images. Specially, we cluster images from the offline training set into regions of similar geometric structure, and model each region (cluster) by learning adaptive bases describing the patches within that cluster using principal component analysis (PCA). This cluster-specific dictionary is then exploited to optimally estimate the underlying HR pixel values using the idea of collaborative sparse coding, in which the similarity between patches in the same cluster is further considered. It conceptually and computationally remedies the limitation of many existing algorithms based on standard sparse coding, in which patches are independently encoded. Experimental results demonstrate the proposed method appears to be competitive with state-of-the-art algorithms.

I. INTRODUCTION

Image super-resolution, which is the art of rescaling a low-resolution (LR) image to a high-resolution (HR) version, has become a very active area of research in image processing. The interest in image super-resolution is born not only in the great practical importance of enhancing resolution of images, such as in the fields of digital photography, computer vision, computer graphics, medical imaging, satellite remote sensing and consumer electronics, but also the important theoretical value of using image super-resolution to understand the validity of different image models in inverse problems. In the last several years, there has been a great deal of work on image super-resolution. In general, image super-resolution techniques can be categorized into three families: interpolation-based methods [1-5], reconstruction-based methods [6][7], and learning-based methods [8][9].

In recent years, the sparse coding based modeling has been proven to be a promising tool for signal representation. It assumes that a signal can be efficiently represented by a sparse linear combination of atoms from a given or learned dictionary. Many sparse based image super-resolution algorithms have been proposed in the literature. Yang *et al.* [10] proposed a joint dictionary training method to learn the dictionaries for high- and low-resolution image patch spaces, and enforce that sparse representations between the

*The corresponding author: dmzhai@jdl.ac.cn

low resolution and high resolution image patch pair with respect to their own dictionaries should be the same. Wang *et al.* [11] proposed to relax this strong regularization of “same sparse representation” by using a mapping function for cross-style image synthesis, and achieved better results than Yang’s work. These methods essentially concatenate the two feature spaces and convert the problem to the standard sparse coding in a single feature space.

One major problem of standard sparse coding is that patches are encoded independently. Therefore, similar patches sometimes admit very different estimates due to the potential instability of sparse decompositions, which can result in practice in noticeable reconstruction artifacts. Some researchers in image super-resolution and other image processing field have found this problem, and proposed some solutions. Mairal *et al.* [12] proposed to combine the non-local means and sparse coding approaches to image restoration into a unified framework where similar patches are decomposed using similar sparsity patterns. Dong *et al.* [13] proposed a centralized sparse representation framework, in which the sparse coding coefficients are forced to be close to their mean values. These methods achieve promising performance for their respective image processing tasks.

In this paper, we propose an alternative solution for addressing the limitation of standard sparse coding in image super-resolution, and present an efficient framework based on hierarchical and collaborative sparse representation (HCSR). Motivated by the observation that natural images typically exhibit multi-modal statistics, we propose a hierarchical sparse coding framework which includes two layers: the first layer encodes individual patches, and the second layer then jointly encodes the set of patches that belong to the same homogeneous subset of image space. We further present a simple solution to achieve such target by identifying optimal sparse representation that is adaptive to the specific statistics of images. Specially, we cluster images from the offline training set into regions of similar geometric structure, and model each region (cluster) by learning an adaptive dictionary describing the patches within that cluster using principal component analysis (PCA). This cluster-specific dictionary is then exploited to optimally estimate the underlying HR pixel values using the framework of collaborative sparse coding, in which the similarity between patches in the same cluster is further considered. Experimental results demonstrate the proposed method appears to be competitive with state-of-the-art algorithms.

The rest of the paper is organized as follows: Section II gives some prior knowledge and introduces the proposed framework. Section III details dictionary learning and optimization for learning sparse representation coefficients. Section IV presents some experimental results and comparative studies. Section V concludes the paper.

II. HIERARCHICAL AND COLLABORATIVE SPARSE CODING

In this section, we will first review the standard sparse coding, which is used for most of existing sparse representation based image super-resolution algorithms. Then, we will introduce the proposed hierarchical and collaborative sparse coding framework.

A. Standard Sparse Coding

Denote $\mathbf{x} \in \mathfrak{R}^{\sqrt{N} \times \sqrt{N}}$ as the image at hand. We divide the image into a set of overlapping blocks with size $\sqrt{d} \times \sqrt{d}$. Such a procedure can be represented as $\mathbf{x}_i = \mathbf{R}_i \mathbf{x}$, where \mathbf{R}_i is the matrix extracting patch \mathbf{x}_i from \mathbf{x} at location i . Each block is then stacked into a vector. In this way, the image \mathbf{x} can be viewed as a collection of vectors

$\{\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$ in a high-dimensional space. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ be the matrix form of the patches set. Let $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_p] \in \mathbb{R}^{d \times p}$ be the dictionary matrix, where each \mathbf{d}_i represents a basis vector in the dictionary. Each sample in \mathbf{X} can be represented as a linear combination of atoms in the dictionary \mathbf{D} plus some perturbation ε , that is, $\mathbf{x}_i = \mathbf{D}\mathbf{a}_i + \varepsilon_i$, $\mathbf{a}_i \in \mathbb{R}^{p \times 1}$. We say that the model is sparse if we can achieve $\|\varepsilon_i\|_2 \ll \|\mathbf{x}_i\|_2$ and $\|\mathbf{a}_i\|_0 \ll d$ simultaneously for all or most $i = 1, \dots, n$.

Seeking the sparsest representation is known to be a NP-hard problem. In order to obtain efficient algorithms, people usually relax the nonconvex ℓ_0 norm to the convex ℓ_1 norm, leading to the following form:

$$\min_{\mathbf{D}, \{\mathbf{a}_1, \dots, \mathbf{a}_n\}} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|^2 + \lambda \sum_{i=1}^n \|\mathbf{a}_i\|_1 \quad (1)$$

where λ is a regularization parameter that controls the tradeoff between the quality of the reconstruction and the sparsity. The above approximation is known as the Lasso [20].

The entire image \mathbf{x} can be sparsely represented by the concatenation of all of sparse codes $\{\mathbf{a}_i\}_{i=1}^n$. We can obtain a very redundant patch-based image representation. Reconstructing the original image from the sparse codes becomes an over-determined system, and a straightforward least-square solution is shown as follows [13][19]:

$$\mathbf{x} \approx \mathbf{D} \circ \mathbf{a} \triangleq \left(\sum_{i=1}^n \mathbf{R}_i^T \mathbf{R}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{R}_i^T \mathbf{D}\mathbf{a}_i \right). \quad (2)$$

where \mathbf{a} is the concatenation of all of sparse codes.

B. Hierarchical Sparse Coding

The basic model of standard sparse coding described above has two major limitations. First, the model can only capture statistical relationships among pixels in each patch, but does not provide any way to capture higher-order relationship that cannot be simply described at the pixel level. Second, the model encodes local patches independently and fails to consider the geometrical structure of the image data space. In this paper, we extend the basic model to a two-layer scheme to model higher-order statistical relationship among patches. The first layer called the patch level encodes individual patches, which is carried out in the same way as standard sparse coding; the second layer called the cluster level encodes jointly the set of patches that belong to the same cluster. For the second layer, we introduce another non-negative dictionary $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q] \in \mathbb{R}^{p \times q}$ to model the statistical correlation among the sparse representation coefficients of local patches in the first layer, that is, $\mathbf{a}_i = \mathbf{U}\mathbf{b}_i + \varsigma_i$, $\mathbf{b}_i \in \mathbb{R}^{q \times 1}$.

The first layer represents the local correlation among pixels within each patch. The second layer packs together similar patches, which may disjoint with each other, to encourage that similar patches share similar sparse decompositions. Therefore, the proposed hierarchical sparse coding framework can be viewed as pooling both local and nonlocal information for robust and accurate signal reconstruction.

We can obtain sparse representations at patch-level and cluster-level simultaneously by using the following unified optimization framework:

$$\begin{aligned} \{\mathbf{a}^*, \mathbf{b}^*\} &= \arg \min_{\mathbf{a}, \mathbf{b}} \sum_{i=1}^n \left\{ \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|^2 + \lambda_1 \|\mathbf{a}_i\|_1 + \lambda_2 \mathbf{a}_i^T f(\mathbf{b}) \mathbf{a}_i \right\} + \gamma \|\mathbf{b}\|_1, \\ \text{s.t. } &\mathbf{b} \geq 0, \end{aligned} \quad (3)$$

where \mathbf{a} and \mathbf{b} are the sparse representation coefficient matrixes for patch-level and cluster-level, respectively. The ℓ_1 norm on each \mathbf{a}_i and on \mathbf{b} encourages sparsity on the representation at both levels. And $f(\mathbf{b})$ is defined as the following inverse diagonal covariance:

$$f(\mathbf{b}) = \left(\sum_{j=1}^q \mathbf{b}_j \text{diag}(\mathbf{a}_j) \right)^{-1}, \quad (4)$$

which is affine in \mathbf{a} .

The above minimization problem can be solved by iteratively alternating optimization: 1) When two dictionaries \mathbf{D} and \mathbf{U} are given, we compute the optimal sparse representation coefficients \mathbf{a} and \mathbf{b} for two levels. This optimization problem is jointly convex in both \mathbf{a} and \mathbf{b} . Therefore, it is convenient to use an alternating optimization procedure to accurately compute the solution by iteratively optimizing \mathbf{a} when \mathbf{b} is fixed, and then optimizing \mathbf{b} when \mathbf{a} is fixed. 2) When \mathbf{a} and \mathbf{b} are derived, we compute the optimal dictionaries \mathbf{D} and \mathbf{U} . In this procedure, \mathbf{D} and \mathbf{U} can be optimized independently.

The hierarchical sparse coding scheme presented above is conceptually appealing. However, its optimization is too complex. Naturally, we have a question: can we find an alternative to achieve the same target but in a simple manner? Let us concatenate the sparse models of patch-level and cluster-level, and we can get the following result:

$$\begin{cases} \mathbf{x}_i = \mathbf{D}\mathbf{a}_i + \varepsilon_i \\ \mathbf{a}_i = \mathbf{U}\mathbf{b}_i + \varsigma_i \end{cases} \Rightarrow \mathbf{x}_i = \mathbf{D}\mathbf{U}\mathbf{b}_i + \eta_i \Rightarrow \mathbf{x}_i = \mathbf{\Phi}\mathbf{b}_i + \eta_i \quad (5)$$

where $\mathbf{\Phi} = \mathbf{D}\mathbf{U} \in \mathbb{R}^{d \times q}$ and $\eta_i = \mathbf{D}\varsigma_i + \varepsilon_i$. The above result is interesting. It tells us in fact two layer sparse coding scheme can be implemented by defining only one dictionary that reflects the geometrical structure of natural images. In this way, we can greatly reduce the complexity of optimization. In the following, we will show how to achieve the target of hierarchical sparse coding by combining the idea of clustering and collaborative sparse coding.

C. Collaborative Sparse Coding

A common observation is that natural images typically exhibit multi-modal statistics as they usually contain many heterogeneous regions with significantly different geometric structures or statistical characteristics [14]. Heterogeneous data can be better represented using a mixture of sparse models, one for each homogeneous subset. Bases for each model are adaptive to the particular homogeneous subset. In this paper, we propose to represent an image by a collection of subspaces, one subspace for each different segment (cluster) of the image. The dimension and basis of each subspace will be chosen adaptively according to the variability and correlation of the data in the corresponding image segment.

In practical design, we borrow the idea of clustering to segment the image into different subspace. Specially, we cluster patches collected from the offline training set into regions with similar geometric structure, and model each region (cluster) by learning a compact sub-dictionary describing the patches within that cluster. These sub-dictionaries are combined together to form an over-complete dictionary. In the test procedure, for each patch to be coded, we adaptively select one sub-dictionary from the trained sub-dictionaries to code it. In practical implementation, we use high-pass results of each patch as features and exploit the simple K-means for clustering.

Formally, for the dictionary Φ which has q atoms, we define groups of atoms through their indices, $G \subseteq \{1, \dots, q\}$. Denote $\mathcal{G} = \{G_1, \dots, G_m\}$ as a partition of $\{1, \dots, q\}$, where m is the number of clusters. For a certain cluster c which includes n_c image patches to be coded, the best sub-dictionary Φ_{G_c} that is most relevant to the given patches is selected. The optimal sparse representation for the cluster c can be obtained through addressing the following optimization problem:

$$\min_{\{\mathbf{s}_1, \dots, \mathbf{s}_k\}} \sum_{i=1}^k \|\mathbf{x}_i - \Phi_{G_c} \mathbf{s}_i\|^2 + \lambda \sum_{i=1}^k \|\mathbf{s}_i\|_1. \quad (6)$$

where k is the sample number in the current cluster c .

Although the procedure of clustering collects patches with similar structure, it cannot ensure patches in the same cluster are with the same geometrical structure. Therefore, the difference of within-class still exists. Moreover, in standard sparse coding, we found similar patches can be encoded as totally different sparse codes, which may bring about the loss of the locality information of the patches to be encoded. To address these problem, we explicitly introduce a regularization term into the optimization problem to preserve the consistency of sparse codes for the similar local patches:

$$\min_S \sum_{i=1}^k \|\mathbf{x}_i - \Phi_{G_c} \mathbf{s}_i\|^2 + \lambda \sum_{i=1}^k \|\mathbf{s}_i\|_1 + \gamma \sum_{i=1}^k \sum_{j=1}^k \|\mathbf{s}_i - \mathbf{s}_j\|^2 W_{ij}, \quad (7)$$

where $S = \{\mathbf{s}_1, \dots, \mathbf{s}_k\}$, W_{ij} measures the similarity between a patches pair $(\mathbf{x}_i, \mathbf{x}_j)$, which is defined as:

$$W_{ij} = \exp \left\{ -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_s^2} \right\}, \quad \sigma_s > 0. \quad (8)$$

We further define the degree of \mathbf{x}_i as $v_i = \sum_{j=1}^k W_{ij}$, and define the degree matrix $V = \text{diag}(v_1, \dots, v_l)$. Define $L = V - W$ as the Laplacian matrix. The regularization term can be written as:

$$\sum_{i=1}^k \sum_{j=1}^k \|\mathbf{s}_i - \mathbf{s}_j\|^2 W_{ij} = \text{Tr}(\mathbf{S} \mathbf{L} \mathbf{S}^T) \quad (9)$$

which can be further written as:

$$\text{Tr}(\mathbf{S} \mathbf{L} \mathbf{S}^T) = \text{Tr} \left(\sum_{i=1}^k \sum_{j=1}^k L_{ij} \mathbf{s}_i \mathbf{s}_j^T \right) = \sum_{i=1}^k \sum_{j=1}^k L_{ij} \mathbf{s}_i^T \mathbf{s}_j \quad (10)$$

where \mathbf{S} is the matrix using $\{\mathbf{s}_i\}_{i=1}^k$ as columns.

With all above definition, the optimization problem can be reformulated as follows:

$$\min_{S=\{\mathbf{s}_1, \dots, \mathbf{s}_k\}} \sum_{i=1}^k \|\mathbf{x}_i - \Phi_{G_c} \mathbf{s}_i\|^2 + \lambda \sum_{i=1}^k \|\mathbf{s}_i\|_1 + \gamma \sum_{i=1}^k \sum_{j=1}^k L_{ij} \mathbf{s}_i^T \mathbf{s}_j. \quad (11)$$

By incorporating the Laplacian regularizer into the standard sparse coding, the sparse codes for local patches are no longer independent. After obtaining the sparse codes, we can use Eq.(2) to reconstruct the original image patches. In the next section, we will show how to obtain an adaptive sub-dictionary and optimal sparse representation coefficients.

III. DICTIONARY LEARNING AND OPTIMIZATION SOLUTION

A. Adaptive Dictionary Learning

Dictionary learning is one of the most important issues in sparsity-based image super-resolution. In the proposed method, we cluster patches collected from the offline training set into regions with similar geometric structure, and model each region (cluster) by learning a compact sub-dictionary. Since the patches in a cluster are similar to each other, the sub-dictionary is no need to be over-complete, but all sub-dictionaries are combined to construct a large over-complete dictionary to characterize all the possible local structures of natural images.

More specially, for a certain cluster c which includes n_c image patches to be coded, we stack the vectors of patches into a matrix denoted by \mathbf{X}_c . Then, we learn a adaptive sub-dictionary Φ_{G_c} that is most relevant to $\mathbf{X}_c \in R^{d \times k}$ by applying principal component analysis (PCA) on \mathbf{X}_c . This is in the same way as [13].

B. Optimization for Sparse Representation Coefficients

In the subsection, we will introduce an optimization method based upon coordinate descent to solve the above optimization problem formulated in Eq.(11). Several approaches have been proposed to solve the problem of this form [16-18].

Since the dictionary Φ_{G_c} is determined previously, it is easy to see the objective function is convex, therefore, we can achieve a global minimum. Instead of optimizing the whole sparse codes matrix S , we optimize each code s_i individually while holding all the remaining sparse representation codes $s_j (j \neq i)$ fixed. When optimizing s_i , we can get the following optimization problem [16]:

$$\begin{aligned} \min_{\mathbf{s}_i} f(\mathbf{s}_i) &= \min_{\mathbf{s}_i} J(\mathbf{s}_i) + \lambda \|\mathbf{s}_i\|_1 \\ &= \min_{\mathbf{s}_i} \|\mathbf{x}_i - \Phi_{G_c} \mathbf{s}_i\|^2 + \gamma L_{ij} \mathbf{s}_i^T \mathbf{s}_j + \mathbf{s}_i^T \mathbf{h}_i + \lambda \sum_{j=1}^q |\mathbf{s}_i^{(j)}| \end{aligned} \quad (12)$$

where $\mathbf{h}_i = 2\gamma(\sum_{j \neq i} L_{ij} \mathbf{s}_j)$, $\mathbf{s}_i^{(j)}$ is the j -th coefficients of s_i .

In the practical implementation, we use the feature-sign search algorithm to solve s_i . And we initialize the sparse codes with the results of standard sparse coding in order to speed up the convergence of sparse codes. The detailed algorithm flow for learning optimal sparse codes are shown in **Algorithm 1**.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, experimental results are presented to demonstrate the performance of the proposed algorithm. For thoroughness and fairness of our comparison study, we exploit some widely used images as test ones. Fig. 1 lists the used six sample images in our experiments. Our algorithm is compared with some representative work in the literature. More specifically, five methods are included in our comparative study: (1) Bicubic interpolation [1]; (2) the sparse representation based method of [10], denoted as CDL; (3) the semi-coupled dictionary learning based method of [11], denoted as SCDL; (4) the centralized sparse representation based method of [13], denoted as CSR; (5) the proposed HCSR method.

In our experiments, the observed low-resolution (LR) image is obtained by first blurring with a blur kernel and then downsampling by a scaling factor, from which the original HR images are reconstructed by the proposed and competing methods. A 6×6 Gaussian

Algorithm 1 Feature-sign Search Algorithm for Learning Sparse coefficients

Input:

- A cluster with n_c image patches $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$;
- The subdictionary Φ_{G_c} , the Laplacian matrix L ;
- The parameters λ and γ ;

Output:

- The optimal sparse code matrix S^* .

Procedure:**For all** i **do****Initialize Step:**

- 1: $\mathbf{s}_i = \vec{0}$, active set $\mathcal{A} = \emptyset$, $\boldsymbol{\theta} = \vec{0}$, where $\theta_j \in \{-1, 0, 1\}$ denotes $\text{sign}(\mathbf{s}_i^{(j)})$;

Activate Step:

- 2: From zero coefficient of \mathbf{s}_i , select $j = \arg \max_j \left| \nabla_i^{(j)} J(\mathbf{s}_i) \right|$. Activate $\mathbf{s}_i^{(j)}$ only if it locally improve the objective function Eq.(13), namely:
 - If $\left| \nabla_i^{(j)} J(\mathbf{s}_i) \right| > \lambda$, then set $\theta_j = -1$, $\mathcal{A} = \{j\} \cup \mathcal{A}$
 - If $\left| \nabla_i^{(j)} J(\mathbf{s}_i) \right| < -\lambda$, then set $\theta_j = 1$, $\mathcal{A} = \{j\} \cup \mathcal{A}$.

Feature-sign Step:

- 3: Let $\hat{\Phi}_{G_c}$ be a submatrix of Φ_{G_c} that only contains columns corresponding to the active set. Let $\hat{\mathbf{s}}_i$ and $\hat{\mathbf{h}}_i$ be subvectors of \mathbf{s}_i and \mathbf{h}_i . Let $\hat{\boldsymbol{\theta}}$ be $\boldsymbol{\theta}$ corresponding to the active set
- 4: Compute the optimal \mathbf{s}_i under the current active set:

$$\hat{\mathbf{s}}_i^{new} = \left(\hat{\Phi}_{G_c}^T \hat{\Phi}_{G_c} + \gamma L_{ii} \mathbf{I} \right)^{-1} \left(\hat{\Phi}_{G_c}^T \mathbf{x}_i - (\lambda \hat{\boldsymbol{\theta}} + \hat{\mathbf{h}}_i) / 2 \right) \quad (13)$$

where \mathbf{I} is the identity matrix.

- 5: Perform a discrete line search on the closed line segment from $\hat{\mathbf{s}}_i$ to $\hat{\mathbf{s}}_i^{new}$: Check the objective value at $\hat{\mathbf{s}}_i^{new}$ and all points where any coefficient changes sign, and update $\hat{\mathbf{s}}_i$ to the point with lowest objective value.

Check the Optimality Conditions Step:

- 6: Condition (a): Check optimality condition for nonzero coefficients: $\left| \nabla_i^{(j)} J(\mathbf{s}_i) \right| + \lambda \text{sign}(\mathbf{s}_i^{(j)}) = 0, \forall \mathbf{s}_i^{(j)} \neq 0$. If condition (a) is satisfied, go to feature-sign step; else check condition (b).
- 7: Condition (b): Check optimality condition for zero coefficients: $\left| \nabla_i^{(j)} J(\mathbf{s}_i) \right| \leq \lambda, \forall \mathbf{s}_i^{(j)} = 0$. If condition (b) is satisfied, go to activate step; else return \mathbf{s}_i as the solution denoted as \mathbf{s}_i^* .

End for

filter with standard deviation of 1.5 is used for blurring, and then downsampling the blurred image by a scaling factor of 3 in both horizontal and vertical directions. Since the original HR images are known in the simulation, we can compare the interpolated results with the true images, and measure the objective and subjective quality of those interpolated images. In the practical experiments, we initialize the HR image \mathbf{x} using the result of Bicubic interpolation.

In our experiments, five training images are selected from the Kodak PhotoCD dataset,



Figure 1. Six sample images in the test set

which are different from the testing images illustrated in Fig.1. All compared methods share the same offline training set. We collect patches from the five training images. The patch size in our implementation is set to 6×6 . Pre-clustering is conducted and the cluster number is set to be 64. The regularization parameters λ and γ are determined by cross-validity. σ_s is set to be 0.5 for computing similarity between patches.

Table I tabulates the quantitative quality comparison with respect to PSNR and SSIM of the five compared methods when applied to the six test images of Fig. 1. It can be observed that for all instances the proposed algorithm consistently works better on PSNR and SSIM than other methods. Compared with interpolation based methods, such as Bicubic, the proposed method can significantly improve the objective quality of generated HR images. The average PSNR gain is 2.66dB. Through the proposed hierarchical sparse coding scheme, our method can efficiently explore the local and nonlocal correlation among patches in images. By further combining the idea of collaborative sparse coding, our method can efficiently reduce the potential instability of sparse decompositions, which often happens in standard sparse coding based methods since patches are encoded individually. Therefore, our method also outperforms the standard sparse coding based methods, such as CDL and SCDL. Compared with these two methods, the average PSNR gains are 2.85dB and 2.1dB, respectively. Compared with the CSR method, which is another solution to address the problem of instability in standard sparse coding, the proposed method works better and the average PSNR gain can be improved 0.19dB.

With respect to SSIM, it can be seen the proposed HCSR method achieves the highest SSIM scores among all of the competing methods for all test images. This demonstrates our method can reconstruct the structures of images better. Given the fact that human visual system (HVS) is the ultimate receiver of the restored images, we also show the subjective comparison results. The generated HR images of compared methods for *Butterfly* are illustrated in Fig. 2. Due to the space limitation, we omit the result of Bicubic method here. It can be clearly observed that the image reconstructed by the CDL method suffers from annoying ringing artifacts. The SCDL method produces clearer edges, however, which suffers from irregular outliers along edges and textures. Our method achieves competitive subjective quality with the CSR method, the produced edges in our method are clean, sharp and consistent. Both the superior subjective and objective qualities on test images convincingly demonstrate the potential of the proposed hierarchical and collaborative sparse coding scheme on image super-resolution.

V. CONCLUSION

In this paper, we presented an efficient image super-resolution algorithm based on hierarchical and collaborative sparse representation (HCSR). Through the proposed hierarchical sparse coding scheme, our method can efficiently explore the local and nonlocal correlation among patches in images. By further combining the idea of collaborative sparse coding, our method can efficiently reduce the potential instability of sparse decom-

Table I
 QUANTITATIVE COMPARISON OF FIVE IMAGE SUPER-RESOLUTION ALGORITHMS ON PSNR (DB) AND SSIM

Method	Bicubic		CDL		SCDL		CSR		HCSR	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
<i>Butterfly</i>	22.77	0.8241	23.08	0.8003	23.34	0.8628	26.87	0.9222	27.37	0.9286
<i>Bike</i>	21.58	0.6853	21.26	0.6998	22.25	0.7279	23.38	0.7868	23.55	0.7949
<i>Flowers</i>	26.18	0.7746	25.81	0.7859	26.69	0.7915	28.31	0.8494	28.42	0.8531
<i>Hat</i>	27.61	0.8061	27.50	0.8201	28.33	0.8168	29.77	0.8576	29.91	0.8596
<i>Plants</i>	27.36	0.8372	27.39	0.8447	28.04	0.8518	30.73	0.9057	30.92	0.9072
<i>Parthenon</i>	24.64	0.6945	23.92	0.6671	24.85	0.7083	25.87	0.7487	25.91	0.7502
Average	25.02	0.7703	24.83	0.7697	25.58	0.7932	27.49	0.8451	27.68	0.8489

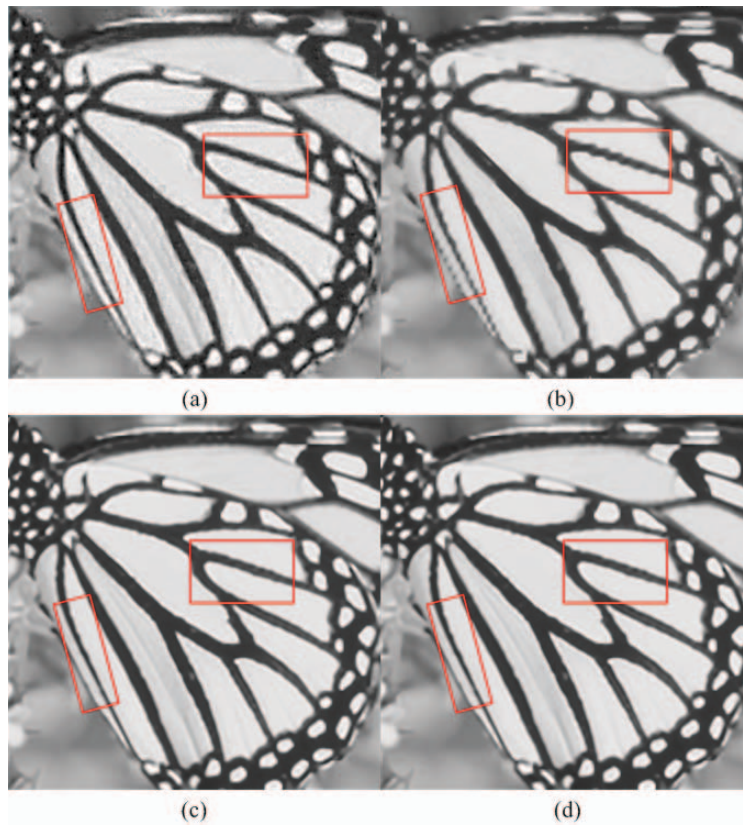


Figure 2. Subjective quality comparison on *Butterfly*: (a) CDL, PSNR: 23.08dB, SSIM: 0.8018; (b) SCDL, PSNR: 23.49dB, SSIM: 0.8628, (c) CSR, PSNR: 26.92dB SSIM: 0.9219, (d) HCSR, PSNR: 27.42dB SSIM: 0.9283.

positions, which often happens in standard sparse coding based methods since patches are encoded independently. Experimental results demonstrate the proposed method appears to be competitive with state-of-the-art algorithms.

VI. ACKNOWLEDGEMENT

This work was supported in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2009CB320905 and by the National

REFERENCES

- [1] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153-1160, Dec. 1981.
- [2] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao and H. Sun, "Transductive Regression with Local and Global Consistency for Image Super-Resolution," in *Proceedings of IEEE Data Compression Conference, DCC2011*, Snowbird, Utah, USA, Mar.29-31, 2011.
- [3] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao and H. Sun, "Image Interpolation via Regularized Local Linear Regression," *IEEE Transaction on Image Processing*, Vol. 20, No. 12, pp. 3455-3469, Dec. 2011.
- [4] X. Zhang, X. Wu, "Image Interpolation by Adaptive 2-D Autoregressive Modeling and Soft-Decision Estimation," *IEEE Transactions on Image Processing*, vol.17, No. 6, pp.887-896, 2008.
- [5] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel Regression for Image Processing and Reconstruction," *IEEE Transactions on Image Processing*, vol. 16, No. 2, pp. 349-366, February 2007.
- [6] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos, "SoftCuts: A Soft Edge Smoothness Prior for Color Image Super-Resolution," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 969-981, May 2009.
- [7] W. Dong, L. Zhang, G. Shi, and X. Wu, "Nonlocal back-projection for adaptive image enlargement," in *Proc. IEEE International Conference on Image Processing*, Oct. 2009.
- [8] W. Freeman, E. Pasztor, and O. Carmichael, "Learning Low-level Vision," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 25-47, 2000.
- [9] K. 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.
- [10] 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.
- [11] S. Wang, L. Zhang, Y. Liang and Q. Pan, "Semi-Coupled Dictionary Learning with Applications to Image Super-Resolution and Photo-Sketch Image Synthesis," in *CVPR 2012*.
- [12] J. Mairal, F. Bach, J. Ponce, G. Sapiro and A. Zisserman, "Non-Local Sparse Models for Image Restoration," in *ICCV 2009*.
- [13] W. Dong, L. Zhang and G. Shi, "Centralized Sparse Representation for Image Restoration," in *ICCV 2011*.
- [14] W. Hong, J. Wright, K. Huang, and Y. Ma, "Multi-Scale Hybrid Linear Models for Lossy Image Representation," *IEEE. Trans. Image Process.*, vol. 15, no. 12, pages 3655-3671, Dec 2006.
- [15] P. Chatterjee and P. Milanfar, "Clustering-based Denoising with Locally Learned Dictionaries," *IEEE Trans. Image Processing* vol. 18, no. 7, pp. 1438-1451, July 2009.
- [16] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," In *NIPS*, Vancouver, British Columbia, Canada, December 4-7, 2006, pages 801-808.
- [17] S. Gao, I. Tsang, L. Chia and P. Zhao, "Laplacian sparse coding, Hypergraph Laplacian sparse Coding, and applications," To appear in *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012.
- [18] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu and D. Cai, "Graph Regularized Sparse Coding for Image Representation," *IEEE. Trans. Image Process.*, vol. 20, no.5, 1327-1336, 2011.
- [19] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, 15(12):3736-3745, Dec. 2006.
- [20] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, Blackwell Publishing for the Royal Statistical Society, 1996, 58, pp. 267-288.