Geometry Constrained Sparse Coding for Single Image Super-resolution

Xiaoqiang Lu, Haoliang Yuan, Pingkun Yan, Yuan Yuan, Xuelong Li
Center for OPTical IMagery Analysis and Learning (OPTIMAL),
State Key Laboratory of Transient Optics and Photonics,
Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences
Xi'an 710119, Shaanxi, P. R. China
{luxq666666, pingkun.yan, y.yuan1, xuelong_li}@opt.ac.cn

Abstract

The choice of the over-complete dictionary that sparsely represents data is of prime importance for sparse coding-based image super-resolution. Sparse coding is a typical unsupervised learning method to generate an over-complete dictionary. However, most of the sparse coding methods for image super-resolution fail to simultaneously consider the geometrical structure of the dictionary and corresponding coefficients, which may result in noticeable super-resolution reconstruction artifacts. In this paper, a novel sparse coding method is proposed to preserve the geometrical structure of the dictionary and the sparse coefficients of the data. Moreover, the proposed method can preserve the incoherence of dictionary entries, which is critical for sparse representation. Inspired by the development on non-local self-similarity and manifold learning, the proposed sparse coding method can provide the sparse coefficients and learned dictionary from a new perspective, which have both reconstruction and discrimination properties to enhance the learning performance. Extensive experimental results on image super-resolution have demonstrated the effectiveness of the proposed method.

1. Introduction

In visual information processing, in order to obtain more information about an image, *high-resolution* (HR) images are always desired [14]. Unfortunately, due to the physical limitation of relevant imaging devices, it is difficult to acquire HR images in some cases. Hence, when physical manners cannot afford high spatial resolution image from an underlying scene, the signal processing methods have to be selected to restore potential information hidden in the source. For this purpose, a great number of attempts, named image *super-resolution* (SR) reconstruction methods [6, 15], have been proposed.

Generally, image SR methods can be classified into three categories: interpolation-based [21], multi-image-based [15], and learning-based [5, 6, 9, 17, 19] methods.

Interpolation-based SR can be simply achieved with conventional interpolation algorithms, such as bilinear, bicubic or other resampling methods [21]. For interpolation methods, much denser pixels in the HR grids can be achieved by applying a base function to obtain the final SR image. However, simple and effective interpolation-based methods are usually prone to yield overly smooth images as the magnification becomes large. The image SR capability of this kind of methods is thus very limited.

Multi-image-based methods can be specialized for multiple *low-resolution* (LR) images [15]. The precondition of this kind of methods needs adequate LR images with sub-pixel shifts. However, these methods are still limited to small increase in spatial resolution [1].

The limitations aforementioned have been broken by learning-based methods [6, 5, 19]. This kind of methods presume that the high-frequency details lost in a LR image can be predicted by learning from a specified training data base. Freeman *et al*. [6] address image SR as the problem of predicting the input LR image into the desired scale. Then, the nearest neighbor (NN) based estimation of high-frequency image is performed and a Markov network is exploited to resolve the compatibility of output patches[10, 7]. A manifold assumption proposed by Chang *et al*. [5] points out that the manifolds of LR image patches and the corresponding HR ones are located in two similar local geometries in their respective feature space. According to this assumption, *neighbor embedding* (NE) [5] partly utilizing *locally linear embedding* (LLE) [16] is proposed to estimate HR image patch by linearly combining these HR counterparts of LR image patches found in the training database. Recently, sparse coding methods [18, 19] are employed to perform image SR. In their works, by enforcing ℓ_1-norm sparsity prior regularization, LR image patches are coded with respect to an over-complete dictionary, and a sparse

vector is obtained to linearly combine corresponding HR counterparts to perform image SR reconstruction.

This paper mainly focuses on studying the sparse coding-based methods for image SR [18, 19]. It can be seen in [5] that learning performance and better reconstruction results can be obtained by preserving local topology structure in data space. Recently, although some sparse coding methods consider the geometrical structure of sparse representation [20, 22], they fail to consider the incoherence of dictionary entries. In fact, the reconstruction results depend on the incoherence between dictionary bases [11]. Hence, it is important to preserve the incoherence of the dictionary when we consider the geometrical structure of data. In this paper, we simultaneously consider the geometrical structure of the dictionary and the corresponding coefficients. In this case, the intrinsic geometrical structure of the data can be captured and the incoherence between dictionary bases can be preserved in sparse decomposition.

Motivated by the recent development in non-local self-similarity and manifold learning, this paper proposes a novel sparse coding method, which simultaneously considers the geometrical structure of the learned dictionary and the corresponding sparse coefficients to capture the intrinsic geometrical structure of the data. The proposed method is a two-step method including computing sparse coefficients and dictionary learning. Given an input data, in the first step, based on the idea of the non-local self-similarity, the proposed method learns the sparse coefficients by taking advantage of the intrinsic geometric detail of the data. In the second step, the intrinsic geometric structure of the cluster centers of the data can be exploited to encode the semantic structure of the basis vectors by using spectral graph techniques. Each cluster centroid denotes the representative mode of the corresponding cluster and the cluster centers are incoherent. The relationship between dictionary entries are determined by mapping the weighted graph of cluster centroids. Thus, the incoherence of dictionary entries can be preserved accordingly. Moreover, the learned dictionary can change smoothly along the geodesics of the manifold. Extensive experimental results on image super-resolution have demonstrated the effectiveness of the proposed method.

The rest of this paper is organized as follows. Section 2 provides a brief review of the original sparse coding method. In Section 3, the proposed work is introduced, as well as the optimization scheme, including learning sparse coefficients and learning the dictionary. The experimental results on image super-resolution and the comparison with other methods are presented in Section 4. Finally, Section 5 concludes this paper.

2. Original Sparse Coding

In this section, the original sparse coding method is briefly reviewed.

Let $X = [x_1, \ldots, x_n] \in \mathbf{R}^{m \times n}$ be the data matrix. Let $D = [d_1, \ldots, d_k] \in \mathbf{R}^{m \times k}$ be the dictionary matrix, where each d_i represents a basis vector in the dictionary. Let $S = [s_1, \ldots, s_n] \in \mathbf{R}^{k \times n}$ be the coefficient matrix, where each s_i represents a sparse coefficient for a data point x_i. Sparse coding aims to make each data point x_i approximate by a sparse combination of those basis elements of the dictionary. The coefficient matrix together with dictionary should best approximate X, which can be represented as $\min_{D,S} \|X - DS\|$. The objective function of sparse coding can be defined as:

$$\min_{D,S} \|X - DS\|^2 + \lambda \|S\|_1$$
$$\text{s.t.} \quad \|d_i\|^2 \leq c, i = 1, 2, \ldots, k, \qquad (1)$$

where λ is the regularization parameter, the term $\|S\|_1$ is to enforce sparsity, and the term $\|d_i\|^2$ constraint on d_i removes the scaling ambiguity.

Although Eq.1 is not convex in both D and S, it is convex in fixing D only or fixing S only. Hence, borrowing the idea of [12, 22], an iterative theme is adopted to minimize the Eq.1 on one variable while fixing the other one.

3. Proposed Work

In this section, a novel sparse coding method is proposed to take advantage of the geometrical structure of the data space. The proposed sparse coding method has two features. Firstly, the learned dictionary and the coefficients can both capture the intrinsic geometrical structure of the data. Secondly, the coefficient for the corresponding dictionary is sparse and the incoherence of dictionary entries can be preserved. In the following, the proposed sparse coding method is adopted for image SR.

3.1. Objective Function

Recently, researches have shown that the geometrical structure of the data can improve the performance of sparse coding [4, 8, 22]. However, these sparse coding methods fail to simultaneously consider the geometrical structure of the data and the incoherence of learned dictionary entries. In this paper, a novel method is proposed to simultaneously consider the geometrical structure of the dictionary and the corresponding coefficients. The proposed method can preserve the incoherence of the dictionary when we consider the geometrical structure of the learned dictionary and the corresponding sparse coefficients to capture the intrinsic geometrical structure of the data.the proposed sparse coding method can be divided into two steps to preserve the geometrical structure of data.

Firstly, one important prior is that the data matrix often contains repetitive structures and patterns [3]. Because sparse decomposition has the potential instability, similar data often have different estimates, which results in noticeable reconstruction artifacts [13]. In this case, it is necessary to exploit the property of non-local self-similarity to stabilize the sparse decompositions. Hence, an assumption can be generated that if a given m-dimensional data point x_j is the j-th most similar (vectorized) data point to x_i in a non-local neighborhood, then the corresponding coefficient s_j is also the j-th most similar (vectorized) coefficients for s_i in a non-local neighborhood. This assumption is represented as non-local self-similar property, which has been successfully applied into image denoising [3]. Based on the above assumption, a non-local self-similarity quadratic constraint is defined as:

$$\sum_{i=1}^{n} \|s_i - \sum_j w_{ji} s_j\|^2, \qquad (2)$$

where w_{ji} is the weight assigned to s_j and s_i represents a sparse coefficient for a data point x_i. Given a set of m-dimensional data points $X = [x_1, \ldots, x_n]$, the data point x_j is selected if it is within the first K ($K = 5$ in our experiments) closest data points to x_i. The weight w_{ji} can be denoted as:

$$w_{ji} = \frac{1}{c_i} \cdot e^{-\frac{\|x_i - x_j\|^2}{h}}, \qquad (3)$$

where h is a parameter enforcing the similarity and c_i is the normalization factor. Eq.2 can be transformed as:

$$\begin{aligned} &\sum_{i=1}^{n} \|s_i - \sum_j w_{ji} s_j\|^2 \\ &= \|S - SW\|^2 = \|(S - SW)^T\|^2 \\ &= \mathrm{Tr}(S(I - W)(I - W)^T S^T) \\ &= \mathrm{Tr}(SMS^T), \end{aligned} \qquad (4)$$

where I is the unit matrix and $M = (I - W)(I - W)^T$. W is defined as:

$$W_{ji} = \begin{cases} w_{ji}, & \text{if } x_j \text{ is within the first } K \text{ closest to } x_i, \\ 0, & \text{otherwise.} \end{cases} \qquad (5)$$

By incorporating Eq.4 into the original ℓ_1 sparse coding, the optimization problem is reformulated as:

$$\begin{aligned} &\min_S \|X - DS\|^2 + \lambda \|S\|_1 + \alpha Tr(SMS^T) \\ &\text{s.t.} \quad \|d_i\|^2 \leq c, i = 1, 2, \ldots, k, \end{aligned} \qquad (6)$$

where α is the regularization parameter.

Secondly, based on the idea of manifold learning, the new dictionary can be learned by exploiting the intrinsic geometric structure of the data. Given a high dimensional data set, sparse coding is equivalent to a large-scale matrix factorization problem that can effectively "compress" the data by finding a learned dictionary and the corresponding coefficient over the given dictionary. In the process of "compress", these cluster centers can be used to represent the original data matrix. It is known that good reconstruction property of the dictionary is critical for sparse representation [11]. We build the relationship between the cluster centroids and the bases in dictionary. Each cluster centroid denotes a representative mode of the corresponding cluster and the cluster centers are incoherent. The relationship between dictionary entries are determined by mapping the weighted graph of cluster centroids. Thus, the incoherence of dictionary entries can be preserved accordingly, which improves reconstruction performance. In this case, it not only makes the dictionary capture the intrinsic geometric structure of the data, but also utilizes the incoherence of the cluster centroid to reduce the coherence of the bases in the dictionary. If the size of the dictionary is k, the data is divided into k clusters accordingly. Let $C = [c_1, \ldots, c_k]$ be the cluster center matrix, where c_i is the center of the i-th class. Our goal is to learn a dictionary by exploiting the cluster centers of data. Thus, an assumption is presented that if a data point c_i is selected as a similar data point to c_j, then d_i and d_j are also close to each other. Based on this assumption, the learned dictionary has powerful capability for reconstruction and discrimination. This assumption is usually regarded as manifold assumption, which can be applied in several image processing applications such as image classification [8] and clustering [22].

Given a set of m-dimensional points $c_1, c_2, \ldots, c_k$, a nearest neighbor graph G can be constructed with k vertices, where each vertex represents a point. Let W' be the weight matrix of G. If c_i is among the K'-nearest neighbors of c_j or c_j is among the K'-nearest neighbors of c_i ($K' = 5$ in our experiments), $W'_{ij} = 1$, otherwise, $W'_{ij} = 0$. The degree of c_i is defined as $B_i = \sum_{j=1}^{k} W'_{ij}$, and $B = \mathrm{diag}(B_1, B_2, \ldots, B_k)$.

Considering mapping the weighted graph G to the coefficients S, an appropriate map is selected by minimizing the objective function [20]:

$$\begin{aligned} &\frac{1}{2} \sum_{ij} \|d_i - d_j\|^2 W'_{ij} \\ &= \sum_i \|d_i\|^2 B'_{ii} + \sum_j \|d_j\|^2 B'_{jj} - 2 \sum_{ij} d_i^T d_j W'_{ij} \\ &= 2Tr(DLD^T), \end{aligned} \qquad (7)$$

where $L = B - W'$ is the Laplacian matrix. By adding the Laplacian regularizer Eq.7 into Eq.6, the final objective

function is formulated as:

$$\min_{D,S} \|X - DS\|^2 + \lambda\|S\|_1 + \alpha\text{Tr}(SMS^T)$$
$$+ \beta\text{Tr}(DLD^T)$$
$$\text{s.t.} \quad \|d_i\|^2 \leq c, i = 1, 2, \ldots, k. \quad (8)$$

Similar to the solution of Eq.1, the proposed method is divided into two steps: learning sparse coefficients S (fixing D), and learning dictionary D (fixing S).

3.2. Computing Sparse Coefficients S

Fixing the dictionary D, Eq.8 becomes:

$$\min_{S} \|X - DS\|^2 + \lambda\|S\|_1 + \alpha\text{Tr}(SMS^T)$$
$$\text{s.t.} \quad \|d_i\|^2 \leq c, i = 1, 2, \ldots, k. \quad (9)$$

Since S may contain values of 0, the traditional unconstrained optimization methods can not solve Eq.9 with ℓ_1-regularization. An optimization method based on coordinate descent is introduced to solve this problem.

Each vector s_i is updated individually, while considering all the other vectors as constant. To optimize over each s_i, the Eq.9 is rewritten in a vector form.

The reconstruction error $\|X - DS\|^2$ can be rewritten as:

$$\sum_{i=1}^{n} \|x_i - Ds_i\|^2. \quad (10)$$

The regularizer $Tr(SMS^T)$ can be rewritten as:

$$\text{Tr}(SMS^T) = \text{Tr}(\sum_{i,j=1}^{n} M_{ij}s_i s_j^T) = \sum_{i,j=1}^{m} M_{ij}s_i^T s_j. \quad (11)$$

Combining Eq.10 and Eq.11, Eq.9 can be rewritten as:

$$\min_{s_1,\ldots,s_n} \sum_{i=1}^{n} \|x_i - Ds_i\|^2 + \alpha \sum_{i,j=1}^{n} M_{ij}s_i^T s_j + \sum_{i=1}^{n} \lambda\|s_i\|_1. \quad (12)$$

Fixing the other vectors $\{s_j\}_{j\neq i}$, the optimization problem to update s_i is represented as:

$$\min_{s_i} f(s_i) = \|x_i - Ds_i\|^2$$
$$+ \alpha M_{ii}s_i^T s_i + s_i^T h_i + \lambda \sum_{j=1}^{k} |s_i^{(j)}|, \quad (13)$$

where $h_i = 2\alpha(\sum_{j\neq i} M_{ij}s_j)$ and $s_i^{(j)}$ is the j-th coefficient of s_i.

According to the feature-sign search method proposed in [12, 22], the subdifferential of the Eq.13 is discussed in the situation for different values of the coefficient $s_i^{(j)}$. Firstly, for simplification, $h(s_i)$ is defined as $\|x_i - Ds_i\|^2 + \alpha M_{ii}s_i^T s_i + s_i^T h_i$, then $f(s_i) = h(s_i) + \lambda \sum_{j=1}^{k} |s_i^{(j)}|$. It is known that making the subdifferential as the zero vector is a necessary condition for a parameter vector to be a local minima. $\nabla_i^{(j)}|s_i|$ is defined as the subdifferentiable value of the j-th coefficient of s_i.

In each feature-sign step [12, 22], the analytical solution $\hat{s}_i^{\text{new}}$ is calculated by the Eq.13 under the current active set and signs. The solution, the active set and the signs can be updated by using an efficient discrete line search between the current solution and $\hat{s}_i^{\text{new}}$. The corresponding algorithm for computing sparse coefficients S is summarized as Algorithm 1.

Algorithm 1: Computing Sparse Coefficients.

Input: The data points $X = [x_1, \ldots, x_n]$, the dictionary D, the matrix M, the parameters λ and α. **For** $1 \leq i \leq n$

1. **Initialize step**:

   $s_i = \overrightarrow{0}, \theta = \overrightarrow{0}$, and active set $\mathcal{A} = \{\}$, where $\theta_j \in \{-1, 0, 1\}$ denotes $\text{sign}(s_i^{(j)})$.

2. **Activate step**:

   From zero coefficient of s_i, choose $j = \arg\max_{j} |\nabla_i^{(j)} h(s_i)|$. Add j to the active set, namely:

   If $\nabla_i^{(j)} h(s_i) > \lambda$, set $s_i^{(j)} = -1, \mathcal{A} = \{j\} \cup \mathcal{A}$.

   If $\nabla_i^{(j)} h(s_i) < -\lambda$, set $s_i^{(j)} = 1, \mathcal{A} = \{j\} \cup \mathcal{A}$.

3. **Feature-sign step**:

   - Let $\hat{D}$ be a submatrix of D that includes only the columns corresponding to the active set. Let $\hat{s}_i$ and $\hat{h}_i$ be subvectors of s_i and h_i. Let $\hat{\theta}$ be θ corresponding to the active set.

   - Calculate the solution via the Eq.13:

     $$\hat{s}_i^{\text{new}} = (D^T D + \alpha M_{ii}I)^{-1}(\hat{D}^T x_i - \frac{(\lambda\hat{\theta} + \hat{h}_i)}{2}). \quad (14)$$

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

   - Remove zero coefficients of $\hat{s}_i$ from the active set and update $\theta = \text{sign}(s_i)$.

4. **Check the optimality conditions step**:

- **Condition (a)**: Check optimality condition for nonzero coefficients: $\nabla_i^{(j)} h(s_i) + \lambda \text{sign}(s_i^{(j)}) = 0, \forall s_i^{(j)} \neq 0$ If condition (a) is not satisfied, go to Step 3 (without any new activation); else check condition (b).

- **Condition (b)**: Check optimality condition for zero coefficients: $|\nabla_i^{(j)} h(s_i)| \leq \lambda, \forall s_i^{(j)} = 0$ If condition (b) is not satisfied, go to step 3; otherwise return s_i as the solution, redenoted as s_i^*.

end for

Output: The optimal coefficient matrix $S^* = [s_1^*, \ldots, s_m^*]$.

3.3. Learning Dictionary D

Fixing the coefficient matrix S, Eq.8 becomes a least square problem with quadratic constraints:

$$\min_D \|X - DS\|^2 + \beta \text{Tr}(DLD^T)$$
$$\text{s.t.} \quad \|d_i\|^2 \leq c, i = 1, 2, \ldots, k. \tag{15}$$

In this paper, a Lagrange dual method [2] is adopted to optimize the dictionary D. Consider the Lagrangian as:

$$L(D, \mu) = \|X - DS\|^2 + \beta \text{Tr}(DLD^T)$$
$$+ \sum_{i=1}^{k} \mu_i(\|d_i\|^2 - c), \tag{16}$$

where $\mu = [\mu_1, \ldots, \mu_k]$ and μ_i is a dual variable. Let $\partial L(D, \mu)/\partial D$ equals to 0, the optimal solution D^* can be gotten as:

$$D^* = XS^T(SS^T + \Lambda + \beta L)^{-1}, \tag{17}$$

where $\Lambda = \text{diag}(\mu)$. Putting Eq.17 into Eq.16, the Lagrange dual function becomes:

$$g(\Lambda) = \text{Tr}(X^T X - 2XS^T(SS^T + \Lambda + \beta L)^{-1} SX^T)$$
$$+ \text{Tr}((SS^T + \Lambda + \beta L)^{-1} SX^T XS^T - c\Lambda)$$
$$= \text{Tr}(X^T X) - c\text{Tr}(\Lambda)$$
$$- \text{Tr}(XS^T(SS^T + \Lambda + \beta L)^{-1} SX^T). \tag{18}$$

This leads to the following Lagrange dual function:

$$\Lambda^* = \max_{\Lambda} \text{Tr}(-XS^T(SS^T + \Lambda + \beta L)^{-1} SX^T - c\Lambda)$$
$$\text{s.t.} \quad \eta_i \geq 0, i = 1, 2, \ldots, p. \tag{19}$$

Eq.19 can be solved by Newtons method or conjugate gradient. After maximizing Eq.19, the optimal $D^* = XS^T(SS^T + \Lambda^* + \beta L)^{-1}$.

3.4. Sparse Coding for Image SR

In this section, the proposed sparse coding algorithm is used for image SR. Let $Y^h = [y_1, \ldots, y_n]$ be the training HR patch matrix and $Z^l = [z_1, \ldots, z_n]$ be the corresponding training LR patch (feature) matrix. Let $T = [t_1, \ldots, t_m]$ be a testing LR image patch matrix and $R = [r_1, \ldots, r_m]$ be the corresponding testing HR ones (unknown). Sparse coding for image SR is summarized in Algorithm 2.

Algorithm 2: Sparse Coding for Image SR.

Input: training dataset Y^h and Z^h, a testing dataset T the regularization parameter λ, α and β.

1. Learning Coupled Dictionaries D_l and D_h, D_l for training LR patch matrix, D_h for HR ones.

   - Use Eq.8 to compute the LR dictionary D_l and the corresponding coefficient matrix W of the LR patch matrix Z^l:

   $$\{D_l, \hat{W}\} = \arg \min_{D,W} \|Z^l - DW\|^2 + \lambda \|W\|_1$$
   $$+ \alpha \text{Tr}(WMW^T) + \beta \text{Tr}(DLD^T)$$
   $$\text{s.t.} \quad \|d_i\|^2 \leq c, i = 1, 2, \ldots, k.$$

   - Utilize the HR patch matrix Y^h and the coefficient matrix $\hat{S}$ to compute the HR dictionary D_h:

   $$D_h = \arg \min_D \|Y^h - D\hat{W}\|^2$$
   $$\text{s.t.} \quad \|d_i\|^2 \leq c, i = 1, 2, \ldots, k.$$

2. For each testing LR patch t_i in T

   - Compute the sparse coefficient $\hat{v}_i$ of t_i from Eq.6:

   $$\hat{v}_i = \arg \min_{v_i} \|t_i - Dv_i\|^2 + \lambda \|v_i\|_1$$
   $$+ \alpha \|v_i - \hat{W}m_i\|^2,$$

   where m_i is the non-local self-similar weight of t_i with respect to the training LR image patch matrix $Z^l = [z_1, \ldots, z_n]$

   - Generate the testing HR patch $r_i = D_h \hat{v}_i$.

Output: The testing HR image $R = \{r_1, \ldots, r_m\}$.

---

4. Experimental Validation

In this section, we verify the performance of the proposed sparse coding method on image SR.

4.1. Training Set and Model Parameters

It is known that different images have different contents, but the micro-structures of different images can be represented by a few of structural elements, such as edges, line segments and elementary features. In order to utilize sparse coding method to learn these micro-structures, a large number of patches need be extracted from some training images. In this paper, training images are collected from [19]. These training images contain different types, such as plants, human faces, architectures and animals and so on.

This paper randomly samples 50000 HR and LR patch pairs from the training images to learn the over-complete dictionary. For the LR images, the size of LR patch is 3×3 (up-sampled to 6×6), with an overlap of 1 pixel between adjacent patches, corresponding to 9×9 patches with overlap of 3 pixels for HR patches. The dictionary size is always fixed as 1024 in all our experiments, which balances between computation complexity and image quality [19].

4.2. Experimental Results on Image SR

In this subsection, we will study non-local self-similar regularization term and graph regularization term to respectively affect the image SR results. For the sake of comparison, we implement the closely related sparse coding method based on the ℓ_1 penalty (or Yang's method) [19]. This paper calls Yang's method as SC. The change of the luminance can provide more sensitivity to the human visual system [7]. Hence, we only apply the SR methods to the luminance component and use the simple bicubic interpolator for the chromatic components.
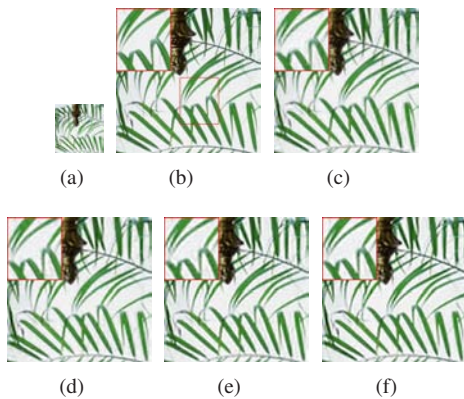

Figure 1. Reconstructed HR images (scaling factor 3) of *Leaves* by different methods. Local magnification in red rectangle is shown in the upper-left corner in each example. (a) LR image; (b) original image; (c) SC; (d) SC-NLSS; (e) SC-Graph; (f) SC-NLSS-Graph.

Firstly, we consider the influence of the non-local self-similar regularization term. Fixing $\beta = 0$, we empirically set $\lambda = 0.1$ and $\alpha = 0.2$. The sparse coding method that only considers the non-local self-similar regularization term is named as SC-NLSS. Secondly, the influence of the graph

regularization term is considered. When α is set by 0, λ and β are empirically set as 0.1 and 0.02 respectively. Similarily, the sparse coding method that only considers the graph regularization term is named as SC-Graph. Finally, the incoherence of the dictionary is measured in the experiments. The correlation coefficients R between dictionary bases can be used for measuring the incoherence of the dictionary. A smaller coefficient R between two entries of the dictionary indicates they are larger incoherence. The largest correlation coefficient R can be computed from all the pairs of the two dictionaries learned from the proposed method and L1 algorithm, which ignores the regularizer terms in Eq.7 and Eq.4. The experimental dictionary size is 144×1024. The experimental results show that the value R of the dictionary obtained by our method is 0.9023, while the value R of the dictionary obtained by L_1 algorithm is 0.9798. This shows that our method learns a better dictionary. In this paper, our proposed method is called by SC-NLSS-Graph.

Figure 2. Reconstructed HR images (scaling factor 3) of *Lena* by different methods. Local magnification in red rectangle is shown in the upper-left corner in each example. (a) LR image; (b) original image; (c) SC; (d) SC-NLSS; (e) SC-Graph; (f) SC-NLSS-Graph.

To objectively assess the quality of the SR reconstruction, PSNR and SSIM [18] are adopted to evaluate the quality of SR reconstruction. Since the SR process is only performed on the luminance component of the color image, we only compare the quantitative difference of this part between the original and the SR results. Table 1 shows the PSNR and SSIM indices of eleven testing images. In Table 1, The PSNR and SSIM values of SC-NLSS, SC-Graph and SC-NLSS-Graph is superior to SC. The average PSNR improvements of SC-NLSS, SC-Graph and SC-NLSS-Graph over SC are 0.3426dB, 0.2384dB and 0.5244dB respectively. The average SSIM improvements of SC-NLSS, SC-Graph and SC-NLSS-Graph over SC are 0.0233, 0.0138 and 0.0257 respectively. These improvements show that the non-local self-similar property on the sparse coefficients has the more influence than the graph property on the basis vectors from the dictionary. It states that using graph

Table 1. PSNR AND SSIM for 3 scale factor. For each column, we have two rows. The first row is PSNR, the second is SSIM.

Images	Bike	Butterfly	Flower	Girl	Hat	Leaves	Lena	Parrots	Parthenon	Plants	Raccoon	Average
SC	23.3606	24.7092	27.7563	32.7706	29.7974	23.9306	30.4856	28.7726	26.1136	31.6618	28.4473	27.9823
	0.7297	0.8242	0.7970	0.8022	0.8351	0.8127	0.8410	0.8831	0.7013	0.8798	0.7296	0.8032
SC-NLSS	23.5482	25.0394	28.1172	33.0018	30.1735	24.6293	30.8427	29.0799	26.2824	32.1635	28.7076	28.3260
	0.7415	0.8400	0.8172	0.8125	0.8478	0.8438	0.8527	0.8916	0.7096	0.8920	0.7437	0.8266
SC-Graph	23.4920	24.8275	28.0896	33.0543	29.9981	24.3865	30.7459	28.9571	26.3368	31.8197	28.7206	28.2207
	0.7457	0.8298	0.8155	0.8161	0.8417	0.8315	0.8567	0.8952	0.7167	0.8862	0.7520	0.8170
SC-NLSS-Graph	**23.7787**	**25.4764**	**28.2994**	**33.1259**	**30.2922**	**24.7833**	**31.0275**	**29.1980**	**26.4438**	**32.3286**	**28.8101**	**28.5067**
	0.7665	**0.8566**	**0.8289**	**0.8188**	**0.8544**	**0.8583**	**0.8638**	**0.9001**	**0.7287**	**0.8994**	**0.7582**	**0.8303**

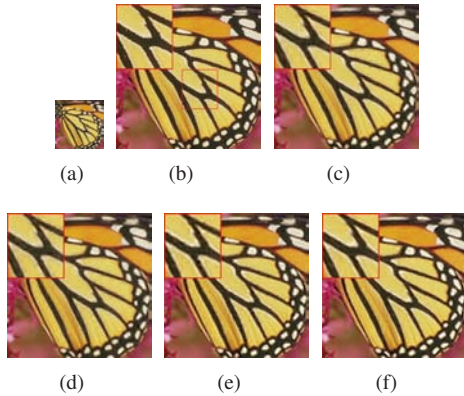
(a)   (b)   (c)

(d)   (e)   (f)

Figure 3. Reconstructed HR images (scaling factor 3) of *Butterfly* by different methods. Local magnification in red rectangle is shown in the upper-left corner in each example. (a) LR image; (b) original image; (c) SC; (d) SC-NLSS; (e) SC-Graph; (f) SC-NLSS-Graph.

property indeed improves the discrimination power of the learned dictionary. However, the property of non-local self-similarity to stablize the sparse decompositions is the key influence for image SR. The proposed method, that is SC-NLSS-Graph, takes full advantage of these two property and obtains the best PSNR and SSIM values.

For visual illustration, we conduct the same experiment on the *Leaves*, *Lena* and *Butterfly* images. In Fig.1, some "ghost" artifacts for the edges can be found by using SC. Since SC-NLSS and SC-Graph fully utilize the geometrical structure of training image patches, they reduce the artifacts and obtain better visual quality. SC-NLSS-Graph simultaneously inherits the advantages of the SC-NLSS and SC-Graph and obtains the best visual quality. Similarly, Fig.2 and Fig.3 give the same conclusion as Fig.1.

4.3. Experimental Results on A 200-Image SR

To more comprehensively test the robustness of the proposed SR method, extensive SR experiments are performed on a large dataset that contains 200 natural image of various contents. To establish this dataset, 200 high-quality natural images are randomly downloaded form the internet. A 200*200 subimage is extracted from each of these 200 image in the experiments. Fig.4 shows some example images in the dataset.

As follows, we perform extensive SR experiments on

200 natural images to compare SC-NLSS-Graph with SC, SC-NLSS and SC-Graph. The average PSNR and SSIM values of the SR images by the testing methods are shown in Table 2. To better illustrate the advantages of our proposed method (SC-NLSS-Graph), we also plot the improvement of PSNR and SSIM values of each image over SC in Fig.5. From Table 2 and Fig.5, we can see that SC-NLSS-Graph achieves excellent performance. It demonstrates that the sparse method based on the non-local self-similar and graph property can effectively deal with image SR.

(a)   (b)   (c)   (d)

Figure 4. Some example images in the established 200-image dataset.

Table 2. Average PSNR and SSIM results of the reconstructed HR images on the 200-image dataset. For each column, the first row is PSNR, and the second is SSIM.

Method	SC	SC-NLSS	SC-Graph	SC-NLSS-Graph
Average	28.5140	28.9220	28.8818	**29.0546**
	0.8287	0.8471	0.8441	**0.8533**

5. Conclusion

This paper propose a novel sparse coding method, which simultaneously consider the geometrical structure of the dictionary and corresponding sparse coefficients. The non-local self-similar and graph method are incorporated as two regularization terms. The purpose of simultaneously imposing these two regularization term is to design a novel sparse coding algorithm for having both reconstruction and discrimination properties, which can enhance the learning performance. Experimental results on image SR show that our proposed method obtains excellent image SR performance.
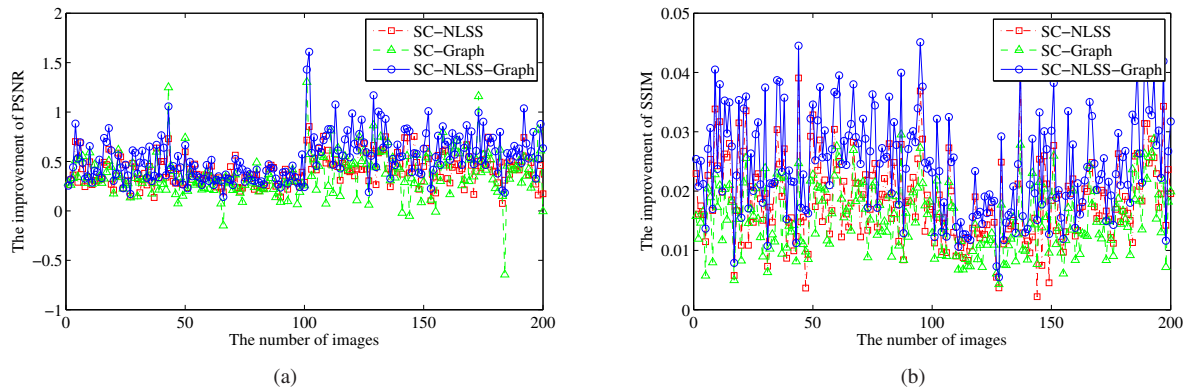
Acknowledges

Figure 5. Comparison with SC-NLSS, SC-Graph, SC-NLSS-Graph, and the results by SC posed as a reference baseline. (a) The improvement of PSNR; (b) the improvement of SSIM.

References

[1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[3] A. Buades, B. Coll, and J. Morel. A non-local algorithm for image denoisng. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[4] D. Cai, H. Bao, and X. He. Sparse concept coding for visual analysis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[5] H. Chang, D. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[6] W. Freeman, T. Jones, and E. pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.

[7] W. Freeman, E. Pasztor, and O. Carmichael. Learning low-level vision. *International Journal of Computer Vision*, 40(1):25–47, 2000.

[8] S. Gao, I. Tsang, and P. Zhao. Local features are not lonely-laplacian sparse coding for image classification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[9] X. Gao, K. Zhang, X. Li, and D. Tao. Joint learning for single-image super-resolution via a coupled constraint. *IEEE Transactions on Image Processing*, 21(2):469–480, 2012.

[10] S. Geman and D. Geman. Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(4):721–741, 1984.

[11] A. Krause and V. Cevher. Submodular dictionary selection for sparse representation. In *Proceedings of International Conference on Machine Learning*, 2010.

[12] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems*, 2007.

[13] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. zisserman. Non-local sparse models for image restoration. In *Proceedings of IEEE International Conference Computer Vision*, 2009.

[14] S. Park, M. Park, and M. Kang. Super-resolution image reconstruction: A technical overview. *IEEE Transactions on Signal Processing*, 20(3):21–36, 2003.

[15] M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generaliizing the nonlocal-means to super-resolution resconstruction. *IEEE Transactions on Image Processing*, 18(1):36–51, 2009.

[16] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[17] Y. Tang, Y. Yuan, P. Yan, and X. Li. Single-image super-resolution via local learning. *International Journal of Machine Learning and Cybernetics*, 6(9):15–23, 2011.

[18] J. Wang, S. Zhu, and Y. Gong. Resolution enhancement based on learning the sparse association of image patches. *Pattern Recognition Letters*, 31(1):1–10, 2009.

[19] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010.

[20] M. Yang, L. Zhang, X. Feng, and D. Zhang. Fisher discrimination dictionary learning for sparse representation. In *Proceedings of IEEE International Conference on computer Vision*, 2011.

[21] L. Zhang and X. Wu. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Transactions on Image Processing*, 15(8):2226–2238, 2006.

[22] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai. Graph regularized sparse coding for image representation. *IEEE Transactions on Image Processing*, 20(5):1327–1336, 2011.