Conformal and Low-Rank Sparse Representation for Image Restoration

Jianwei Li, Xiaowu Chen*, Dongqing Zou, Bo Gao, Wei Teng
State Key Laboratory of Virtual Reality Technology and Systems
School of Computer Science and Engineering, Beihang University, Beijing, China

Abstract

Obtaining an appropriate dictionary is the key point when sparse representation is applied to computer vision or image processing problems such as image restoration. It is expected that preserving data structure during sparse coding and dictionary learning can enhance the recovery performance. However, many existing dictionary learning methods handle training samples individually, while missing relationships between samples, which result in dictionaries with redundant atoms but poor representation ability. In this paper, we propose a novel sparse representation approach called conformal and low-rank sparse representation (CLRSR) for image restoration problems. To achieve a more compact and representative dictionary, conformal property is introduced by preserving the angles of local geometry formed by neighboring samples in the feature space. Furthermore, imposing low-rank constraint on the coefficient matrix can lead more faithful subspaces and capture the global structure of data. We apply our CLRSR model to several image restoration tasks to demonstrate the effectiveness.

1. Introduction

Sparse representation has been proven to be a promising model for image restoration, such as image super-resolution [33] and image denoising [9]. The sparse model is an emerging and powerful method to describe signals based on the sparsity and redundancy of their representations [9, 21]. The model suggests that there exists a dictionary which can reconstruct the signals. Each signal can be represented by a sparse linear combination of atoms in the dictionary.

In earlier sparsity-based image restoration methods, dictionaries are analytically predefined (e.g., wavelets [22]). After that, the emergence of new methods [9, 21] approves that the learned dictionary from training samples have better performance in image restoration. Therefore, the quality of the learned dictionary becomes the key point for image restoration problem.

Many approaches have been devoted to learning appropriate dictionaries. Some early dictionary learning methods, such as K-SVD algorithm [1], focus on reconstruction power of the dictionary, and have been applied to image restoration [9]. But these methods are dependent on a large training dataset. Besides, the dictionary size is fixed, which may result in dictionary redundancy, making it difficult to do the signal decompositions. Some other dictionary learning methods add discrimination constraints to make the dictionary smaller [36, 34]. However, these methods only train samples individually or consider the discrimination between classes, while ignoring local relationships between samples or structure of the data manifold, which may result in redundant dictionaries with poor representation ability. All these problems may lead to bad restoration results due to the bad dictionary.

Some sparsity-based image restoration methods imply that preserving the data structure during sparse coding by building relationships of samples can enhance the recovery performance [19]. The relationships between samples can be analyzed from two aspects. From local perspective, local samples have similar features and form a local subspace, reflecting the affinities between each other. From global perspective, samples with similar features are linearly related, and thus they lie on a low-dimensional latent space. Therefore, the key problem for dictionary based image restoration is how to embed these two relationships into dictionary learning and sparse representation.

In this paper, we propose a conformal and low-rank sparse representation (CLRSR) approach for image restoration. The work from Wright *et al*. [31] implies that the images (or their features) lying on a high dimensional manifold exhibit degenerate structure and form its own inner structure. This inner structure indeed is a kind of local relationships. Some researches [16, 10] suggest that the data inner structures can be better modelled through Conformal Eigenmaps [27], which projects data from a high dimensional space to a low dimensional manifold while preserving the angles formed by neighboring samples. These

*corresponding author (email: chen@buaa.edu.cn)

angle relationships computed by Conformal Eigenmaps model the inner structures of data, which are called as the *conformal property*. By considering the conformal property in dictionary learning, the local geometric relationships in the samples can be preserved. As a result, all the samples can find their corresponding embeddings in the dictionary space with similar local structures, yielding a *representative* dictionary. Moreover, locally neighboring samples are able to use the same dictionary atoms for their decompositions. Thus the redundancy of the dictionary can be eliminated by deleting these unnecessary atoms, resulting in a *compact* dictionary.

To involve the global structure of data, we enforce the coefficient matrix to be *low-rank*. The samples extracted from image/video are relevant to each other, thus these samples lie on low-rank subspaces. With the representative dictionary, samples having similar features will have similar sparse representations, resulting in similar coding coefficients. Therefore, the coefficient matrix A is expected to be low-rank, too. Imposing low-rank constraint on the coefficient matrix can lead more faithful subspace and capture the global structure of data [17]. Finally, introducing global structure can complement the local structure by conformal property and contribute to dictionary learning.

The contributions of this paper include: 1) a novel conformal and low-rank sparse representation approach is proposed to involve the local and global structure information of data to obtain a compact and representative dictionary; 2) our approach can be applied to image restoration tasks including image super-resolution and image denoising, and extended to image editing problems such as image/video recoloring. Various experiments and comparisons demonstrate that our approach can learn a better dictionary with competitive performance compared to the previous methods.

2. Related Work

Sparse representation and dictionary learning have been developed rapidly and received much attention in recent years. Here we briefly review the techniques most related to our work.

Many classic sparse representation methods have been proposed to learn a good dictionary. The K-SVD algorithm [1] is a representative dictionary learning method applied to image processing. It focuses on efficiently learning an over-complete dictionary which can maximize the reconstruction power. In [14], Lee *et al*. sped up the sparse coding algorithm by solving the optimization function in a new way. Mairal *et al*. [20] proposed an online dictionary learning method based on stochastic approximations to handle large datasets efficiently. These methods most focus on the original sparse coding problems and improve efficiency of the learning algorithm.

Learning a compact and representative dictionary is very important for sparse representation. Qiu *et al*. [23] adopted the rule of maximization of mutual information to learn a compact and discriminative dictionary for action attributes. Siyahjani *et al*. [28] learned a context aware dictionary to improve the recognition and localization of multiple objects in images. Other methods improve the dictionary performance by separating the particularity and commonality [13] and adding discrimination constraints [36, 34]. Most of these methods consider the discrimination between classes, while seldom involving relationships between samples or the local information in the data space.

Due to the power of the sparse representation technique and good effectiveness of the ℓ_1 norm minimization algorithm, sparse representation has been applied to many computer vision and image processing tasks such as image restoration [21, 8]. Elad *et al*. [9] used the K-SVD algorithm to learn a dictionary to reconstruct the noisy images. Mairal *et al*. [21] extended the sparsity-based restoration model to color images. Yang *et al*. [33] proposed a sparse representation based image super-resolution method by jointly learning two dictionaries for the low and high resolution images. Semi-coupled dictionary learning [30] and double sparsity regularized manifold learning [19] are also proposed to achieve good performance in image super-resolution. Chen *et al*. [6] introduced the sparse representation technique to the edit propagation task. Sparse representation and dictionary learning can also be applied to face recognition [32] and classification [35].

3. Conformal and Low-Rank Constraints on Sparse Representation

Sha and Saul [27] proposed a spectral method for nonlinear dimensionality reduction through conformal transformations. Specifically, a d-dimensional embedding is first computed from m bottom eigenvectors by LLE with $m > d$; then the conformal mapping is applied to these embeddings via optimizing the degree of neighborhood similarity. The solution to the optimization yields the underlying manifold dimensions. Essentially, the conformal transformations constrain that the local angles formed by neighboring samples maintain unchanged during dimensional reduction. With this constraint, this method could reveal a more faithful subspace which nonlinearly embedded in high dimensional space. Sparse representation can be regarded as a member of the manifold learning family. In fact, sparse representation is to build dictionaries of atoms or subspaces that provide efficient representations of classes of signals [29]. Therefore, the conformal constraint for manifold learning holds for sparse representation too. Moreover, in sparse representation, training data lying on a high dimensional manifold exhibits degenerate structure and forms its own inner structure [31]. This inner structure

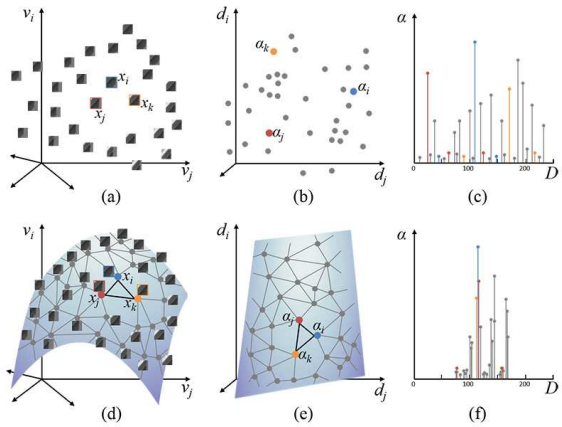


Figure 1. An illustration of the principle of conformal and low-rank sparse representation. (a) is the input data set, (b) shows the coefficients distribution in the *dictionary space* without the conformal and low-rank constraints, and (c) shows the corresponding relationship between dictionary and coefficients, while (e) and (f) are the coefficients distribution and relationship with the conformal and low-rank constraints, respectively. (d) is the input data with the computed conformal structure.

can be treated as one of local geometric structures which can be described by conformal constraints. Thus preserving the conformal constraints in sparse representation brings more local information than traditional sparse coding. For example, as shown in Fig. 1(d, e), the local structure (x_i, x_j, x_k) in input dataset is preserved for the corresponding coding coefficients $(\alpha_i, \alpha_j, \alpha_k)$.

Low-rank property has achieved excellent results in matrix completion [5, 11] and been applied to many image processing problems [18, 7]. The basic observation of these works is that similar patches extracted from images lie on a very low-dimensional subspace. The low-rank subspace constraint can capture the underlying structure of the patches. We enforce low-rank constraint on the coefficient matrix to acquire global structure of the coding coefficients and enable more robust dictionary learning.

The conformal and low-rank constraints can make the learned dictionary more compact. Let $\mathcal{M}$ be the high dimensional data manifold, and $\mathcal{N}$ be the space of coding coefficients. The sparse coding algorithm is a mapping process from $\mathcal{M}$ to $\mathcal{N}$, as shown in Fig. 1. The samples $(x_i, x_j, x_k) \in \mathcal{M}$ are mapped to $(\alpha_i, \alpha_j, \alpha_k) \in \mathcal{N}$. In the space $\mathcal{N}$, the dictionary atoms d_i can be seen as the (orthogonal) axes, and the sparse coding coefficients α are the coordinate values in the dictionary space, as shown in Fig. 1(b, e). If the conformal constraints are added into sparse representation, the coding coefficients can preserve the local geometric properties, as shown in Fig. 1(e). On the other hand, in the global perspective, low-rank constraint on the coefficient matrix makes more faithful geometric structure in dictionary space, making a sparser

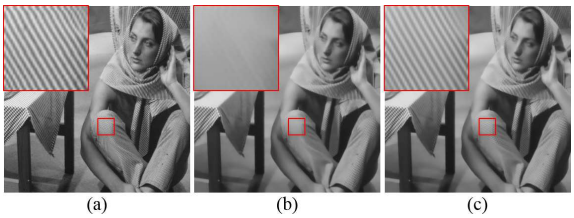


Figure 2. The reconstruction results of image "*Barbara*". (a) is the source image, (b) and (c) are the reconstruction results without and with conformal and low-rank constraints, respectively.

representation. Therefore samples in a local subspace will share some atoms, leading to a compact dictionary, as shown in Fig. 1(f). In contrast, coding coefficients are scattered without such constraints (Fig. 1(b)), resulting in a redundant dictionary (Fig. 1(c)).

The conformal and low-rank constraints could improve the representation ability of the learned dictionary. The conformal property describes the local geometric relationships of a sample with its K-nearest neighbors. The work [27] shows that the conformal constraints contribute to reveal a more faithful subspace. Furthermore, due to the low-rank property of the coefficient matrix, low-rank constraint can help with optimizing subspaces in a global view. Consequently, preserving the conformal and low-rank constraints in sparse representation will contribute to build dictionaries which provide more efficient representations. As shown in Fig. 2, (b) and (c) are the reconstruction results without and with conformal and low-rank constraints, respectively. The dictionary learned from the source image is only with 16 atoms. Under the same settings, the dictionary with conformal and low-rank constraints can reconstruct the input data (a) better than that without such constraints, showing a better representation ability. The small size also reveals the compactness of the learned dictionary.

The conformal constraint is better than other local geometric descriptors such as the classic proximity relations used by LLE [25] and Laplacian eigenmaps [3]. As pointed out by [27], the proximity relationships can hardly capture precise geometric properties, thus these relationships cannot reflect the underlying subspace. Besides, the proximity relationships are somewhat unpredictably dependent on parameters such as the number of nearest neighbors and boundary conditions. Therefore, the corresponding dictionaries learnt with the constraints of proximity relationships are less representative in sparse representation. In contrast, conformal constraints can maximally preserve the local geometric properties. With low-rank constraint capturing the global structure property, we can achieve more faithful dictionaries. We demonstrate the effectiveness of conformal and low-rank constraints through comparing with proximity relationships [19] for sparse representation.

4. The CLRSR Approach

Given an input sample set $X = [x_1, x_2, ..., x_N]$, the over-complete dictionary $D = [d_1, d_2, ..., d_m]$ and sparse decomposition coefficients $A = [\alpha_1, \alpha_2, ..., \alpha_N]$ can be learnt by solving the sparse representation problem

$$\min_{D,\alpha} \sum_i \|x_i - D\alpha_i\|_2^2 + \lambda \sum_i \|\alpha_i\|_1, \tag{1}$$

where λ is a scalar constant. However, the traditional sparse representation cannot capture the structure of the data X. To achieve this goal, we consider the conformal constraint and low-rank constraint to preserve the local and global structure of the data.

4.1. Conformal Constraint

We involve the local geometric information of the data X by adding a conformal term $f(\alpha)$ into (1), leading to

$$\min_{D,\alpha} \sum_i \|x_i - D\alpha_i\|_2^2 + \lambda_1 \sum_i \|\alpha_i\|_1 + \lambda_2 f(\alpha). \tag{2}$$

Inspired by conformal eigenmaps [27], let $g : z_i = g(x_i)$ be a mapping between two discrete point sets $\mathbf{X}$ and $\mathbf{Z}$. Suppose that the point x_i and two of its K-nearest neighbors x_j and x_k were sampled from input data, as well as their corresponding mapping points z_i, z_j and z_k from the final embedding space. To achieve the conformal mapping, the triangle formed by x points should be similar to that formed by z points. Thus we have

$$\frac{\|x_i - x_j\|^2}{\|z_i - z_j\|^2} = \frac{\|x_i - x_k\|^2}{\|z_i - z_k\|^2} = \frac{\|x_j - x_k\|^2}{\|z_j - z_k\|^2} = s_i, \tag{3}$$

where s_i represents the scale ratio of these triangles from original manifold to embedding.

In order to find maximally angle-preserving embedding, the z points can be solved by minimizing the cost function

$$\sum_{j,k \in N_i} \left(\|x_j - x_k\|^2 - s_i \|z_j - z_k\|^2\right)^2, \tag{4}$$

where N_i involves the neighbors of x_i, including x_i itself. This means that every two points in the same neighborhood set are connected, forming a locally complete graph.

We have the observation that sparse coding is a mapping process from samples in feature space to the coding coefficients in dictionary manifold. Specifically, each sample x_i in dense feature space has its embedding α_i in sparse dictionary space due to sparse coding technique. We introduce the conformal property into sparse representation to preserve the local structure information of the input data and consequently, problem (4) turns to our conformal term as

$$f(\alpha_i) = \sum_{j,k \in N_i} \left(\|x_j - x_k\|^2 - s_i \|\alpha_j - \alpha_k\|^2\right)^2. \tag{5}$$

By substituting (5) into (2), we obtain the conformal sparse representation cost function

$$\begin{aligned} \min_{D,\alpha,S} & \sum_i \|x_i - D\alpha_i\|_2^2 + \lambda_1 \sum_i \|\alpha_i\|_1 \\ & + \lambda_2 \sum_i \sum_{j,k \in N_i} \left(\|x_j - x_k\|^2 - s_i \|\alpha_j - \alpha_k\|^2\right)^2, \end{aligned} \tag{6}$$

where $S = [s_1, s_2, ..., s_N]$ represents all the scales.

4.2. Low-Rank Constraint

The samples extracted from the same scene are similar to each other and lie on a very low-dimensional subspace. The neighboring samples are combined into a data matrix, and this matrix can be approximated by a matrix with a very low rank [18]. This low-rank property is also preserved in the coefficient matrix. Thus we introduce $rank(A)$, the low-rank constraint of coefficient matrix, to capture the global structure of data. However, minimizing $rank(A)$ is a NP-hard problem. As a common practice, it can be converted to minimizing $\|A\|_*$, where $\|\cdot\|_*$ means the nuclear norm of the matrix (sum of the singular values) [5]. We add the low-rank constraint to (6), and get the final CLRSR problem

$$\begin{aligned} \min_{D,A,S} & \sum_i \|x_i - D\alpha_i\|_2^2 + \lambda_1 \sum_i \|\alpha_i\|_1 + \lambda_3 \|A\|_* \\ & + \lambda_2 \sum_i \sum_{j,k \in N_i} \left(\|x_j - x_k\|^2 - s_i \|\alpha_j - \alpha_k\|^2\right)^2. \end{aligned} \tag{7}$$

4.3. Optimization

We first convert problem (7) to the following equivalent problem by introducing an auxiliary variable P:

$$\begin{aligned} \min_{D,A,S,P} & \sum_i \|x_i - D\alpha_i\|_2^2 + \lambda_1 \sum_i \|\alpha_i\|_1 + \lambda_3 \|P\|_* \\ & + \lambda_2 \sum_i \sum_{j,k \in N_i} \left(\|x_j - x_k\|^2 - s_i \|\alpha_j - \alpha_k\|^2\right)^2, \\ & \text{s.t.} \quad A = P. \end{aligned} \tag{8}$$

This problem can be solved by augmented Lagrange multiplier (ALM) method [15], which minimizes the following augmented Lagrange function:

$$\begin{aligned} \min_{D,A,S,P} & \sum_i \|x_i - D\alpha_i\|_2^2 + \lambda_1 \sum_i \|\alpha_i\|_1 + \lambda_3 \|P\|_* \\ & + \lambda_2 \sum_i \sum_{j,k \in N_i} \left(\|x_j - x_k\|^2 - s_i \|\alpha_j - \alpha_k\|^2\right)^2 \\ & + tr[Y(A - P)] + \frac{\mu}{2} \|A - P\|_F^2, \end{aligned} \tag{9}$$

where Y is Lagrange multiplier, and $\mu > 0$ is a penalty parameter. Observe that there are four variables (D, A, S, P) to be solved in (9). We separate the objective function

Algorithm 1 Optimization of CLRSR problem (7).

Input: Data X, parameters λ_1, λ_2 and λ_3.
Initialize: $D = random(d, m), A = P = 0, Y = 0, S = 1, \mu = 10^{-6}$.
while not converged **do**

1. Fix the others and update sparse coefficients A by

$$A = \arg\min_{\alpha} \sum_i \|x_i - D\alpha_i\|_2^2 + \lambda_1 \sum_i \|\alpha_i\|_1 + \lambda_2 \sum_i \sum_{j,k \in N_i} \left(\|x_j - x_k\|^2 - s_i \|\alpha_j - \alpha_k\|^2\right)^2 + \frac{\mu}{2} \|A - (P - Y/\mu)\|_F^2 .$$

2. Fix the others and update P by

$$P = \arg\min_{P} \frac{\lambda_3}{\mu} \|P\|_* + \frac{1}{2} \|P - (A + Y/\mu)\|_F^2 .$$

3. Fix the others and update D by

$$D = \arg\min_{D} \sum_i \|x_i - D\alpha_i\|_2^2 .$$

4. Fix the others and update s_i one by one through

$$s_i = \frac{\sum_{j,k \in N_i} \|x_j - x_k\|^2 \cdot \|\alpha_j - \alpha_k\|^2}{\sum_{j,k \in N_i} \left(\|\alpha_j - \alpha_k\|^2\right)^2}.$$

5. Update the multiplier Y by

$$Y = Y + \mu * (A - P).$$

6. Update μ by $\mu = \min(1.1\mu, 10^{10})$.
7. Check the convergence condition: $A - P \to 0$.
end while
return D.

into four subproblems. In each subproblem, only one variable is tackled while the others fixed. The program runs through these steps iteratively until converging. The details is outlined in Algorithm 1.

In the first step, we use the Iterative Projection Method (IPM) [24] to solve the problem. We rewrite the equation as

$$A = \arg\min_{\alpha} F(\alpha) + 2\tau \sum_i \|\alpha_i\|_1, \tag{10}$$

where

$$F(\alpha) = \sum_i \|x_i - D\alpha_i\|_2^2 + \frac{\mu}{2} \|A - (P - Y/\mu)\|_F^2 + \lambda_2 \sum_i \sum_{j,k \in N_i} \left(\|x_j - x_k\|^2 - s_i \|\alpha_j - \alpha_k\|^2\right)^2$$

and $\tau = \lambda_1/2$. Let $\tilde{A} = \text{vec}(A)$ denote the vector obtained by concatenating all α_i in A, i.e., $\tilde{A} = \left[\alpha_1^T, \alpha_2^T, ..., \alpha_N^T\right]^T$.

Algorithm 2 Sparse coding using IPM.

Require: $\sigma, \tau > 0$
Initialize: $\tilde{A}^{(1)} = 0, t = 1$.
while not converged **do**
$t = t + 1$,

$$\tilde{A}^{(t)} = \mathbf{S}_{\frac{\tau}{\sigma}} \left(\tilde{A}^{(t-1)} - \frac{1}{2\sigma} \nabla F \left(\tilde{A}^{(t-1)} \right) \right),$$

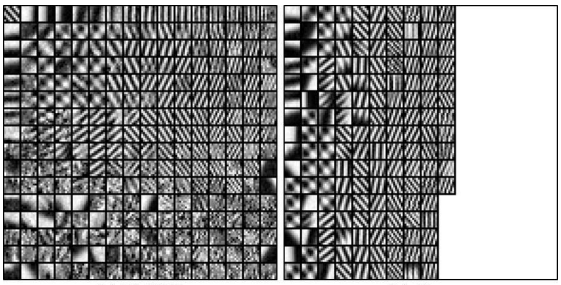
where $\nabla F \left(\tilde{A}^{(t-1)} \right)$ is the derivative of $F(A)$ w.r.t. $\tilde{A}^{(t-1)}$, that is

$$\nabla F(A) = 2(D^T DA - D^T X) + \mu(A - (P - Y/\mu)) + 4\lambda_2 \sum_i \sum_{j,k \in N_i} s_i(\alpha_k - \alpha_j) \left(\|x_j - x_k\|^2 - s_i \|\alpha_j - \alpha_k\|^2\right)^2 ,$$

and $\mathbf{S}_{\tau/\sigma}$ is a soft-thresholding operator defined as

$$\mathbf{S}_{\frac{\tau}{\sigma}}(\alpha) = \text{sign}(\alpha) \cdot \max\{(|\alpha| - \tau/\sigma), 0\}.$$

end while
return $\tilde{A}^{(t)}$.



(a) K-SVD (b) Ours

Figure 3. The dictionaries learned from the image "*Barbara*" with noise level $\sigma = 20$ using K-SVD and our approach. The initial size of dictionary is 256. Our approach learns a smaller dictionary (155 atoms) by removing the redundant atoms, and has a better representation with less noise.

The memory usage of $\tilde{A}$ depends on the dictionary size and the number of training samples. For dictionary size of 1024 and 100k samples, the memory cost is about 390MB. Following IPM described in [24], the problem (10) can be solved by Algorithm 2.

Step 2 can be solved by the singular value thresholding operator [4]. In step 3, we require that each column of the dictionary is a unit vector, i.e., $d_j^T d_j = 1$ for all j. It is a quadratic programming problem which can be solved by using a one-by-one update algorithm [35]. In this step, if one atom in the dictionary was not used by any sample for decomposition during sparse coding, it will be deleted from the dictionary. This appears when there exist similar dictionary atoms which can replace each other. Thus deleting the similar atoms can make the dictionary more compact and representative.

Table 1. PSNR(dB) and SSIM values of the super-resolution results (scaling factor = 2)

Image	*Yel. Butt.*	*Foreman*	*Hat*	*Hexagon*	*House*	*Leaves*	*Parrot*	*Red Butt.*	*Starfish*	Average
Bicubic[12]	27.47 0.9156	36.73 0.9527	31.75 0.8942	23.45 0.8166	34.11 0.9125	27.46 0.9222	31.39 0.9394	24.62 0.8689	30.25 0.9092	29.69 0.9035
SCDL[30]	29.21 0.8528	37.07 0.7631	32.71 0.6897	21.31 0.6672	32.80 0.6028	28.35 0.8857	32.05 0.7398	24.65 0.8182	30.50 0.8408	29.85 0.7622
DSRML[19]	30.05 0.9457	38.80 0.9651	33.46 0.9228	24.58 0.8428	35.47 0.9268	30.42 0.9579	33.47 0.9566	26.53 0.9058	32.32 0.9389	31.68 0.9292
Ours	**30.72** **0.9526**	**39.15** **0.9665**	**33.81** **0.9260**	**24.85** **0.8566**	**35.74** **0.9284**	**30.76** **0.9613**	**33.67** **0.9579**	**26.97** **0.9117**	**32.51** **0.9406**	**32.02** **0.9335**

Figure 4. Image super-resolution results of "*Red Butterfly*" (scaling factor: 2). From left to right: low resolution image, high resolution ground-truth, reconstructed results by Bicubic [12], SCDL [30], DSRML [19] and our approach.

4.4. The Learned Dictionary

The incoherence of the dictionary can indicate its representation power. We use the correlation coefficients R between dictionary atoms to measure the incoherence of the dictionary [36]. The correlation coefficient R is defined as

$$R(d_i, d_j) = \frac{cov(d_i, d_j)}{\sqrt{cov(d_i, d_i) * cov(d_j, d_j)}}, \qquad (11)$$

where $cov(d_i, d_j)$ represents the covariance of atoms d_i, d_j in the dictionary. Smaller correlation coefficient R indicates larger incoherence. We compute the largest R of all the atom pairs to represent the incoherence of the dictionary, i.e., $R_D = max_{i \neq j} R(d_i, d_j), d_i, d_j \in D$. We compare the correlation incoherence between the approaches with and without the conformal and low-rank constraints. We learn the dictionary with size 64×256 from ten different natural images. The value R_D of the dictionary learned by the proposed approach is 0.8219, while without the conformal and low-rank constraints the value R_D is 0.8721. This shows that the proposed approach learns a more representative dictionary.

The proposed approach can remove the redundant atoms in the dictionary, and the size of the final dictionary may be smaller than the initial one. Figure 3 shows an example of learned dictionaries from K-SVD and the proposed approach. The initial size of dictionary is 256, while it reduces to 155 after training by the proposed approach. We can see that our approach gets a more compact dictionary with better representation and less noise.

5. Applications and Comparisons

We apply our conformal and low-rank sparse representation approach to image restoration problems including image super-resolution and image denoising. It is also extended to image editing problems such as video recoloring. It's important to select appropriate parameters for different applications. In the proposed approach, there are four main parameters K, λ_1, λ_2 and λ_3. Like most methods, we build a validation set for parameter selection. We optimize each parameter while fixing others, until all parameters achieve the optimal values. The specific values of each application in this paper will be given below. In general, the time complexity of our approach is lower than SCDL [30] method, comparable to DSRML [19] method, and a little higher than K-SVD [9] method.

5.1. Image Super-resolution

Image super-resolution aims to reconstruct a high resolution (HR) image from its corresponding low resolution (LR) image. We follow the super-resolution framework of Yang *et al*. [33] to do our experiments. The dictionary consists of two coupled dictionaries of HR images and LR images. The two dictionaries are trained together so that the sparse representation for each pair of patches are the same. Our CLRSR approach brings the structure information in the data space to make the coupled dictionaries more compact and representative. As a result, for image super-resolution, the coupled dictionaries can better represent the local and global structure of the HR and LR patches respectively, constructing better HR images.

As same as previous super-resolution methods [30, 19], we transform the images to YCbCr color space and only perform on the luminance component. The size of the dictionary is 1024. The patch size is set to 5×5. The parameters $K, \lambda_1, \lambda_2, \lambda_3$ are set to be 5, 2, 0.005, 0.005, respectively. The training images are selected from [33]. We

Table 2. PSNR(dB) and SSIM values of the super-resolution results (scaling factor = 3)

Image	*Yel. Butt.*	*Foreman*	*Hat*	*Hexagon*	*House*	*Leaves*	*Parrot*	*Red Butt.*	*Starfish*	Average
Bicubic[12]	24.12 0.8208	33.84 0.9158	29.30 0.8262	21.33 0.6572	31.27 0.8651	23.49 0.8009	28.17 0.8898	21.64 0.7513	27.09 0.8175	26.69 0.8161
SCDL[30]	24.32 0.7125	33.21 0.6413	28.98 0.4863	18.77 0.4060	29.17 0.4395	22.53 0.7366	27.55 0.5904	21.06 0.6415	26.18 0.6789	25.75 0.5926
DSRML[19]	25.44 0.8505	35.42 0.9308	30.33 0.8526	21.73 0.6450	32.44 0.8806	25.18 0.8580	29.40 0.9101	22.85 0.7957	28.12 0.8515	27.88 0.8416
Ours	**26.07** **0.8717**	**35.79** **0.9337**	**30.62** **0.8575**	**21.83** **0.6994**	**32.67** **0.8844**	**25.32** **0.8649**	**29.64** **0.9136**	**23.26** **0.8074**	**28.19** **0.8542**	**28.16** **0.8541**

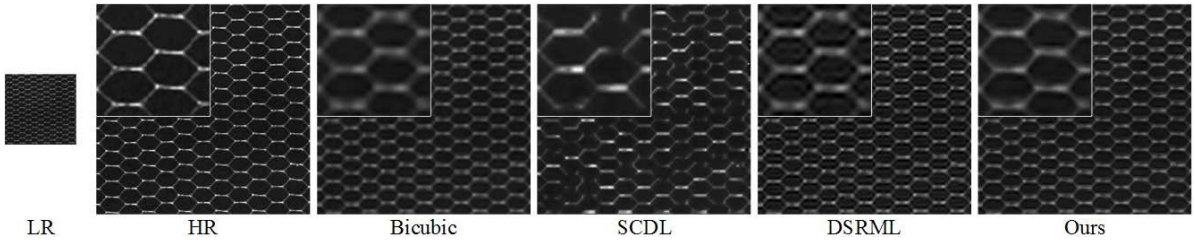


Figure 5. Image super-resolution results of "*Hexagon*" (scaling factor: 3). From left to right: low resolution image, high resolution ground-truth, reconstructed results by Bicubic [12], SCDL [30], DSRML [19] and our approach.

make comparisons with the existing image super-resolution approaches, including bicubic [12], SCDL [30] and DSRML [19]. The test images are from [19] and [30], and some are from internet.

As most of previous methods did, we first do the image super-resolution with scaling factor 2. The Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity (SSIM) results are listed in Table 1. The values shown in the table are calculated only on the luminance channel. From this table we can see that the proposed approach outperforms other methods. The PSNR is about 0.34dB higher than DSRML [19] in average. Fig. 4 shows one example of super-resolution results by these methods. From the example we can notice that SCDL method smoothed the edges well, but sometimes generated artifacts on some spots. The DSRML method generated blurred result. In contrast, our approach can well preserve these local structures to some extent.

We also do image super-resolution with scaling factor 3. The results are shown in Table 2. Our CLRSR approach outperforms other methods in terms of PSNR and SSIM values. A texture example in Fig. 5 shows the advantage of our approach in preserving the structure of image. While the SCDL method cannot preserve the structure of the hexagons, and the DSRML method generated small hexagon shadows in it. Figure 6 illustrates the comparisons in different dictionary sizes. Our approach achieves better results even with small dictionary size. It means that our learned dictionary has a higher compactness.

We also validate the contributions of each constraint towards the final results by setting each λ value to zero. Here we test with scaling factor 2. With all constraints working, the average PSNR of the test images is 32.02. By turning off conformal constraint, the average PSNR decreases to 31.83. By turning off low-rank constraint, the average PSNR becomes 31.58. We can see that low-rank constraint contributes more than conformal constraint.

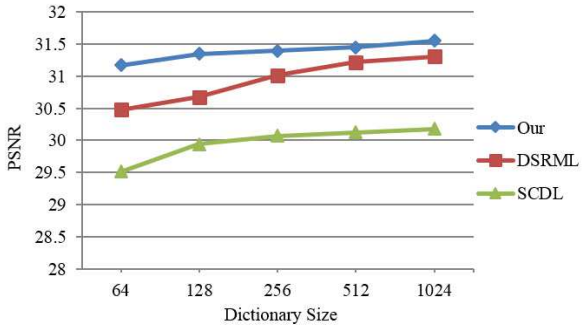


Figure 6. Comparison of image super-resolution with different dictionary sizes using different methods. The PSNR is the mean value of all the test images used in the paper.

5.2. Image Denoising

We also evaluate our approach on image denoising with different noise levels. Many researchers have reported that sparse representations are very effective toward image denoising [1]. The dictionary is first learnt on the patches set sampled from the noisy image. Then we employ the OMP method [26] to compute the coding coefficients to reconstruct the image. During dictionary learning process, we set $\lambda_1 = 0.03\sigma$ as the sparsity factor for all experiments, where σ is the standard deviation of Gaussian noise. Other parameters K, λ_2, λ_3 are set to be 5, 0.001, 0.001, respectively. The patch size is set to 8×8. The dictionary is initialized to DCT basis with dimension of 256. We make comparisons with K-SVD denoising method [9] and Bao's proximal method [2] in different noise levels. The test images are selected from these compared methods. Table 3

Table 3. PSNR(dB) values of the denoised results.

Image	*Boat*				*Fingerprint*				*Hill*			
σ	10	15	20	25	10	15	20	25	10	15	20	25
K-SVD [9]	33.65	31.63	30.28	29.09	32.34	30.04	28.42	27.16	33.35	31.35	30.01	29.11
Proximal [2]	33.57	31.62	30.20	29.16	32.35	29.97	28.28	27.03	33.31	31.29	30.02	29.06
Ours	**33.68**	**31.81**	**30.44**	**29.34**	**32.40**	**30.11**	**28.49**	**27.34**	**33.40**	**31.47**	**30.22**	**29.23**
Image	*Lena*				*Man*				*Peppers*			
σ	10	15	20	25	10	15	20	25	10	15	20	25
K-SVD [9]	35.49	33.62	32.26	31.18	33.59	31.48	30.04	28.97	34.76	33.25	32.10	31.24
Proximal [2]	35.41	33.57	32.25	31.19	33.47	31.43	30.02	29.00	34.64	33.22	32.14	31.18
Ours	**35.56**	**33.72**	**32.42**	**31.37**	**33.66**	**31.60**	**30.22**	**29.22**	**34.83**	**33.41**	**32.39**	**31.52**

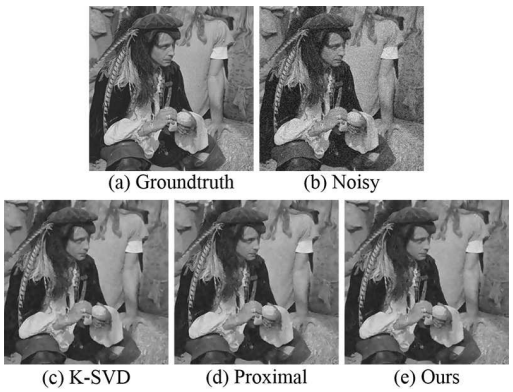


Figure 7. Denoising comparison of image "*man*" with K-SVD [9] and proximal method [2] in noise level $\sigma = 25$.

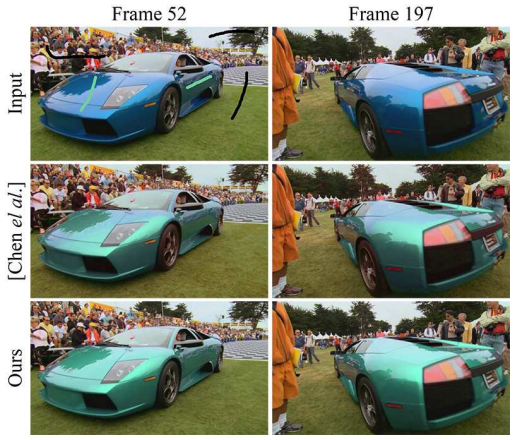


Figure 8. Comparison results on video recoloring. From top to bottom: input video, recoloring results from [6] and our approach.

shows PSNR values of the results. Figure 7 shows one example of the comparison images. From the table we can see that our approach is better than others, especially in higher noise levels.

5.3. Video Recoloring

Our approach can also be applied to image editing tasks such as video recoloring. We have the observation that editing problem can be regarded as reconstruction problem. According to some requirements, some content of the image is modified to reconstruct a new one. Here we use the learned dictionary to achieve the editing and reconstruction. Taking image/video recoloring as an example, users can change object's color by drawing strokes with desired colors, or editing the color palettes of the video or image. Firstly, we learn a dictionary in RGB color space from the original image. When a user makes some color edits, the corresponding dictionary atoms will be changed and the final recolored image is reconstructed by the new dictionary. Please refer to [6] for more details.

In this experiment, the parameters $K, \lambda_1, \lambda_2, \lambda_3$ are set to be 5, 0.05, 0.001, 0.001, respectively. The dictionary size is initialized to 300. Our recoloring results are compared to Chen's method [6] for visual effect. As shown in Fig. 8, the user would like to change the color of the blue car by drawing green strokes on it. Our approach achieves good reconstruction results, while artifacts appear on the results of Chen *et al*. [6], which you can see that some areas of the car still remain blue. This indicates that our approach can get a more representative dictionary, which can cover all the input data and get better image reconstruction.

6. Conclusions

In this paper, we proposed a novel conformal and low-rank sparse representation approach for image restoration. By considering conformal property, the local structure of data can be preserved. In the meanwhile, low-rank constraint on the coefficient matrix can reveal more faithful subspaces during sparse representation. With these constraints, our approach can get a more representative and compact dictionary. It was applied to various applications such as image super-resolution, image denoising and video recoloring. We evaluated our approach by comparing to many related methods on various datasets. Our approach runs not very fast due to the additional computation for conformal relationships, especially for large dataset. However, our program is far from optimizing and we will develop our program in parallel for more efficiency.

Acknowledgement. We thank the reviewers for their valuable feedback. This work is supported in part by grants from NSFC (61325011) & (61421003), 863 Program (2013AA013801), and SRFDP (20131102130002).

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322, 2006. 1, 2, 7

[2] C. Bao, H. Ji, Y. Quan, and Z. Shen. ℓ_0 norm based dictionary learning by proximal methods with global convergence. In *CVPR*, pages 3858–3865, 2014. 7, 8

[3] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003. 3

[4] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010. 5

[5] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009. 3, 4

[6] X. Chen, D. Zou, J. Li, X. Cao, Q. Zhao, and H. Zhang. Sparse dictionary learning for edit propagation of high-resolution images. In *CVPR*, pages 2854–2861, June 2014. 2, 8

[7] Y.-L. Chen and C.-T. Hsu. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *ICCV*, pages 1968–1975, 2013. 3

[8] W. Dong, D. Zhang, and G. Shi. Centralized sparse representation for image restoration. In *ICCV*, pages 1259–1266, 2011. 2

[9] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15(12):3736–3745, 2006. 1, 2, 6, 7, 8

[10] T. Igarashi, T. Moscovich, and J. F. Hughes. As-rigid-as-possible shape manipulation. *ACM Trans. Graph.*, 24(3):1134–1141, July 2005. 1

[11] R. Keshavan, A. Montanari, and S. Oh. Matrix completion from noisy entries. In *NIPS*, pages 952–960, 2009. 3

[12] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(6):1153–1160, 1981. 6, 7

[13] S. Kong and D. Wang. A dictionary learning approach for classification: separating the particularity and the commonality. In *ECCV*, pages 186–199. 2012. 2

[14] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems*, pages 801–808, 2006. 2

[15] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *Technical Report*, UILU-ENG-09-2215, 2009. 4

[16] Y. Lipman, D. Levin, and D. Cohen-Or. Green coordinates. *ACM Trans. Graph.*, 27(3):78:1–78:10, Aug. 2008. 1

[17] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In *ICML*, pages 663–670, 2010. 2

[18] S. Lu, X. Ren, and F. Liu. Depth enhancement via low-rank matrix completion. In *CVPR*, pages 3390–3397, 2014. 3, 4

[19] X. Lu, Y. Yuan, and P. Yan. Image super-resolution via double sparsity regularized manifold learning. *IEEE TCSVT*, 23(12):2022–2033, Dec 2013. 1, 2, 3, 6, 7

[20] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research*, 11:19–60, 2010. 2

[21] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Trans. Image Process.*, 17(1):53–69, 2008. 1, 2

[22] S. Mallat. *A wavelet tour of signal processing*. Academic press, 1999. 1

[23] Q. Qiu, Z. Jiang, and R. Chellappa. Sparse dictionary-based representation and recognition of action attributes. In *ICCV*, pages 707–714, Nov 2011. 2

[24] L. Rosasco, A. Verri, M. Santoro, S. Mosci, and S. Villa. Iterative projection methods for structured sparsity regularization. *CSAIL Technical Reports*, MIT-CSAIL-TR-2009-050(CBCL-282), Oct 2009. 5

[25] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000. 3

[26] R. Rubinstein, M. Zibulevsky, and M. Elad. Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit. *CS Technion*, 40(8):1–15, 2008. 7

[27] F. Sha and L. K. Saul. Analysis and extension of spectral methods for nonlinear dimensionality reduction. In *ICML*, pages 784–791, 2005. 1, 2, 3, 4

[28] F. Siyahjani and G. Doretto. Learning a context aware dictionary for sparse representation. In *ACCV*, pages 228–241. 2013. 2

[29] I. Tosic and P. Frossard. Dictionary learning. *IEEE Signal Processing Magazine*, 28(2):27–38, March 2011. 2

[30] S. Wang, L. Zhang, Y. Liang, and Q. Pan. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *CVPR*, pages 2216–2223, June 2012. 2, 6, 7

[31] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, June 2010. 1, 2

[32] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210–227, 2009. 2

[33] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Trans. Image Process.*, 19(11):2861–2873, 2010. 1, 2, 6

[34] M. Yang, L. Zhang, X. Feng, and D. Zhang. Fisher discrimination dictionary learning for sparse representation. In *ICCV*, pages 543–550, Nov 2011. 1, 2

[35] M. Yang, L. Zhang, J. Yang, and D. Zhang. Metaface learning for sparse representation based face recognition. In *ICIP*, pages 1601–1604, 2010. 2, 5

[36] Q. Zhang and B. Li. Discriminative k-svd for dictionary learning in face recognition. In *CVPR*, pages 2691–2698, 2010. 1, 2, 6