

# Group-based Sparse Representation for Image Compressive Sensing Reconstruction with Non-Convex Regularization

Zhiyuan Zha<sup>a</sup>, Xinggan Zhang<sup>a</sup>, Qiong Wang<sup>a</sup>, Lan Tang<sup>a,b</sup>, Xin Liu<sup>c</sup>

<sup>a</sup> School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China.

<sup>b</sup> National Mobile Commun. Research Lab., Southeast University, Nanjing 210023, China.

<sup>c</sup> The Center for Machine Vision and Signal Analysis, University of Oulu, 90014, Finland.

---

## Abstract

Patch-based sparse representation modeling has shown great potential in image compressive sensing (CS) reconstruction. However, this model usually suffers from some limits, such as dictionary learning with great computational complexity, neglecting the relationship among similar patches. In this paper, a group-based sparse representation method with non-convex regularization (GSR-NCR) for image CS reconstruction is proposed. In GSR-NCR, the local sparsity and nonlocal self-similarity of images is simultaneously considered in a unified framework. Different from the previous methods based on sparsity-promoting convex regularization, we extend the non-convex weighted  $\ell_p$  ( $0 < p < 1$ ) penalty function on group sparse coefficients of the data matrix, rather than conventional  $\ell_1$ -based regularization. To reduce the computational complexity, instead of learning the dictionary with a high computational complexity from natural images, we learn the principle component analysis (PCA) based dictionary for each group. Moreover, to make the proposed scheme tractable and robust, we have developed an efficient iterative shrinkage/thresholding algorithm to solve the non-convex optimization problem. Experimental results demonstrate that the proposed method outperforms many state-of-the-art techniques for image CS reconstruction.

---

<sup>☆</sup>Fully documented templates are available in the elsarticle package on CTAN.

*Keywords:* Image CS reconstruction, group sparse representation, nonlocal self-similarity, non-convex weighted  $\ell_p$  minimization, iterative shrinkage/thresholding algorithm.

---

## 1. Introduction

Compressive sensing (CS) [1, 2, 3], which aims to recover signals from fewer measurements than suggested by the Nyquist sampling theory, is based on the hypothesis that the signals in question have compressible representations. The most attractive aspect of CS-based compression is that the sampling and compression are conducted simultaneously, and almost all computational cost is derived from the decoder stage, and thus, leading to a low computational cost of the encoder stage. Due to the superior property of CS, it has been widely applied to various areas, such as MRI image [4], remoting sensing [5], single-pixel camera [6] and sensor networks [7].

In the theory of CS, if a signal is sparse in some transform domain, it is often sampled by the random projection and reconstructed by solving the  $\ell_0$  minimization problem with the prior information which usually makes up the regularization terms. However, due to the fact that  $\ell_0$  minimization is a difficult combinatorial optimization problem, solving this problem is NP-hard. For this reason, it has been proposed to replace the  $\ell_0$  norm by its convex  $\ell_1$  counterpart to make the optimization easy. For instance, Candès *et al.* [1] proposed that solving  $\ell_1$  minimization problem can recover a  $K$ -sparse signal  $\mathbf{X} \in \mathbb{R}^N$  from  $M = O(K \log(N/K))$  random measurements. To solve the above  $\ell_1$  minimization problem, many CS reconstruction algorithms have been proposed, such as linear programming [3], gradient projection sparse reconstruction [8], match pursuit [9] and iterative thresholding [10].

As a basic image inverse problem in the filed of image restoration, maybe the hottest topic is image CS reconstruction, which has attracted a lot of research interest in the past few years [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38]. Image CS reconstruction

aims to reconstruct high quality image from fewer measurements, which may even be far below the traditional Nyquist sampling rate. Due to the ill-posed nature of image CS reconstruction, it has been well-known that the prior knowl-  
 30 edge of images plays a critical role in improving the performance of image CS reconstruction algorithms. Therefore, how to design an effective regularization term to describe the image priors is vital for image CS reconstruction tasks.

Early regularization models mainly consider the prior on the level of pixels, such as Tikhonov regularization [39] and total variation (TV) regularization  
 35 [19, 21, 22], utilize the local structure patterns of an image and high effectiveness to preserve image edges and recover the smooth regions. However, some undesirable properties are produced, including smearing out the image details and over-smoothing the images.

Another popular prior is based on image patch, which has shown promising  
 40 performance in image CS reconstruction [15, 18, 24, 26]. The well-known work is sparse representation-based model [40, 41], which assumes that image patch can be precisely encoded as a sparse linear combination of basic elements. These elements, called atoms, compose a dictionary [42, 43]. The dictionary is usually learned from a natural image dataset [12]. Compared with the traditional  
 45 analytically designed dictionaries, such as DCT [14, 44] and wavelet [29], dictionaries learned directly from images are superior to be adapted to image local structures [40, 41], and thus could improve the sparsity which results in better performance. For example, Zhang *et al.* [24] proposed a method for image CS reconstruction using adaptively learned sparsifying basis via  $\ell_0$  minimization. Zha  
 50 *et al.* [26] proposed an adaptive sparse nonlocal regularization (ASNR) model for image CS reconstruction. However, two main problems are still existing for patch-based sparse representation model. First, it is computationally expensive to learn an off-the-shelf dictionary. Second, this sparse representation-based model usually neglects the correlations between sparsely coded patches.

55 Image patches that have similar patterns can be spatially far from each other and thus can be collected in the whole image. The nonlocal self-similarity (NSS) prior characterizes the repetitiveness of textures and structures reflected by nat-

60 ural images within nonlocal regions, which can be exploited to retain edges and sharpness effectively. The seminal work of nonlocal means (NLM) denoising [45] has motivated a wide range of studies on NSS and a flurry of NSS-based methods have been proposed for image CS reconstruction [14, 29, 46]. For instance, Zhang *et al.* [46] proposed a nonlocal total variation (NLTV) regularization model for image CS reconstruction. Zhang *et al.* [14] proposed a framework via collaborative sparsity, which enforces local 2D sparsity and nonlocal 3D sparsity simultaneously, in an adaptive hybrid space-transform domain. Nasser *et al.* [29] proposed a new technique for high-fidelity image CS reconstruction via joint adaptive sparsity regularization (JASR) in transform domain.

Recent advances have suggested that, by exploiting the NSS prior and clustering similar patches, group-based sparse representation has shown great potential in various image inverse problems [47, 48, 49, 50]. In this paper, we propose a new method for image CS reconstruction, using group-based sparse representation framework with non-convex regularization (GSR-NCR). The GSR offers a powerful mechanism of combining local sparsity and NSS of images simultaneously. Unlike the previous sparsity-promoting convex regularization methods, we extend the non-convex weighted  $\ell_p$  ( $0 < p < 1$ ) penalty function on group sparse coefficients of the data matrix, rather than conventional  $\ell_1$ -based regularization. In order to reduce the computational complexity, we learn the principle component analysis (PCA) based dictionary for each group to substitute for the dictionary with a high computational complexity learned from natural images. In addition, to make the optimization tractable, an efficient iterative shrinkage/thresholding algorithm is adopted to solve the non-convex optimization problem. Experimental results show that the proposed method can outperform many existing state-of-the-art image CS reconstruction methods.

The reminder of this paper is organized as follows. Section 2 briefly introduces CS theory, patch-based sparse representation modeling and group-based sparse representation modeling. Section 3 presents the modeling of group-based sparse representation with non-convex regularization (GSR-NCR) for image CS reconstruction and develops an iterative shrinkage/thresholding algorithm to

solve the proposed GSR-NCR model. Section 4 presents the experimental re-  
90 sults. Finally, some conclusions are given in Section 5.

## 2. Background and related work

### 2.1. Compressive Sensing

Compressive sensing (CS) has attracted considerable attention from signal/image processing communities [1, 2, 3]. In the theory of CS,  $\mathbf{X} \in \mathbb{R}^N$  is a finite length signal.  $\mathbf{X}$  is said to be sparse if  $\mathbf{X}$  can be represented as a superposition of a small number of vectors taken from a known sparsifying transform domain basis  $\Psi$ , such that  $\boldsymbol{\theta} = \Psi^T \mathbf{X}$  contains only a small set of non-zero entries. The number of significant elements within the coefficient vector  $\boldsymbol{\theta}$  is regarded as the quantitative criteria of the sparsity of  $\mathbf{X}$  in  $\Psi$ . To be concrete, one seeks the perfect reconstruction of a signal  $\mathbf{X}$  from its  $M$  randomized linear measurements, i.e.,  $\mathbf{Z} = \phi \mathbf{X}$ , where  $\mathbf{Z} \in \mathbb{R}^M$ ,  $\phi \in \mathbb{R}^{M \times N}$  represents the random projection matrix and satisfies  $M < N$ . The goal of CS recovery is to reconstruct  $\mathbf{X}$  from  $\mathbf{Z}$  with subrate being  $S = M/N$ , which is usually formulated as the following  $\ell_0$  minimization problem,

$$\arg \min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_0, \quad s.t. \quad \mathbf{Z} = \phi \Psi \boldsymbol{\theta} \quad (1)$$

where  $\|\cdot\|_0$  is  $\ell_0$ -norm, counting the non-zero entries of  $\boldsymbol{\theta}$ .

However, since  $\|\cdot\|_0$  norm minimization is discontinuous and an NP-hard problem, it is usually relaxed to the convex  $\ell_1$ -norm minimization. Therefore, Eq. (1) can be rewritten as the following unconstrained optimization problem,

$$\boldsymbol{\theta} = \arg \min_{\boldsymbol{\theta}} \left( \frac{1}{2} \|\mathbf{Z} - \phi \Psi \boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1 \right) \quad (2)$$

where  $\lambda$  is regularization parameter. According to [1], CS is capable of recover-  
95 ing a  $K$ -sparse signal  $\mathbf{X}$  (with highly probability) from  $\mathbf{Z}$  of size  $M$ , where the number of random measurements satisfies  $M = O(K \log(N/K))$ .

## 2.2. Patch-based Sparse Representation

Traditional patch-based sparse representation model has been proven to be very effective in image CS reconstruction [15, 18, 24, 26]. It assumes that each image patch can be precisely modeled as a sparse linear combination of basic elements [40, 41]. These elements are called atoms and they compose a dictionary [42, 43]. Mathematically, for an image  $\mathbf{X} \in \mathbb{R}^N$ , let  $\mathbf{x}_i = \mathbf{R}_i \mathbf{X}$ ,  $i = 1, 2, \dots, n$  denotes an image patch of size  $\sqrt{m} \times \sqrt{m}$  extracted at location  $i$ , where  $\mathbf{R}_i$  is the matrix extracting patch  $\mathbf{x}_i$  from  $\mathbf{X}$  at location  $i$ . Given a dictionary  $\mathbf{D} \in \mathbb{R}^{m \times M}$ ,  $m \leq M$ , the sparse representation processing of each patch  $\mathbf{x}_i$  is to discover a sparse vector  $\boldsymbol{\alpha}_i$  such that  $\mathbf{x}_i = \mathbf{D}^{-1} \mathbf{x}_i$ , where  $\boldsymbol{\alpha}_i$  is a sparse vector whose entries are mostly zero or close to zero. Then the whole image  $\mathbf{X}$  can be reconstructed by averaging all the reconstructed patches  $\{\mathbf{x}_i\}$ , which can be expressed as

$$\mathbf{X} \approx \mathbf{D} \boldsymbol{\alpha} = \left( \sum_{i=1}^n \mathbf{R}_i^T \mathbf{R}_i \right)^{-1} \left( \sum_{i=1}^n \mathbf{R}_i^T \mathbf{D} \boldsymbol{\alpha}_i \right) \quad (3)$$

where  $\boldsymbol{\alpha}$  denotes the concatenation of all  $\boldsymbol{\alpha}_i$ , that is,  $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \dots, \boldsymbol{\alpha}_n^T]^T$ , which is the patch-based redundant sparse representation for  $\mathbf{X}$ .

Now, we merge Eq. (3) into Eq. (2), the patch-based sparse representation scheme for image CS reconstruction is formulated as

$$\boldsymbol{\alpha}_i = \arg \min_{\boldsymbol{\alpha}_i} \sum_{i=1}^n \left( \frac{1}{2} \|\mathbf{z}_i - \phi \mathbf{D} \boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1 \right) \quad (4)$$

100 where  $\mathbf{D}$  replaces  $\boldsymbol{\Psi}$  in Eq. (2), standing for a learning dictionary, and  $\boldsymbol{\alpha}_i$  is a patch-based sparse representation coefficient for each patch  $\mathbf{x}_i$  over the dictionary  $\mathbf{D}$ .  $\mathbf{z}_i$  is the linear measurements of each patch  $\mathbf{x}_i$ .

However, there exists two main issues for patch-based sparse representation model. On the one hand, since dictionary learning is a large-scale and highly  
 105 non-convex problem, it is computationally expensive to solve the sparsity optimization problem. On the other hand, the patch-based sparse representation model usually assumes the independence between sparsely coded patches, which takes no account of the correlation of similar patches in essence [47, 48, 49, 50].

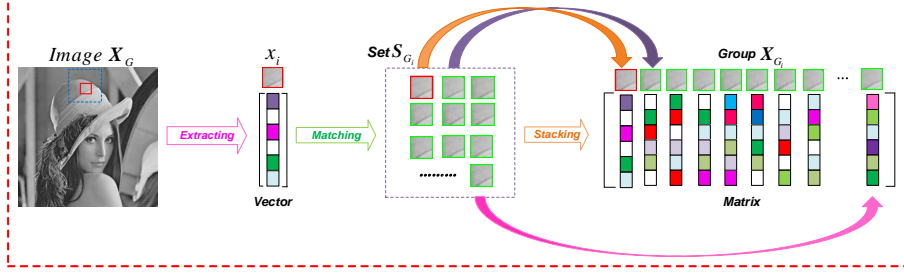


Figure 1: Illustrations for the group construction. Extract each exemplar patch vector  $\mathbf{x}_i$  from image  $\mathbf{X}_G$ . For each patch  $\mathbf{x}_i$ , denote  $\mathbf{S}_{G_i}$  is the set, which composed of its most  $c$  similar patches. Stack all the patches in  $\mathbf{S}_{G_i}$  to construct the data matrix to generate the group, denoted by  $\mathbf{X}_{G_i}$ .

### 2.3. Group-based Sparse Representation

Recent studies have shown that structured or group sparsity can offer more promising performance for image restoration tasks [47, 48, 49, 50]. Since the unit of our proposed sparse representation model is group, this section will give briefs to introduce how to construct the groups. Specifically, as shown in Fig. 1, image  $\mathbf{X}_G$  with size  $N$  is divided into  $n$  overlapped patches  $\mathbf{x}_i$  of size  $\sqrt{m} \times \sqrt{m}, i = 1, 2, \dots, n$ . Then for each exemplar patch  $\mathbf{x}_i$ , denoted by small red square in Fig. 1, within the  $H \times H$  sized searching window (big blue square), its most similar  $c$  patches (small green squares) are selected to form a set  $\mathbf{S}_{G_i}$ . Since then, all the patches in  $\mathbf{S}_{G_i}$  are stacked into a matrix  $\mathbf{X}_{G_i} \in \mathbb{R}^{m \times c}$ , which contains every element of  $\mathbf{S}_{G_i}$  as its column, i.e.,  $\mathbf{X}_{G_i} = \{\mathbf{x}_{G_i,1}, \mathbf{x}_{G_i,2}, \dots, \mathbf{x}_{G_i,c}\}$ . The matrix  $\mathbf{X}_{G_i}$  consisting of all the patches with similar structures is called as a group, where  $\mathbf{x}_{G_i,c}$  denotes the  $c$ -th similar patch (column form) of the  $i$ -th group. Finally, similar to patch-based sparse representation [40, 41], given a dictionary  $\mathbf{D}_{G_i}$ , which is often learned from each group, such as DCT [14], PCA-based dictionary [51]. Therefore, in image CS reconstruction, similar to Eq. (4), each group  $\mathbf{X}_{G_i}$  can be sparsely represented as  $\boldsymbol{\alpha}_{G_i} = \mathbf{D}_{G_i}^{-1} \mathbf{X}_{G_i}$  and

solved by the following  $\ell_1$ -norm minimization problem,

$$\boldsymbol{\alpha}_{G_i} = \arg \min_{\boldsymbol{\alpha}_{G_i}} \sum_{i=1}^n \left( \frac{1}{2} \|\mathbf{Z}_{G_i} - \boldsymbol{\phi} \mathbf{D}_{G_i} \boldsymbol{\alpha}_{G_i}\|_F^2 + \lambda \|\boldsymbol{\alpha}_{G_i}\|_1 \right) \quad (5)$$

where  $\mathbf{Z}_{G_i}$  is the linear measurements of each group  $\mathbf{X}_{G_i}$ . Note that  $\ell_1$ -norm is imposed on each column of  $\boldsymbol{\alpha}_{G_i}$ , which also holds true for the following derivation with  $\ell_1$ -norm on matrix.

### 3. Image CS reconstruction using group-based sparse representation model with non-convex weighted $\ell_p$ Minimization

115

Typical patch-based sparse representation methods for image CS reconstruction usually suffer from a common drawback that the dictionary learning with great computational complexity and neglecting the relationships among similar patches [47, 48, 49, 50]. The sparsity-promoting convex  $\ell_1$  minimization is usually regarded as a standard scheme for recovering a sparse signal. However, a fact that cannot be ignored is that,  $\ell_1$  minimization is hard to achieve the desired sparsity solution in some practical problems, such as image inverse problems [52]. Based on the fact above, this paper proposes a new method for image CS reconstruction using group-based sparse representation with non-convex weighted  $\ell_p$  minimization. To make the optimization tractable, an iterative shrinkage/thresholding (IST) algorithm [53] is developed to solve the above non-convex weighted  $\ell_p$  minimization problem efficiently.

120

125

#### 3.1. Modeling of the Proposed Image CS Reconstruction

To obtain sparsity solution more accurately, inspired by the success of  $\ell_p$  ( $0 < p < 1$ ) sparse optimization [54, 55, 56] and our previous work [52], we apply the non-convex weighted  $\ell_p$  ( $0 < p < 1$ ) penalty function on group sparse coefficients of the data matrix to replace the convex  $\ell_1$  norm. To be concrete, different from Eq. (5), the proposed group-based sparse representation for image CS reconstruction with non-convex weighted  $\ell_p$  minimization is formulated as

$$\boldsymbol{\alpha}_G = \arg \min_{\boldsymbol{\alpha}_G} \frac{1}{2} \|\mathbf{Z}_G - \boldsymbol{\phi} \mathbf{D}_G \boldsymbol{\alpha}_G\|_2^2 + \|\mathbf{W}_G \cdot \boldsymbol{\alpha}_G\|_p \quad (6)$$



where  $\cdot$  represents the dot product and  $\mathbf{W}_G$  is a weight assigned to  $\boldsymbol{\alpha}_G$ . The weight  $\mathbf{W}_G$  will enhance the representation capability of group sparse coefficient  $\boldsymbol{\alpha}_G$ . Note that here we abused the  $p$ -norm for matrix and  $\ell_p$ -norm is imposed on each column of  $\boldsymbol{\alpha}_{G_i}$ .

### 3.2. Solving the Non-convex Weighted $\ell_p$ Minimization by the Iterative Shrinkage/Thresholding (IST) Algorithm

Solving the objective function of Eq. (6) is very difficult, since it is a large scale non-convex optimization problem. To make the proposed scheme tractable and robust, in this paper we adopt the iterative shrinkage/thresholding (IST) algorithm [53] to solve Eq. (6). We will briefly introduce IST algorithm. More specifically, consider the following general optimization problem,

$$\min_{\mathbf{u} \in \mathbb{R}^N} f(\mathbf{u}) + g(\mathbf{u}) \quad (7)$$

where  $f(\mathbf{u})$  is a smooth convex function with gradient, which is Lipschitz continuous.  $g(\mathbf{u})$  is a continuous convex function which is possibly non-smooth. The IST algorithm to solve Eq. (7) with a constant step  $\rho$  is formulated as

$$\mathbf{z}^{(k+1)} = \mathbf{u}^{(k)} - \rho \nabla f(\mathbf{u}^{(k)}) \quad (8)$$

$$\mathbf{u}^{(k+1)} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{u} - \mathbf{z}^{(k+1)}\|_2^2 + \lambda g(\mathbf{u}) \quad (9)$$

where  $k$  denotes the iteration number. Then, by invoking IST algorithm, the proposed non-convex weighted  $\ell_p$  minimization problem Eq. (6) with the constraint  $\mathbf{u}_G = \mathbf{D}_G \boldsymbol{\alpha}_G$  can be rewritten as

$$\mathbf{u}_G^{(k)} = \mathbf{D}_G^{(k)} \boldsymbol{\alpha}_G^{(k)} \quad (10)$$

$$\mathbf{Y}_G^{(k+1)} = \mathbf{u}_G^{(k)} - \rho \boldsymbol{\phi}^T (\boldsymbol{\phi} \mathbf{u}_G^{(k)} - \mathbf{Z}_G) \quad (11)$$

$$\boldsymbol{\alpha}_G^{(k+1)} = \arg \min_{\boldsymbol{\alpha}_G} \frac{1}{2} \|\mathbf{D}_G \boldsymbol{\alpha}_G - \mathbf{Y}_G^{(k+1)}\|_2^2 + \|\mathbf{W}_G \cdot \boldsymbol{\alpha}_G\|_p \quad (12)$$

135 Obviously, the crux for solving Eq. (6) is translated into solving Eq. (12). Next, we will show that there is an efficient solution to Eq. (12). To avoid confusion, the subscribe  $k$  may be omitted for conciseness.

However, due to the complex structure of  $\|\mathbf{W}_G \cdot \boldsymbol{\alpha}_G\|_p$ , it is difficult to solve Eq. (12), Let  $\mathbf{X}_G = \mathbf{D}_G \boldsymbol{\alpha}_G$ , Eq. (12) can be rewritten as

$$\boldsymbol{\alpha}_G = \arg \min_{\boldsymbol{\alpha}_G} \frac{1}{2} \|\mathbf{X}_G - \mathbf{Y}_G\|_2^2 + \|\mathbf{W}_G \cdot \boldsymbol{\alpha}_G\|_p \quad (13)$$

To enable a tractable solution of Eq. (13), in this paper, a general assumption is made, with which even a closed-form solution can be achieved. Specifically, 140  $\mathbf{Y}_G$  can be regarded as some type of noisy observation of  $\mathbf{X}_G$ , and then the assumption is made that each element of  $\mathbf{E} = \mathbf{X}_G - \mathbf{Y}_G$  follows an independent zero-mean distribution with variance  $\sigma^2$ . The following conclusion can be proved with this assumption.

**Theorem 1** Define  $\mathbf{X}_G, \mathbf{Y}_G \in \Re^N$ ,  $\mathbf{X}_{G_i}, \mathbf{Y}_{G_i} \in \Re^{m \times c}$ , and  $\mathbf{e}(j)$  as each element of error vector  $\mathbf{e}$ , where  $\mathbf{e} = \mathbf{X}_G - \mathbf{Y}_G, j = 1, \dots, N$ . Assume that  $\mathbf{e}(j)$  follows an independent zero mean distribution with variance  $\sigma^2$ , and thus for any  $\varepsilon > 0$ , we can represent the relationship between  $\frac{1}{N} \|\mathbf{X}_G - \mathbf{Y}_G\|_2^2$  and  $\frac{1}{K} \sum_{i=1}^n \|\mathbf{X}_{G_i} - \mathbf{Y}_{G_i}\|_F^2$  by the following property,

$$\lim_{\substack{N \rightarrow \infty \\ K \rightarrow \infty}} \mathbf{P}\{\left|\frac{1}{N} \|\mathbf{X}_G - \mathbf{Y}_G\|_2^2 - \frac{1}{K} \sum_{i=1}^n \|\mathbf{X}_{G_i} - \mathbf{Y}_{G_i}\|_F^2\right| < \varepsilon\} = 1 \quad (14)$$

where  $\mathbf{P}(\bullet)$  represents the probability and  $K = m \times c \times n$ . The detailed proof 145 of *Theorem 1* can be seen in our previous work [57].

Therefore, based on *Theorem 1*, we have the following equation with a very large probability (restricted 1) at each iteration,

$$\frac{1}{N} \|\mathbf{X}_G - \mathbf{Y}_G\|_2^2 = \frac{1}{K} \sum_{i=1}^n \|\mathbf{X}_{G_i} - \mathbf{Y}_{G_i}\|_F^2 \quad (15)$$

Based on Eqs. (13) and (15), we have

$$\begin{aligned}
& \min_{\alpha_G} \frac{1}{2} \|\mathbf{X}_G - \mathbf{Y}_G\|_2^2 + \|\mathbf{W}_G \cdot \alpha_G\|_p \\
&= \min_{\alpha_{G_i}} \sum_{i=1}^n \left( \frac{1}{2} \|\mathbf{X}_{G_i} - \mathbf{Y}_{G_i}\|_F^2 + \frac{K}{N} \|\mathbf{W}_{G_i} \cdot \alpha_{G_i}\|_p \right) \\
&= \min_{\alpha_{G_i}} \sum_{i=1}^n \left( \frac{1}{2} \|\mathbf{Y}_{G_i} - \mathbf{D}_{G_i} \alpha_{G_i}\|_F^2 + \frac{K}{N} \|\mathbf{W}_{G_i} \cdot \alpha_{G_i}\|_p \right) \\
&= \min_{\alpha_{G_i}} \sum_{i=1}^n \left( \frac{1}{2} \|\mathbf{Y}_{G_i} - \mathbf{D}_{G_i} \alpha_{G_i}\|_F^2 + \tau \|\mathbf{W}_{G_i} \cdot \alpha_{G_i}\|_p \right)
\end{aligned} \tag{16}$$

where  $\tau = K/N$ . Clearly, Eq. (16) can be regarded as a sparse representation problem by solving  $n$  sub-problems for all the group  $\mathbf{X}_{G_i}$ .

Note that, dictionary learning are often learned from images, but we have only the linear measurements  $\mathbf{Z}_G$ . Thus, we need to generate a initial image from the linear measurements  $\mathbf{Z}_G$ . In this paper, we first use the Multi-hypothesis block-based compressive sensing (MH-BCS) method [33] to generate the initial image  $\mathbf{X}_G$ . To adapt to the local image structures, instead of learning an over-complete dictionary for each group as in [47], we learn the principle component analysis (PCA) based dictionary [51] for each group  $\mathbf{Y}_{G_i}$ . The proposed PCA-based dictionary learning method is efficient and convenient since it only requires one PCA decomposition operator for each group  $\mathbf{Y}_{G_i}$ , rather than learning the dictionary from natural image dataset with a high computational complexity.

Due to the fact that each dictionary  $\mathbf{D}_{G_i}$  is orthogonal, Eq. (16) is equal to the following formula:

$$\begin{aligned}
\hat{\alpha}_{G_i} &= \min_{\alpha_{G_i}} \sum_{i=1}^n \left( \frac{1}{2} \|\gamma_{G_i} - \alpha_{G_i}\|_F^2 + \tau \|\mathbf{W}_{G_i} \cdot \alpha_{G_i}\|_p \right) \\
&= \min_{\tilde{\alpha}_{G_i}} \sum_{i=1}^n \left( \frac{1}{2} \|\tilde{\gamma}_{G_i} - \tilde{\alpha}_{G_i}\|_2^2 + \tau \|\tilde{\mathbf{w}}_{G_i} \cdot \tilde{\alpha}_{G_i}\|_p \right)
\end{aligned} \tag{17}$$

where  $\mathbf{Y}_i = \mathbf{D}_{G_i} \gamma_{G_i}$  and  $\mathbf{X}_i = \mathbf{D}_{G_i} \alpha_{G_i}$ .  $\tilde{\alpha}_{G_i}$ ,  $\tilde{\gamma}_{G_i}$  and  $\tilde{\mathbf{w}}_{G_i}$  denote the vectorization of the matrix  $\alpha_{G_i}$ ,  $\gamma_{G_i}$  and  $\mathbf{W}_{G_i}$ , respectively.

To obtain the solution of Eq. (17) effectively, in this paper, the generalized soft-thresholding (GST) algorithm [56] is adopted to solve Eq. (17). Specifically,

given  $p$ ,  $\tilde{\gamma}_{G_i}$  and  $\tilde{\mathbf{w}}_{G_i}$ , there exists a specific threshold,

$$\tau_p^{GST}(\tilde{w}_{G_{i,j}}) = (2\tilde{w}_{G_{i,j}}(1-p))^{\frac{1}{2-p}} + \tilde{w}_{G_{i,j}}p(2\tilde{w}_{G_{i,j}}(1-p))^{\frac{p-1}{2-p}} \quad (18)$$

where  $\tilde{\gamma}_{G_{i,j}}$ ,  $\tilde{\alpha}_{G_{i,j}}$  and  $\tilde{w}_{G_{i,j}}$  are the  $j$ -th element of  $\tilde{\gamma}_i$ ,  $\tilde{\alpha}_i$  and  $\tilde{\mathbf{w}}_i$ , respectively. Then, if  $\tilde{\gamma}_{G_{i,j}} < \tau_p^{GST}(\tilde{w}_{G_{i,j}})$ ,  $\tilde{\alpha}_{G_{i,j}} = 0$  is the global minimum. Otherwise, the optimum will be achieved at non-zero point. According to [56], for any  $\tilde{\gamma}_{G_{i,j}} \in (\tau_p^{GST}(\tilde{w}_{G_{i,j}}), +\infty)$ , Eq. (17) has one unique minimum  $\mathbf{S}_p^{GST}(\tilde{\gamma}_{G_{i,j}}; \tilde{w}_{G_{i,j}})$ , which can be obtained by solving the following equation,

$$\mathbf{S}_p^{GST}(\tilde{\gamma}_{G_{i,j}}; \tilde{w}_{G_{i,j}}) - \tilde{\gamma}_{G_{i,j}} + \tilde{w}_{G_{i,j}}p(\mathbf{S}_p^{GST}(\tilde{\gamma}_{G_{i,j}}; \tilde{w}_{G_{i,j}}))^{p-1} = 0 \quad (19)$$

The complete description of the GST algorithm is shown in Table 1. For more details about the GST algorithm, please refer to [56].

Table 1: Generalized Soft-Thresholding (GST)

<b>Input:</b> $\tilde{\gamma}_{G_{i,j}}, \tilde{w}_{G_{i,j}}, p, J$ .	
1.	$\tau_p^{GST}(\tilde{w}_{G_{i,j}}) = (2\tilde{w}_{G_{i,j}}(1-p))^{\frac{1}{2-p}} + \tilde{w}_{G_{i,j}}p(2\tilde{w}_{G_{i,j}}(1-p))^{\frac{p-1}{2-p}};$
2.	<b>If</b> $ \tilde{\gamma}_{G_{i,j}}  \leq \tau_p^{GST}(\tilde{w}_{G_{i,j}})$
3.	$\mathbf{S}_p^{GST}(\tilde{\gamma}_{G_{i,j}}; \tilde{w}_{G_{i,j}}) = 0;$
4.	<b>else</b>
5.	$k = 0, \tilde{\alpha}_{G_{i,j}}^{(k)} =  \tilde{\gamma}_{G_{i,j}} ;$
6.	Iterate on $k = 0, 1, \dots, J$
7.	$\tilde{\alpha}_{G_{i,j}}^{(k+1)} =  \tilde{\gamma}_{G_{i,j}}  - \tilde{w}_{G_{i,j}}p(\tilde{\alpha}_{G_{i,j}}^{(k)})^{p-1};$
8.	$k \leftarrow k + 1;$
9.	$\mathbf{S}_p^{GST}(\tilde{\gamma}_{G_{i,j}}; \tilde{w}_{G_{i,j}}) = \text{sgn}(\tilde{\gamma}_{G_{i,j}})\tilde{\alpha}_{G_{i,j}}^k;$
10.	<b>End</b>
<b>Output:</b> $\mathbf{S}_p^{GST}(\tilde{\gamma}_{G_{i,j}}; \tilde{w}_{G_{i,j}})$ .	

Each weight  $\mathbf{W}_{G_i}$  is assigned to group sparse coefficient  $\alpha_{G_i}$ , large values of each  $\alpha_{G_i}$  usually include major edge and texture information. This implies that to reconstruct  $\mathbf{X}_{G_i}$  from its degraded one, we should shrink large values less, while shrinking smaller ones more [58]. Inspired by [59], the weight  $\mathbf{W}_{G_i}$

of each group is set as  $\tilde{\mathbf{w}}_{G_i} = [\tilde{w}_{G_i,1}, \tilde{w}_{G_i,2}, \dots, \tilde{w}_{G_i,j}]$  and we have

$$\tilde{w}_{G_i,j} = \frac{2\sqrt{2}\sigma^2}{(\tilde{\delta}_{G_i} + \epsilon)} \quad (20)$$

where  $\tilde{\delta}_{G_i}$  denotes the estimated variance of each group sparse coefficient  $\tilde{\gamma}_{G_i}$ , and  $\epsilon$  is a small constant. Obviously, it can be seen that each value of weight  $\mathbf{W}_{G_i}$  is inverse proportion to each value of  $\gamma_{G_i}$  [58]. In light of all derivations, the complete description of the proposed image CS reconstruction using group-based sparse representation via non-convex weighted  $\ell_p$  minimization is given in Table 2.

---

Table 2: The proposed GSR-NCR method for Image CS reconstruction.

---

**Input:** The observed measurement  $\mathbf{Y}_G$ , the measurement matrix  $\phi$ .

**Initialization:** Estimate an initial image  $\mathbf{X}_G^{(0)}$  using a MH-BCS method [33] and

set parameters  $m, c, \rho, p, \sigma, \epsilon, H, J$ ;

**For**  $k = 1, 2, \dots, Max\_iter$  **do**

Update  $\mathbf{Y}_G^{k+1}$  computing by Eq. (11).

Generating the groups  $\mathbf{Y}_{G_i}$  by searching similar patches from  $\mathbf{Y}_G$ .

**For** each group  $\mathbf{Y}_{G_i}$  **do**

Constructing dictionary  $\mathbf{D}_{G_i}^{k+1}$  for  $\mathbf{Y}_{G_i}$  by PCA operator.

Update  $\gamma_{G_i}^{k+1}$  by  $\gamma_{G_i} = \mathbf{D}_{G_i}^{-1} \mathbf{Z}_{G_i}$ .

Update  $\mathbf{W}_{G_i}^{k+1}$  by Eq. (20).

Update  $\alpha_{G_i}^{k+1}$  computing by Table 1.

Get the estimation  $\mathbf{X}_{G_i}^{k+1} = \mathbf{D}_{G_i}^{k+1} \alpha_{G_i}^{k+1}$ .

**End for**

Aggregate all group  $\mathbf{X}_{G_i}^{k+1}$  to form the recovered image  $\hat{\mathbf{X}}_G^{k+1}$ .

**End for**

**Output:**  $\hat{\mathbf{X}}_G^{k+1}$ .

---

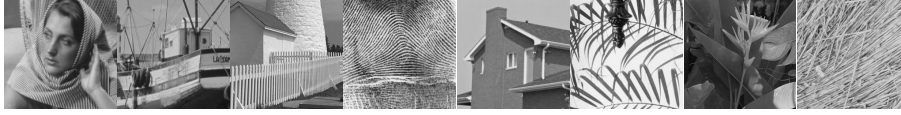


Figure 2: All test images. From left to right: Barbara, boats, Fence, F.print, House, Leaves, plants, staw.

#### 4. Experimental Results

170 In this section, we will report the experimental results of the proposed GSR-NCR for image CS reconstruction. All the experimental images are shown in Fig. 2. To evaluate the quality of the restored images, the PSNR and the recently proposed powerful perceptual quality metric FSIM [60] are calculated. **The source code of the proposed can be downloaded at: <https://drive.google.com/open?id=0B0wKhHwcknCjYmttcFY3Zy1Kd0E>.**  
175

##### 4.1. Parameter Setting

We generate the CS measurements at the block level by using a Gaussian random projection matrix to test images, i.e., the block-based CS reconstruction with block size of  $32 \times 32$ . The parameters are set as follows. The size of each patch  $\sqrt{m} \times \sqrt{m}$  is set to be  $7 \times 7$ . Similar patch numbers  $c = 60$ , the search  
180 window size  $H = 20$ ,  $\sigma = \sqrt{2}$ ,  $\epsilon = 10^{-14}$ ,  $J = 2$ .  $(\rho, p)$  are set to  $(0.3, 0.5)$ ,  $(1.5, 0.95)$  and  $(1.5, 0.95)$  when  $0.2N$ ,  $0.3N$  and  $0.4N$ , respectively.

##### 4.2. Performance Comparison with the State-of-the-Art methods

We have compared the proposed GSR-NCR against six other competing approaches including BCS [11], BM3D-CS [13], ADS-CS [15], SGSR [34], ALSB  
185 [24] and MRK [36]. Note that ADS-CS and ALSB are patch-based sparse representation methods for image CS reconstruction. The PSNR and FSIM results by the competing CS reconstruction methods are shown in Table 3 and Table 4, respectively. It can be seen that the proposed GSR-NCR performs competi-  
190 tively compared to other methods. In terms of PSNR, the proposed GSR-NCR

achieves 7.89dB, 2.85dB, 1.11dB, 1.20dB, 3.22dB and 2.72dB improvement on average over the BCS, BM3D-CS, ADS-CS, SGSR, ALSB and MRK, respectively. Meanwhile, based on the FSIM, the proposed GSR-NCR achieves 0.1031, 0.0339, 0.0123, 0.0063, 0.0106 and 0.0316 improvement on average over the BCS, BM3D-CS, ADS-CS, SGSR, ALSB and MRK, respectively. The visual comparisons of the image CS reconstruction are shown in Figs. 3 - 6. It can be seen that the BCS, BM3D-CS, ADS-CS, SGSR, ALSB and MRK methods still suffer from some undesirable artifacts or over-smooth phenomena. By contrast, the proposed GSR-NCR not only removes most of the visual artifacts, but also preserves large-scale sharp edges and small-scale fine image details more effectively.

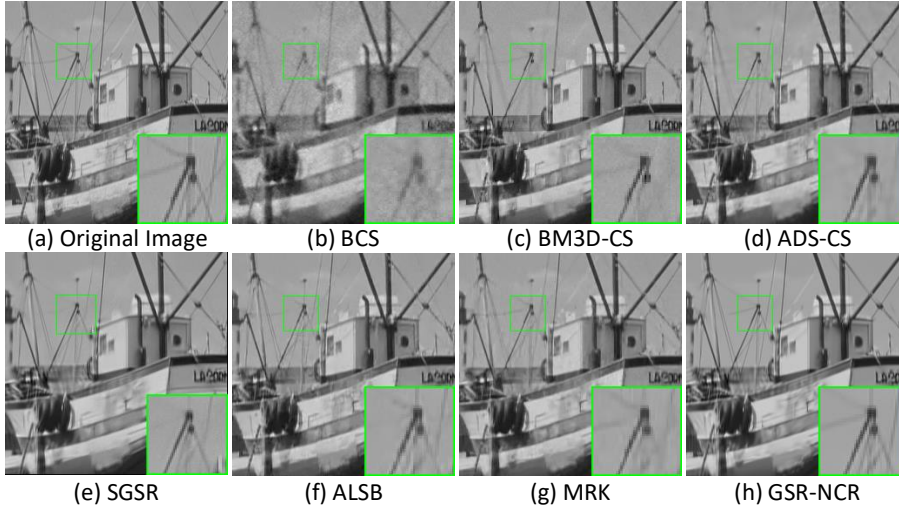


Figure 3: CS reconstructed image *boats* with  $0.2N$  measurements. (a) Original image; (b) BCS [11] (PSNR=27.05dB, FSIM=0.8654); (c) BM3D-CS [13] (PSNR=31.02dB, FSIM=0.9314); (d) ADS-CS [15] (PSNR=33.15dB, FSIM=0.9508); (e) SGSR [34] (PSNR=32.43dB, FSIM=0.9468); (f) ALSB [24] (PSNR= 32.96dB, FSIM=0.9514); MRK [36] (PSNR=32.38dB, FSIM=0.9476); GSR-NCR (PSNR=**33.31dB**, FSIM=**0.9526**).



Figure 4: CS reconstructed image *Barbara* with  $0.2N$  measurements. (a) Original image; (b) BCS [11] (PSNR=22.24dB, FSIM=0.8443); (c) BM3D-CS [13] (PSNR=28.82dB, FSIM=0.9072); (d) ADS-CS [15] (PSNR=32.27dB, FSIM=0.9498); (e) SGSR [34] (PSNR=33.44dB, FSIM=0.9615); (f) ALSB [24] (PSNR= 30.72dB, FSIM=0.9324); MRK [36] (PSNR=27.99dB, FSIM=0.9135); GSR-NCR (PSNR=**33.93dB**, FSIM=**0.9642**).



Table 3: PSNR (dB) Comparisons of BCS [11], BM3D-CS [13], ADS-CS [15], SGSR [34], ALSB [24], MRK [36] and the Proposed GSR-NCR.

<b>Ratio</b>	<b>Method</b>	<i>Barbara</i>	<i>boats</i>	<i>Fence</i>	<i>F.print</i>	<i>House</i>	<i>Leaves</i>	<i>plants</i>	<i>straw</i>	<b><i>Average</i></b>
0.2	BCS	22.24	27.05	21.57	18.50	30.54	21.12	30.67	20.69	24.30
	BM3S-CS	28.82	31.02	26.87	19.37	35.01	28.13	34.98	20.04	28.03
	ADS-CS	32.27	33.15	28.37	22.70	35.76	27.88	35.45	23.75	29.92
	SGSR	33.44	32.43	<b>29.42</b>	23.60	35.81	28.79	34.64	<b>24.54</b>	30.33
	ALSB	30.72	32.96	28.41	<b>23.69</b>	36.08	27.15	32.16	24.33	24.33
	MRK	27.99	32.38	22.20	20.54	36.36	27.75	<b>35.99</b>	23.02	28.28
	GSR-NCR	<b>33.93</b>	<b>33.31</b>	29.10	23.66	<b>36.57</b>	<b>29.03</b>	35.72	24.42	<b>30.72</b>
0.3	BCS	25.59	28.91	23.24	19.96	32.85	23.16	32.81	22.19	26.09
	BM3D-CS	33.01	34.04	30.67	23.01	36.88	32.52	38.30	22.37	31.35
	ADS-CS	35.81	36.35	31.29	25.33	38.21	32.55	38.45	26.58	33.07
	SGSR	35.91	35.22	31.56	25.84	37.37	33.00	37.20	27.34	32.93
	ALSB	35.00	36.42	30.83	25.84	38.34	31.08	38.05	26.61	32.77
	MRK	32.64	34.97	24.44	24.21	38.35	32.37	39.06	25.52	31.45
	GSR-NCR	<b>37.19</b>	<b>37.27</b>	<b>32.26</b>	<b>26.35</b>	<b>39.38</b>	<b>34.95</b>	<b>40.10</b>	<b>27.58</b>	<b>34.38</b>
0.4	BCS	27.10	30.56	24.81	21.67	34.65	25.07	34.77	23.71	27.79
	BM3D-CS	35.92	36.69	33.84	25.47	38.08	35.87	41.18	24.38	33.93
	ADS-CS	38.34	38.79	34.02	27.32	40.30	35.94	40.77	28.80	35.54
	SGSR	37.70	37.41	33.35	27.85	38.99	35.83	39.23	29.63	35.00
	ALSB	37.19	38.92	32.83	27.70	40.25	34.57	40.66	28.54	35.08
	MRK	36.17	37.20	26.63	26.83	40.04	35.53	41.64	27.69	33.97
	GSR-NCR	<b>39.23</b>	<b>39.65</b>	<b>34.39</b>	<b>28.53</b>	<b>41.12</b>	<b>38.55</b>	<b>42.48</b>	<b>30.06</b>	<b>36.75</b>

Table 4: FSIM Comparisons of BCS [11], BM3D-CS [13], ADS-CS [15], SGSR [34], ALSB [24], MRK [36] and the Proposed GSR-NCR.

Ratio	Method	<i>Barbara</i>	<i>boats</i>	<i>Fence</i>	<i>F.print</i>	<i>House</i>	<i>Leaves</i>	<i>plants</i>	<i>straw</i>	<i>Average</i>
0.2	BCS	0.8443	0.8654	0.7653	0.7355	0.9011	0.7531	0.8973	0.7606	0.8153
	BM3S-CS	0.9072	0.9314	0.8325	0.8184	0.9498	0.9231	0.9450	0.7604	0.8835
	ADS-CS	0.9498	0.9508	0.9181	0.8976	0.9423	0.9015	0.9458	0.8704	0.9220
	SGSR	0.9615	0.9468	<b>0.9398</b>	0.9208	0.9502	0.9381	0.9431	<b>0.8856</b>	0.9357
	ALSB	0.9324	0.9514	0.9275	<b>0.9226</b>	0.9563	0.9089	0.9145	0.8830	0.9246
	MRK	0.9135	0.9476	0.7765	0.8397	<b>0.9586</b>	0.9169	<b>0.9555</b>	0.8418	0.8938
	GSR-NCR	<b>0.9642</b>	<b>0.9526</b>	0.9377	0.9224	0.9508	<b>0.9431</b>	0.9505	0.8852	<b>0.9383</b>
0.3	BCS	0.8782	0.8997	0.8345	0.8149	0.9299	0.8018	0.9276	0.8266	0.8641
	BM3D-CS	0.9587	0.9630	0.9572	0.9111	0.9690	0.9601	0.9714	0.8322	0.9403
	ADS-CS	0.9733	0.9728	0.9521	0.9408	0.9667	0.9550	0.9697	0.9220	0.9565
	SGSR	0.9762	0.9684	0.9600	0.9482	0.9648	0.9676	0.9654	0.9316	0.9603
	ALSB	0.9729	0.9746	0.9557	0.9475	0.9732	0.9511	0.9736	0.9226	0.9589
	MRK	0.9611	0.9687	0.8415	0.9225	0.9727	0.9598	0.9768	0.9040	0.9384
	GSR-NCR	<b>0.9816</b>	<b>0.9783</b>	<b>0.9664</b>	<b>0.9534</b>	<b>0.9795</b>	<b>0.9799</b>	<b>0.9819</b>	<b>0.9351</b>	<b>0.9695</b>
0.4	BCS	0.9068	0.9248	0.8807	0.8747	0.9490	0.8422	0.9479	0.8748	0.9001
	BM3D-CS	0.9777	0.9805	0.9758	0.9452	0.9781	0.9803	0.9855	0.8848	0.9635
	ADS-CS	0.9837	0.9835	0.9726	0.9608	0.9803	0.9763	0.9816	0.9487	0.9734
	SGSR	0.9836	0.9793	0.9728	0.9653	0.9759	0.9799	0.9777	0.9570	0.9739
	ALSB	0.9827	0.9840	0.9702	0.9638	0.9824	0.9738	0.9840	0.9479	0.9736
	MRK	0.9795	0.9802	0.8979	0.9539	0.9819	0.9783	0.9873	0.9377	0.9620
	GSR-NCR	<b>0.9879</b>	<b>0.9867</b>	<b>0.9784</b>	<b>0.9702</b>	<b>0.9862</b>	<b>0.9894</b>	<b>0.9892</b>	<b>0.9609</b>	<b>0.9811</b>

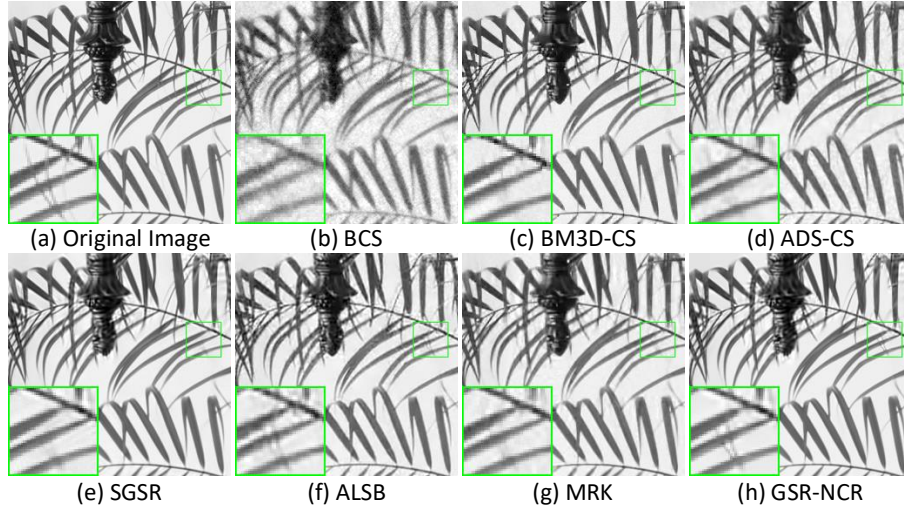


Figure 5: CS reconstructed image *Leaves* with  $0.2N$  measurements. (a) Original image; (b) BCS [11] (PSNR=21.12dB, FSIM=0.7531); (c) BM3D-CS [13] (PSNR=28.13dB, FSIM=0.9231); (d) ADS-CS [15] (PSNR=27.88dB, FSIM=0.9015); (e) SGSR [34] (PSNR=28.79dB, FSIM=0.9381); (f) ALSB [24] (PSNR= 27.15dB, FSIM=0.9089); MRK [36] (PSNR=27.75dB, FSIM=0.9169); GSR-NCR (PSNR=**29.03dB**, FSIM=**0.9431**).

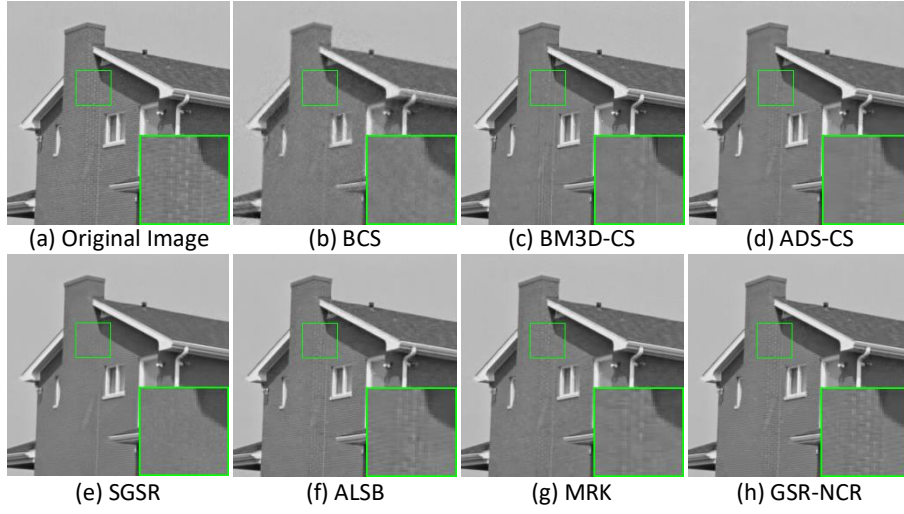


Figure 6: CS recovered *House* images with  $0.3N$  measurements. (a) Original image; (b) BCS [11] (PSNR=32.85dB, FSIM=0.9299); (c) BM3D-CS [13] (PSNR=36.88dB, FSIM=0.9690); (d) ADS-CS [15] (PSNR=38.21dB, FSIM=0.9667); (e) SGSR [34] (PSNR=37.37dB, FSIM=0.9648); (f) ALSB [24] (PSNR= 38.34dB, FSIM=0.9732); MRK [36] (PSNR=38.35dB, FSIM=0.9727); GSR-NCR (PSNR=**39.38dB**, FSIM=**0.9795**).

#### 4.3. Effect of the number of the best matched patches

We have discussed how to select the best matching patch numbers  $c$  for the performance of the proposed GSR-NCR. Specifically, to investigate the sensitivity of our method against  $c$ , two experiments were conducted with respect to different  $c$ , ranging from 20 to 160, in the case of  $0.2N$  and  $0.3N$  measurements, respectively. The results with different  $c$  are shown in Fig. 7. It can be seen that all the curves are almost flat, showing the performance of the proposed GSR-NCR scheme is insensitive to  $c$ . The best performance of each case was usually achieved with  $c$  in the range  $[40, 80]$ . Therefore, in this paper  $c$  was empirically set to be 60.

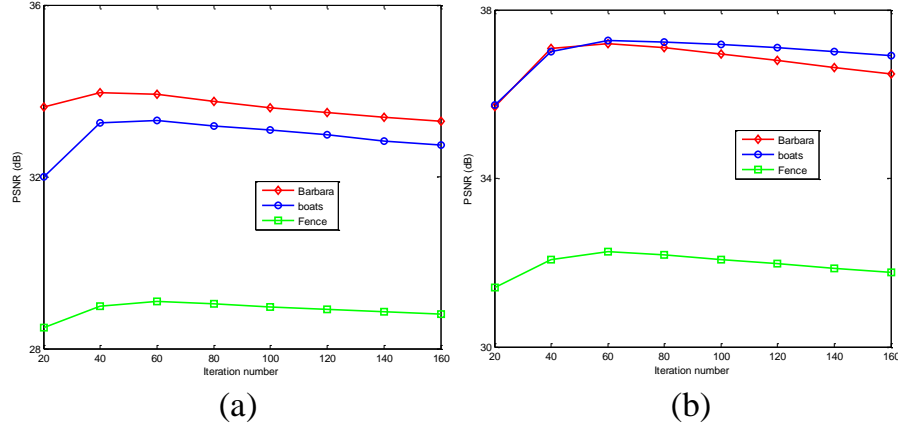


Figure 7: Performance comparison with different matched patch numbers  $c$  for image CS reconstruction. (a) PSNR results achieved by different  $c$  in the case of  $0.2N$  measurements. (b) PSNR results achieved by different  $c$  in the case of  $0.3N$  measurements.

#### 4.4. Convergence analysis

Since the proposed GSR-NCR model (Eq. (6)) is non-convex, it is difficult to give its theoretical proof for global convergence. Here, we only provide empirical evidence to illustrate the good convergence of the proposed CS reconstruction method. Fig. 8 illustrates the convergent performance of the proposed GSR-NCR. It shows the curves of the PSNR values versus the iteration numbers

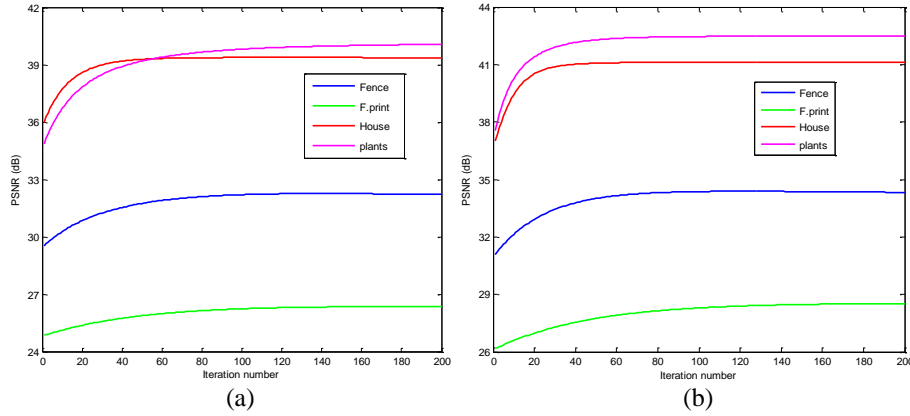


Figure 8: Convergence analysis of the proposed GSR-NCR. (a) PSNR results versus iteration numbers for image CS reconstruction with  $0.3N$  measurements; (b) PSNR results versus iteration numbers for image CS reconstruction with  $0.4N$  measurements.

for four test images with  $0.3N$  and  $0.4N$  measurements, respectively. One can observe that with the increase of the iteration numbers, the PSNR curves gradually increase and ultimately become flat and stable, showing good stability of the proposed non-convex GSR-NCR model.

## 5. Conclusion

In this paper, we proposed a efficient method for image CS reconstruction using group-based sparse representation model, which is able to more accurately enforce the local sparsity and nonlocal self-similarity of images simultaneously in a unified framework. Different from the typical sparsity-promoting convex  $\ell_1$  minimization methods, we extended the non-convex weighted  $\ell_p$  ( $0 < p < 1$ ) penalty function on group sparse coefficients of the data matrix to replace the convex  $\ell_1$ -norm. To reduce the computational complexity, we learned the principle component analysis (PCA) based dictionary for each group to substitute for the dictionary with a high computational complexity learned from natural image dataset. Furthermore, to make the proposed model tractable and robust, an efficient iterative shrinkage/thresholding algorithm was adopted to

solve the non-convex minimization problem. Experimental results have shown  
235 that the proposed method not only outperforms many state-of-the-art methods  
both quantitatively and qualitatively, but also results in fine stability.

## 6. Acknowledge

This work was supported by the NSFC (61571220, 61462052, 61502226 and  
61601362). In addition, the first author Zhiyuan Zha thanks to God for meeting  
240 Dr. Jie Zhang of Beihang University. Jie Zhang is the best beautiful girl in  
Zhiyuan's heart and Zhiyuan wants to love Jie all his life.

## References

- [1] E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact  
signal reconstruction from highly incomplete frequency information, IEEE  
245 Transactions on information theory 52 (2) (2006) 489–509.
- [2] D. L. Donoho, Compressed sensing, IEEE Transactions on information the-  
ory 52 (4) (2006) 1289–1306.
- [3] E. J. Candès, M. B. Wakin, An introduction to compressive sampling, IEEE  
signal processing magazine 25 (2) (2008) 21–30.
- 250 [4] M. Lustig, D. Donoho, J. M. Pauly, Sparse mri: The application of com-  
pressed sensing for rapid mr imaging, Magnetic resonance in medicine 58 (6)  
(2007) 1182–1195.
- [5] J. Ma, F.-X. Le Dimet, Deblurring from highly incomplete measurements  
for remote sensing, IEEE Transactions on Geoscience and Remote Sensing  
255 47 (3) (2009) 792–802.
- [6] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kel-  
ly, R. G. Baraniuk, Single-pixel imaging via compressive sampling, IEEE  
signal processing magazine 25 (2) (2008) 83–91.

- [7] W. U. Bajwa, J. D. Haupt, A. M. Sayeed, R. D. Nowak, Joint source–  
 260 channel communication for distributed estimation in sensor networks, *IEEE Transactions on Information Theory* 53 (10) (2007) 3629–3653.
- [8] M. A. Figueiredo, R. D. Nowak, S. J. Wright, Gradient projection for s-  
 parse reconstruction: Application to compressed sensing and other inverse  
 problems, *IEEE Journal of selected topics in signal processing* 1 (4) (2007)  
 265 586–597.
- [9] J. A. Tropp, A. C. Gilbert, Signal recovery from random measurements via  
 orthogonal matching pursuit, *IEEE Transactions on information theory*  
 53 (12) (2007) 4655–4666.
- [10] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm  
 270 for linear inverse problems with a sparsity constraint, *Communications on pure and applied mathematics* 57 (11) (2004) 1413–1457.
- [11] S. Mun, J. E. Fowler, Block compressed sensing of images using directional  
 transforms, in: *Image Processing (ICIP), 2009 16th IEEE International Conference on*, IEEE, 2009, pp. 3021–3024.
- [12] J. Yang, J. Wright, T. S. Huang, Y. Ma, Image super-resolution via sparse  
 275 representation, *IEEE transactions on image processing* 19 (11) (2010) 2861–2873.
- [13] K. Egiazarian, A. Foi, V. Katkovnik, Compressed sensing image recon-  
 struction via recursive spatially adaptive filtering, in: *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, Vol. 1, IEEE, 2007,  
 280 pp. I–549.
- [14] J. Zhang, D. Zhao, C. Zhao, R. Xiong, S. Ma, W. Gao, Image compressive  
 sensing recovery via collaborative sparsity, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 2 (3) (2012) 380–391.



- 285 [15] W. Dong, G. Shi, X. Li, L. Zhang, X. Wu, Image reconstruction with locally adaptive sparsity and nonlocal robust regularization, *Signal Processing: Image Communication* 27 (10) (2012) 1109–1122.
- [16] X. Yuan, S. Pang, Structured illumination temporal compressive microscopy, *Biomedical Optics Express* 7 (2016) 746–758.
- 290 [17] C. Yan, H. Xie, D. Yang, J. Yin, Y. Zhang, Q. Dai, Supervised hash coding with deep neural network for environment perception of intelligent vehicles, *IEEE Transactions on Intelligent Transportation Systems* 19 (1) (2018) 284–295. doi:10.1109/TITS.2017.2749965.
- [18] Y. Shen, J. Li, Z. Zhu, W. Cao, Y. Song, Image reconstruction algorithm from compressed sensing measurements by dictionary learning, *Neurocomputing* 151 (2015) 1153–1162.
- 295 [19] J. M. Bioucas-Dias, M. A. Figueiredo, A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration, *IEEE Transactions on Image processing* 16 (12) (2007) 2992–3004.
- 300 [20] X. Yuan, Generalized alternating projection based total variation minimization for compressive sensing, in: *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 2539–2543.
- [21] Y. Wang, J. Yang, W. Yin, Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, *SIAM Journal on Imaging Sciences* 1 (3) (2008) 248–272.
- 305 [22] S. Becker, J. Bobin, E. J. Candès, NESTA: A fast and accurate first-order method for sparse recovery, *SIAM Journal on Imaging Sciences* 4 (1) (2011) 1–39.
- [23] X. Yuan, J. Yang, P. Lhull, X. Liao, G. Sapiro, D. J. Brady, L. Carin, Adaptive temporal compressive sensing for video, in: *2013 IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 14–18.
- 310

- [24] J. Zhang, C. Zhao, D. Zhao, W. Gao, Image compressive sensing recovery using adaptively learned sparsifying basis via  $l_0$  minimization, *Signal Processing* 103 (2014) 114–126.
- 315 [25] X. Yuan, Y. Pu, Parallel lensless compressive imaging via deep convolutional neural networks, *Opt. Express* 26 (2) (2018) 1962–1977. doi:10.1364/OE.26.001962.  
URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-26-2-1962>
- 320 [26] Z. Zha, X. Liu, X. Zhang, Y. Chen, L. Tang, Y. Bai, Q. Wang, Z. Shang, Compressed sensing image reconstruction via adaptive sparse nonlocal regularization, *The Visual Computer* 34 (1) (2018) 117–137. doi:10.1007/s00371-016-1318-9.  
URL <https://doi.org/10.1007/s00371-016-1318-9>
- 325 [27] W. Dong, G. Shi, X. Li, Y. Ma, F. Huang, Compressive sensing via nonlocal low-rank regularization, *IEEE Transactions on Image Processing* 23 (8) (2014) 3618–3632.
- [28] X. Yuan, S. Pang, Compressive video microscope via structured illumination, in: 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 1589–1593.
- 330 [29] N. Eslahi, A. Aghagolzadeh, Compressive sensing image restoration using adaptive curvelet thresholding and nonlocal sparse regularization, *IEEE Transactions on Image Processing* 25 (7) (2016) 3126–3140.
- [30] N. Eslahi, A. Aghagolzadeh, S. M. H. Andargoli, Image/video compressive sensing recovery using joint adaptive sparsity measure, *Neurocomputing* 200 (2016) 88–109.
- 335 [31] Y. Yan, J. Sun, H. Li, Z. Xu, et al., Deep admm-net for compressive sensing mri, in: *Advances in Neural Information Processing Systems*, 2016, pp. 10–18.

- 340 [32] X. Yuan, X. Liao, P. Llull, D. Brady, L. Carin, Efficient patch-based approach for compressive depth imaging, *Appl. Opt.* 55 (27) (2016) 7556–7564.
- [33] C. Chen, E. W. Tramel, J. E. Fowler, Compressed-sensing recovery of images and video using multihypothesis predictions, in: *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, IEEE, 2011, pp. 1193–1198.
- 345 [34] J. Zhang, D. Zhao, F. Jiang, W. Gao, Structural group sparse representation for image compressive sensing recovery, in: *Data Compression Conference (DCC), 2013*, IEEE, 2013, pp. 331–340.
- 350 [35] X. Yuan, P. Llull, X. Liao, J. Yang, D. J. Brady, G. Sapiro, L. Carin, Low-cost compressive sensing for color video and depth, in: *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3318–3325. doi:10.1109/CVPR.2014.424.
- [36] T. N. Canh, K. Q. Dinh, B. Jeon, Multi-scale/multi-resolution kronecker compressive imaging, in: *Image Processing (ICIP), 2015 IEEE International Conference on*, IEEE, 2015, pp. 2700–2704.
- 355 [37] B. Wen, S. Ravishankar, Y. Bresler, Fristflipping and rotation invariant sparsifying transform learning and applications, *Inverse Problems* 33 (074007) (2017) 074007.
- 360 [38] X. Yuan, T.-H. Tsai, R. Zhu, P. Llull, D. J. Brady, L. Carin, Compressive hyperspectral imaging with side information, *IEEE Journal of Selected Topics in Signal Processing* 9 (6) (2015) 964–976.
- [39] A. N. Tikhonov, A. Goncharsky, V. Stepanov, A. G. Yagola, *Numerical methods for the solution of ill-posed problems*, Vol. 328, Springer Science & Business Media, 2013.
- 365

- [40] M. Aharon, M. Elad, A. Bruckstein, k-svd: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Transactions on signal processing* 54 (11) (2006) 4311–4322.
- [41] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Transactions on Image processing* 15 (12) (2006) 3736–3745.
- [42] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, in: *Proceedings of the 26th annual international conference on machine learning*, ACM, 2009, pp. 689–696.
- [43] Q. Zhang, B. Li, Discriminative k-svd for dictionary learning in face recognition, in: *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on, IEEE, 2010, pp. 2691–2698.
- [44] B. Wen, S. Ravishankar, Y. Bresler, Structured overcomplete sparsifying transform learning with convergence guarantees and applications, *International Journal of Computer Vision* 114 (2-3) (2015) 137–167.
- [45] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 2, IEEE, 2005, pp. 60–65.
- [46] X. Zhang, M. Burger, X. Bresson, S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, *SIAM Journal on Imaging Sciences* 3 (3) (2010) 253–276.
- [47] J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman, Non-local sparse models for image restoration, in: *Computer Vision, 2009 IEEE 12th International Conference on*, IEEE, 2009, pp. 2272–2279.
- [48] J. Zhang, D. Zhao, W. Gao, Group-based sparse representation for image restoration, *IEEE Transactions on Image Processing* 23 (8) (2014) 3336–3351.

- [49] W. Dong, G. Shi, Y. Ma, X. Li, Image restoration via simultaneous sparse coding: Where structured sparsity meets gaussian scale mixture, International Journal of Computer Vision 114 (2-3) (2015) 217–232.
- [50] Z. Zha, X. Liu, Z. Zhou, X. Huang, J. Shi, Z. Shang, L. Tang, Y. Bai, Q. Wang, X. Zhang, Image denoising via group sparsity residual constraint, in: Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on, IEEE, 2017, pp. 1787–1791.
- [51] W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally centralized sparse representation for image restoration, IEEE Transactions on Image Processing 22 (4) (2013) 1620–1630.
- [52] Z. Zha, X. Liu, X. Huang, H. Shi, Y. Xu, Q. Wang, L. Tang, X. Zhang, Analyzing the group sparsity based on the rank minimization methods, in: 2017 IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 883–888. doi:10.1109/ICME.2017.8019334.
- [53] A. Beck, M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, IEEE Transactions on Image Processing 18 (11) (2009) 2419–2434.
- [54] Q. Lyu, Z. Lin, Y. She, C. Zhang, A comparison of typical  $\ell_p$  minimization algorithms, Neurocomputing 119 (2013) 413–424.
- [55] R. Chartrand, B. Wohlberg, A nonconvex admm algorithm for group sparsity with sparse groups, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, IEEE, 2013, pp. 6009–6013.
- [56] W. Zuo, D. Meng, L. Zhang, X. Feng, D. Zhang, A generalized iterated shrinkage algorithm for non-convex sparse coding, in: Proceedings of the IEEE international conference on computer vision, 2013, pp. 217–224.

- [57] Z. Zha, X. Zhang, Y. Wu, Q. Wang, L. Tang, A comparative study for  
 420 the weighted nuclear norm minimization and nuclear norm minimization,  
 arXiv preprint arXiv:1608.04517.
- [58] E. J. Candes, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted  $\ell_1$   
 minimization, *Journal of Fourier analysis and applications* 14 (5-6) (2008)  
 877–905.
- 425 [59] S. G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image  
 denoising and compression, *IEEE Transactions on image processing* 9 (9)  
 (2000) 1532–1546.
- [60] L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: A feature similarity index for  
 image quality assessment, *IEEE transactions on Image Processing* 20 (8)  
 430 (2011) 2378–2386.