J. Vis. Commun. Image R. 36 (2016) 28–39

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Denoising by low-rank and sparse representations ☆

Mansour Nejati [a], Shadrokh Samavi [a,b,*], Harm Derksen [c], Kayvan Najarian [d]

[a] Department of Electrical and Computer Engineering, Isfahan University of Technology, Iran
[b] Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada
[c] Department of Mathematics, University of Michigan, Ann Arbor, USA
[d] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, USA

ARTICLE INFO

Article history:
Received 17 April 2015
Accepted 6 January 2016
Available online 11 January 2016

Keywords:
Image denoising
Sparse representation
Low-rank matrix recovery
Nonlocal self-similarity
Random matrix theory
Dictionary learning
Rank minimization
Optimal singular value shrinkage

ABSTRACT

Due to the ill-posed nature of image denoising problem, good image priors are of great importance for an effective restoration. Nonlocal self-similarity and sparsity are two popular and widely used image priors which have led to several state-of-the-art methods in natural image denoising. In this paper, we take advantage of these priors and propose a new denoising algorithm based on sparse and low-rank representation of image patches under a nonlocal framework. This framework consists of two complementary steps. In the first step, noise removal from groups of matched image patches is formulated as recovery of low-rank matrices from noisy data. This problem is then efficiently solved under asymptotic matrix reconstruction model based on recent results from random matrix theory which leads to a parameter-free optimal estimator. Nonlocal learned sparse representation is adopted in the second step to suppress artifacts introduced in the previous estimate. Experimental results, demonstrate the superior denoising performance of the proposed algorithm as compared with the state-of-the-art methods.

1. Introduction

This paper addresses the problem of image denoising in which the goal is to reconstruct the latent clean image $\mathbf{x}$ from its noise-degraded observation $\mathbf{y} = \mathbf{x} + \mathbf{w}$ where $\mathbf{w}$ is additive white Gaussian noise (AWGN) with zero mean and standard deviation σ. This problem is a classical yet active topic in low level image processing that serves as an important pre-processing step for many vision applications and provides a convenient test-bed over which various statistical image modeling methods can be assessed [1].

As image denoising is typically an ill-posed problem, the solution might not be unique. Therefore, natural image priors are widely used in order to regularize the possible solution spaces into the desired one. In fact, image priors are of utmost importance for an effective noise removal; hence various image priors have been developed [1]. A very strong prior for natural images is the nonlocal self-similarity (NSS) of small image patches within the image. The nonlocal means algorithm [2] is the first attempt to explicitly exploit nonlocal self-similarity for image denoising. This influential work has unleashed a flood of studies on nonlocal image restoration. Hence, many variants of nonlocal means [3–10] and several advanced nonlocal image denoising algorithms [11–20] have been developed.

Another well-known prior model is sparse representation which assumes that the clean signal can be well approximated by a linear combination of few basis elements – or atoms – from a set called a dictionary [21]. The idea of sparsity is traced back to the late 1980s when the sparsity of the wavelet coefficients was considered as an appropriate prior knowledge of natural images [21,22], leading to the famous shrinkage algorithm [23]. Evolution of this idea and development of overcomplete dictionaries introduced efficient and provably effective algorithms based on greedy pursuit [24,25] or convex optimization [26] to compute signal representations over arbitrary overcomplete dictionaries. Specializing the dictionary atoms in representing the intended signal and producing data adaptive dictionaries have motivated a wide range of investigations on dictionary learning during the last decade, resulting in several state-of-the-art dictionary learning algorithms such as K-SVD [27], online dictionary learning (ODL) [28], and recursive least squares dictionary learning (RLS-DL) [29].

By using the learned sparse representation models, promising results have been obtained in image and video denoising [22,30–32]. These methods produce patch-wise estimates and the final denoising results are calculated by aggregating the multiple estimates for pixels lying on the patch overlaps. Despite good denoising performance compared to point-wise estimators such as nonlocal means algorithm [2], these methods do not take into

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author at: Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada.
E-mail address: samavi@mcmaster.ca (S. Samavi).

http://dx.doi.org/10.1016/j.jvcir.2016.01.004
1047-3203/

account the patch redundancy within image. It has been shown that joint use of sparsity and nonlocal self-similarity priors provides stronger image model, as exhibited in the state-of-the-art denoising methods BM3D [15] and LSSC [16]. The sparsifying dictionaries employed in these methods are kept fixed for all groups of similar image patches. However, sub-dictionaries adapted to each group could be used for better image modeling.

More recently, low-rank approximation for extracting low-dimensional structures in data has attracted attention in science and engineering resulting in an explosion of research in its theory and algorithms. Low-rank matrix approximation, for recovering of a low-rank matrix from its corrupted observation, appears in very many applications in various fields including computer vision, machine learning, signal processing, and bioinformatics. For instance, use of low-rank approximation can be found in applications such as face recognition [33], background modeling [34], medical image reconstruction [35,36], image alignment [37], video denoising [38], and image restoration [20,39] among others. As the generalization of sparse structures to correlated signals, low-rank approximation provides an effective approach toward modeling of nonlocal self-similarities in natural images.

In this paper, we develop a nonlocal image denoising approach in which two steps of low-rank approximation and sparse representation are employed. The first step of our algorithm is built upon the methodology of patch grouping and collaborative filtering where the proposed low-rank regularized collaborative filtering is applied. Indeed, noise removal from groups of matched image patches is formulated as low-rank matrix denoising. Based on recent results from random matrix theory, this problem is solved under asymptotic matrix reconstruction model leading to an optimal singular value shrinkage operator. In the second step, nonlocal learned sparse representation model is adopted to improve the shortcomings of the first step in flat image areas and to reduce artifacts around edges. This sparse model exploits the nonlocal redundancies to obtain a more accurate estimate of the original image.

In summary, our main contributions are as follows: (a) We exploit both sparsity and low-rank priors within a nonlocal denoising framework. (b) For the patch-based image denoising, we introduce a low-rank matrix estimator based on an optimal singular value shrinker, which does not require any threshold tuning. We prove that this shrinkage function can be applied to obtain optimal solution of weighted rank minimization problem with Frobenius norm data fidelity. (c) We apply nonlocal sparse representation model using a sparsifying overcomplete dictionary learned from the first-step estimate. The nonlocal redundancy is exploited to modify the initial sparse representations and produce a more accurate estimate of the original image.

The experimental results demonstrate that our proposed algorithm, called Sparse and Low-rank Representation based Denoising (SLRD), has superior performance compared with the state-of-the-art methods in both peak signal-to-noise ratio and visual quality.

The rest of the paper is organized as follows. In Section 2 we elaborate on the details of low-rank matrix recovery in presence of noise and its integration into a nonlocal denoising scheme. Nonlocal sparse model as second part of our algorithm is also described in this section. Experimental results, comparison with the state-of-the-art methods, and objective assessments are presented in Section 3. Finally, Section 4 concludes the paper.

2. Proposed denoising scheme

Motivated by the great success of sparse representation and dictionary learning in various image restoration applications and considerable progress of low-rank approximation in recent years, we developed a new image denoising algorithm by considering both concepts of sparse representation and low-rank models. The denoising procedure is accomplished in two successive steps of the low-rank representation of nonlocal similarities and the sparse representation with respect to a learned overcomplete dictionary. The details of these steps are presented in the following subsections.

2.1. Step 1: Low-rank representation

In recent years there has been a surge of interest in data approximation by low-dimensional models such as sparsity, low-rank structures, and manifolds [40]. The basic idea of proposed Low-rank Representation based Denoising (LRD) approach is to approximate true noise-free image patches by low-rank modeling of image nonlocal similarities. In other words, the image patches are grouped by block matching, such that the patches in each group share similar underlying image structures. Thus, stack of these similar patches into a matrix, with vectorized image patches as its columns, forms a noisy version of an approximately low-rank matrix. This low-rank characteristic of image nonlocal similarities is demonstrated in Fig. 1. Based on this observation, recovery of original image patches can be cast as the problem of low-rank matrix recovery. This is done with the goal of estimating the latent low-rank matrix $\mathbf{X}$ from its noisy observation $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{X} + \sigma\mathbf{W} \tag{1}$$

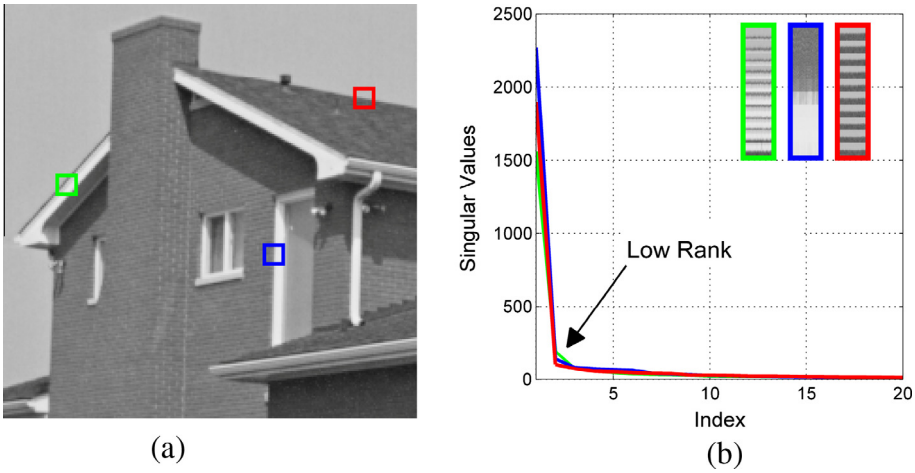

Fig. 1. Low-rank characteristic of natural image blocks due to the nonlocal self-similarities. (a) "House" image, (b) singular values of three matrices formed from similar patches corresponding to the reference ones shown in (a). These matrices have been shown on top of singular values plot.

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m] \in \mathbb{R}^{n\times m}$ is a patch matrix composed by m similar $\sqrt{n}\times\sqrt{n}$ patches from noisy image and $\mathbf{W}$ denotes the noise matrix with i.i.d. entries $W_{i,j} \sim \mathcal{N}(0,1)$.

A natural approach for approximating a low-rank matrix from noisy data consists in truncating, or hard-thresholding of the singular values of the observed matrix $\mathbf{Y}$. Let $g : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}$ denotes the matrix recovery scheme. Let $\mathbf{Y} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}' = \sum_{i=1}^{\min(n,m)} \lambda_i \mathbf{u}_i \mathbf{v}_i'$ be the singular value decomposition (SVD) of $\mathbf{Y}$, in which $\mathbf{u}_i \in \mathbb{R}^n$ and $\mathbf{v}_i \in \mathbb{R}^m$, $i = 1, \ldots, \min(n,m)$ are the left and right singular vectors of $\mathbf{Y}$ corresponding to the singular value λ_i. Then the hard-thresholding scheme can be expressed as

$$g_\tau^H(\mathbf{Y}) = \sum_{i=1}^{\min(n,m)} \lambda_i \mathbf{1}(\lambda_i \geqslant \tau)\mathbf{u}_i\mathbf{v}_i' \tag{2}$$

where $\tau > 0$ is a threshold parameter and $\mathbf{1}(\cdot)$ is the indicator function. A popular alternative estimator applies a soft thresholding rule to the singular values of observation matrix:

$$g_\tau^S(\mathbf{Y}) = \sum_{i=1}^{\min(n,m)} (\lambda_i - \tau)_+ \mathbf{u}_i\mathbf{v}_i' \tag{3}$$

where $(x)_+ = \max(x, 0)$. In this estimation strategy, which is referred to as singular value thresholding (SVT) [41], all the singular values are shrunk toward zero by the same value τ. In practice, hard and soft thresholding schemes require appropriate selection of threshold level and optimal choice of this parameter has been suggested in some recent works [42,43]. However, it has been shown that these low-rank recovery schemes are sub-optimal in the matrix denoising problem, and in general, they may not provide a good estimate of the underlying signal matrix to be recovered [44,45].

In our denoising algorithm we use low-rank matrix recovery method of Shabalin and Nobel [45] who developed an optimal singular value shrinker in the mean square error (MSE) sense. Under an asymptotic framework, they designed a new matrix recovery method by analyzing the effects of noise on singular value decomposition of signal matrix $\mathbf{X}$ using recent results from random matrix theory [46–48]. In such asymptotic framework, it is assumed that the rank of $\mathbf{X}$ and its non-zero singular values remains fixed as the matrix dimensions n and m grow, at the same rate, to infinity such that the matrix aspect ratio converges to $\frac{n}{m} \to \beta$. The asymptotic observation model for n-by-m matrices is considered as follows [45]:

$$\mathbf{Y} = \mathbf{X} + \frac{1}{\sqrt{m}}\mathbf{W} \tag{4}$$

where the noise matrix $\mathbf{W}$ has i.i.d. entries $W_{i,j} \sim \mathcal{N}(0,1)$. Given a noisy matrix $\mathbf{Y}$ obeying the above model, an estimate $\widehat{\mathbf{X}}$ of the original matrix $\mathbf{X}$ based on singular value shrinkage has the general form of:

$$\widehat{\mathbf{X}} = g(\mathbf{Y}) = \sum_{i=1}^{\min(n,m)} \eta(\lambda_{\mathrm{Y},i})\mathbf{u}_{\mathrm{Y},i}\mathbf{v}_{\mathrm{Y},i}' \tag{5}$$

where $\eta : [0,\infty) \to [0,\infty)$ is the shrinkage function applied on singular values of $\mathbf{Y}$. In the above equation and in what follows, the subscripts Y and X are used to respectively indicate singular values and singular vectors of matrices $\mathbf{Y}$ and $\mathbf{X}$.

Ideally, in the recovery problem, we would like to find an estimate $\widehat{\mathbf{X}}$ closest to the original signal matrix $\mathbf{X}$ in terms of mean square error (MSE); that is, we should seek a shrinkage function η that minimize the Frobenius norm loss (MSE),

$$L(\mathbf{X}, \widehat{\mathbf{X}}) = \|\mathbf{X} - \widehat{\mathbf{X}}\|_F^2 = \left\| \sum_{i=1}^{r} \lambda_{\mathrm{X},i}\mathbf{u}_{\mathrm{X},i}\mathbf{v}_{\mathrm{X},i}' - \sum_{i=1}^{\min(n,m)} \omega_i \mathbf{u}_{\mathrm{Y},i}\mathbf{v}_{\mathrm{Y},i}' \right\|_F^2 \tag{6}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, first sum is the SVD of $\mathbf{X}$ with $r = rank(\mathbf{X})$, and the second sum is the estimator based on singular value shrinkage in which $\omega_i = \eta(\lambda_{\mathrm{Y},i})$. By using recent advances in random matrix theory, Shabalin and Nobel [45] have provided some results regarding asymptotic behavior of singular values and singular vectors of model (4). These results show that the singular values of $\mathbf{X}$ smaller than $\sqrt[4]{\beta}$ cannot be asymptotically recovered from the observed matrix $\mathbf{Y}$. Since the asymptotic relation $\lambda_{\mathrm{Y},i} \leqslant 1 + \sqrt{\beta}$ implies $\lambda_{\mathrm{X},i} \leqslant \sqrt[4]{\beta}$, we can restrict the second sum in (6) to the top r_0 singular values of $\mathbf{Y}$ where $r_0 = \#\{j : \lambda_{\mathrm{Y},j} > 1 + \sqrt{\beta}\}$ and for other singular values we set the corresponding ω_i to zero. It has also been shown that left singular vectors $\mathbf{u}_{\mathrm{X},i}$ and $\mathbf{u}_{\mathrm{Y},j}$, and right singular vectors $\mathbf{v}_{\mathrm{X},i}$ and $\mathbf{v}_{\mathrm{Y},j}$ for $i = 1, \ldots, r$, $j = 1, \ldots, r_0$ and $i \neq j$ are asymptotically orthogonal. Based on these properties, we can rewrite (6) as:

$$L(\mathbf{X}, \widehat{\mathbf{X}}) = \sum_{i=1}^{r_0} \left\| \omega_i \mathbf{u}_{\mathrm{Y},i}\mathbf{v}_{\mathrm{Y},i}' - \lambda_{\mathrm{X},i}\mathbf{u}_{\mathrm{X},i}\mathbf{v}_{\mathrm{X},i}' \right\|_F^2 + \sum_{i=r_0+1}^{r} \lambda_{\mathrm{X},i}^2 \tag{7}$$

By expanding the i-th term of the first sum, differentiating with respect to ω_i, and setting this result equal to zero, the optimal shrinker can be obtained:

$$\omega_i^* = \eta^*(\lambda_{\mathrm{Y},i}) = \lambda_{\mathrm{X},i}\langle \mathbf{u}_{\mathrm{X},i}, \mathbf{u}_{\mathrm{Y},i}\rangle\langle \mathbf{v}_{\mathrm{X},i}, \mathbf{v}_{\mathrm{Y},i}\rangle \tag{8}$$

The derived optimal shrinker η^* depends on the singular values and singular vectors of unobserved matrix $\mathbf{X}$ and thus cannot be directly measured. Fortunately, based on the asymptotic relations between the singular values and singular vectors of the signal and observed matrices, these terms are estimated as:

$$\hat{\lambda}_{\mathrm{X},i}^2 = \frac{1}{2}\left[(\lambda_{\mathrm{Y},i}^2 - 1 - \beta) + \sqrt{(\lambda_{\mathrm{Y},i}^2 - 1 - \beta)^2 - 4\beta} \right], \tag{9}$$

$$\hat{\theta}_i^2 = \left(1 - \frac{\beta}{\hat{\lambda}_{\mathrm{X},i}^4}\right) \Big/ \left(1 + \frac{\beta}{\hat{\lambda}_{\mathrm{X},i}^2}\right), \tag{10}$$

$$\hat{\phi}_i^2 = \left(1 - \frac{\beta}{\hat{\lambda}_{\mathrm{X},i}^4}\right) \Big/ \left(1 + \frac{1}{\hat{\lambda}_{\mathrm{X},i}^2}\right), \tag{11}$$

where $\hat{\lambda}_{\mathrm{X},i}$, $\hat{\theta}_i$ and $\hat{\phi}_i$ are the estimates of $\lambda_{\mathrm{X},i}$, $\langle \mathbf{u}_{\mathrm{X},i}, \mathbf{u}_{\mathrm{Y},i}\rangle$, and $\langle \mathbf{v}_{\mathrm{X},i}, \mathbf{v}_{\mathrm{Y},i}\rangle$ respectively. The complete definition for optimal singular value shrinkage function can be written as

$$\eta^*(\lambda_{\mathrm{Y},i}) = \begin{cases} \hat{\lambda}_{\mathrm{X},i}\hat{\theta}_i\hat{\phi}_i & \lambda_{\mathrm{Y},i} > 1 + \sqrt{\beta} \\ 0 & \lambda_{\mathrm{Y},i} \leqslant 1 + \sqrt{\beta} \end{cases} \tag{12}$$

and the low-rank matrix recovery method is defined by the following equation.

$$g^*(\mathbf{Y}) = \sum_{i=1}^{\min(n,m)} \eta^*(\lambda_{\mathrm{Y},i})\mathbf{u}_{\mathrm{Y},i}\mathbf{v}_{\mathrm{Y},i}' \tag{13}$$

Fig. 2 compares the hard and soft thresholding operators with the optimal shrinkage function of (12) for threshold value of $\tau = 2$ and aspect ratio of $\beta = 1$. This figure reveals that the optimal estimator in MSE yields a nonconvex shrinker in which the amount of shrinkage varies with the singular values. The larger singular values are shrunk less than the smaller ones It should be noted that the above matrix recovery method is calibrated for the model $\mathbf{Y} = \mathbf{X} + \mathbf{W}/\sqrt{m}$. To apply this method to n-by-m data matrix $\mathbf{Y}$ from the general model $\mathbf{Y} = \mathbf{X} + \sigma\mathbf{W}$, we should use the following estimator

$$g_{m,\sigma}^*(\mathbf{Y}) = \sqrt{m}\sigma \cdot g^*\left(\frac{\mathbf{Y}}{\sqrt{m}\sigma}\right) \tag{14}$$

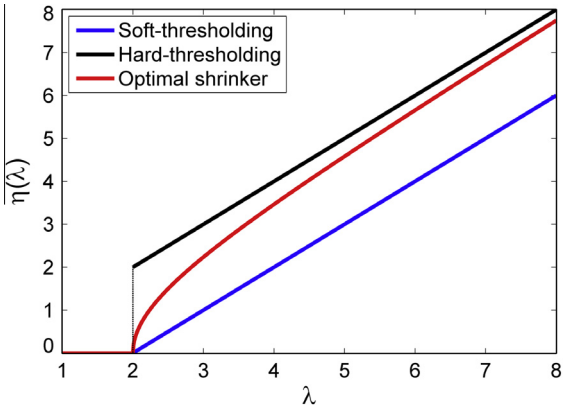

Fig. 2. Comparison of hard-thresholding, soft-thresholding, and optimal shrinkage operators.

Using this low-rank estimator, we then develop an image denoising algorithm by exploiting the image nonlocal self-similarity. The algorithm is started by dividing the input noisy image $\mathbf{y}$ into a set of overlapping patches $\{\mathbf{y}_i\}$, where $\mathbf{y}_i \in \mathbb{R}^n$ is the i-th patch in vectorized form. A patch matrix $\mathbf{Y} = [\mathbf{y}_{i_1}, \ldots, \mathbf{y}_{i_m}] \in \mathbb{R}^{n\times m}$ is created for each exemplar patch $\mathbf{y}_i$ by stacking its m nearest neighbors in Euclidean space. As explained above, the latent clean patch matrix $\mathbf{X}$ corresponding to the observation data $\mathbf{Y}$ is assumed to be a low-rank matrix. Thus, the estimation of $\mathbf{X}$ can be formulated as a low-rank minimization problem:

$$\widehat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{Y}-\mathbf{X}\|_F^2 + \tau\|\mathbf{X}\|_* \tag{15}$$

where the first term is data fidelity function and τ is a positive constant. Also, $\|\mathbf{X}_*\|$ denotes the nuclear norm of the matrix $\mathbf{X}$ which is defined as the sum of its singular values, i.e. $\|\mathbf{X}\|_* = \sum_{i=1}^{r}\lambda_{\mathrm{X},i}$, $r = \min(n,m)$. The nuclear norm is the tightest convex surrogate of rank operator and enforces low-rank in solution [41]. It was proved in [41] that the solution to the problem in (15) can be easily obtained by applying the soft-thresholding operator on the singular values of observation matrix. That is

$$\widehat{\mathbf{X}} = \mathbf{U}g_\tau^s(\mathbf{\Sigma})\mathbf{V}' \tag{16}$$

where $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$ is SVD of $\mathbf{Y}$ with $\mathbf{\Sigma} = \operatorname{diag}(\{\lambda_i\}_{1\leqslant i\leqslant r})$, and $g_\tau^s(\cdot)$ is the soft-thresholding operator defined in (3). As $\mathbf{\Sigma}$ is a diagonal matrix, $g_\tau^s(\mathbf{\Sigma})$ is equivalent to soft-thresholding of each diagonal element λ_i in $\mathbf{\Sigma}$:

$$g_\tau^s(\mathbf{\Sigma}) = \operatorname{diag}(\{(\lambda_i-\tau)_+\}_{1\leqslant i\leqslant r}) \tag{17}$$

This method treats all singular values equally by shrinking them with the same amount. In fact, nuclear norm regularization in problem of (15) penalizes all singular values the same. However, it is more appropriate to be able to give different weights according to our prior knowledge of the problem. Furthermore, from the optimal shrinker described above, we know that the amount of optimal shrinkage varies with the singular values. Therefore, we employ weighted nuclear norm regularization in order to improve the flexibility of the low-rank minimization problem as follows:

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y}-\mathbf{X}\|_F^2 + \sum_{i=1}^{\min(n,m)} w_i\lambda_{\mathrm{X},i} \tag{18}$$

where each singular value $\lambda_{\mathrm{X},i}$ is assigned a non-negative weight $w_i \geqslant 0$. This minimization problem is not convex in general, which can make it difficult to solve. But for a certain ordering of weights, there exists a closed form solution.

Lemma 1 [49, Theorem 2.3]. *For any $\mathbf{Y} \in \mathbb{R}^{n\times m}$ and weights satisfying $0 \leqslant w_1 \leqslant \cdots \leqslant w_r$ $(r = \min(n,m))$, a globally optimal solution to the optimization problem in (18) is given by*

$$\widehat{\mathbf{X}} = g_{\boldsymbol{w}}^s(\mathbf{Y}) = \mathbf{U}g_{\boldsymbol{w}}^s(\mathbf{\Sigma})\mathbf{V}' \tag{19}$$

where $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$ is SVD of $\mathbf{Y}$ and $g_{\boldsymbol{w}}^s(\mathbf{\Sigma}) = \operatorname{diag}(\{(\lambda_i - w_i)_+\}_{1\leqslant i\leqslant r})$ is the generalized soft-thresholding operator with weight vector $\boldsymbol{w} = [w_1, \ldots, w_r]$.

According to Lemma 1 when the weights are in a non-increasing order, such that weight decreases with singular value, the optimal solution of problem in (18) can be efficiently obtained. This is the estimate of the original patch matrix. Fortunately, such a constraint is met if we compute the weights $\{w_i\}$ using the optimal singular value shrinkage function η^* defined in (12). By plugging (9)–(11) into the definition of η^*, subtracting it from an identity singular value shrinker and after some simplifications, we have

$$w^*(\lambda) = \begin{cases} \frac{\lambda^2 - \sqrt{(\lambda^2-\beta-1)^2-4\beta}}{\lambda} & \lambda \geqslant 1+\sqrt{\beta} \\ 1+\sqrt{\beta} & \lambda < 1+\sqrt{\beta} \end{cases} \tag{20}$$

where $w^*(\lambda)$ gives the optimal weight corresponding to the singular value λ. By calculating the derivative of $w^*(\lambda)$ with respect to λ, for the case that $\lambda \geqslant 1+\sqrt{\beta}$, we obtain

$$\frac{\partial w^*(\lambda)}{\partial\lambda} = \frac{(\beta-1)^2 - \lambda^4}{\lambda^2\sqrt{(\lambda^2-\beta-1)^2-4\beta}} + 1 \quad \text{for } \lambda \geqslant 1+\sqrt{\beta} \tag{21}$$

It is easy to show that the above expression is negative for all $\lambda \geqslant 1+\sqrt{\beta}$. Thus, $w^*(\lambda)$ is a strictly decreasing function of λ over an interval $[1+\sqrt{\beta}, \infty)$ and its maximum value is $w^*(1+\sqrt{\beta}) = 1+\sqrt{\beta}$. Considering singular values smaller than $1+\sqrt{\beta}$, the optimal weighting function $w^*(\lambda)$ generates a weakly decreasing weight sequence that satisfies the constraint of Lemma 1.

In order to use this weighting function in the weighted nuclear norm minimization problem, we need to scale the data. This is due to the observation model of (4) for which the optimal shrinkage function is calibrated. Hence, we first compute $\dot{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{m}\sigma}$ and the weights $w_i = w^*(\dot{\lambda}_i)$, $i = 1, \ldots, \min(n,m)$ using the singular values of $\dot{\mathbf{Y}}$. We then use the generalized soft-thresholding operator defined in (19) to calculate $\widehat{\dot{\mathbf{X}}} = g_{\boldsymbol{w}}^s(\dot{\mathbf{Y}})$. Finally, the estimate $\widehat{\mathbf{X}}$ of the original patch matrix is obtained as $\widehat{\mathbf{X}} = \sqrt{m}\sigma\widehat{\dot{\mathbf{X}}}$. One of the advantages of this method is that it is parameter-free and does not require any threshold tuning.

After doing so for all patch matrices, the denoised patches are aggregated to obtain the whole image estimate. This procedure is iterated for several rounds to enhance the denoising outputs. Inspired by work [50], we adopt the following strategy at the start of each iteration to update the noisy image that the algorithm gets as an input.

$$\mathbf{y}^{(t)} = \hat{\mathbf{x}}^{(t-1)} + \delta(\mathbf{y} - \hat{\mathbf{x}}^{(t-1)}) \tag{22}$$

where t indicates the iteration number, $\hat{\mathbf{x}}$ is the estimate of original image, and δ is a small positive constant that controls the amount of noise fed back to the next iteration. This strategy leads to performance improvement by better recovery of fine scale image structures [50].

Due to the change of noise level in different iterations, we re-estimate it from residual image (defined as the difference between the noisy image and its denoised version) as follows

$$\hat{\sigma}^{(t)} = \gamma\sqrt{\sigma^2 - \frac{1}{N}\|\mathbf{y}-\mathbf{y}^{(t)}\|_2^2} \tag{23}$$

where γ is used as a scaling factor for the re-estimation of noise variance and N is the total number of image pixels. The proposed low-rank representation based denoising (LRD) algorithm is summarized in Algorithm 1.

Algorithm 1. Image denoising by LRD

- Input: Noisy image $\mathbf{y}$
- Initialization: $\hat{\mathbf{x}}^{(0)} = \mathbf{y}$.
- Iterate on $t = 1, \ldots, J$
  Iterative regularization: $\mathbf{y}^{(t)} = \hat{\mathbf{x}}^{(t-1)} + \delta(\mathbf{y} - \hat{\mathbf{x}}^{(t-1)})$.
  Estimation of noise level $\hat{\sigma}^{(t)}$ via Eq. (23).
  for each patch $\mathbf{y}_i$ in $\mathbf{y}^{(t)}$ **do**
    Create patch matrices $\mathbf{Y}_i$ by block matching.
    Compute $\dot{\mathbf{Y}}_i = \frac{\mathbf{Y}_i}{\sqrt{m}\sigma}$ and SVD of $\dot{\mathbf{Y}}_i$.
    Compute weights for singular values of $\dot{\mathbf{Y}}_i$ via Eq. (20).
    Get the estimate $\widehat{\mathbf{X}}_i = \sqrt{m}\sigma g^s_{\mathbf{w}}(\dot{\mathbf{Y}}_i)$.
  end for
  Aggregate all denoised patches $\widehat{\mathbf{X}}_i$ to reconstruct $\hat{\mathbf{x}}^{(t)}$.
- Output: $\hat{\mathbf{x}}^{(J)}$

2.2. Step 2: Sparse representation

Sparse representation models usually work in a patch-based framework where an image is redundantly represented by sparse coefficients of its overlapping patches. Then the recovery of image would consist of averaging the sparse representation of all overlapping patches. Let $\mathbf{x}_i = \mathbf{R}_i\mathbf{x}$ denote a $\sqrt{n} \times \sqrt{n}$ patch at location i in image $\mathbf{x}$ which is extracted and ordered as a column vector by the patch extraction operator $\mathbf{R}_i$. The operator $\mathbf{R}_i$ is a binary matrix that extracts $\mathbf{x}_i$ and converts it to a column vector. Sparse decomposition of the image patches $\{\mathbf{x}_i\}$ over a dictionary $\mathbf{D} \in \mathbb{R}^{n\times K}$ results in sparse representations $\{\boldsymbol{\alpha}_i\}$, $\boldsymbol{\alpha}_i \in \mathbb{R}^K$. From these representations, the whole image can be reconstructed by aggregating all of the $\mathbf{D}\boldsymbol{\alpha}_i$ patches as follows [22]

$$\mathbf{x} \approx \mathbf{D} \circ \boldsymbol{\alpha} \triangleq \left(\sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1}\left(\sum_i \mathbf{R}_i^T\mathbf{D}\boldsymbol{\alpha}_i\right) \tag{24}$$

where $\circ$ denotes the patch aggregation operator and $\boldsymbol{\alpha}$ is the concatenation of all $\boldsymbol{\alpha}_i$. Consider a noisy observation $\mathbf{y}$ of a clean image $\mathbf{x}$ of the form $\mathbf{y} = \mathbf{x} + \mathbf{w}$ where $\mathbf{w}$ is assumed to be zero mean i.i.d. Gaussian noise with standard deviation σ. We aim to find sparse representations $\hat{\boldsymbol{\alpha}}_i$ from patches of $\mathbf{y}$ that are close enough to original ones, such that $\hat{\mathbf{x}} = \mathbf{D} \circ \hat{\boldsymbol{\alpha}}$ would be a good estimate of $\mathbf{x}$. A commonly used approach for denoising a patch is to obtain sparse decomposition of that patch by solving the minimization problem of

$$\hat{\boldsymbol{\alpha}}_i = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \mu\|\boldsymbol{\alpha}\|_p \tag{25}$$

where $p = 0, 1$ and $\|\cdot\|_p$ is sparsity-inducing norm serving as local sparsity regularization term. Once $\hat{\boldsymbol{\alpha}}_i$ is found, the denoising result is obtained by $\hat{\mathbf{x}}_i = \mathbf{D}\hat{\boldsymbol{\alpha}}_i$. However, recovering of true sparse representation from a noisy patch is very challenging and using only the local sparse representation model (25) may not lead to an accurate enough estimate [18].

In natural images, there are usually recurrent patterns, in the form of small image patches, across image. These nonlocal redundancies can be exploited to improve the sparse representation model. State-of-the-art denoising methods, such as BM3D [15] and LSSC [16], are built upon such considerations. Following this idea, we propose a nonlocal sparse model in which the sparse representation of similar patches helps us to get a more accurate estimate of the original signal. On the other hand, dictionary plays a critical role for the success of sparse representation modeling and data-adaptive learned overcomplete dictionaries are the most popular ones in recent years. In contrast to [17,18], where the authors propose to train orthogonal dictionaries, in our approach we learn a sparsifying overcomplete dictionary which provides more flexibility in description of data in the representation domain. Given a set of training image patches $\{\mathbf{x}_i\}_{i=1}^N$, learning a sparsifying dictionary $\mathbf{D}$ is typically formulated as

$$\min_{\mathbf{D}\in\mathcal{D},\{\boldsymbol{\alpha}_i\}_{i=1}^N} \sum_{i=1}^{N}\left(\frac{1}{2}\|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \mu\|\boldsymbol{\alpha}_i\|_p\right) \tag{26}$$

where $\mathcal{D}$ is usually defined as the set of all dictionaries with unit column-norms. To solve this problem, several algorithms have been developed in recent years [27–29]. In our denoising method, we adopt the mini-batch variant of online dictionary learning (ODL) algorithm [28] which utilizes an online optimization algorithm based on stochastic approximations to minimize the ℓ_1-regularized dictionary learning problem. It has been shown in [28] that ODL has faster convergence than batch alternatives and it can produce better dictionaries for denoising as compared to the popular K-SVD algorithm [27]. Furthermore, we use the denoised image resulted from the previous step to train a dictionary instead of using the input noisy image. This can enhance the quality of the trained dictionary atoms, which subsequently improves the denoising results.

Given the learned dictionary, we first find the initial sparse representations $\boldsymbol{\alpha}_i^{(0)}$ of all overlapping patches by solving the minimization problem of (25) with $p = 0$. This choice of p leads to ℓ_0 pseudo norm regularization term which has usually better reconstruction performance than its convex ℓ_1 counterpart [16]. As ℓ_0 minimization problem is NP-hard in general, greedy approaches are employed to get an approximate solution. One of the most widely used greedy approach is the orthogonal matching pursuit (OMP) that successively selects the best atom of dictionary minimizing the representation error, until a stopping criterion is satisfied. In our algorithm we employ a modified version of OMP, referred to as generalized OMP (GOMP) [51], which allows selection of multiple atoms per iteration. Simultaneous selection of multiple atoms reduces the number of iterations. Moreover, we empirically found that identification of two atoms at each step yields better denoising results.

It is known that natural images have local and nonlocal correlations. These correlations can be partly observed in sparse representation domain too. Hence, after computing the sparse representations $\boldsymbol{\alpha}_i^{(0)}$ based on (25), we exploit the representation of nonlocal similar patches to produce better estimate of the original image. The nonlocal sparse model adopted here is a modification of (25) in which the local sparsity regularization term is replaced by a nonlocal one:

$$\hat{\boldsymbol{\alpha}}_i = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \mu\|\boldsymbol{\alpha} - \boldsymbol{\kappa}_i\|_1 \tag{27}$$

Here $\boldsymbol{\kappa}_i$ is defined as the weighted average of initial sparse representations associated with patches similar to the patch $\mathbf{y}_i$. Suppose that Ω_i is a set containing indices of patches similar to $\mathbf{y}_i$ (including i). Then $\boldsymbol{\kappa}_i$ for i-th patch is computed as follows

$$\boldsymbol{\kappa}_i = \sum_{j\in\Omega_i} w_{i,j}\boldsymbol{\alpha}_j^{(0)} \tag{28}$$

where the weights $w_{i,j}$ depend on the similarity between i-th patch and other patches of Ω_i as expressed in Eq. (29):

$$w_{i,j} = \frac{1}{A_i} \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|_2^2}{h}\right), \quad \forall j \in \Omega_i \tag{29}$$

where $A_i = \sum_{j\in\Omega_i} \exp(-\|\mathbf{y}_i - \mathbf{y}_j\|_2^2/h)$ is a normalization factor. Also, the smoothing parameter h is a positive number controlling the exponential decay of weights with increasing of the patch-wise Euclidean distances.

According to the above mentioned nonlocal sparse model, the new sparse representations $\hat{\boldsymbol{\alpha}}_i$ are obtained by solving the ℓ_1-minimization problem in Eq. (27). This problem is convex and can be efficiently solved using iterative soft-thresholding (IST) algorithm [52]. Based on the idea of surrogate functions of [52], iterative solution to the problem (27) is given by:

$$\boldsymbol{\alpha}_i^{(t+1)} = S_\tau(\mathbf{v}_i^{(t)} - \boldsymbol{\kappa}_i) + \boldsymbol{\kappa}_i, \tag{30}$$

where

$$\mathbf{v}_i^{(t)} = \boldsymbol{\alpha}_i^{(t)} - \frac{1}{c}\mathbf{D}^T(\mathbf{D}\boldsymbol{\alpha}_i^{(t)} - \mathbf{y}_i) \tag{31}$$

and $S_\tau(\cdot) = \text{sign}(\cdot).(|\cdot| - \tau)_+$ is the soft-thresholding operator, $\tau = \frac{\mu}{c}$, and c is a constant to guarantee the strict convexity of surrogate function and it is selected so that $\mathbf{D}^T\mathbf{D} < c$ [52]. After J iterations, all estimated patches $\hat{\mathbf{x}}_i = \mathbf{D}\boldsymbol{\alpha}_i^{(J)}$ are averaged to obtain the final denoised image. We initialize the IST algorithm with sparse representations $\boldsymbol{\alpha}_i^{(0)}$ obtained by error-constrained GOMP. By doing this, a few iterations of IST is needed to obtain the improved sparse representations based on nonlocal sparse model (27). A complete description of denoising procedure in the second step of our algorithm is summarized in Algorithm 2. We call this step as sparse representation based denoising (SRD). Our complete image denoising algorithm is referred to as SLRD, which applies both of LRD and SRD methods in succession in order to produce the final denoising result. The outline of SLRD is presented in Algorithm 3.

Algorithm 2. Image denoising by SRD

- Input: Noisy image $\mathbf{y}$ and dictionary $\mathbf{D}$
- Patch estimate:
  for each patch $\mathbf{y}_i$ in $\mathbf{y}$ **do**
    Compute $\boldsymbol{\alpha}_i^{(0)}$ by solving Eq. (25) using GOMP.
    Find similar patches.
    Compute $\boldsymbol{\kappa}_i$ using Eqs. (28) and (29).
    Iterate on $t = 1, \ldots, J$
      Compute $\mathbf{v}_i^{(t-1)}$ via Eq. (31).
      Update representation: $\boldsymbol{\alpha}_i^{(t)} = S_\tau(\mathbf{v}_i^{(t-1)} - \boldsymbol{\kappa}_i) + \boldsymbol{\kappa}_i$.
  end for
- Aggregate all denoised patches $\hat{\mathbf{x}}_i = \mathbf{D}\boldsymbol{\alpha}_i^{(J)}$ to reconstruct $\hat{\mathbf{x}}$.
- Output: denoised image $\hat{\mathbf{x}}$.

Algorithm 3. Image denoising via SLRD

- Input: Noisy image $\mathbf{y}$
- Step 1: Apply **LRD** on $\mathbf{y}$ to obtain basic estimate $\tilde{\mathbf{x}}$.
- Iterative regularization: $\tilde{\mathbf{y}} = \tilde{\mathbf{x}} + \delta(\mathbf{y} - \tilde{\mathbf{x}})$.
- Update noise level: $\tilde{\sigma} = \gamma\sqrt{\sigma^2 - \|\mathbf{y} - \tilde{\mathbf{y}}\|_2^2/N}$.
- Learn dictionary $\mathbf{D}$ from overlapping patches of $\tilde{\mathbf{y}}$ via ODL.
- Step 2: Apply **SRD** on $\tilde{\mathbf{y}}$ to obtain final estimate $\hat{\mathbf{x}}$.
- Output: $\hat{\mathbf{x}}$

Fig. 3. The 12 test images used in the experiments.

Table 1
The basic parameter settings of SLRD.

Noise level	$\sigma \leqslant 15$		$15 < \sigma \leqslant 30$		$30 < \sigma \leqslant 60$		$60 < \sigma$	
Parameters	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2
n	6×6	5×5	7×7	6×6	8×8	7×7	10×10	9×9
m	50	20	60	20	70	20	100	20
J	6	4	7	4	10	5	14	5
K	–	256	–	256	–	512	–	512
δ	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
γ	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34

Table 2
Denoising performance of LRD algorithm based on PSNR (dB) with three different SVD-based low-rank matrix recovery methods. Columns *O*, *H* and *S* respectively show results based on singular value optimal shrinkage, hard thresholding, and soft thresholding. Best results are in bold.

σ	10			30			50			100		
Method	*O*	*H*	*S*	*O*	*H*	*S*	*O*	*H*	*S*	*O*	*H*	*S*
C. Man	**34.38**	34.12	33.74	**28.58**	28.37	27.95	**26.45**	26.03	25.62	**23.43**	22.97	22.72
House	**36.96**	36.86	36.42	**32.53**	32.29	31.80	**30.45**	30.03	29.64	**26.98**	26.47	26.11
Monarch	**34.89**	34.62	34.02	**28.80**	28.55	28.02	**26.24**	25.73	25.43	**22.81**	22.12	21.85
Parrot	**33.77**	33.60	33.29	**28.35**	28.08	27.71	**26.18**	25.70	25.42	**23.30**	22.75	22.48
Peppers	**34.91**	34.69	34.23	**29.45**	29.33	28.65	**26.94**	26.56	26.18	**23.67**	23.09	22.74
Straw	**31.62**	31.45	30.68	25.37	**25.45**	24.42	22.87	**22.91**	21.99	19.53	**19.72**	18.91
Barbara	**35.40**	35.04	34.72	**30.30**	30.13	29.53	**27.78**	27.62	27.02	**24.50**	24.29	23.76
Boat	**34.01**	33.85	33.58	**29.16**	29.11	28.57	**26.91**	26.71	26.33	**24.08**	23.85	23.55
Couple	**34.09**	33.94	33.58	**28.91**	28.86	28.31	**26.61**	26.48	26.02	**23.47**	23.34	22.98
Hill	**33.75**	33.65	33.45	**29.21**	29.19	28.86	**27.29**	27.21	26.92	**24.59**	24.49	24.21
Lena	**36.05**	35.81	35.61	**31.46**	31.39	30.93	**29.30**	29.15	28.78	**26.36**	26.03	25.88
Man	**34.18**	34.08	33.63	**28.94**	28.92	28.52	**26.90**	26.75	26.48	**24.33**	24.11	23.94
Average	**34.50**	34.31	33.91	**29.26**	29.14	28.61	**26.99**	26.74	26.32	**23.92**	23.60	23.26

Table 3
PSNR (dB) improvements obtained using both of sparse representation and low-rank models in SLRD algorithm compared to LRD which applies low-rank representation model alone.

σ	10			30			50			100		
Method	LRD	SLRD	Improve.	LRD	SLRD	Improve.	LRD	SLRD	Improve.	LRD	SLRD	Improve.
C. Man	34.38	34.44	0.06	28.58	28.76	0.18	26.45	26.64	0.19	23.43	23.64	0.21
House	36.96	37.05	0.09	32.53	32.69	0.16	30.45	30.67	0.22	26.98	27.17	0.19
Monarch	34.89	34.95	0.06	28.80	28.93	0.13	26.24	26.42	0.18	22.81	23.05	0.24
Parrot	33.77	33.78	0.01	28.35	28.39	0.04	26.18	26.27	0.09	23.30	23.46	0.16
Peppers	34.91	35.02	0.11	29.45	29.55	0.10	26.94	27.08	0.14	23.67	23.84	0.17
Straw	31.62	31.64	0.02	25.37	25.44	0.07	22.87	22.91	0.04	19.53	19.67	0.14
Barbara	35.40	35.38	-0.02	30.30	30.34	0.04	27.78	27.85	0.07	24.50	24.54	0.04
Boat	34.01	34.09	0.08	29.16	29.27	0.11	26.91	27.04	0.13	24.08	24.22	0.14
Couple	34.09	34.17	0.08	28.91	29.02	0.11	26.61	26.74	0.13	23.47	23.62	0.15
Hill	33.75	33.83	0.08	29.21	29.31	0.10	27.29	27.41	0.12	24.59	24.75	0.16
Lena	36.05	36.10	0.05	31.46	31.51	0.05	29.30	29.38	0.08	26.36	26.48	0.12
Man	34.18	34.24	0.06	28.94	29.06	0.12	26.90	27.02	0.12	24.33	24.45	0.12
Average	34.50	34.56	0.06	29.26	29.36	0.10	26.99	27.12	0.13	23.92	24.07	0.15

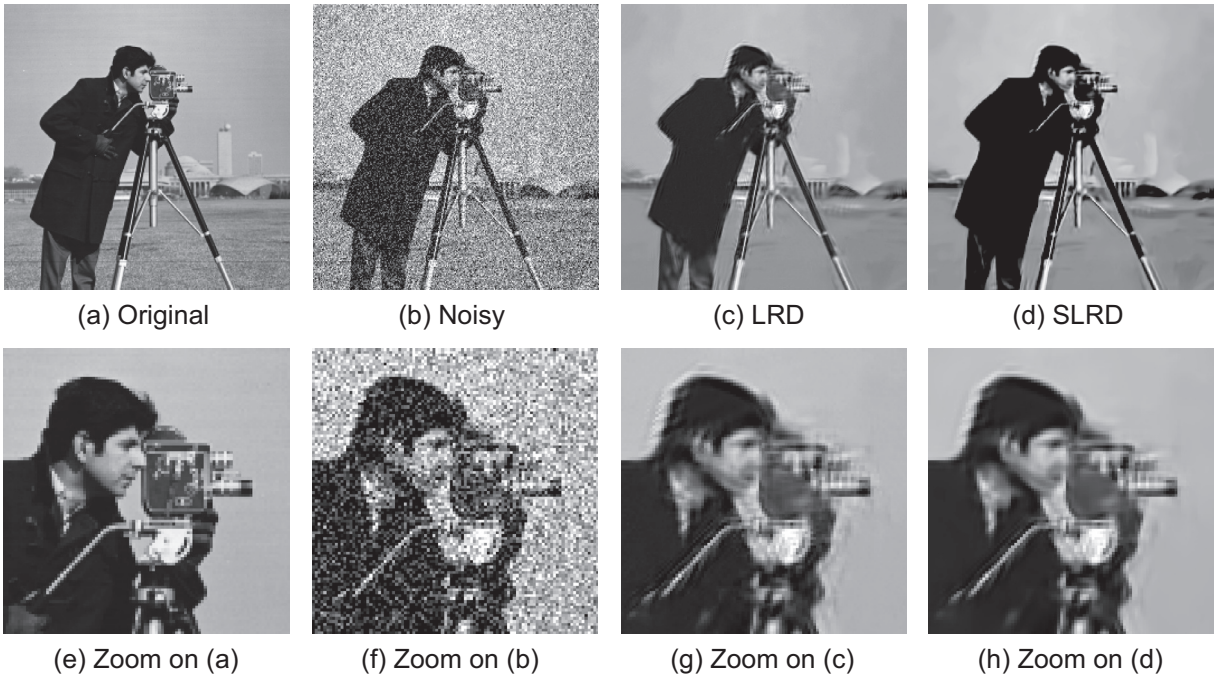

Fig. 4. Comparison between the denoising results of LRD and SLRD for noisy "Cameraman" image at $\sigma = 50$ noise level. Note that the SLRD which exploits both low-rank approximation and sparse representation based denoising models, produces visually more pleasant results with fewer artifacts than LRD.

3. Experiments

In this section, we present our experimental results obtained by the proposed SLRD algorithm to verify its denoising performance. As done in other papers [15,16,18,20], we experimented with a set of 12 commonly used benchmark images shown in Fig. 3, at low, moderate, and high noise levels. We discuss the parameter setting of SLRD algorithm in Section 3.1. In Section 3.2 we compare the optimal singular value shrinker with two popular hard and soft thresholding operators. The evaluations of basic and final estimates, which are obtained respectively after the first and second steps of SLRD, are presented in Section 3.3. We also compare SLRD algorithm with recently developed state-of-the-art denoising methods. In our quantitative evaluations, peak signal-to-noise ratio (PNSR) is used as performance measure. The Matlab code of our SLRD algorithm can be downloaded at http://mansournejati.ece.iut.ac.ir/content/slrd.

3.1. Parameter setting

The proposed algorithm has several parameters including patch size n, number of similar patches m used to create patch matrices in first step and to compute nonlocal sparse representations in second step of SLRD algorithm. Also, iterative regularization parameter δ, noise re-estimate scale factor γ, number of dictionary atoms K, smoothing parameter h, and iteration number J are among other control variables. In our experiments, the two parameters δ and γ are fixed to 0.2 and 0.34 respectively which are observed to be appropriate, in terms of denoising performance, for all test images across all noise levels. The smoothing parameter h is chosen based on an empirical rule of $h = 7.5\sigma\sqrt{n}$. Experimentally we found this to be close to optimal for different images and noise variances.

The patch size, cardinality of similar patch groups, and number of iterations in the first step of SLRD are set depending on the noise level as follows: $n = 6 \times 6$ and $m = 50$ for $\sigma \leqslant 15$; $n = 7 \times 7$ and $m = 60$ for $15 < \sigma \leqslant 30$; $n = 8 \times 8$ and $m = 70$ for $30 < \sigma \leqslant 60$; and, $n = 10 \times 10$ and $m = 100$ for $60 < \sigma$. Number of iterations is also set to 6, 7, 10, and 14 respectively, on the mentioned four noise levels. These parameters were tuned to produce the best results in terms of PSNR. As the amount of noise is reduced after the first step, the best patch size for the second step of SLRD would thus be smaller than that of the first step. By using GOMP for initializing the IST, we found empirically that a few iterations of the IST algorithm are sufficient to obtain improved sparse

Table 4
PSNR comparison of proposed SLRD with five current state-of-the-art methods BM3D [15], LSSC [16], NCSR [18], SAIST [20], and SGHP [19]. Best results are in bold.

σ	5						10					
Method	BM3D	LSSC	NCSR	SAIST	SGHP	SLRD	BM3D	LSSC	NCSR	SAIST	SGHP	SLRD
C. Man	38.29	38.32	38.26	38.36	37.37	**38.52**	34.19	34.24	34.18	34.30	33.78	**34.44**
House	39.83	39.93	39.94	39.92	39.46	**40.09**	36.72	36.95	36.80	36.66	36.30	**37.05**
Monarch	38.21	38.57	38.49	38.76	37.65	**38.91**	34.12	34.46	34.51	34.76	34.08	**34.95**
Parrot	37.86	37.87	37.82	37.96	37.42	**38.03**	33.57	33.61	33.56	33.66	33.43	**33.78**
Peppers	38.12	38.19	38.11	38.17	37.90	**38.37**	34.68	34.80	34.68	34.82	34.55	**35.02**
Straw	35.37	35.91	35.83	36.03	35.24	**36.10**	30.92	31.38	31.45	31.60	31.17	**31.64**
Barbara	38.31	38.48	38.37	38.55	37.58	**38.72**	34.98	34.99	35.00	35.24	34.31	**35.38**
Boat	37.28	37.35	37.34	37.25	37.30	**37.44**	33.92	34.03	33.91	33.91	33.81	**34.09**
Couple	37.52	37.45	37.50	37.46	37.45	**37.67**	34.04	34.01	34.00	33.96	33.87	**34.17**
Hill	37.13	37.17	37.18	37.13	37.19	**37.30**	33.62	33.68	33.69	33.65	33.59	**33.83**
Lena	38.72	38.69	38.74	38.67	38.75	**38.94**	35.93	35.86	35.85	35.90	35.63	**36.10**
Man	37.82	37.89	37.85	37.92	37.82	**38.05**	33.98	34.11	34.05	34.12	33.93	**34.24**
Average	37.87	37.99	37.95	38.02	37.59	**38.18**	34.22	34.34	34.31	34.38	34.04	**34.56**
σ	15						30					
C. Man	31.92	32.06	32.01	32.08	31.77	**32.17**	28.64	28.64	28.59	28.36	28.59	**28.76**
House	34.93	**35.28**	35.05	34.87	34.78	35.25	32.10	32.34	32.06	32.31	31.91	**32.69**
Monarch	31.86	32.13	32.26	32.47	31.91	**32.64**	28.36	28.20	28.47	28.65	28.28	**28.93**
Parrot	31.38	31.40	31.37	31.44	31.33	**31.63**	28.12	27.97	28.04	28.12	28.13	**28.39**
Peppers	32.70	32.82	32.66	32.85	32.57	**33.04**	29.28	29.26	29.10	29.24	29.05	**29.55**
Straw	28.63	28.93	29.09	29.20	28.88	**29.26**	24.94	25.15	25.13	25.41	25.08	**25.44**
Barbara	33.11	33.01	33.06	33.35	32.66	**33.47**	29.81	29.64	29.62	30.14	29.49	**30.34**
Boat	32.14	32.21	32.08	32.09	32.01	**32.32**	29.12	29.07	28.94	28.98	28.93	**29.27**
Couple	32.11	32.10	32.00	32.01	31.89	**32.26**	28.87	28.78	28.57	28.72	28.54	**29.02**
Hill	31.86	31.91	31.88	31.85	31.82	**32.08**	29.16	29.10	28.97	29.06	28.95	**29.31**
Lena	34.27	34.17	34.12	34.21	34.00	**34.41**	31.26	31.22	31.06	31.27	30.93	**31.51**
Man	31.93	32.07	31.98	32.05	31.94	**32.22**	28.86	28.88	28.78	28.81	28.78	**29.06**
Average	32.24	32.34	32.30	32.37	32.13	**32.56**	29.04	29.02	28.94	29.09	28.89	**29.36**
σ	50						100					
C. Man	26.13	26.37	26.15	26.15	26.25	**26.64**	23.08	23.11	22.93	23.09	22.40	**23.64**
House	29.70	29.99	29.62	30.17	29.46	**30.67**	25.88	25.77	25.57	26.54	24.97	**27.17**
Monarch	25.82	25.82	25.76	26.10	25.77	**26.42**	22.52	22.23	22.12	22.61	21.60	**23.05**
Parrot	25.90	25.82	25.71	25.95	25.95	**26.27**	22.96	22.76	22.53	23.04	21.80	**23.46**
Peppers	26.68	26.79	26.54	26.73	26.58	**27.08**	23.39	23.18	22.84	23.32	22.30	**23.84**
Straw	22.41	22.67	22.49	22.86	22.37	**22.91**	19.59	19.51	19.41	19.56	19.39	**19.67**
Barbara	27.23	27.06	26.99	27.51	26.31	**27.85**	23.62	23.62	23.19	24.07	22.61	**24.54**
Boat	26.78	26.78	26.67	26.63	26.68	**27.04**	23.97	23.91	23.68	23.80	23.40	**24.22**
Couple	26.46	26.36	26.19	26.30	26.15	**26.74**	23.51	23.28	23.15	23.21	22.96	**23.62**
Hill	27.19	27.15	26.99	27.04	27.02	**27.41**	24.58	24.43	24.36	24.29	24.15	**24.75**
Lena	29.05	28.96	28.90	29.01	28.70	**29.38**	25.95	25.93	25.72	25.93	25.14	**26.48**
Man	26.81	26.72	26.67	26.68	26.72	**27.02**	24.22	23.97	24.02	24.01	23.69	**24.45**
Average	26.68	26.71	26.56	26.76	26.50	**27.12**	23.61	23.48	23.29	23.62	22.87	**24.07**

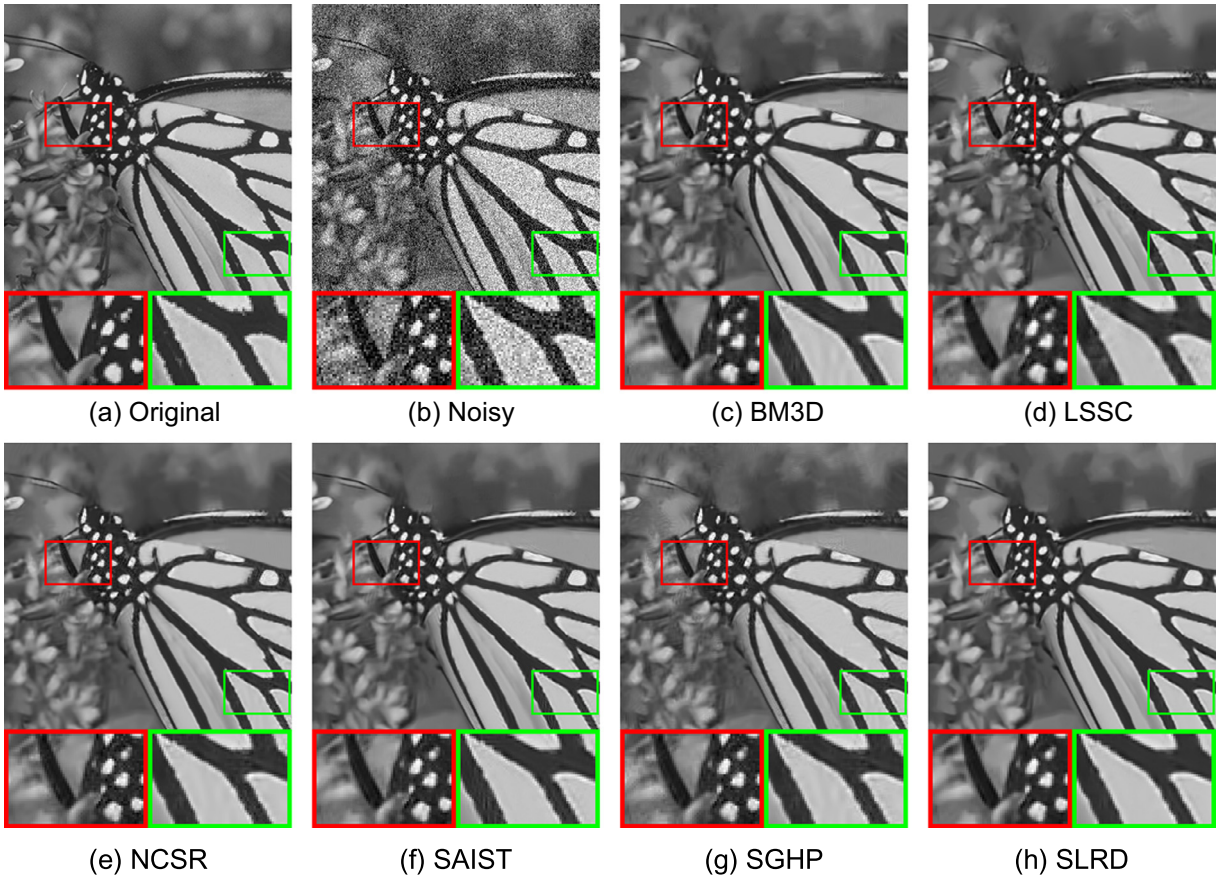

Fig. 5. Comparison of denoising results on "Monarch" image contaminated by AWGN of $\sigma = 30$. (a) Original image, (b) noisy image, (c) BM3D [15] (PSNR = 28.36 dB), (d) LSSC [16] (PSNR = 28.20 dB), (e) NCSR [18] (PSNR = 28.47 dB), (f) SAIST [20] (PSNR = 28.65 dB), (g) SGHP [19] (PSNR = 28.28 dB), and (h) SLRD (PSNR = **28.93** dB).

representations based on nonlocal sparse model (27). For the number of dictionary atoms we tested several settings and chose $K = 256$ for $\sigma \leqslant 30$ and $K = 512$ for other noise levels. By experience we observed that increasing K beyond these choices does not significantly improve denoising performance while the computational cost and memory requirements are increased. The summary of parameter settings for the proposed SLRD algorithm is reported in Table 1.

3.2. Assessment of optimal shrinker

Hard thresholding and soft thresholding of singular values are popular methods in low-rank approximation problems. However, these methods are sub-optimal in low-rank matrix denoising. In our algorithm we use the matrix recovery method of (13) that is based on an optimal singular value shrinkage function. We compare this optimal shrinker with hard and soft thresholding operators according to their denoising performance within our nonlocal patch-based denoising framework (Algorithm 1). The PSNR comparison over different noise levels are reported in Table 2. It should be mentioned that in this experiment, the threshold parameter of hard and soft thresholding is optimally selected according to recent works [42,43] respectively. From Table 2, we can see that the optimal shrinkage function leads to better denoising performance compared with the hard and soft thresholding ones.

3.3. Results and discussion

We evaluate the proposed denoising algorithm through experiments on 12 widely used images at different noise levels. To validate the effectiveness of combined use of low-rank and nonlocal sparse representation models, we first assess the denoising results of LRD alone and compare it with the results of SLRD. By doing so, we can assess how much improvement is achieved by the second step of the algorithm. In Table 3, the PSNR results of the SLRD algorithm are compared to that of LRD for different values of noise standard deviation. It can be observed that SLRD leads to higher PSNR values than LRD in almost all cases, especially at high noise levels. In fact, application of nonlocal sparse representation model in the second step of SLRD, improves the output denoising results of low-rank model. On the average, SLRD provides 0.06–0.15 dB PSNR improvement for $\sigma = 10$–100 over the LRD method. Such an improvement is worthy of attention since LRD alone outperforms several state-of-the-arts such as the benchmark BM3D method [15].

For illustrative purposes, we show an example of the denoising outputs by LRD and SLRD on "Cameraman" image, corrupted by AWGN of $\sigma = 50$ in Fig. 4. In this example, the SLRD provides an improved PSNR of 0.19 dB over the LRD algorithm and this improvement is also visually apparent. As can be seen from the close-up views of Fig. 4(g) and (h), SLRD leads to better restoration with fewer artifacts than LRD in regions around the camera and the man's head. The artifacts observed in the result of LRD may be due to the shortage of similar patches in such regions. This would degrade the quality of low-rank based collaborative estimates for matched patches. Learned sparse image models as in the SRD can handle such situations by exploiting the redundancy between overlapping patches. In addition, LRD tends to produce a very weak noise-like pattern in flat areas of image when the noise level is moderate or high (i.e. $\sigma > 30$). This arises from the fact that the noise in a group of overlapping similar patches is partially correlated which can lead to reconstruction of incorrect low-rank

patterns as the output of low-rank estimator. These artifacts are drastically reduced after the second step of our SLRD algorithm.

We have also compared the proposed SLRD algorithm with several state-of-the-art image denoising methods, namely block-matching with 3D transform (BM3D) [15], learned simultaneous sparse coding (LSSC) [16], spatially adaptive iterative singular-value thresholding (SAIST) [20], nonlocally centralized sparse representation (NCSR) [18], and segmentation-based gradient histogram preservation (SGHP) method [19]. The denoising results of all competing methods are obtained using the source codes or executables provided by their authors, and all experiments are performed with the same noise realization for each choice of σ. The PSNR results of these state-of-the-art techniques and the proposed one for the 12 test images at different noise levels are reported in Table 4 where the best results are bolded. As can be seen from Table 4, the highest PSNR performance in almost all cases belongs to the proposed SLRD and our method consistently outperforms all the other ones at all noise levels. In terms of average PSNR results, SLRD surpasses the benchmark method of BM3D by 0.31–0.46 dB at all six noise levels. The maximum improvement compared to BM3D is for the "House" image at $\sigma = 100$, where 1.3 dB improvement in PSNR is obtained. SLRD also achieves noticeable PSNR improvements over the SAIST method, as its closest competitor, especially for high values of σ.

For visual evaluation, the denoising results of SLRD and all competing methods at three different noise levels are shown in Figs. 5–7. In Figs. 5 and 6 we compare the denoising outputs for two typical images ("Monarch" and "House") corrupted with medium noise levels of $\sigma = 30$ and $\sigma = 50$, respectively. Close-up views are shown in the bottom of each result for better visualization. It can be observed from these figures that the results of BM3D, LSSC and SGHP suffer from artifacts in smooth areas and around edges which this situation is more severe for SGHP. In the results of NCSR one can also observe the undesirable visual artifacts in some regions. NCSR tends to blur the edges such as roof borders and edges of chimney in "House" image more than other methods (see Fig. 6(e)). The best visual quality at medium noise strengths, in general, is achieved by SLRD and SAIST. However, as shown in the close-up views of Figs. 5 and 6, SLRD leads to better restoration of edges and fewer artifacts than SAIST.

Fig. 7 shows an example of denoising results with strong noise contamination ($\sigma = 100$). As demonstrated in this figure, the visual quality improvement of our method over the other ones is noticeable in this extreme case. In particular, SLRD considerably outperforms all competing methods in preservation of both edge structures and smooth features of the image, and gives the most visually pleasant denoising result of "Barbara" image which has much fewer artifacts than the other methods. Overall, it can be concluded from the experiments, that by both visual comparison and objective assessment, the proposed method shows better denoising performance compared with the state-of-the-art methods. In our MATLAB implementation, SLRD takes about 80 s to denoise a 256×256 image, with a noise level of $\sigma = 20$, on a 3.40 GHz Intel Core i7 CPU. Compared with SLRD, the running time of competing methods is about 210 s for LSSC, 164 s for NCSR, 20 s for SAIST, and 152 s for SGHP. The fastest method is BM3D (about 0.7 s) but it should be noted that, unlike the other compared methods, it is implemented with compiled C++ mex-function.

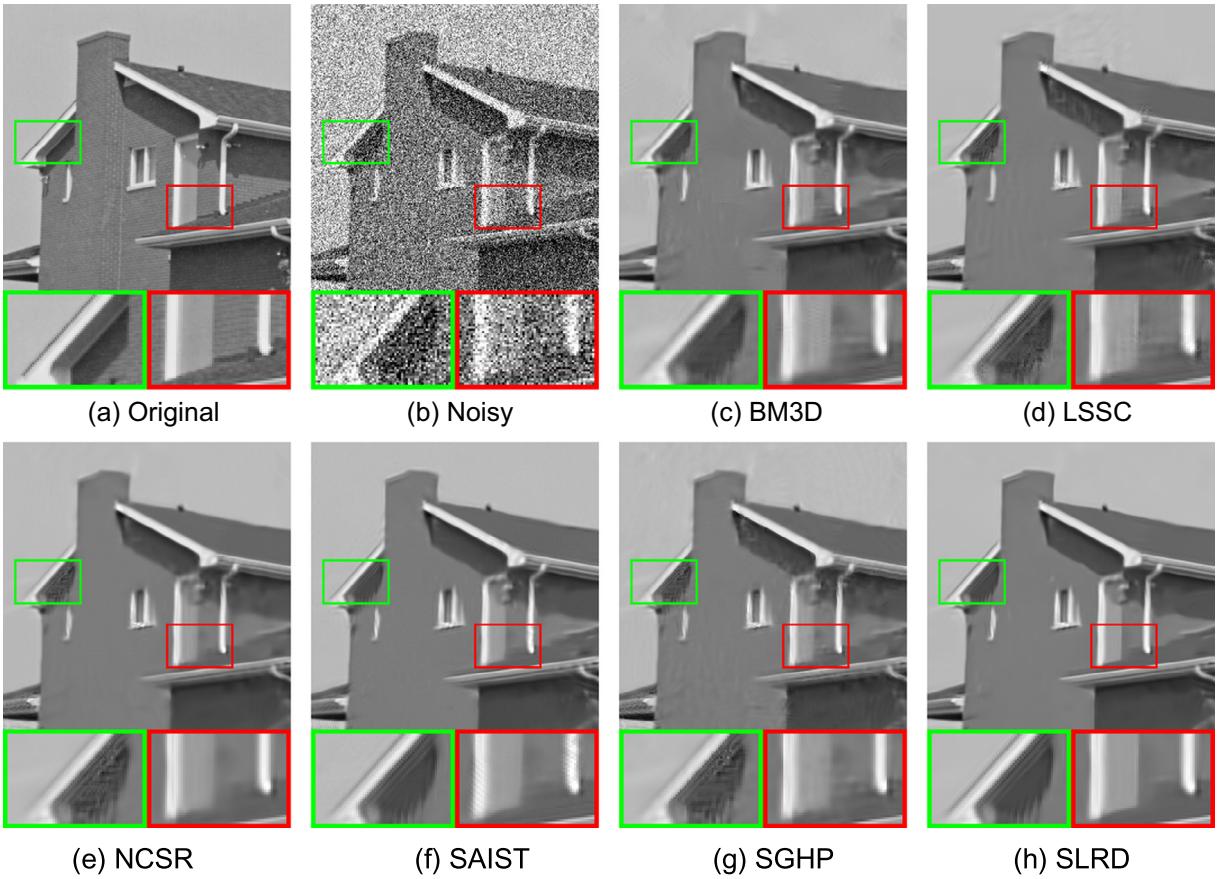

Fig. 6. Comparison of denoising results on "House" image contaminated by AWGN of $\sigma = 50$. (a) Original image, (b) noisy image, (c) BM3D [15] (PSNR = 29.70 dB), (d) LSSC [16] (PSNR = 29.99 dB), (e) NCSR [18] (PSNR = 29.62 dB), (f) SAIST [20] (PSNR = 30.17 dB), (g) SGHP [19] (PSNR = 29.46 dB), and (h) SLRD (PSNR = **30.67** dB).

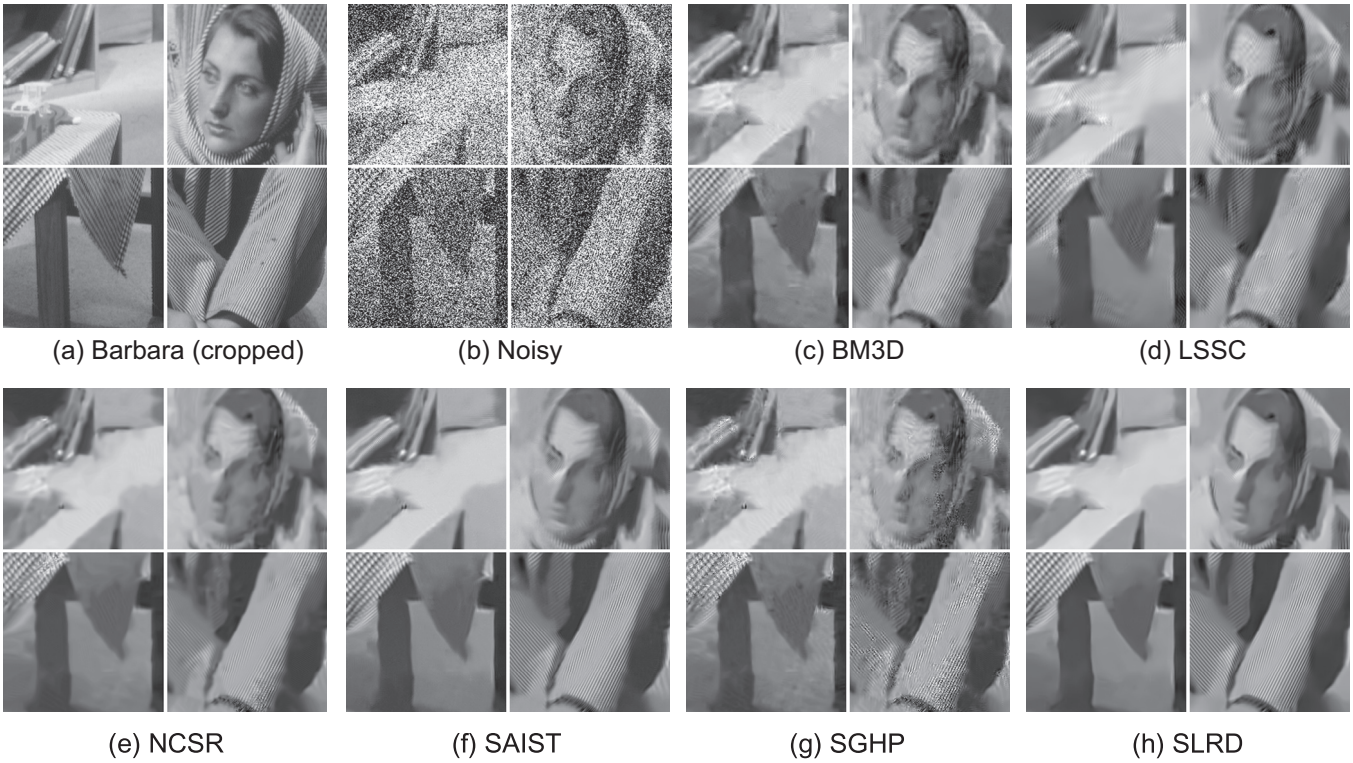

Fig. 7. Denoising of "Barbara" image contaminated by AWGN of $\sigma = 100$. The cropped regions of results obtained by (c) BM3D [15] (PSNR = 23.62 dB), (d) LSSC [16] (PSNR = 23.62 dB), (e) NCSR [18] (PSNR = 23.19 dB), (f) SAIST [20] (PSNR = 24.07 dB), (g) SGHP [19] (PSNR = 22.61 dB), and (h) SLRD (PSNR = **24.54** dB) are compared.

4. Conclusion

In this paper, we presented a new image denoising algorithm which makes use of sparsity and low-rank priors within a nonlocal patch-based denoising framework. The denoising procedure is accomplished in two successive steps based on the low-rank approximation of nonlocal self-similarities and the sparse representation using a learned overcomplete dictionary, respectively. Thanks to recent advances in random matrix theory and low-rank matrix recovery, it was shown that there is an optimal SVD-based estimator to the problem of reduced-rank matrix recovery from its noisy observation which does not require any threshold tuning. Using this efficient SVD-based scheme, we developed an iterative low-rank regularized collaborative filtering as the first step of our denoising algorithm. Nonlocal learned sparse model is adopted in the second step to further improve the denoising performance. In this step, the image nonlocal redundancy is exploited to modify the initial sparse representations and produce a more accurate estimate of the original image. From experiments carried out with a collection of standard test images, it was demonstrated that the proposed algorithm not only leads to superior PSNR performance compared with the state-of-the-art methods, but also provides better restoration of both the smooth and the texture/edge regions and produces much less visual artifacts than other approaches. It should be noted that the denoising performance of our algorithm has shown convincing improvements over the benchmark BM3D method at all tested noise levels.

References

[1] W. Zuo, L. Zhang, C. Song, D. Zhang, Texture enhanced image denoising via gradient histogram preservation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Portland, OR, June 2013, pp. 1203–1210.
[2] A. Buades, B. Coll, J.M. Morel, A non-local algorithm for image denoising, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., vol. 2, June 2005, pp. 60–65.
[3] M. Mahmoudi, G. Sapiro, Fast image and video denoising via non-local means of similar neighborhoods, IEEE Signal Process. Lett. 12 (12) (2005) 839–842.
[4] J. Wang, Y. Guo, Y. Ying, Y. Liu, Q. Peng, Fast non-local algorithm for image denoising, in: Proc. IEEE Int. Conf. Image Process., Atlanta, GA, October 2006, pp. 1429–1432.
[5] C. Kervrann, J. Boulanger, Optimal spatial adaptation for patch-based image denoising, IEEE Trans. Image Process. 15 (10) (2006) 2866–2878.
[6] C. Kervrann, J. Boulanger, Local adaptivity to variable smoothness for exemplar-based image regularization and representation, Int. J. Comput. Vis. 79 (1) (2008) 45–69.
[7] B. Goossens, Q. Luong, A. Pizurica, W. Philips, An improved non-local denoising algorithm, in: Proc. Int. Workshop on Local and Non-Local Approximation in Image Processing (LNLA 2008), Switzerland, August 2008, pp. 143–156.
[8] D. Van De Ville, M. Kocher, SURE-based non-local means, IEEE Signal Process. Lett. 16 (11) (2009) 973–976.
[9] C.A. Deledalle, V. Duval, J. Salmon, Non-local methods with shape-adaptive patches (NLM-SAP), J. Math. Image Vis. 43 (2) (2012) 103–120.
[10] Y. Wu, B. Tracey, P. Natarajan, J.P. Noonan, Probabilistic non-local means, IEEE Signal Process. Lett. 20 (8) (2013) 763–766.
[11] P. Chatterjee, P. Milanfar, Clustering-based denoising with locally learned dictionaries, IEEE Trans. Image Process. 18 (7) (2009) 1438–1451.
[12] P. Chatterjee, P. Milanfar, Patch-based near-optimal image denoising, IEEE Trans. Image Process. 21 (4) (2012) 1635–1649.
[13] D. Zoran, Y. Weiss, From learning models of natural image patches to whole image restoration, in: Proc. IEEE Int. Conf. Comput. Vis., Barcelona, Spain, November 2011, pp. 479–486.
[14] L. Zhang, W. Dong, D. Zhang, G. Shi, Two-stage image denoising by principal component analysis with local pixel grouping, Pattern Recogn. 43 (4) (2010) 1531–1549.
[15] K. Dabov, A. Foi, V. Katkovnik, K.O. Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering, IEEE Trans. Image Process. 16 (8) (2007) 2080–2095.
[16] J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman, Non-local sparse models for image restoration, in: Proc. IEEE Int. Conf. Comput. Vis., Kyoto, Japan, September/October 2009, pp. 2272–2279.
[17] W. Dong, X. Li, L. Zhang, G. Shi, Sparsity-based image denoising via dictionary learning and structural clustering, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., June 2011, pp. 457–464.
[18] W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally centralized sparse representation for image restoration, IEEE Trans. Image Process. 22 (4) (2013) 1620–1630.
[19] W. Zuo, L. Zhang, C. Song, D. Zhang, H. Gao, Gradient histogram estimation and preservation for texture enhanced image denoising, IEEE Trans. Image Process. 23 (6) (2014) 2459–2472.
[20] W. Dong, G. Shi, X. Li, Nonlocal image restoration with bilateral variance estimation: a low-rank approach, IEEE Trans. Image Process. 22 (2) (2013) 700–711.

[21] M. Elad, M.A.T. Figueiredo, Y. Ma, On the role of sparse and redundant representations in image processing, Proc. IEEE 98 (6) (2010) 972–982.
[22] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, IEEE Trans. Image Process. 15 (12) (2006) 3736–3745.
[23] D.L. Donoho, De-noising by soft thresholding, IEEE Trans. Inform. Theory 41 (3) (1995) 613–627.
[24] S. Mallat, Z. Zhang, Matching pursuit in a time-frequency dictionary, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.
[25] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: Conference Record of the 27th Asilomar Conference on Signals, Systems and Computers, November 1993, pp. 40–44.
[26] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1998) 33–61.
[27] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[28] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, in: Proc. 26th Annu. Int. Conf. Mach. Learn., 2009, pp. 689–696.
[29] K. Skretting, K. Engan, Recursive least squares dictionary learning algorithm, IEEE Trans. Signal Process. 58 (4) (2010) 2121–2130.
[30] J. Mairal, G. Sapiro, M. Elad, Learning multiscale sparse representations for image and video restoration, SIAM Multiscale Model. Simul. 7 (1) (2008) 214–241.
[31] J. Mairal, M. Elad, G. Sapiro, Sparse representation for color image restoration, IEEE Trans. Image Process. 17 (1) (2008) 53–69.
[32] M. Protter, M. Elad, Image sequence denoising via sparse and redundant representations, IEEE Trans. Image Process. 18 (1) (2009) 27–35.
[33] C. Chen, C. Wei, Y. Wang, Low-rank matrix recovery with structural incoherence for robust face recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., June 2012, pp. 2618–2625.
[34] E. Candes, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, J ACM 58 (3) (2011) 1–37.
[35] S.G. Lingala, Y. Hu, E. DiBella, M. Jacob, Accelerated dynamic MRI exploiting sparsity and low-rank structure: kt SLR, IEEE Trans. Med. Imaging 30 (5) (2011) 1042–1054.
[36] R. Otazo, E. Candès, D.K. Sodickson, Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components, Magn. Reson. Med. (2014).
[37] Y. Peng, A. Ganesh, J. Wright, W. Xu, Y. Ma, RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2233–2246.
[38] H. Ji, C. Liu, Z. Shen, Y. Xu, Robust video denoising using low rank matrix completion, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., San Francisco, CA, June 2010, pp. 1791–1798.
[39] S. Wang, L. Zhang, Y. Liang, Nonlocal spectral prior model for low-level vision, in: Computer Vision – ACCV 2012, Lecture Notes in Computer Science, vol. 7726, Springer, Heidelberg, 2013, pp. 231–244.
[40] M.A. Davenport, M.F. Duarte, Y.C. Eldar, G. Kutyniok, Introduction to compressed sensing, in: Compressed Sensing: Theory and Applications, Cambridge University Press, 2012.
[41] J.F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. 20 (4) (2010) 1956–1982.
[42] M. Gavish, D.L. Donoho, The optimal hard threshold for singular values is 4/sqrt(3), IEEE Trans. Inform. Theory 60 (8) (2014) 5040–5053.
[43] E. Candes, C. Sing-Long, J. Trzasko, Unbiased risk estimates for singular value thresholding and spectral estimators, IEEE Trans. Signal Process. 61 (19) (2013) 4643–4657.
[44] R.R. Nadakuditi, OptShrink: an algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage, IEEE Trans. Inform. Theory 60 (5) (2014) 3002–3018.
[45] A.A. Shabalin, A.B. Nobel, Reconstruction of a low-rank matrix in the presence of Gaussian noise, J. Multivar. Anal. 118 (2013) 67–76.
[46] J. Baik, J.W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, J. Multivar. Anal. 97 (6) (2006) 1382–1408.
[47] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, Stat. Sinica 17 (4) (2007) 1617–1642.
[48] B. Nadler, Finite sample approximation results for principal component analysis: a matrix perturbation approach, Ann. Stat. 36 (6) (2008) 2791–2817.
[49] K. Chen, H. Dong, K.-S. Chan, Reduced rank regression via adaptive nuclear norm penalization, Biometrika (2013).
[50] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, Multiscale Model. Simul. 4 (2) (2005) 460–489.
[51] J. Wang, S. Kwon, B. Shim, Generalized orthogonal matching pursuit, IEEE Trans. Signal Process. 60 (12) (2012) 6202–6216.
[52] I. Daubechies, M. Defriese, C. DeMol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Commun. Pure Appl. Math. 57 (11) (2004) 1413–1457.