

Image Denoising via Robust Simultaneous Sparse Coding

Lei Li

School of Technology, Beijing Forestry University, Beijing 100083, P. R. China
Email:Lilei7760@163.com

Jiangming Kan

School of Technology, Beijing Forestry University, Beijing 100083, P. R. China
Email:kanjm@bjfu.edu.cn

Wenbin Li

School of Technology, Beijing Forestry University, Beijing 100083, P. R. China
Email:leewb@bjfu.edu.cn

Abstract—Simultaneous sparse coding (SSC) has shown great potential in image denoising, because it exploits dependencies of patches in nature images. However, imposing joint sparsity might neglect the slight difference between patches. In this paper, we propose an image denoising algorithm based on robust simultaneous sparse coding (RSSC). In our algorithm, the sparse coefficient matrix is decomposed into two parts. One coefficient matrix is imposed on the joint sparse regularizer which exploits self-similarities of image patches while the other matrix is imposed by the elementwise sparse regularizer which considers the subtle differences between patches. Experiments on the benchmark data show the superior performance over the state-of-art algorithms.

Index Terms—Image denoising, robust simultaneous sparse coding, regularization, accelerated proximal gradient (APG)

I. INTRODUCTION

In recent years, affordable hardware has made it possible for digital cameras to capture images of very high resolution. However, images are often corrupted by noise during the procedures of both image acquisition and transmission. An efficient denoising algorithm becomes very important to the performance of image processing techniques. Hence, denoising of images remains one of the most fundamental tasks of image processing. As follows, for an origin image $\mathbf{x} \in \mathbf{R}^N$ and its degraded image \mathbf{y} the problem of denoising can mathematically be defined as the following observation model:

$$y_i = x_i + \eta_i \quad (1)$$

where x_i is the original pixel intensity of the y_i pixel observed as after being corrupted by zero mean independent identically distributed additive noise η_i .

recently, Many proposed denoising methods are based on image patches. Decomposing the origin image into overlapping patches, the data model can be written as

$$\mathbf{y}_i = \mathbf{x}_i + \boldsymbol{\eta}_i \quad (2)$$

where \mathbf{x}_i is the original image patch intensities with the i -th pixel at its center written in a vectorized format and \mathbf{y}_i is the observed patch corrupted by a noise vector $\boldsymbol{\eta}_i$. Denoising an image is thus solving the inverse problem to estimate pixel intensities \mathbf{x}_i .

In the past several decades, image denoising has been extensively studied and many algorithms [1] [2] [12] [13] have been proposed, leading to state-of-the-art performances. Of these various approaches, Non-Local Means (NLM) algorithm [1] recently proposed by Buades *et al.* assumes that the noise is zero-mean and uncorrelated across locations. Thanks to the presence of repeating structures in a given image, performing a weighted averaging of pixels with similar neighborhoods can suppress the noise. Inspired by the idea of NLM, Dabov *et al.* [2] proposed an effective algorithm named BM3D. BM3D performs denoising by utilizing similar patches across the image in the transform domain rather than in the origin image space.

Recently, using redundant representations and sparsity for denoising of signals have drawn a lot of research attention [3] [4] [14]. In [3], Elad and Aharon applied the K-SVD algorithm to learn the optimal over-completed dictionaries for the observed noisy image. Each image patch can be sparsely represented by the learned dictionary and denoising is carried out by coding each patch as a linear combination of only a few atoms in the dictionary. Mairal *et al.* [4] extended the K-SVD algorithm to the color image denoising.

However, in [3] [4], image patches are sparsely represented independently, which ignores self-similarities

Corresponding author: Jiangming Kan, Email: kanjm@bjfu.edu.cn, Tel: 8610-62337736.

in natural images as NLM did. Mairal *et al.* [5] argued exploiting similarities can improve the denoising performance. And he grouped image patches into a few groups and argued patches in the same group have similar sparse decomposition. And the simultaneous sparse coding (SSC) is applied to learn the sparse representation matrix. Experiments show the superior performance over the K-SVD algorithm. Dong *et al.* [6] propose a new denoising algorithm based on clustering-based sparse representation (CSR), which incorporates the dictionary learning and structural clustering into a unified variational framework.

However, there is a key problem that similar patches corrupted by noises might not share the same structure and imposing joint sparse regularizer as done in [5] may degrade the denoising performance [7] [8]. To overcome this problem, in this paper, we propose our image denoising algorithm based on robust simultaneous sparse coding (RSSC). As shown in Fig.1, in our algorithm, the sparse coefficient matrix \mathbf{A} is decomposed into two parts \mathbf{P} , \mathbf{Q} . A joint sparse regularizer is imposed on \mathbf{P} , corresponding to the shared structure while an elementwise sparse regularizer is imposed on \mathbf{Q} which corresponds to the non-shared features. The joint-sparsity exploits the similarities between patches while the elementwise sparsity considers the differences of patches. Experiments with images corrupted by synthetic noise show that the proposed method outperforms the state of the art algorithms in image denoising.

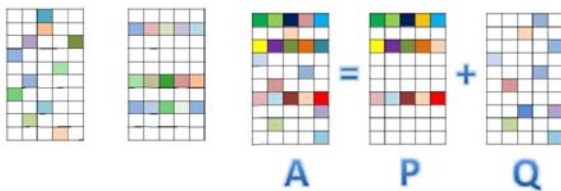


Figure 1. Illustration of sparsity vs. joint sparsity vs. robust joint sparsity: color squares represent nonzero values in the coefficient matrix.

The rest of this paper is organized as follows. In Sec.2, our new image denoising algorithm is described in detail. Experimental results and comparison with other state-of-the-art approaches are presented in Sec.3. Our work is summarized in Sec.4.

II. ROBUST SPARSE REPRESENTATION MODEL

A. Image Denoising Model

Various image restoration tasks based on sparse coding can be formulated into the following minimization problem:

$$\mathbf{\alpha}_i = \arg \min_{\mathbf{\alpha}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{\alpha}_i\|_2^2 + \gamma \|\mathbf{\alpha}_i\|_1 \quad (3)$$

where $\mathbf{y}_i = \mathbf{R}_i \mathbf{x}$ denotes an degraded image patch of size $\sqrt{n} \cdot \sqrt{n}$ extracted at location i , \mathbf{R}_i is the matrix

extracting patch \mathbf{y}_i from \mathbf{y} at location i and $\mathbf{D} \in \mathbf{R}^{n \times K}$ is a fixed dictionary and K is its size, $\mathbf{\alpha}_i \in \mathbf{R}^K$ is the reconstruction coefficients, γ is the penalty to control the sparsity of $\mathbf{\alpha}_i$. After determining sparse coefficients $\mathbf{\alpha}_i$, we estimate the original image \mathbf{x} by solving the following over-determined system:

$$\mathbf{x} = \arg \min_{\mathbf{x}} \lambda \|\mathbf{y} - \mathbf{x}\|_2^2 + \sum_{ij} \|\mathbf{R}_{ij} \mathbf{x} - \mathbf{D}\mathbf{\alpha}_{ij}\|_2^2 \quad (4)$$

This above quadratic equation has a straightforward solution:

$$\mathbf{x} = \left(\lambda \mathbf{I} + \sum_{ij} \mathbf{R}_{ij}^T \mathbf{R}_{ij} \right)^{-1} \left(\lambda \mathbf{y} + \sum_{ij} \mathbf{R}_{ij}^T \mathbf{D}\mathbf{\alpha}_{ij} \right) \quad (5)$$

A major drawback of (3) is that image patches are sparsely represented independently and dependencies between patches are ignored. The self-similarities among patches can be used to improve learned sparse models [5]. The basic idea is to group a set of similar patches $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m] \in \mathbf{R}^{n \times m}$ and imposing a grouped-sparsity regularizer on the matrix $\mathbf{A} = [\mathbf{\alpha}_1, \mathbf{\alpha}_2, \dots, \mathbf{\alpha}_m]$ and (3) is reformulated as follows:

$$\mathbf{A} = \arg \min_{\mathbf{A}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_{p,q} \quad (6)$$

where m is the number of similar patches and $\|\cdot\|_{p,q}$ is the pseudo-matrix norm defined as

$$\|\mathbf{A}\|_{p,q} = \sum_{i=1}^n \|\mathbf{\alpha}^i\|_q^p \quad (7)$$

where $\mathbf{\alpha}^i = [\alpha_{i1}, \dots, \alpha_{im}]$ denotes the i -th row of matrix \mathbf{A} . In this paper, we choose the pair (p, q) with the values (1, 2). From (6), we can see that simultaneous sparse coding encourages similar patches to be represented by the same atoms and this can exploit the dependencies between patches.

However, due to image patches corrupted by noise, similar patches may not fall into a single shared structure [7] and block ℓ_{12} regularization might perform worse than simple separate elementwise ℓ_{11} regularization. Therefore, to deal with this problem, we propose our image denoising model based on robust sparse coding (RSC) as follows:

$$\mathbf{P}, \mathbf{Q} = \arg \min_{\mathbf{P}, \mathbf{Q}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}(\mathbf{P} + \mathbf{Q})\|_F^2 + \lambda_1 \|\mathbf{P}\|_{1,2} + \lambda_2 \|\mathbf{Q}\|_{1,1} \quad (8)$$

where \mathbf{P} is a group sparse component which reflects the same structure shared by similar patches and \mathbf{Q} is an elementwise component which reflects the difference between patches. That is, we decompose the coefficient matrix into two components and impose different regularizer on them. This exploits the similarities between patches while considering their differences.

B. Optimization Algorithm

In this section, we show how to solve the robust image denoising model in (8) efficiently. Denote

$$\begin{aligned} F(\mathbf{P}, \mathbf{Q}) &= \frac{1}{2} \|\mathbf{Y} - \mathbf{D}(\mathbf{P} + \mathbf{Q})\|_F^2 \\ G(\mathbf{P}, \mathbf{Q}) &= \lambda_1 \|\mathbf{P}\|_{1,2} + \lambda_2 \|\mathbf{Q}\|_{1,1} \end{aligned} \quad (9)$$

where $F(\mathbf{P}, \mathbf{Q})$ is the empirical loss function and $G(\mathbf{P}, \mathbf{Q})$ is the regularization term. Obviously, the object function in Problem (8) is a composite model, which is consist of a differential term $F(\mathbf{P}, \mathbf{Q})$ and a non-differential term $G(\mathbf{P}, \mathbf{Q})$. And it can easily be proved $F(\mathbf{P}, \mathbf{Q})$ is jointly convex and $G(\mathbf{P}, \mathbf{Q})$ is also convex with respect to all their variables. Therefore, the global solution can be obtained. We propose a method based on the accelerated proximal gradient (APG) methods [9] [10] to solve the optimization problem.

In the APG method [9] [10], at every iteration k , we need optimize the following problem:

$$(\mathbf{P}^k, \mathbf{Q}^k) = \arg \min_{\mathbf{P}, \mathbf{Q}} T_{\mathbf{P}^k, \mathbf{Q}^k, L_k}(\mathbf{P}, \mathbf{Q}) + G(\mathbf{P}, \mathbf{Q}) \quad (10)$$

where

$$\begin{aligned} T_{\mathbf{P}^k, \mathbf{Q}^k, L_k}(\mathbf{P}, \mathbf{Q}) &= \frac{1}{2} T(\mathbf{P}^k, \mathbf{Q}^k) + \left\langle \frac{\partial T(\mathbf{P}, \mathbf{Q})}{\partial \mathbf{P}}, \mathbf{P} - \mathbf{P}^k \right\rangle \\ &+ \frac{L_k}{2} \|\mathbf{P} - \mathbf{P}^k\|_F^2 + \left\langle \frac{\partial T(\mathbf{P}, \mathbf{Q})}{\partial \mathbf{Q}}, \mathbf{Q} - \mathbf{Q}^k \right\rangle + \frac{L_k}{2} \|\mathbf{Q} - \mathbf{Q}^k\|_F^2 \end{aligned} \quad (11)$$

L is the Lipschitz constant of $\frac{\partial T(\mathbf{P}, \mathbf{Q})}{\partial \mathbf{P}}$ and $\frac{\partial T(\mathbf{P}, \mathbf{Q})}{\partial \mathbf{Q}}$. And the initial $L = L_0 = 100$ and is

updated $L_k = \tau L_{k-1}$ given $\tau > 1$. The composite object function in Problem (8) has two nonsmooth functions. However, because they are separable, Problem (10) can be solved efficiently as the following two separate problems:

$$\begin{aligned} \mathbf{P}^{k+1} &= \frac{1}{2} \left\| \mathbf{P} - \left(\mathbf{P}^k - \frac{1}{L_k} \nabla_{\mathbf{P}} F \right) \right\|_F^2 + \frac{\lambda_1}{L_k} \|\mathbf{P}\|_{1,2} \\ \mathbf{Q}^{k+1} &= \frac{1}{2} \left\| \mathbf{Q} - \left(\mathbf{Q}^k - \frac{1}{L_k} \nabla_{\mathbf{Q}} F \right) \right\|_F^2 + \frac{\lambda_2}{L_k} \|\mathbf{Q}\|_{1,1} \end{aligned} \quad (12)$$

where $\nabla_{\mathbf{P}} F$ and $\nabla_{\mathbf{Q}} F$ is the partial derivatives of $F(\mathbf{P}, \mathbf{Q})$ with respect to \mathbf{P} and \mathbf{Q} at $(\mathbf{P}^k, \mathbf{Q}^k)$. The above problems admit closed form solutions:

$$\begin{aligned} (\mathbf{P}^{k+1})^i &= \max \left(0, 1 - \frac{\lambda_1}{L_k \left\| (\mathbf{u}^k)^i \right\|} \right) (\mathbf{u}^k)^i, \forall i = 1 : n \\ \mathbf{Q}^{k+1} &= \text{sign}(\mathbf{V}^k) \max \left(0, 1 - \frac{\lambda_2}{L_k} \right) \mathbf{V}^k \end{aligned} \quad (13)$$

Where $\mathbf{U}^k = \mathbf{P}^k - \frac{1}{L_k} \nabla_{\mathbf{P}} F$, $\mathbf{V}^k = \mathbf{Q}^k - \frac{1}{L_k} \nabla_{\mathbf{Q}} F$.

In [11], it has been proved the APG method achieves an $O\left(\frac{1}{k^2}\right)$ residual from the optimal solution after k iterations. Finally, the algorithm for solving Problem (8) is given in Algorithm1.

Algorithm 1 Optimization for Problem (8)

-
- Input: \mathbf{D} and \mathbf{Y} .
 - Output: Optimal solution $(\mathbf{P}^*, \mathbf{Q}^*)$.
 - (1). Set $\mathbf{P}^1 = \mathbf{P}^0 = \mathbf{0}$, $\mathbf{Q}^1 = \mathbf{Q}^0 = \mathbf{0}$, $t_0 = 0$, $t_1 = 1$, $k = 1$, $L = L_0 = 100$, $\eta = 1.05$, $\lambda_1 = 0.3$, $\lambda_2 = 0.01$.
 - (2). while not converged do
 - (3). Compute the proximal points:

$$\begin{aligned} \mathbf{P}_v^k &= \mathbf{P}^k + \frac{t^{k-1} - 1}{t^k} (\mathbf{P}^k - \mathbf{P}^{k-1}); \\ \mathbf{Q}_v^k &= \mathbf{Q}^k + \frac{t^{k-1} - 1}{t^k} (\mathbf{Q}^k - \mathbf{Q}^{k-1}); \end{aligned}$$
 - (4). Calculate the gradient $\nabla_{\mathbf{P}} F, \nabla_{\mathbf{Q}} F$;
 - (5). Calculate $\mathbf{P}^{k+1}, \mathbf{Q}^{k+1}$ via (13);
 - (6). If $F_{\mathbf{P}^{k+1}, \mathbf{Q}^{k+1}} > T_{\mathbf{P}^k, \mathbf{Q}^k, L_k}$, update $L_k = \tau L_{k-1}$ and go to Step 5;
 - (7). Stepsize update:

$$t^{k+1} = \frac{1 + \sqrt{4(t^k)^2 + 1}}{2}$$
 - (8). end while
-

III. IMAGE DENOISING EXPERIMENTS

We have implemented the proposed RSSC denoising algorithm under MATLAB. In all experiments, the block-size $n = 64$, the penalties on sparsity λ_1, λ_2 are set as 0.5, 0.01 respectively. The dictionary is trained on patches from nature images and is shown in Fig 2. And the size of the dictionary $K = 256$.

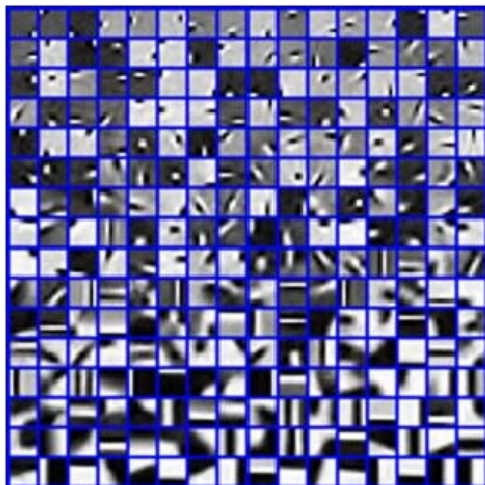


Figure 2. The dictionary used in our denoising algorithm

We first compare the proposed RSSC denoising algorithm and two leading methods for removing additive white Gaussian noise: K-SVD [3], BM3D [2]. The denoising results of all benchmark schemes are generated from the source codes or executables released by their authors. The PSNR performance of three competing denoising algorithms are reported in Table 1 (the highest PSNR value is in *Italic*).

From Table 1, we can see that our RSSC outperforms the other two algorithms on all tested images except that

BM3D achieves better performances on Barbara at the noise level $\theta_n = 15, 25$. Especially, when images are corrupted by $\theta_n = 35$ level Gauss noise, our algorithm achieves the best results on all test images in the three ones. Therefore, we can conclude that RSSC outperforms the other two benchmark methods.

The visual comparison of the denoising methods is shown in Figs. 3-8. Fig. 5 presents visual results of different methods for Peper image with the noise level $\theta_n = 25$. We can observe that all methods can remove the noise effectively and our RSSC which exploits the similarities between patches and considering their differences provides an important improvement over the K-SVD method which considers each patch independently. The proposed method also outperforms over the BM3D method.

Fig. 6 gives the results of our algorithm on the corrupted image House. We can see that our RSSC method shows better visually quality result. The K-SVD method generates many visually disturbing artifacts in the denoised image. The BM3D method loses many details compared our RSSC method.

VI. CONCLUSION

In this paper, we propose a robust image denoising algorithm based on the robust simultaneous sparse coding. We decompose the sparse coefficient matrix into two halves. Joint sparsity regularizer is imposed on one matrix and elementwise sparsity regularizer is imposed on another. This makes our algorithm exploit the dependencies between patches while considering the differences between them. Experiments on the benchmark data show the superior performance over the state-of-art algorithms.

TABLE 1.
QUANTITATIVE COMPARASION. THE GAUSS NOISE LEVEL $\theta_n=15, 25, 35$ AND THE PSNR ARE CHOSEN AS THE PERFORMANCE MEASURE.
BEST RESULTS ARE SHOWED IN *ITALIC*.

θ_n	Method	House	Peppers	Barbara	Boat	Lena
15	KSVD	34.33	32.11	34.53	34.49	34.98
	BM3D	36.14	33.36	35.35	34.82	35.33
	RSSC	<i>37.13</i>	<i>35.20</i>	35.09	<i>35.13</i>	<i>35.48</i>
25	KSVD	32.08	29.71	30.38	30.29	30.35
	BM3D	32.03	30.14	<i>30.68</i>	29.82	30.03
	RSSC	<i>32.30</i>	<i>30.31</i>	30.59	<i>30.43</i>	<i>30.77</i>
35	KSVD	31.08	28.91	28.94	29.09	29.15
	BM3D	31.13	29.18	29.47	29.11	29.21
	RSSC	<i>31.20</i>	<i>29.45</i>	<i>29.59</i>	<i>29.33</i>	<i>29.40</i>

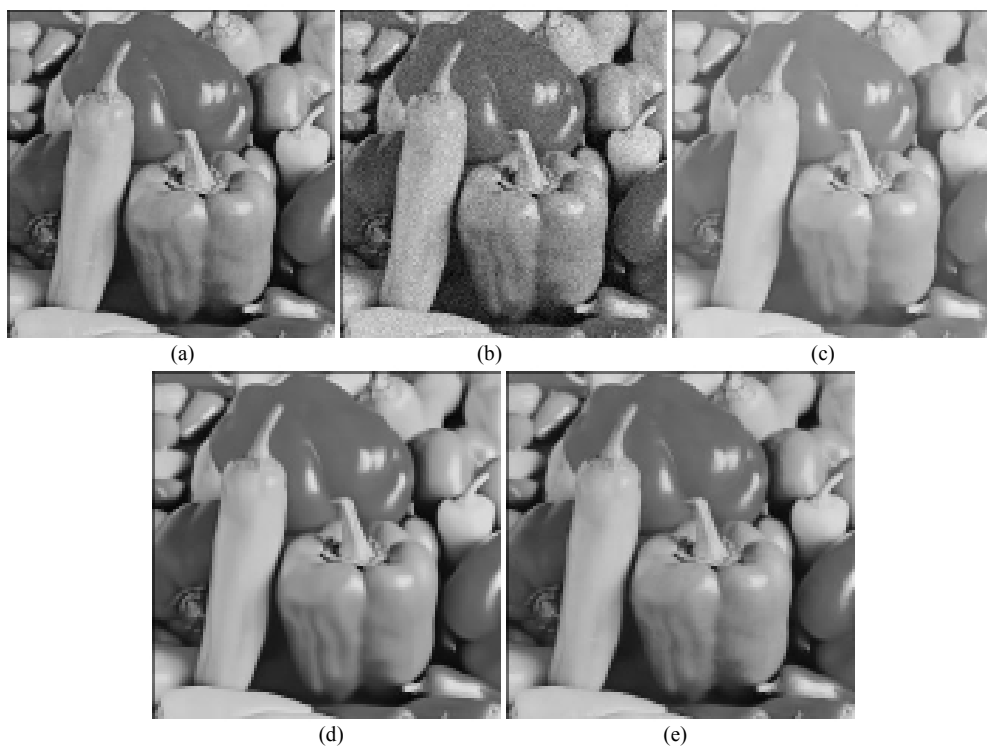


Figure 3. Denoising performance comparison for Barbara image at Gauss noise level $\sigma_n = 15$.
From the left to right, the original image, noising image, KSVD, BM3D, RSSC.

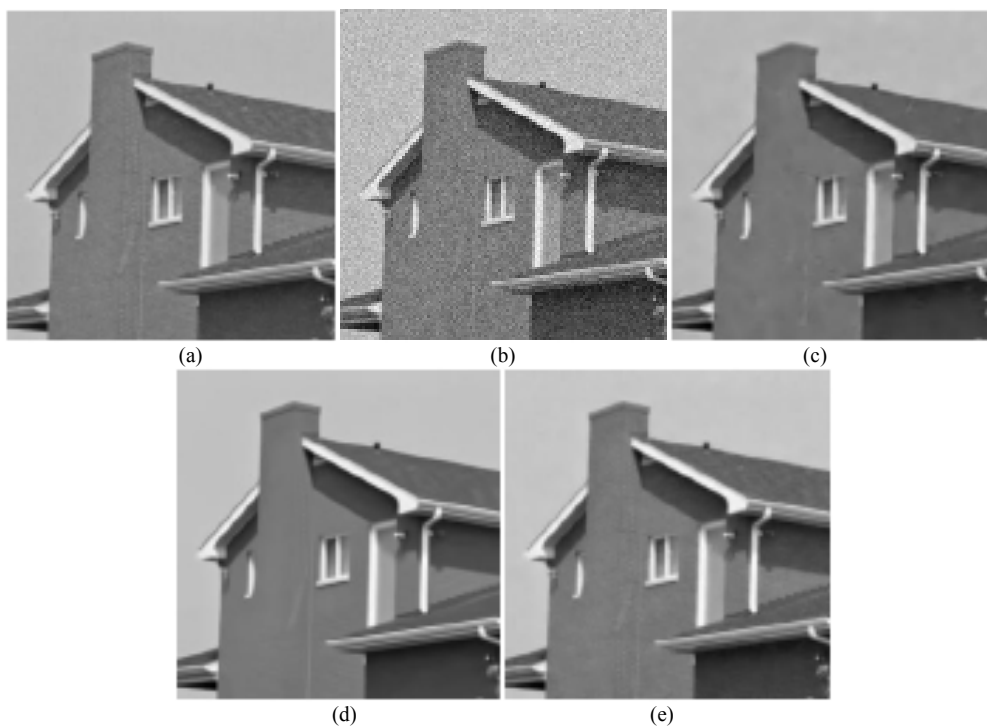


Figure 4. Denoising performance comparison for Boat image at Gauss noise level $\sigma_n = 15$.
From the left to right, the original image, noising image, KSVD, BM3D, RSSC.

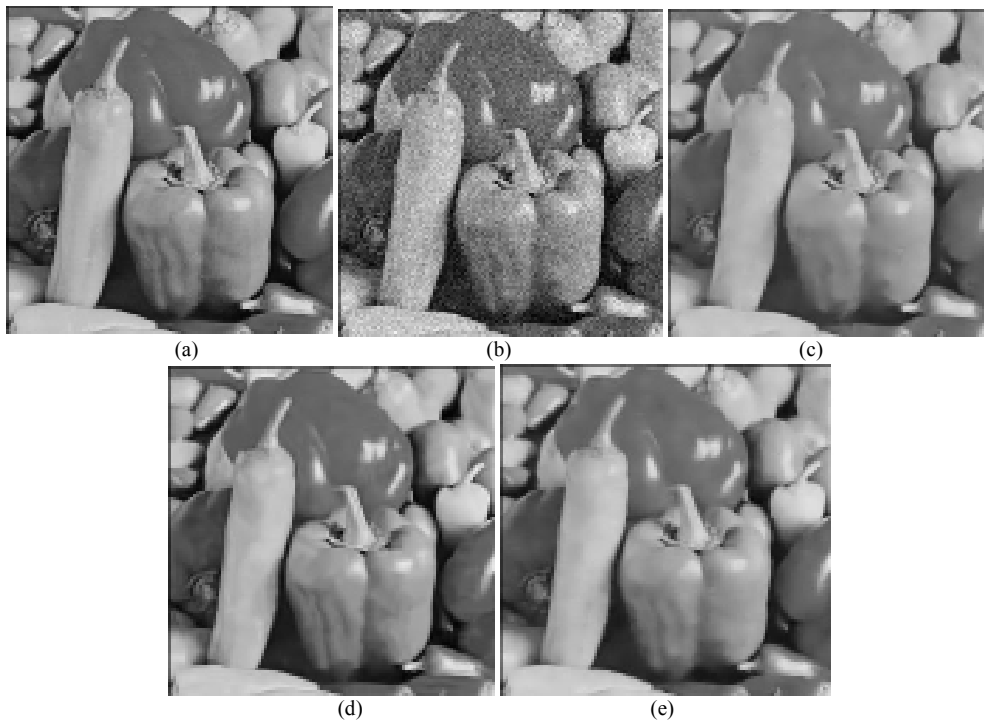


Figure 5. Denoising performance comparison for Peppers image at Gauss noise level $\sigma_n = 25$.
From the left to right, the original image, noising image, KSVD, BM3D, RSSC.

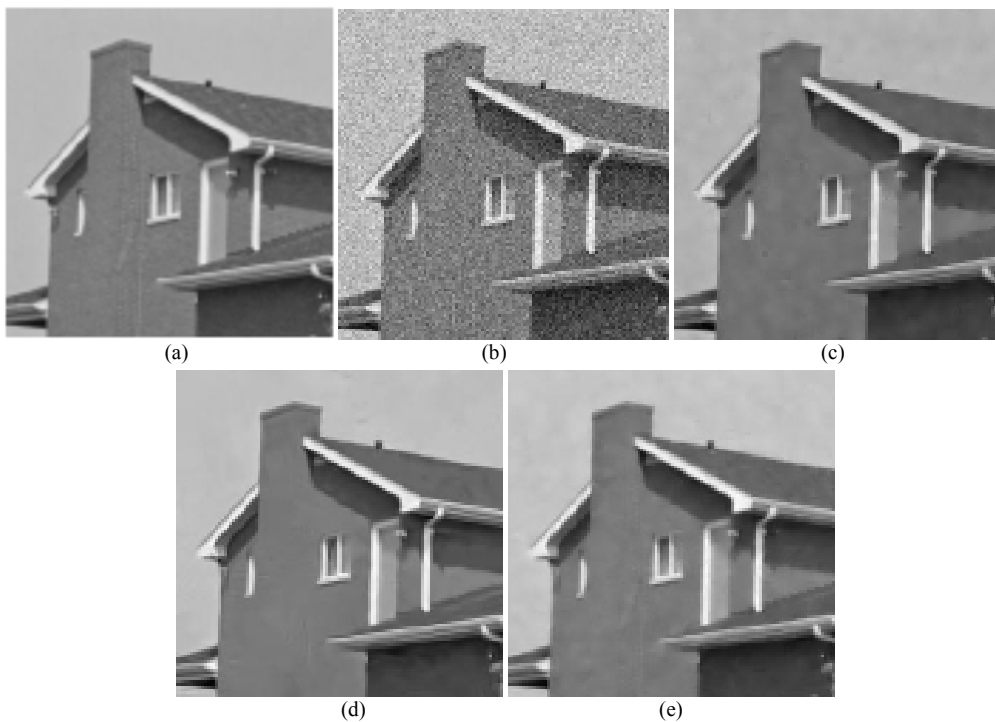


Figure 6. Denoising performance comparison for House image at Gauss noise level $\sigma_n = 25$.
From the left to right, the original image, noising image, KSVD, BM3D, RSSC.

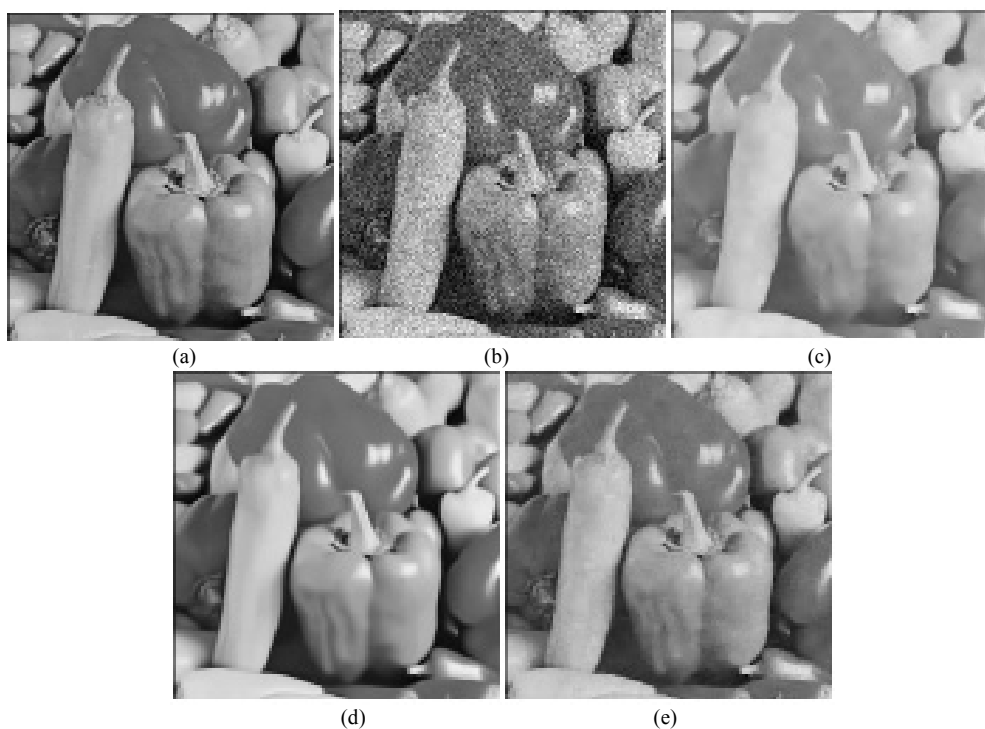


Figure 7. Denoising performance comparison for Peppers image at Gauss noise level $\sigma_n = 35$.
From the left to right, the original image, noising image, KSVD, BM3D, RSSC

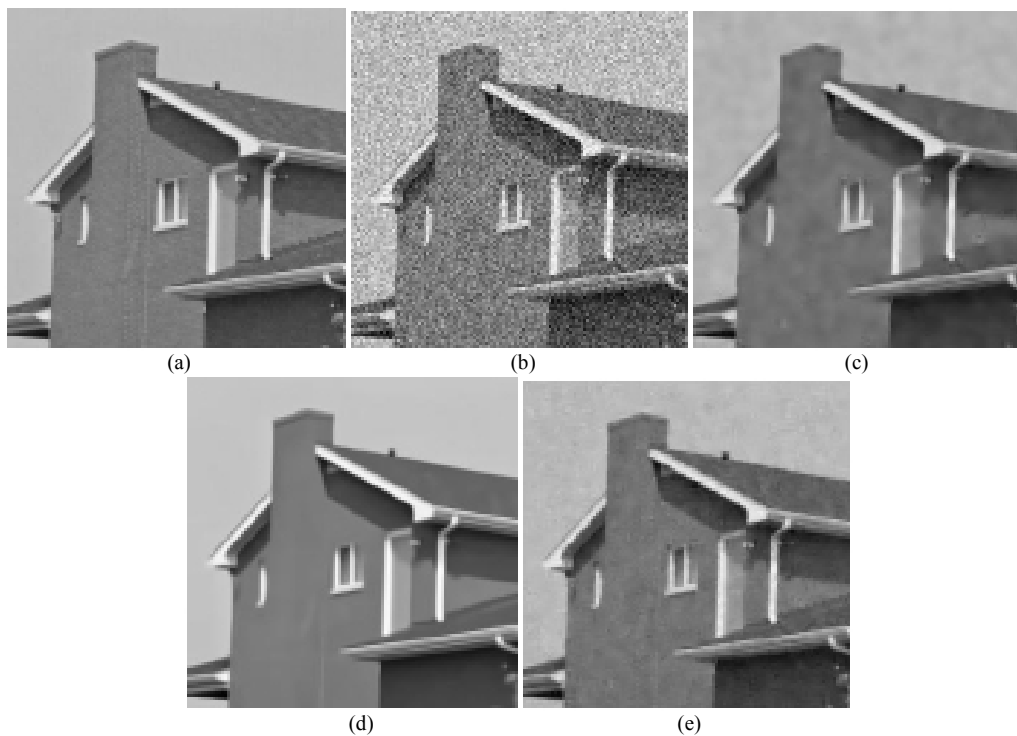


Figure 8. Denoising performance comparison for House image at Gauss noise level $\sigma_n = 35$.
From the left to right, the original image, noising image, KSVD, BM3D, RSSC

ACKNOWLEDGMENT

This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. BLYX200905 and TD2013-4) and National Natural Science Foundation of China (Grant No. 30901164).

REFERENCES

[1] A. Buades, B. Coll, and J. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[3] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans Image Processing*, vol. 15, no. 12, pp.3736–3745, 2006.

[4] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Computer Vision, IEEE 12th International Conference on*. IEEE, 2009, pp. 2272–2279.

[6] W. Dong, X. Li, D. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Computer Vision and Pattern Recognition(CVPR)*, 2011 IEEE Conference on, 2011, pp. 457–464.

[7] A. Jalali, P. Ravikumar, S. Sanghavi, and C. Ruan, "A dirty model for multi-task learning," *Advances in Neural Information Processing Systems*, vol. 23, pp. 964–972, 2010.

[8] S. Negahban and M. J. Wainwright, "Joint support recovery under high-dimensional scaling: Benefits and perils of l_1 -regularization," *Advances in Neural Information Processing Systems*, vol. 21, pp. 1161–1168, 2008.

[9] P. Tseng, "On Accelerated Proximal Gradient Methods for Convex-Concave Optimization," unpublished.

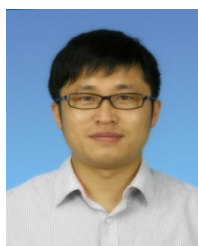
[10] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *International Conference on Machine Learning*. ACM, 2009.

[11] P. Gong, J. Ye, and C. Zhang, "Robust multi-task feature learning," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA: ACM, pp. 895–903. 2012.

[12] G. Liu, J. Liu, Q. Wang, W. He, "The Translation Invariant Wavelet-based Contourlet Transform for Image Denoising," *Journal of Multimedia*, North America, 7, jun. 2012.

[13] H. Fan, Y. Wang, J. Li, "Image Denoising Algorithm Based on Dyadic Contourlet Transform," *Journal of Software*, North America, 6, jun. 2011.

[14] J. Hu, Y. Pu, J. Zhou, "A Novel Image Denoising Algorithm Based on Riemann-Liouville Definition," *Journal of Computers*, North America, 6, jul. 2011.



Lei Li was born in JiLin Province, China, in 1980. He studies in Technology Beijing Forestry University, and separately gained a bachelor's degree in 2005 and a master's degree in 2008. Now, he is a doctorate candidate in the same university. His research interests mainly include information security, image processing image compression.



Jiangming Kan received in PhD degree in forestry engineering from Beijing Forestry University, P.R.China in 2009. Currently, he is an associate professor in Beijing Forestry University. His research interests include computer vision and intelligent control.



intelligent.

Wenbin Li received M.S. and Ph.D. degrees in Shizuoka University and Ehime University, Japan, in 1987, and 1990, respectively. Starting 1992, he was a faculty with the School of Technology, Beijing Forestry University and was promoted to be a professor in 1996, Ph.D. supervisor. His current research interests include forest machinery automation and