Image Denoising via Improved Sparse Coding

Xiaoqiang Lu[1]
luxq666666@gmail.com

Haoliang Yuan[2]
haoliangyuan@gmail.com

Pingkun Yan[1]
pingkun.yan@opt.ac.cn

Yuan Yuan[1]
yuany@opt.ac.cn

Luoqing Li[2]
lilq@hubu.edu.cn

Xuelong Li[1]
xuelong_li@opt.ac.cn

[1] Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi, P. R. China 710119

[2] Faculty of Mathematics and Computer Science, Hubei University, Wuhan, Hubei, P. R. China 430062

Abstract

This paper presents a novel dictionary learning method for image denoising, which removes zero-mean independent identically distributed additive noise from a given image. Choosing noisy image itself to train an over-complete dictionary, the dictionary trained by traditional sparse coding methods contains noise information. Through mathematical derivation of equation, we found that a lower bound of dictionary is related with the level of noise in dictionary learning. The proposed idea is to take advantage of the noise information for designing a sparse coding algorithm called improved sparse coding (ISC), which effectively suppresses the noise influence for training a dictionary. This denoising framework utilizes the effective method, which is based on sparse representations over trained dictionaries. Acquiring an over-complete dictionary by ISC mainly includes three stages. Firstly, we utilize K-means method to group the noisy image patches. Secondly, each dictionary is trained by ISC in corresponding class. Finally, an over-complete dictionary is merged by these dictionaries. Theory analysis and experimental results both demonstrate that the proposed method yields excellent performance.

1 Introduction

Image denoising is not only an old image processing problem but also a hot research field that goes on to attract researchers with a goal to do better restoration in the present of noise. With the modern science and technology development, acquisition hardware is easy to deal with images with wonderful high resolution and at high shutter speeds. But these factors also lead that image capturing devices are prone to corruption by noise. Effective image denoising technique can help camera manufacturers solve this problem. This also makes image denoising still a hot problem continuing research. As follows, the problem of denoising can mathematically be shown from the observation model

$$y = x + \eta, \tag{1}$$

where x is the original image, η denotes the noise with distribution unknown. Our goal is to design a denoising algorithm which can remove the noise from y. Recently, many denoising algorithms [1-6] work on image patches, so it is useful to formulate a patch-based image model

$$y_i = x_i + \eta_i, \tag{2}$$

where x_i is the original patch intensity written in a vectorized form, y_i is the vectorized noisy image patch and η_i is a vectorized noise patch.

Although the denoising model of (2) is linear in the unknown image, some successful methods [1-6] related with denoising problem have been adopted to this inverse problem, leading to state-of-the-art performance. Of these various approaches, recently the proposed Non-Local Means (NLM) [1] algorithm assumes that the noise is uncorrelated across locations and zero-mean, making use of the repetitive structures in an image. Kervrann and Boulanger [2] improve on the performance of NLM by adding an adaptive search window along with a revised weight calculation formula. Takeda *et al.* [3] propose a framework for denoising problem by utilizing steering kernel regression (SKR) where the kernels are designed to make the similarity of image pixels in a neighbourhood. Another successful denosing method is called K-SVD [4], which assumes the image patches can be sparse representation. Elad and Aharon take advantage of this assumption to learn an effective over-complete dictionary for the noisy image patches such that each noisy patch can be represented as a linear combination of only a few atoms among the dictionary. Chatterjee and Milanfar [5] present a clustering-based framework called K-LLD which is inspired by the approaches in [3] and [4] to perform denoising. The denoising methods stated above are based on spatial domain. In [6], BM3D proposed by Dabov *et al.* also makes use of the idea of NLM by finding similar patches in an image and grouping them together. The denoising process in BM3D is, however, done by transform domain.

Utilizing sparse representation framework for image denosing has acquired a state-of-the-art performance. Choosing a suitable dictionary is important for this framework. The concept of dictionary learning that generates sparse representation for a set of training image patches has been studied in a series of work [7-14]. In this work, the dictionary is trained by using patches from the noisy image itself. It is important to design a suitable dictionary for image denoising in the process of dictionary learning. Based on the image denoising theory analysis, we propose a novel dictionary learning method called improved sparse coding (ISC), which can effectively restrain the noise among the dictionary. Through mathematical derivation of equation, we find that a lower bound of dictionary is related with the level of noise in dictionary learning process. Inspired by the idea that the data of similar type can be effectively represented by a complete dictionary, a remedy to the above problem, however, is simple and also intuitive. That is to use a cluster method. The K-means algorithm has been tested to be an effective method. By employing K-means algorithm, ISC can be used to train a complete dictionary in each class, utilizing these complete dictionary for getting an over-complete dictionary. Our main contribution is to find that a lower bound of dictionary is related with the level of noise, and design a dictionary learning method with the purpose of image denoising. We get through two aspects to test that our method is reasonable: 1) giving detailed algorithm theory analysis that contains strictly mathematical derivation; 2) a lot of experimental results also test our conclusion.

The rest of the paper is organized as follows. Section 2 describes ISC algorithm and gives detailed theory analysis. Section 3 shows that ISC is applied in our denoising

algorithm. In Section 4, we show some experimental results that test the effectiveness of this algorithm. Section 5 summarizes and concludes our work.

2 Improved Sparse Coding

In this section, we first briefly review a traditional dictionary learning algorithm. Given only unlabelled data, it discovers basis functions that search higher-level features in the data. Our algorithm named improved sparse coding (ISC) mainly improves the algorithm in [15] proposed by Honglak *et al.*. Let $X \in R^{N \times L}$ be the input matrix (each column is an input vector), let $D \in R^{N \times M}$ be the basis matrix also called dictionary (each column is a coefficient vector), and let $S \in R^{M \times L}$ be the coefficient matrix (each column is a coefficient vector). The optimization model about sparse coding is

$$\widehat{D} = arg \min_{D,S} \|X - DS\|_2^2 + \lambda \|S\|_1$$
$$s.t. \quad \|d_i\|_2^2 \leq b, \quad i = 1, \cdots, M. \tag{3}$$

In the model (3), L_1 norm is to guarantee sparsity, and L_2 norm limitation on the columns of D can remove the scaling ambiguity, d_i denotes the i-th column of D. This particular formulation has been extensively studied [7, 15, 16].

This paper considers the noise influence, thus we consider adding additive restraints for getting a suitable dictionary for denoising. The input vector x_i can be represented by $D\alpha_i$, *i.e.,* $x_i = D\alpha_i$. So (2) is rewritten by

$$y_i = D\alpha_i + \eta_i. \tag{4}$$

For simplicity, D is assumed to be an invertible matrix, *i.e.,* $D \in R^{N \times N}$. This assumption doesn't affect the theory analysis. So we get the noise coefficient as

$$\hat{\alpha}_i = D^{-1}y_i = \alpha_i + D^{-1}\eta_i, \tag{5}$$

where $D^{-1}\eta_i$ is considered as the noise term, its covariance matrix C is

$$C = Cov(D^{-1}\eta_i) = \mathrm{E}[D^{-1}\eta_i\eta_i^T(D^{-1})^T]$$
$$= D^{-1}\mathrm{E}[\eta_i\eta_i^T](D^{-1})^T, \tag{6}$$

where $\mathrm{E}[\cdot]$ denotes the expected value, T represents transposition. In this paper, we consider two different noises.

1) Additive noise. If the noise is constant and uncorrelated, $\mathrm{E}[\eta_i\eta_i^T]$ is a diagonal matrix, and can be represented by $\sigma^2 I$, *i.e.*

$$\mathrm{E}[\eta_i\eta_i^T] = \sigma^2 I, \tag{7}$$

where σ is the standard deviation of the noise, I denotes unit matrix. The equation (6) can be rewritten as

$$C_1 = \sigma^2 D^{-1}(D^{-1})^T. \tag{8}$$

2) Multiplicative noise. This covariance of this kind of noise is direct proportion to original image, *i.e.,*

$$\mathrm{E}[\eta_i\eta_i^T] = Kdiag[y_i], \tag{9}$$

where K is constant. Hence, we get

$$C_2 = KD^{-1} diag[y_i](D^{-1})^T. \tag{10}$$

If we assume that the noises of different type are statistically irrelevant, $E[\eta_i \eta_i^T]$ can be divided into two parts

$$C = C_1 + C_2 = \sigma^2 D^{-1}(D^{-1})^T + KD^{-1}diag[y_i](D^{-1})^T. \tag{11}$$

In order to measure the level of image noise, we need consider the sum of the diagonal element of the covariance matrix

$$Tr[C] = \sigma^2 Tr[D^{-1}(D^{-1})^T] + KTr[D^{-1} diag[y_i](D^{-1})^T], \tag{12}$$

where Tr denotes the trace operation. We define a variable R called relative noise level, and its form is

$$R = Tr[C]/\sigma^2 = Tr[D^{-1}(D^{-1})^T] + KTr[D^{-1} diag[y_i](D^{-1})^T]/\sigma^2. \tag{13}$$

Then according to the inequality

$$\min[y_i]Tr[D^{-1}(D^{-1})^T] \le Tr[D^{-1} diag[y_i](D^{-1})^T] \le \max[y_i]Tr[(D^{-1}(D^{-1})^T], \tag{14}$$

there exists a constant ρ such that

$$Tr[D^{-1} diag[y_i](D^{-1})^T] = \rho Tr[D^{-1}(D^{-1})^T]. \tag{15}$$

Then

$$R = Tr(D^{-1}(D^{-1})^T + K\rho Tr[D^{-1}(D^{-1})^T]/\sigma^2$$
$$= (K\rho/\sigma^2 + 1)Tr[(D^{-1}(D^{-1})^T]. \tag{16}$$

Let $\hat{R} = min_D R$ and $\hat{D} = argmin_D R$, then

$$Tr(\hat{D}^{-1}(\hat{D}^{-1})^T) = \hat{R}/(K\rho/\sigma^2 + 1).$$

Utilizing singular value decomposition method, we get $D = U\Lambda V^T$. Then

$$\hat{R}/(K\rho/\sigma^2 + 1) = Tr(\hat{D}^{-1}(\hat{D}^{-1})^T) = Tr[V\Lambda^{-2}V^T] = Tr[\hat{D}^{-2}]$$
$$= \sum_{i=1}^N \gamma_i^{-2} \ge 1/\sum_{i=1}^N \gamma_i^2 = 1/Tr[\hat{D}\hat{D}^T]. \tag{17}$$

Where γ_i is the eigenvalue of $\hat{D}$. Thus, we can get a restraint condition as

$$Tr[\hat{D}\hat{D}^T] \ge (K\rho/\sigma^2 + 1)/\hat{R}. \tag{18}$$

The inequality (18) has an equivalent form

$$\|D\|_2^2 \ge (K\rho/\sigma^2 + 1)/\hat{R}. \tag{19}$$

From the formulation (19), we find that a lower bound of dictionary is related with the level of noise. Thus we assume each columns of the dictionary also has a lower bound related with noise, *i.e.*

$$\|d_i\|_2^2 \ge a, \quad i = 1, \cdots, M, \tag{20}$$

where a is a constant that is dictated by σ. So the optimization model called improved sparse coding is proposed as

$$\hat{D} = arg\min_D \|X - DS\|_2^2 + \lambda\|S\|_1,$$

$$s.t. \quad a \leq \|d_i\|_2^2 \leq b, \quad i = 1, \cdots, M. \tag{21}$$

The optimization performs in an alternative manner over D and S.

1) Initialize D with a Gaussian random matrix where each column is unit normalized.
2) Fix D, update S by

$$\hat{S} = arg\min_S \|X - DS\|_2^2 + \lambda\|S\|_1, \tag{22}$$

which can be solved by feature-sign search algorithm [14].

3) Fix S, update D by

$$\hat{D} = arg\min_D \|X - DS\|_2^2,$$

$$s.t. \quad a \leq \|d_i\|_2^2 \leq b, \quad i = 1, \cdots, M, \tag{23}$$

which is a Quadratically Constrained Quadratic Programming that can be solved in many optimization packages.

4) Iterate between 2) and 3) until converge.

3 Denoising Algorithm

In this section, we utilize the denoising framework of [5]，which proposes denoising algorithm by introducing how sparsity and redundancy are brought to use. First of all, we will describe the general idea of the proposed algorithm with an illustration shown in Figure 1.
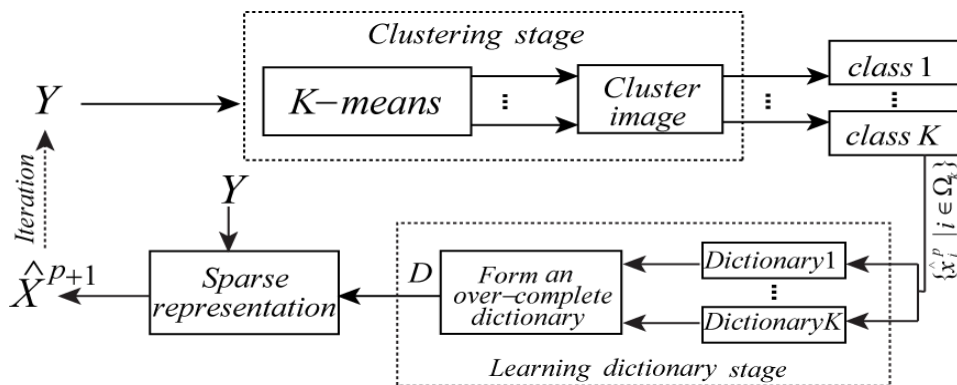

Figure 1: Illustration of sparse representation for image denoising.

Let Y be a noisy image. Because the denosing framework of [5] is based on the requirement of sparse representation, we need to get an over-complete dictionary to

sparsely represent each noisy image patch, this process leads to denoising result. While the dictionary generated by ISC algorithm is complete. With regard to this problem, we use K-means algorithm to divide the noisy image Y into $\{Y_1, \cdots, Y_K\}$. Since the elements from the same cluster have similar geometric structure, the corresponding complete dictionaries $\{D_1, \cdots, D_K\}$ are trained by ISC algorithm.

$$D_j = arg\min_D \left\| Y_j - DS \right\|_2^2 + \lambda \|S\|_1,$$

$$s.t. \qquad a \leq \|d_i\|_2^2 \leq b, \quad i = 1, \cdots, M, \quad j = 1, \cdots, K. \tag{24}$$

At last, an over-complete dictionary D is generated by combining these complete dictionaries.

As follows, the image denoising model on the image patches is formed by solving

$$\{\hat{\alpha}_{ij}, \hat{X}\} = arg\min_{\alpha_{ij}, X} \lambda_1 \|X - Y\|_2^2 + \sum_{ij} \left\| D\alpha_{ij} - R_{ij}X \right\|_2^2 + \sum_{ij} \mu_{ij} \left\| \alpha_{ij} \right\|_0. \tag{25}$$

Where Y is the noisy image and its denoised version X, R_{ij} is a matrix that extracts the (ij) block from the image, and each patch x_{ij} can be represented by $x_{ij} = R_{ij}X$.

In model (25), the first term is the log-likelihood global force that requires the proximity between the measured image Y and X. Given a restraint, this penalty would have a limitation of $\|X - Y\|_2^2 \leq Const \cdot \sigma^2$. The second and third terms are the image prior which guarantees every patch x_{ij} in corresponding location has a sparse representation with bounded error.

When an over-complete dictionary D is known, the proposed framework suggests that each noisy image patch could be sparsely represented over this dictionary D, *i.e.*, the solution of

$$\hat{\alpha}_{ij} = arg\min_\alpha \|\alpha\|_0 \quad s.t. \quad \left\| D\alpha - R_{ij}X \right\|_2^2 \leq T(\sigma) \tag{26}$$

is indeed sparse. The notation $\|\alpha\|_0$ denotes the count of the nonzero entries in α, $T(\sigma)$ is constant dictated by σ, which stands for standard deviation of a Gaussian noise. In this paper, we will make use of the orthonormal matching pursuit (OMP) to get the sparse coefficient. Given all $\hat{\alpha}_{ij}$ and fixed those, we turn to update X. Returning to (25), we need to solve

$$\hat{X} = arg\min_X \lambda_1 \|X - Y\|_2^2 + \sum_{ij} \left\| D\hat{\alpha}_{ij} - R_{ij}X \right\|_2^2. \tag{27}$$

This is a common quadratic term which has a closed solution of the form

$$\hat{X} = \left(\lambda_1 I + \sum_{ij} R_{ij}^T R_{ij} \right)^{-1} \left(\lambda_1 Y + \sum_{ij} R_{ij}^T D\hat{\alpha}_{ij} \right). \tag{28}$$

All it says is that averaging of the denoised patches is to be completed, with some relaxation gotten by averaging with the noisy image. Because the invertible matrix in the above expression is a diagonal one, hence the calculation of (28) can be also done on a pixel-by-pixel basis.

By now, we have known that the denoising algorithm needs for sparse coding of small patches, and averaging of their results. As follows, due to minimization of (25) is our goal, then this process can continue. Given the updated X, we can use iterated scheme for going on denoising via sparse representation. While the iterative process is applied by only one time, this is because following iterations need knowledge of the new noise standard

deviation σ, which is unknown after first updating X. Finally, we propose our denoising algorithm below.

Algorithm 1: Image denoising via sparse representation

Objective: Estimate the denoised image X of a noisy image Y form white and additive Gaussian white noise with standard deviation σ.

Input: n: block size; k: dictionary size;

        λ, λ_1: Lagrange multipliers; K: number of classes.

Output: Denoised image $\hat{X}$.

Initialization: Set $X = Y$;

1) Use K-means method to divide X into K classes, *i.e.*, $\{X_1, \cdots, X_K\}$;

2) Use ISC to train dictionary in corresponding class, generating each complete dictionary as $\{D_1, \cdots, D_K\}$:

$$D_j = \arg\min_{D,S} \left\| X_j - DS \right\|_2^2 + \lambda \|S\|_1$$

    s.t. $a \le \|d_i\|_2^2 \le b, \quad i = 1, \cdots, M, \quad j = 1, \cdots, K.$

3) Merge these complete dictionaries $\{D_1, \cdots, D_K\}$ into an over-complete dictionary D, *i.e.*, $D = [D_1, \cdots, D_K]$.

4) Apply OMP method to approximate the solution of

$$\alpha_{ij} = arg\min_{\alpha} \|\alpha\|_0 \quad s.t. \quad \left\| D\alpha - R_{ij}X \right\|_2^2 \le T(\sigma).$$

5) Set:

$$\hat{X} = \left(\lambda_1 I + \sum_{ij} R_{ij}^T R_{ij} \right)^{-1} \left(\lambda_1 X + \sum_{ij} R_{ij}^T D\alpha_{ij} \right).$$

End

---

4 Results

In this section, the experimental results with our methods are shown on several test images, which contain animals, fruit and people. The tested noise level of each input image is all the same as sparse coding method [14] used in denoising experiments, so as to guarantee a fair comparison.

In all experiments, the denoising process includes a sparse coding of each patch of size 9×9 pixels from the noisy image. The number of classes K is chosen heuristically to lie between 5 and 10, our result is obtained using K=10. The size of these dictionaries are all 81×81 in every class, hence the over-complete dictionary used is of size 81×810. In other words, proposed algorithm assumes the knowledge of σ, this assumption is the same as [4]. Applying OMP method, atoms are accumulated till the average error passes the threshold, which is empirically chosen to be $T(\sigma) = (1.15 \cdot \sigma)^2$.

Experimental results are based on simulated noise case. For this, the noisy images are created from original noise-free images. The noise is modelled to be zero-mean additive white Gaussian. For the simulated noise experiments, the parameters for our method are tuned to get the best possible results. Some parameters, such as $b = 1$, $\lambda = 0.02$ and $\lambda_1 = 50/\sigma$, can be fixed. One of important parameters for our method is a for the

dictionary learning process. In Figure 2 and 3, the noisy parameter is set by $\sigma = 30$ and the value of a is 0.6.

PSNR	*peppers*	*baboon*	*Parrot*	*Cameraman 1*	*Cameraman 2*
Sparse coding	27.6333	22.7561	25.7651	26.0114	29.3847
Our method	**27.7037**	**22.8115**	**25.8245**	**26.1159**	**29.4619**

Table 2: PSNR (dB) results of different algorithms.
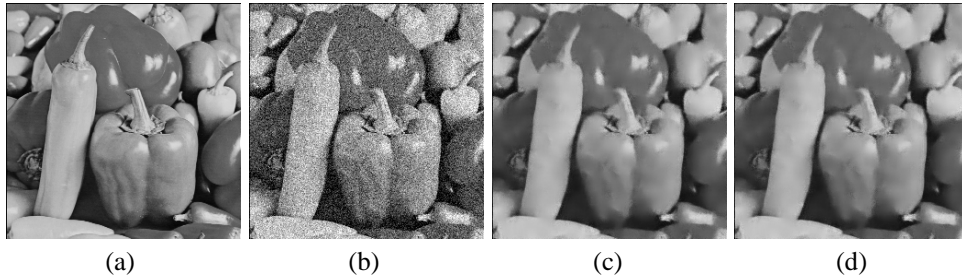

(a)   (b)   (c)   (d)

Figure 2: Comparison of denoising results on *peppers* image corrupted by additive white Gaussian noise of standard deviation 30: (a) Original image; (b) noisy image; (c) sparse coding; (d) our method.
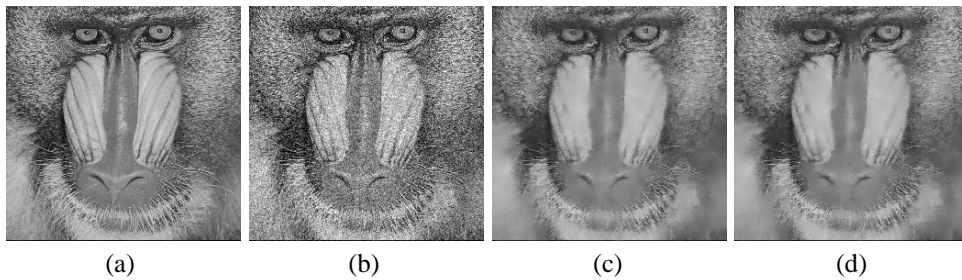

(a)   (b)   (c)   (d)

Figure 3: Comparison of denoising results on *baboon* image corrupted by additive white Gaussian noise of standard deviation 30: (a) Original image; (b) noisy image; (c) sparse coding; (d) our method.

As follows, to verify proposed method is suitable for dealing with strong noise case, Figure 4 and 5 show two results with the standard deviation $\sigma = 35$, a is tuned by 0.8. Section 2 has proved that proposed learning dictionary method can deal with some hybrid

noises. Proposed method is able to effectively denoise even for such a hybrid noise in Figure 6. The hybrid noise includes additive white Gaussian noise of standard deviation $\sigma = 20$ and speckle noise of density $d = 1 \times 10^{-3}$. The value of a is set by 0.8.

To evaluate the quality of denoised image, comparing the results will adopt both objective and subjective assessment criterions. Objectively, PSNR is good criterion to evaluate difference between the original images and the denoised results. Table 2 demonstrates the PSNR of two different algorithms for five test images. In terms of PSNR, our method outperforms sparse coding method. Subjectively, according to the results of each Figure, it is shown that the proposed method is able to obtain better visual quality than sparse coding method.

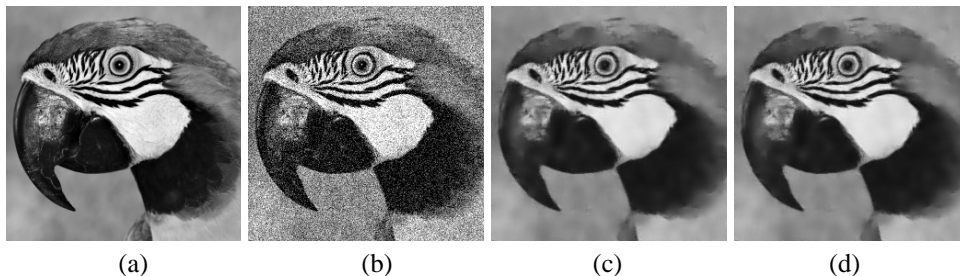| (a) | (b) | (c) | (d) |

Figure 4: Comparison of denoising results on *parrot* image corrupted by additive white Gaussian noise of standard deviation 35: (a) Original image; (b) noisy image; (c) sparse coding; (d) our method.

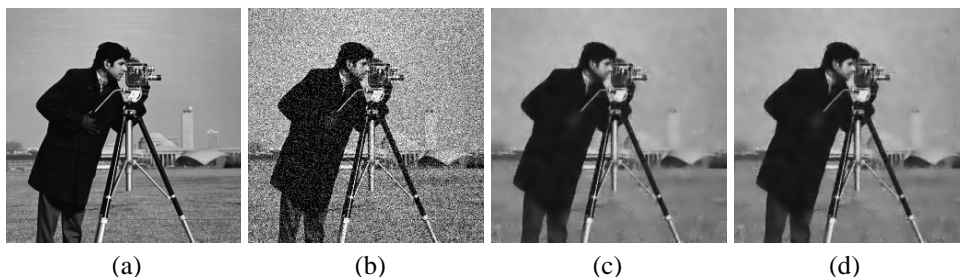| (a) | (b) | (c) | (d) |

Figure 5: Comparison of denoising results on *cameraman 1* image corrupted by additive white Gaussian noise of standard deviation 35: (a) Original image; (b) noisy image; (c) sparse coding; (d) our method.

5 Concluding Remarks

This work takes advantage of the framework of sparse representation, leading to state-of-the-art performance. The proposed method is based on local operations and includes sparse representation of each image block under one fixed over-complete dictionary. The content of dictionary is important for the denoising process. In this paper, we have shown that a dictionary learning on patches of the noisy image itself. Our main contribution is to design a novel dictionary learning method called improved sparse coding, which effectively

restrains the effect of noise. Lots of experimental results show that our method is capable of dealing with denosing process, even in the case of simulated as well as the hybrid noise.

There are still several research directions that we need to consider, such as choosing the number of class, optimizing the parameters, replacing the OMP by a better pursuit technique, and more. As follows, one direction we consider is how to use our dictionary learning method to deal with hybrid noise more effectively.

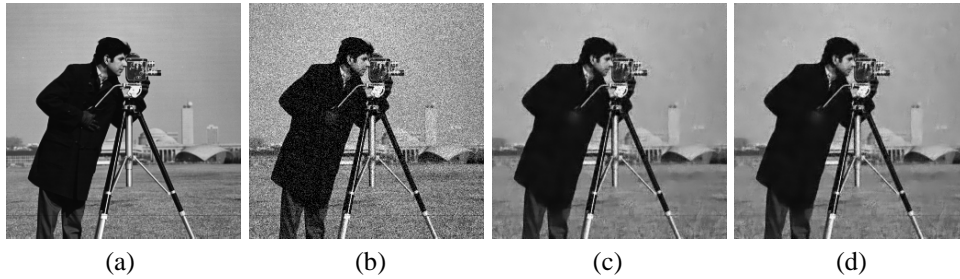| (a) | (b) | (c) | (d) |

Figure 6: Comparison of denoising results on *cameraman 2* image corrupted by additive white Gaussian noise of standard deviation 20 and speckle noise of density 1×10^{-3}: (a) Original image; (b) noisy image; (c) sparse coding; (d) our method.

6 Acknowledgement

References

[1] A. Buades, B. Coll, and J. M. Morel. A review of image denoising methods, with a new one. *Multiscale Model. Simul*, 4(2): 490-530, 2005.

[2] C. Kervrann and J. Boulanger.Optimal spatial adaptation for patch-based image denoisng. *IEEE Trans. Image Process.*, 15(10): 2866-2878, Oct. 2006.

[3] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. *IEEE Trans. Image Process.*, 16(2): 349-366, Feb. 2007.

[4] M. Elad and M. Aharon. Image denosing via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.,* 15(12): 3736-3745, Dec. 2006.

[5] P.Chatterjee and P. Milanfar. Clustering-based denoising with locally learned dictionaries. *IEEE Trans. Image Process.*, 18(7): 1438-1451, Jul. 2009.

[6] K.Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8): 2080-2095, Aug. 2007.

[7] B. Olshausen and D. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vis. Res*. 37: 311-325, 1997.

[8] K.Engan, S. O. Aase, and J. H. Hakon-Husoy. Method of optimal directions for frame design. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 5: 2443-2446, 1999.

[9] K. Kreutz-Delgado and B. D. Rao. Focuss-based dictionary learning algorithms. In *Wavelet Applications in Signal and Image Processing VIII*, 2000.

[10] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. Lee, and T. J. Sejnowski. Dictionary learning algorithms for sparse representation. *Neur. Comput.*, 15(2): 349-396, 2003.

[11] M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neur. Comput.*, 12: 337-365, 2000.

[12] L. Lesage, R. Gribonval, F. Bimbot, and L. Benaroya.Learning unions of orthonormal bases with thresholded singular value decomposition. In *IEEE Intl Conf. Acoustics, Speech, and Signal Processing*, 2005.

[13] Y. Tang, P. Yan, Y. Yuan, X. Li, Single-Image Super-resolution via Local Learning, *International Journal of Machine Learning and Cybernetics*, 2(1): 15-23, 2011.

[14] M. Aharon, M. Elad, and A. M. Bruckstein. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. *J. Linear Algebra Appl.*, 416(1): 48-67, Jul. 2006.

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

[16] J. F. Murray and A. C. Bovik. Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1): 324-335, 1993.