IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 13, 2017, accepted October 6, 2017, date of publication October 13, 2017,
date of current version November 7, 2017.

*Digital Object Identifier 10.1109/ACCESS.2017.2762760*

# Image Denoising via Sparse Representation Over Grouped Dictionaries With Adaptive Atom Size

**LINA JIA[1,2], SHENGTAO SONG[1], LINHONG YAO[1,3], HANTAO LI[1], QUAN ZHANG[1], YUNJIAO BAI[1], AND ZHIGUO GUI[1]**

[1]Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data, North University of China, Taiyuan 030051, China
[2]Department of Electronic Information Engineering, Shanxi University, Taiyuan 030013, China
[3]School of Science, North University of China, Taiyuan 030051, China

Corresponding author: Zhiguo Gui (gzgtg@163.com)

This work was supported in part by the National Key Scientific Instrument and Equipment Development Project of China under Grant 2014YQ24044508, in part by the National Natural Science Foundation of China under Grant 61671413, in part by the Natural Science Foundation of Shanxi Province under Grant 2015011046, in part by the Shanxi Province Science Foundation for Youths under Grant 201601D021080, and in part by the Opening Project of State Key Laboratory of Explosion Science and Technology (Beijing Institute of Technology) under Grant KFJJ13-11M.

**ABSTRACT** The classic K-SVD based sparse representation denoising algorithm trains the dictionary only with one fixed atom size for the whole image, which is limited in accurately describing the image. To overcome this shortcoming, this paper presents an effective image denoising algorithm with the improved dictionaries. First, according to both geometrical and photometrical similarities, image patches are clustered into different groups. Second, these groups are classified into the flat category, the texture category, and the edge category. In different categories, the atom sizes of dictionaries are designed differently. Then, the dictionary of each group is trained with the atom size determined by the category that the group belongs to and the noisy level. Finally, the denoising method is presented by using sparse representation over the constructed grouped dictionaries with adaptive atom size. Experimental results show that the proposed method achieves better denoising performance than related denoising algorithms, especially in image structure preservation.

**INDEX TERMS** Adaptive dictionary learning, image denoising, K-SVD, non-local grouping, sparse representation.

## I. INTRODUCTION

Images often contain noises, which may be caused by sensor imperfection, poor illumination, communication errors and so on. Hence, denoising remains an important research problem in image processing. The problem of denoising can mathematically be shown as estimating the latent clean image $\boldsymbol{X}$ from the noise degraded observation model:

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{V} \tag{1}$$

where $\boldsymbol{V}$ is additive white Gaussian noise (AWGN) with zero-mean and standard deviation $\sigma$.

Being the simplest possible and important inverse problem, noise removal has been widely studied. In the past decades, plenty of researchers tried to address this problem from many points of view, such as partial differential equations [1], [2], transform domain methods [3]–[8], morphological analysis [9]–[11], spatial adaptive filters [12]–[15] and so on.

Among the large number of methods for denoising, the local and non-local self-similarities of natural images are used widely. One of the earlier methods of this kind is the SUSAN filter [13], which was proposed by S. M. Smith and J. M. Brady. The authors proposed denoising by weighted averaging pixels similar in intensity within a local neighborhood. Subsequently, Tomasi and Manduchi [14] proposed bilateral filter (BF) which determines similarities between pixels according to the spacial distance and intensity distance. Indeed, BF is a generalization of SUSAN filter. Based on the fact that natural images often contain large amount of repetitive structures, Buades *et al.* [16] proposed a non-local means (NLM), which searches the similar pixels in non-local neighborhoods. However, the accuracy of identifying similar pixels in these methods is decreasing as the noise level is increasing. In [15], Takeda *et al.* proposed a signal-dependent steering kernel regression (SKR) framework for image denoising, which is proved to be much more robust under high noise levels.

Sparse representation of images, known as another successful method for denoising, has aroused growing interest

of researchers in recent years [17]–[21]. And it has been recognized that dictionary plays an important role in the sparse representation model. Dictionaries that use prespecified transform matrix such as wavelets, curvelets [5], contourlets [6], [7], bandelets [8] are relatively simple and have been applied in various applications successfully. However, this kind of dictionary is lack of generality and adaptability. Learning dictionary can adapt its content through learning from either a given set of image examples or the image itself. In many excellent dictionary learning algorithms, K-SVD [18] is known to be a classical and widely used one. However, it trains only a dictionary with fixed atom size for the whole image, which is limited in describing images accurately.

Considering the excellent performance of the above mentioned two kinds of methods, many researchers combined corresponding models of image self-similarities with sparse representation. Mairal *et al.* [19] proposed simultaneous sparse coding as a framework, which is achieved by jointly decomposing groups of similar signals on subsets of the learned dictionary. Chatterjee and Milanfar [22] proposed a novel image denoising method through clustering the image into regions of similar geometric structure and learning dictionaries by principle component analysis (PCA). In [23], they further improved denoising performances by using similar patches to estimate the different filter parameters of their proposed patch-based Wiener filter. In [24], a PCA-based scheme was proposed for image denoising by using a moving window to calculate the local statistics, from which the local PCA transformation matrix was estimated. However, this scheme applies PCA directly to the noisy image without data selection and many noise residual and visual artifacts will appear in the denoised outputs. Zhang *et al.* [20] overcame the shortcomings of literature [24] by using local pixel grouping (LPG) algorithm which can ensure only the sample blocks with similar contents are used for PCA transform estimation. In literature [21], the authors presented a sparse representation based image restoration algorithm by exploiting the image non-local self-similarity to obtain good estimates of the sparse coding coefficients. However, the atom sizes of the dictionaries used in all the above mentioned references are fixed. An earlier attempt towards adaptive block size selection can be found in [25], where each pixel is estimated pointwise using Local Polynomial Approximation (LPA). Sahoo and Lu [26] proposed an adaptive window selection procedure for local sparse approximation, according to which image patches are clustered into different groups. Due to the different block sizes in different groups, the size of the dictionary for each group is different too. Sahoo [27] provided detailed description of some parameters and further improved their algorithm in [26]. In addition, [27] adopted a sequential generalization of K-means (SGK) [28] to learn dictionaries instead of K-SVD, which is used in [26]. The performance of SGK denoising is similar or comparable to the K-SVD denoising, which is shown in [29].

In this paper we propose an effective sparse representation based denoising method, in which the grouped dictionaries with adaptive atom sizes (GDwAAS) are employed. In the proposed GDwAAS, a new feature detector is designed by making use of both photometrical and geometrical similarities in images. The former is captured by steering kernel regression (SKR) coefficients and the latter by pixel values of image patches. According to the proposed feature detector, image patches are grouped into several clusters. With such a procedure, only similar patches are used for the following dictionary learning, which can improve the accuracy of dictionary. In order to further improve the adaptivity of dictionary, the patch groups are classified into three different genres, then the atom size of each dictionary is selected according to the genre the group belongs to and the noise level.

Although the GDwAAS and the algorithm in [26] and [27] both use adaptive dictionary atom sizes for different clusters, the former is different from [26] and [27]. References [26] and [27] firstly select the optimal block size for each patch at each location of the image according to the rule of Minimum Mean Square Error (MMSE), and then cluster the image patches according to their block sizes. Whereas, our algorithm firstly group image patches with the same block size into different clusters according to the proposed feature detector, and then select atom size adaptively for each group according to the category the group belongs to and the noise level. Our proposed algorithm not only uses the self-similarities in images but also the features of image patches of different types (texture, edge and flat), which are both helpful to improve the accuracy of the dictionaries and the denoising performance.

The contributions of this paper are summarized as follows:

First, we designed a new feature detector by making use of both photometrical and geometrical similarities in images. With this feature detector, searching for similar blocks are more accurate than the commonly used Euclidean distance.

Second, we proposed to use different dictionaries with different atom sizes adaptively according to the category the group belongs to and the noise level. Experiments have shown that image blocks belonging to different types (flat, edge and texture) can achieve the best denoising results with different patch sizes.

The rest of the paper is structured as follows. Section II briefly introduces the sparse representation model and the procedure of SKR. Section III presents the proposed GDwAAS denoising algorithm in detail. Section IV analyzes parameters selection, presents the experimental results, and compares the proposed method with some related algorithms. Section V concludes the paper.

## II. RELATED WORKS

### A. SPARSE REPRESENTATION

In the theory of sparse representation, a signal $\boldsymbol{x} \in R^n$ can be mathematically represented as $\boldsymbol{x} = \boldsymbol{Da}$, where $\boldsymbol{D} \in R^{d^2 \times K} (d^2 < K)$ is an over-complete dictionary with its columns $\boldsymbol{d_1}, \boldsymbol{d_2} \cdots \boldsymbol{d_K}$ as dictionary atoms, $d \times d$ is the

size of the dictionary atoms, $K$ is the number of atoms in $\boldsymbol{D}$, and most entries of the representation coefficients $\boldsymbol{a}$ are zero. This theory has been widely studied and applied in various applications in image processing, such as compression [30], separation [31], inpainting [32], demosaicking [21] and so on. One of the most basic and successful applications can be attributed to M. Elad and M. Aharon, who proposed a classic image denoising model based on sparse representation [17] with their previous proposed dictionary learning algorithm [18]. This model can be described as the following expression:

$$\left\{\hat{\boldsymbol{D}}, \hat{\boldsymbol{a}}_{\boldsymbol{ij}}, \hat{\boldsymbol{X}}\right\} = \arg\min_{\boldsymbol{D},\boldsymbol{a}_{\boldsymbol{ij}},\boldsymbol{X}} \lambda||\boldsymbol{X}-\boldsymbol{Y}||_2^2 + \sum_{i,j}\mu_{ij}||\boldsymbol{a}_{\boldsymbol{ij}}||_0 + \sum_{i,j}||\boldsymbol{D}\boldsymbol{a}_{\boldsymbol{ij}} - \boldsymbol{R}_{\boldsymbol{ij}}\boldsymbol{X}||_2^2 \quad (2)$$

In this model, $\boldsymbol{X}$ is the latent clean image, $\boldsymbol{Y}$ is the observed noised image according to (1), and $\hat{\boldsymbol{X}}$ is the estimation of $\boldsymbol{X}$. $||\boldsymbol{a}_{\boldsymbol{ij}}||_0$ is a pseudo norm that counts the number of non-zero entries in $\boldsymbol{a}_{\boldsymbol{ij}}$, $\boldsymbol{R}_{\boldsymbol{ij}}$ is the matrix extracting patch $\boldsymbol{x}_{\boldsymbol{ij}}$ from $\boldsymbol{X}$, i.e., $\boldsymbol{x}_{\boldsymbol{ij}} = \boldsymbol{R}_{\boldsymbol{ij}}\boldsymbol{X}$ denotes an image patch of size $\sqrt{n}\times\sqrt{n}$ in the *ith* row and *jth* column. Here we can see that the size of the dictionary atoms is the same as the image patch size, that is $d = \sqrt{n}$.

On the right side of (2), the first term is the log-likelihood global force that demands the proximity between $\boldsymbol{X}$ and $\boldsymbol{Y}$. $\lambda$ is a parameter that controls this fidelity term. When the stand deviation $\sigma$ of noisy is small, $\boldsymbol{Y}$ is close to $\boldsymbol{X}$, and $\lambda$ should be larger. In contrast, with the increasing of $\sigma$, $\boldsymbol{Y}$ is getting away from $\boldsymbol{X}$, and $\lambda$ should be smaller. The second and the third terms are used to makes sure that in the constructed image, every patch in every location has a sparse representation with bounded error. $\mu_{ij}$ is used to get a balance between the sparse approximation error of $\boldsymbol{x}_{\boldsymbol{ij}}$ and the sparsity of $\boldsymbol{a}_{\boldsymbol{ij}}$.

In order to solve (2), the block-coordinate minimization algorithm is used and three main steps are included.

Firstly, sparse coding. We start with an initialization $\boldsymbol{X} = \boldsymbol{Y}$ and assume the dictionary $\boldsymbol{D}$ is known and fixed, the sparse decomposition of the sparse decomposition of $\boldsymbol{x}_{\boldsymbol{ij}}$ can be obtained by solving an $l_0$-minimization problem, formulated as:

$$\hat{\boldsymbol{a}}_{\boldsymbol{ij}} = \arg\min_{\boldsymbol{a}} \mu_{ij}||\boldsymbol{a}||_0 + ||\boldsymbol{D}\boldsymbol{a} - \boldsymbol{x}_{\boldsymbol{ij}}||_2^2 \quad (3)$$

The most used and effective methods are the matching pursuit (MP) and the orthogonal matching pursuit (OMP) algorithms.

Secondly, update dictionary. The over-complete dictionary has a great effect on the denoising performance, which can be predefined or learned. However, researches show that the dictionaries learned from images can improve much the sparse representation performance since they can better characterize the image structures [18], [19], [32]. K-SVD [18] is one of the best dictionary learning algorithms, which is used in [17]. The data used to train dictionary can be image patches taken from good quality images, or the corrupted image itself. In this paper we use the latter. Authors can see [18] for the details of K-SVD.

By alternately solving problem in the first two steps several times, we will eventually get the training dictionary $\hat{\boldsymbol{D}}$ and all of the final sparse coefficients $\hat{\boldsymbol{a}}_{\boldsymbol{ij}}$ of the image over the dictionary $\hat{\boldsymbol{D}}$.

Finally, the denoised image is obtained through the following expression:

$$\hat{\boldsymbol{X}} = (\lambda\boldsymbol{I} + \sum_{ij}\boldsymbol{R}_{\boldsymbol{ij}}^{\boldsymbol{T}}\boldsymbol{R}_{\boldsymbol{ij}})^{-1}(\lambda\boldsymbol{Y} + \sum_{ij}\boldsymbol{R}_{\boldsymbol{ij}}^{\boldsymbol{T}}\boldsymbol{D}\hat{\boldsymbol{a}}_{\boldsymbol{ij}}) \quad (4)$$

where $\boldsymbol{I}$ is an identity matrix and $\hat{\boldsymbol{X}}$ is the denoised image.

### B. STEERING KERNEL REGRESSION (SKR)

Kernel regression is a well-studied non-parametric method in statistics and signal processing. It can be used to address image processing problems like denoising, interpolation and deblurring [33]–[36]. Classical kernel regression estimation is a linear combination of local data. Although this method has good properties and is easy to understand due to linearity, it brings some limitations. Adaptive kernel regression methods introduce non-linearity by considering both the spatial distance and the radiometric distances. Bilateral Kernel [36] is one of adaptive kernels, which uses space kernel and gray kernel respectively, however, it ignores the relationship between the spatial position and the pixel value. Steering Kernel [15] overcomes the shortcoming of Bilateral Kernel by using estimation of local gradient. In SKR, a robust estimate of the gradient is taken into consideration in analyzing the radiometric similarity of two pixels in a neighborhood. This information is then used to determine the shape and size of a canonical kernel, resulting in elongated, elliptical contours spread along the directions of the local edge structure. With these locally adapted kernels, the image restoration and reconstruction are effected most strongly along the edges, resulting in strong preservation of edges and details in the final output. In particular, choosing a Gaussian kernel, the steering kernel in this particular case can be expressed as:

$$w_{i,j} = \frac{\sqrt{\det(\boldsymbol{C}_j)}}{2\pi h^2}\exp\{-\frac{(\boldsymbol{p}_i - \boldsymbol{p}_j)^T\boldsymbol{C}_j(\boldsymbol{p}_i - \boldsymbol{p}_j)}{2h^2}\} \quad (5)$$

where $w_{i,j}$ describes the correlation between the *ith* and *jth* pixels. $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ denote the location of the *ith* and the *jth* pixels, respectively. $h$ is a global smoothing parameter which controls the support of the steering kernel. The matrix $\boldsymbol{C}_j$ denotes the symmetric gradient covariance matrix formed from the estimated vertical and horizontal gradients of the *jth* pixel, and $\boldsymbol{C}_j$ can effectively spread the kernel function along the underlying local edges. It can be expressed mathematically as follow:

$$\boldsymbol{C}_j = \gamma_j\boldsymbol{U}_{\boldsymbol{\theta}_j}\boldsymbol{\Lambda}_j\boldsymbol{U}_{\boldsymbol{\theta}_j}^{\boldsymbol{T}} \quad (6)$$

$$\boldsymbol{U}_{\boldsymbol{\theta}_j} = \begin{bmatrix}\cos\theta_j & \sin\theta_j \\ -\sin\theta_j & \cos\theta_j\end{bmatrix} \quad (7)$$

$$\boldsymbol{\Lambda}_j = \begin{bmatrix}\sigma_j & 0 \\ 0 & \sigma_j^{-1}\end{bmatrix} \quad (8)$$

where $\boldsymbol{U}_{\theta_j}$ represents the rotation matrix that aligns the Gaussian to the direction of the underlying local edge, and it can be described in (7). $\boldsymbol{\Lambda_j}$ denotes the elongation operator, which is described in (8). $\gamma_j$ is a scaling parameter.

In order to demonstrate the robustness of the SKR coefficients in expressing image structures, an experiment is done in the house image. According to (5), we calculate the weights $w_{i,j}$ for each selected pixel in its neighborhood $(9\times9)$ with the *ith* pixel at center and draw its contour. The result is shown in Fig. 1, which intuitively illustrates the behaviors of steering kernels at various image structures of house. In Fig. 1, the superiority of SKR in representing the underlying structures of images is shown especially in the edge regions. In addition, similar underlying structure can be seen in several locations with different intensities. What is more, despite the presence of noise, the structure of steering kernel weights in the same location remains almost unchanged in the original and noisy images.

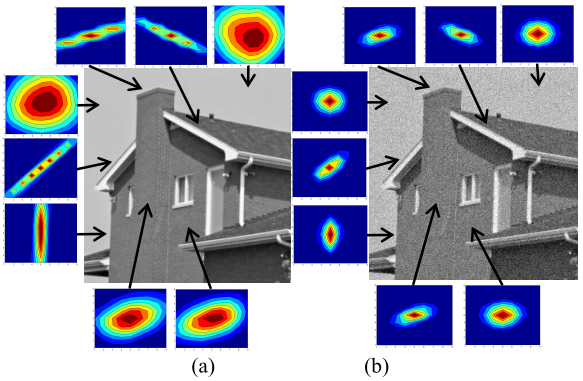



**FIGURE 1.** Steering kernel weights formed from (a) the original House image; (b) noisy image corrupted by AWGN of standard deviation 15.

## III. GDwAAS DENOISING ALGORITHM

Although the classic image denoising model based on sparse representation is effective, it is limited in accurately describing the image due to the single dictionary with fixed atom size. In order to overcome this shortcoming and improve the denoising performance, we proposed a GDwAAS denoising algorithm, in which there are multiple dictionaries with adaptive atom sizes. This is achieved through works in two aspects. On one hand, exploiting the non-local self-similarities, image patches that are similar both geometrically and photometrically are clustered into the same group. In our works, the geometrical information is described by the SKR coefficients, whereas the photometrical information is described by the pixel value. Based on the grouping results, a data-adaptive dictionary can be learned for each group. On the other hand, groups are classified into three different categories: flat category, texture category and edge category. According to the group type and the noise level, the atom size of each dictionary can be designed adaptively.

The GDwAAS denoising algorithm includes three steps: grouping, adaptive atom size selection of dictionaries and denoising. Firstly we identify regions of similar structure in the given image by a new feature detector and group them together. Secondly we classify these patch groups in step 1 into three categories and select appropriate atom sizes for each group. Finally we process image denoising using sparse representation model with learned dictionaries whose atom sizes have been selected in the step 2. The overall block diagram of the GDwAAS algorithm is shown in Fig. 2.

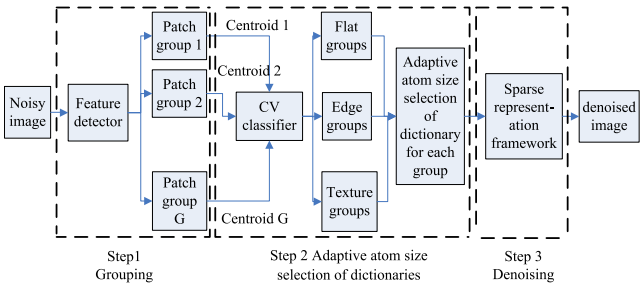



**FIGURE 2.** Overall block diagram of the GDwAAS algorithm.

### A. GROUPING

Suppose the clean image is $\boldsymbol{X} \in R^{M\times N}$, $M$ is the width and $N$ is the height. According to the image degradation model in (1), the observed noisy image is $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{V}$. Select a $\sqrt{n}\times\sqrt{n}$ patch with the pixel $y_i$ at center and realign it to a vector denoted by $\boldsymbol{y_i} = [y_1, y_2 \cdots, y_n]^T$. In order to be pair for each pixel, we extend the boundaries of image $\boldsymbol{Y}$. Hence there are totally $M \times N$ vectors, each of which is used for describing the local information of the centered pixel. Then the image $\boldsymbol{Y}$ can be denoted as $\boldsymbol{Y} = [\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_{M\times N}}]$. Let $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ to be the associated noiseless vectors of $\boldsymbol{y_i}$ and $\boldsymbol{y_j}$ respectively. We can calculate the distance between $\boldsymbol{y_i}$ and $\boldsymbol{y_j}$ by

$$\begin{aligned} d_{i,j} &= ||\boldsymbol{y_i} - \boldsymbol{y_j}||_2 = \sum_{k=1}^{n} (y_i(k) - y_j(k))^2 \\ &\approx \sum_{k=1}^{n} (x_i(k) - x_j(k))^2 + 2\sigma^2 \end{aligned} \tag{9}$$

In (9), we used the fact that noise is white and uncorrelated with signal. if

$$d_{i,j} < T + 2\sigma^2 \tag{10}$$

where $T$ is a preset threshold, then the *ith* and the *jth* patches are thought to be photometrically similar.

If directly use (9) and (10) to search similar patches, two parameters $T$ and $\sigma$ need to be estimated. In practice, we can use classification method without any parameters to find the similar patches. Among the large number of classification algorithms, K-Means proves to be one of the simplest and most effective unsupervised methods. So K-Means is adopted for grouping in this paper.

(9) is actually the Euclidean distance between $\boldsymbol{y_i}$ and $\boldsymbol{y_j}$. The smaller the $d_{i,j}$is, the more similar $\boldsymbol{y_i}$ is to $\boldsymbol{y_j}$. However, Euclidean distance is not always a good choice for block similarity measure. For example, in Fig. 3, (a) is the reference


**FIGURE 3.** Patches used to verify the effectiveness of different methods for measuring similarity between two patches. (a) is the reference patch, (b) and (c) are the candidate patches.

patch and (b) and (c) are the candidate patches. From Table 1, we might conclude that (b) is closer to (a) than (c) according to photometrical similarity. However, by visual comparison, (a) and (c) have the same edge structure, grouping them into the same cluster is beneficial to preserve structure better.

**TABLE 1.** Photometrical similarity(PS), geometric similarity(GS) and the proposed feature similarity(FS) between images in Fig. 3.

| | PS | GS | FS |
|---|---|---|---|
| (a) and (b) | 20.8166 | 16.3357 | 10.6851 |
| (a) and (c) | 162.2803 | 0.4204 | 1.0650 |

In order to solve the deficiency of Euclidean distance in judging similarity, in this paper we make use of geometric similarity(GS) in images. We select the weight coefficients of SKR as a geometric feature to do grouping. For one thing, SKR as introduced in section II, has the ability to align the Gaussian kernel to the underlying edge orientation and strength. For the other, the steering kernel is robust to the presence of noise. Therefore, it can be used as a feature to recognize the underlying geometric structure. For each pixel $y_i$, we calculate the steering kernel weight coefficients $w_{i,j}$ in a patch of size $\sqrt{n}\times\sqrt{n}$ with $y_i$ at center, then vectorize it into $\boldsymbol{w_i} = [w_{i,1}, w_{i,2}\cdots w_{i,n}]^T$ which is used as the feature to identify GS. The GS of the patches in Fig. 3 are shown in the third column of Table 1. We can see that the GS between (a) and (b) is much higher than (a) and (c). Fig. 4 is the grouping results of original House image and its degraded images in the presence of AWGN with a standard deviation of 15, using geometric features extracted by SKR coefficients. Note that the five clusters of the noisy house image are largely in keeping with those obtained from their noise-free images respectively.

Taking into account the respective advantages of photometrical and geometrical similarity in clustering, in this paper we combine the two features into one joint feature, denoted as:

$$\boldsymbol{f_i} = [t\boldsymbol{y_i^T}, (1-t)\boldsymbol{w_i^T}]^T \tag{11}$$

where $t$ is a factor to adjust the weights between $\boldsymbol{y_i^T}$ and $\boldsymbol{w_i^T}$ according to different images. Finally the new feature $\boldsymbol{f_i}$ is used to do patch grouping by K-Means in this paper. For convenience, we denote the similarity measured by the proposed feature $\boldsymbol{f_i}$ as feature similarity (FS). The FS of the patches in Fig. 3 are shown in the fourth column of Table 1.

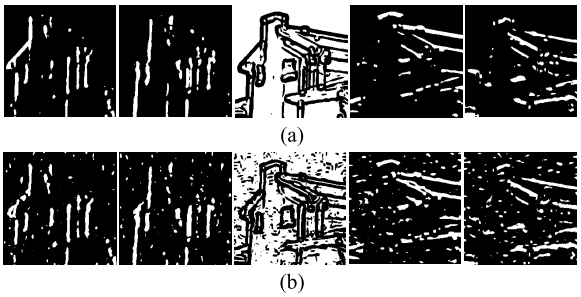



**FIGURE 4.** Performance of grouping using the steering kernel weights as features on the House image. (a) original House image; (b) noisy House image corrupted by AWGN of standard deviation 15.

Hence, the whole noisy image can be thought to be composed of a union of such clusters:

$$\boldsymbol{Y} = \bigcup_{g=1}^{G} \{\boldsymbol{y_i} | i \in \Omega_g\} \tag{12}$$

where G is the number of clusters. Examples of the proposed grouping algorithm are displayed in Fig. 5.

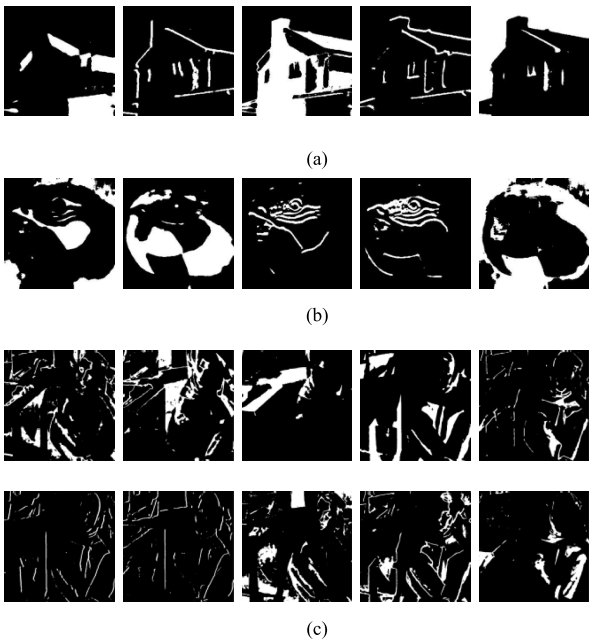



**FIGURE 5.** Performance of grouping using the proposed feature on the noisy House, Parrot and Barbara images corrupted by AWGN of standard deviation 25 respectively. (a) House image; (b) Parrot image; (c) Barbara image.

### B. ADAPTIVE ATOM SIZE SELECTION OF DICTIONARY

Once the image is classified into different groups, each of which has the similar features as we discussed above, we should learn a dictionary that can well describe the patches for each group. That is to say, the dictionary of each group should have the ability of preserving the underlying structures, details and connections of the patches in this group. The problem here can be described mathematically as:

$$\boldsymbol{D_g} = \arg\min_{\boldsymbol{D_g}} ||\boldsymbol{Y_g} - \boldsymbol{D_g A_g}||_2^2 \tag{13}$$

where $\boldsymbol{Y_g} = [\boldsymbol{y_1}, \boldsymbol{y_2} \cdots \boldsymbol{y_m}]$ is the matrix of patches in the *gth* group and $m$ is the total number of patches in the *gth* group. The matrix $\boldsymbol{D_g}$ denotes the learning dictionary of the *gth* group. $\boldsymbol{A_g} = [\boldsymbol{a_1}, \boldsymbol{a_2} \cdots \boldsymbol{a_m}]$ represents the sparse coding coefficients.

In this part we propose an adaptive atom size selection method of dictionary, in order to further improve the adaptability of the dictionary and the denoising performance. As we know, nature images consist of flat, edge and texture regions in most instances. For example, the background areas with the pixel values changing slowly in the House image are regarded as flat region, the outline of the house with the pixel values changing sharply are edge region and the brick wall of the house with the pixel values changing periodically are texture region. Instinctively, if we can distinguish them and deal with each of them using dictionaries with different atom sizes, the dictionary will be more appropriate and the denoising performance will be improved. So there are two problems to be solved. The first problem is how to distinguish the different regions in images? The second problem is how to select proper atom sizes for regions of different types. Next, we will address these two issues in turn.

To address the first problem, two elements are needed. One is the data that to be distinguished and the other is the distinguish criterion.

In our work the centroid of each group, which is the mean vector of that group, is use to do the discrimination. Denote the centroid of group $\boldsymbol{Y_g}$ as $\boldsymbol{y_{gc}}$, then,

$$\boldsymbol{y_{gc}} = (\boldsymbol{y_1} + \boldsymbol{y_2} + \cdots + \boldsymbol{y_m})/m \tag{14}$$

where $m$ is the total number of patches in the group $\boldsymbol{Y_g}$.

There are three reasons for our choice. The first reason is that all patches in the same group are similar. Thus, the centroid can well describe the whole information of that group. The second reason is that the noise discussed in this paper is zero-mean. So the centroid is robust although in the presence of noises due to the average operation in (14). The last reason is that it can be obtained easily in the process of grouping in the first step of our proposed GDwAAS algorithm.

For the distinguish criterion, we use one of mathematical statistics-coefficient of variation (CV), which is a standardized measure of dispersion of a probability distribution or frequency distribution. In the field of image processing, CV can be used for measuring the homogeneity of a region in an image. The homogeneity of a region is defined as the ratio of the standard deviation to the mean of the region. Mathematically,

$$cv(i) = \sqrt{\frac{\mathrm{var}(P)}{\mathrm{ave}^2(P)}} = \sqrt{\frac{\sum_{\mathrm{j}\in\mathrm{P}} (y_j - \mathrm{ave}(P))^2/|P|}{\mathrm{ave}^2(P)}} \tag{15}$$

where $P$ is a square region with the *ith* pixel $y_i$ at center in an image and $|P|$ is the total number of pixels in $P$. $y_j$ is the pixel value at position $j$ in $P$. $ave(P) = \sum_{j\in P} y_j/|P|$ represents the mean pixel value in $P$ and $var(P)$ denotes the mean variance of all pixels in $P$.

From the definition of CV in (15), we can see that the larger the CV value is, the smaller the homogeneity is. On the contrary, the smaller the CV value is, the greater the homogeneity is. In other words, in flat region $cv(i)$ is small, whereas in the texture and edge regions $cv(i)$ is large. Furthermore, experiments show that texture regions have larger $cv(i)$ values than edge regions in general.

Here we can answer the problem how to distinguish the different regions in images. Firstly, we calculate the centroid $\boldsymbol{y_{gc}}$ using (14) for each group $\boldsymbol{Y_g}(g = 1, 2, \cdots G)$, which is derived from the first step of our algorithm. Secondly, $\boldsymbol{y_{gc}}$ is reshaped to a $\sqrt{n} \times \sqrt{n}$ patch, and then the CV of this patch is calculated according to (15). We denote it as $cv_g$. Finally, the $cv_g(g = 1, 2, \cdots G)$ are classified into three classes. The groups with large $cv_g$ belong to texture category, whereas the groups with small $cv_g$ belong to flat category, and the others are edge category.

Following, the problem is how to decide the atom size for each group? Intuitively, for flat regions, bigger atom size of dictionary are preferred in order to remove noisy. And for texture and edge regions, smaller atom sizes of dictionaries are preferred due to their ability to describe details in images. However, with the increase of noise, it is hard to distinguish between signal and noise in small sized atoms even in visual perspective. Hence bigger atom sizes are suitable for denoising in the case of high noise. Nevertheless, the atom sizes are not the bigger the better for two reasons. For one thing, bigger atom sizes will increase computational complexity greatly. For the other, too big atom sizes will result in insufficient image patches to train over-complete dictionary. Thus, there exists a trade off. In this paper we verify our conjecture through some experiments. As we know that House image is rich in edge and flat regions, and Barbara is rich in texture, so in this paper we use House and Barbara to do the following experiments.

To evaluate the quality of the denoised images objectively, two evaluation criteria are used in this paper. One is the peak-signal-to-noise ratio (PSNR), which is the most commonly and widely used objective evaluation standard of image quality. Given two images $X$ and $Y$, both of size $M \times N$, the PSNR between $X$ and $Y$ is defined by:

$$PSNR(X, Y) = 10 \log_{10}(255^2/MSE) \tag{16}$$

where

$$MSE(X, Y) = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij} - y_{ij})^2 \tag{17}$$

From (16) and (17), it shows that the smaller the mean square error (MSE) of $X$ and $Y$ is, the higher the PSNR value is. At the other end of the scale, a small value of the PSNR implies high numerical differences between images.

The other is the structural similarity (SSIM) [37], which is considered to be correlated with the quality perception of the

**TABLE 2.** The K-SVD based denoising results of image House corrupted by AWGN of different level($\sigma = 15, 30, 50$) using different atom sizes. The left is PSNR value and the right is SSIM value.

| atom size | 15 | | 30 | | 50 | |
|---|---|---|---|---|---|---|
| 5*5 | 33.3956 | 0.8533 | 29.5675 | 0.7540 | 26.4089 | 0.6711 |
| 7*7 | 34.2415 | 0.8776 | 30.9559 | 0.8211 | 27.0718 | 0.7471 |
| 9*9 | **34.3898** | **0.8795** | 31.3969 | 0.8376 | **27.0943** | 0.7630 |
| 11*11 | 34.3439 | 0.8770 | **31.4511** | **0.8390** | 26.8437 | 0.7637 |
| 13*13 | 34.2351 | 0.8749 | 31.2593 | 0.8364 | 26.6966 | **0.7650** |
| 15*15 | 34.0718 | 0.8727 | 31.0932 | 0.8355 | 26.3018 | 0.7591 |
| 17*17 | 33.9655 | 0.8717 | 30.9211 | 0.8337 | 25.9557 | 0.7541 |
| 19*19 | 33.8530 | 0.8710 | 30.6821 | 0.8320 | 25.6498 | 0.7484 |

**TABLE 3.** The K-SVD based denoising results of image Barbara corrupted by AWGN of different level($\sigma = 15, 30, 50$) using different atom sizes. The left is PSNR value and the right is SSIM value.

| atom size | 15 | | 30 | | 50 | |
|---|---|---|---|---|---|---|
| 5*5 | 31.4873 | 0.8860 | 27.6473 | 0.7813 | 24.1410 | 0.6480 |
| 7*7 | 32.2496 | 0.9068 | 28.4434 | 0.8176 | **24.4902** | **0.6819** |
| 9*9 | **32.4182** | **0.9099** | **28.4837** | **0.8212** | 24.3292 | 0.6778 |
| 11*11 | 32.3166 | 0.9080 | 28.2957 | 0.8174 | 24.0730 | 0.6684 |
| 13*13 | 32.2131 | 0.9070 | 28.0350 | 0.8108 | 23.7724 | 0.6575 |
| 15*15 | 32.0800 | 0.9051 | 27.7980 | 0.8045 | 23.4787 | 0.6461 |
| 17*17 | 31.9377 | 0.9032 | 27.5522 | 0.7983 | 23.1751 | 0.6352 |
| 19*19 | 31.8273 | 0.9017 | 27.3662 | 0.7938 | 22.9197 | 0.6269 |

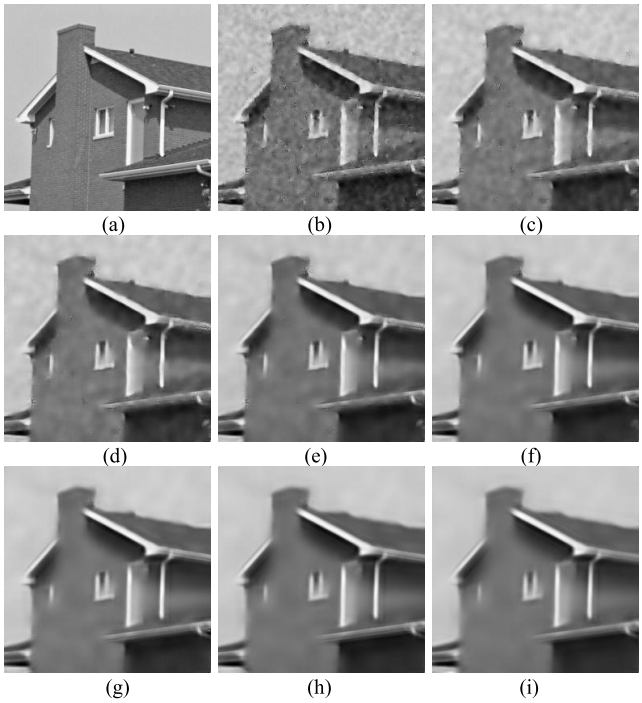



**FIGURE 6.** Comparison of K-SVD based denoising results of image House corrupted by AWGN of standard deviation 50 using different atom sizes. (a)clean, (b)-(i) are the denoising results using atom size of 5,7,9,11,13,15,17 and 19 respectively.

human visual system (HVS). It is commonly used to measure the similarity between two images.

The SSIM between windows $x$ and $y$ of size $W \times W$ is defined as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (18)$$

where $\mu_x$ is the average of $x$, $\mu_y$ is the average of $y$, $\sigma_x^2$ is the variance of $x$, $\sigma_y^2$ is the variance of $y$, $\sigma_{xy}$ is the covariance of $x$ and $y$, The positive constants $c_1$ and $c_2$ are used to avoid a null denominator.

For the given two images $X$ and $Y$, the mean of SSIM (MSSIM) indices to evaluate the overall image quality is given by

$$MSSIM(X, Y) = \frac{1}{tol\_num} \sum_{i=1}^{tol\_num} SSIM(x_i, y_i) \quad (19)$$

where $tol\_num$ is the total number of windows in image $X$ or $Y$. The values of the MSSIM index are in [-1,1]. A value of 0 means no correlation between images, and 1 means that $X = Y$.

*Experiment 1:* corrupted images under different noise levels ($\sigma = 15, 30, 50$) are denoised by K-SVD based denoising method, using different atom sizes (5,7,9,11,13,15,17,19). The denoising results of House and Barbara are shown in Table 2 and Table 3 respectively. From Table 2, we can observe that for $\sigma = 15$ and 30, the denoising results reach their best when the atom size is 9 and 11 respectively. For the high level noise ($\sigma = 50$), the best PSNR and SSIM results are at atom size 9 and 11 respectively. The denoising results of House ($\sigma = 50$) using different atom sizes are shown in Fig.6, in which we can obviously see that although in Table 1 the PSNR value achieves its best at atom size 9, the denosing results in Fig. 6 (d) are visually worse than Fig.6(f). We can also observe that in the small atom size, edge is relatively clear and contrast is large, but there is still residual noise as shown in Fig. 6(b) and Fig. 6(c). With the increase of atom size, flat areas in House become more and more smooth, however, the edge is blurring, and the contrast is decreased. From Table 3, it can be seen that for $\sigma = 15$, 30 and 50, the denoising results reach their best when the atom size is 9, 9 and 7 respectively. This does not seem to agree with our previous conjecture. However it is because that the Barbara image is rich in texture region, so even under high noisy level, objective indicators are still maximum on small atom size. Meanwhile, we can visually see in Fig.7 that denoising image with small size retain more details than big atom size. For example, Fig.7(b) and Fig.7(i). The former is noisy but retains more details and the latter is smooth but loses too many details.

*Experiment 2:* In order to further explain our idea, we extract image patches of size $8 \times 8$ from the clean House image, and classify these patches into three groups according to the patch variance $var$: texture patch ($var \geq 0.02$), edge patch ($0.002 \leq var < 0.02$) and flat patch ($var \leq 0.002$). Then, extract the corresponding patches from the eight denoised images respectively, calculate the PSNR and SSIM


**FIGURE 7.** Comparison of K-SVD based denoising results of image Barbara corrupted by AWGN of standard deviation 50 using different atom sizes. (a)clean, (b)-(i) are the denoising results using atom size of 5,7,9,11,13,15,17 and 19 respectively.

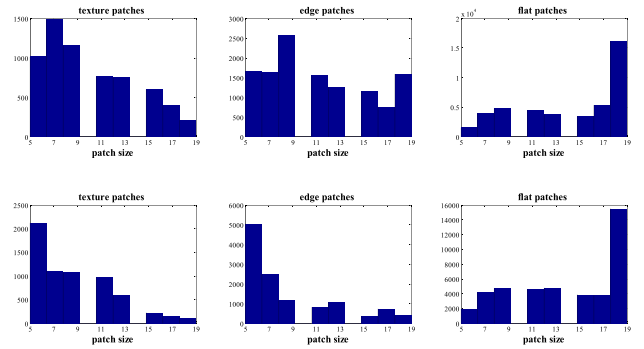



**FIGURE 8.** Distribution of the best PSNR value of each patch in the three different types (texture, edge and flat) in House corrupted by AWGN of standard deviation 30 (top) and 50 (bottom) respectively.

and select the atom size that can performs the best result for each patch. The results are drawn as a bar graph which is shown in Fig.8. From this bar graph, it can be seen that most texture patches reach their best denoisng results under small atom size, and most flat patches prefer big atom size to get their best performance. The average PSNR and SSIM are calculated for each group under different atom sizes and shown in Fig. 9, from which we can get the same conclusion: for texture and edge patches, small atom size can preserve more details, and for flat patches, big atom size are suitable for smoothness.

*Experiment 3:* we select some edge regions, flat regions and texture regions from House and Barbara to do the same work as experiment 2. The selected regions are shown

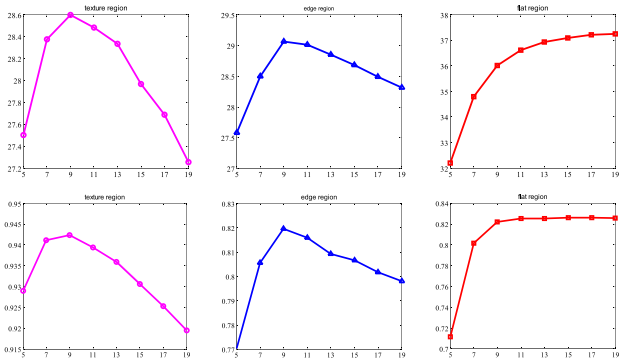



**FIGURE 9.** Average PSNR(top) and SSIM(bottom) of different group (from left to right: texture, edge and flat) in House corrupted by AWGN of standard deviation 30 by using different patch sizes (5,7,9,11,13,15,17,19).

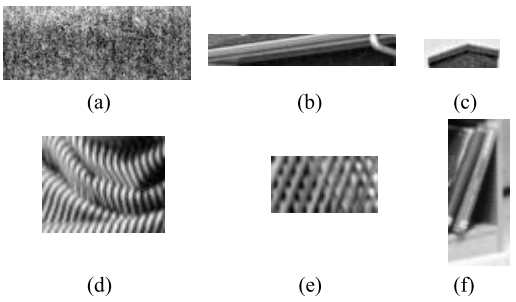



**FIGURE 10.** The sample regions in House and Barbara. (a)-(c) are flat region, edge region 1 and edge region 2 in House respectively. (d)-(f) are texture region 1, texture region 2 and edge region in Barbara respectively.

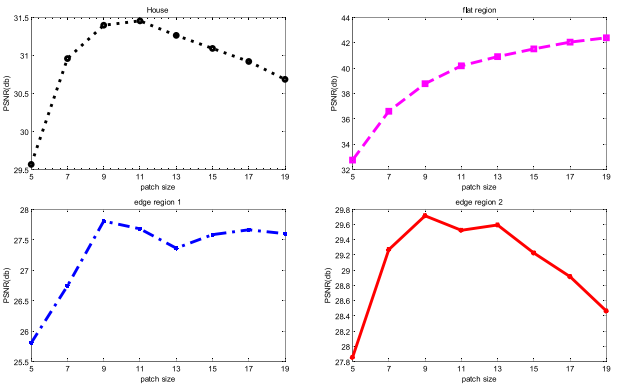



**FIGURE 11.** Comparison of PSNR using different patch sizes on image House corrupted by AWGN of standard deviation 30.

in Fig.10. The results are shown in Fig.11-14. In Fig.11, for the whole House corrupted by AWGN of standard deviation of 30, it reaches its best PSNR and SSIM value when the patch size is 11. For the flat region which is shown in Fig.10(a), the bigger the patch size is, the higher the PSNR value is, but the SSIM values almost no longer increase starting from patch size 15 which can be seen in Fig. 12. And from Fig. 15, we can visually see that the bigger the patch size is, the smoother the denoised image is. What's more, the denosied images using small patch sizes, such as 5,7,9,11, are not only noisy but also bring artificial traces like clouds

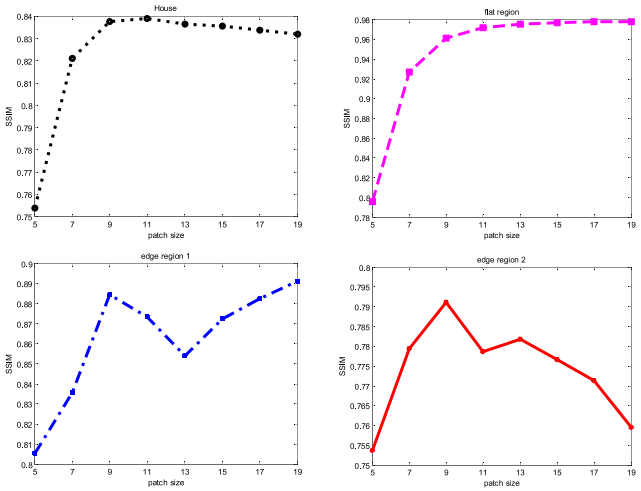



**FIGURE 12.** Comparison of SSIM using different patch sizes on image House corrupted by AWGN of standard deviation 30.

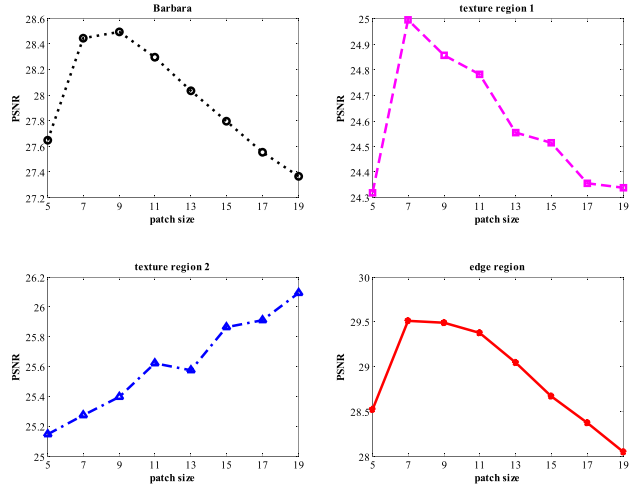



**FIGURE 13.** Comparison of PSNR using different patch sizes on image Barbara corrupted by AWGN of standard deviation 30.

which can seriously affect visual effects. For the edge regions which are shown in Fig. 10(b) and Fig. 10(c), the best results are at patch size 9, which is smaller than the whole image. In Fig. 13-14, although at the same noise level as House, Barbara reaches its best PSNR and SSIM value at patch size 9. It may be because Barbara is rich in texture. For the strong texture region, for example Fig.10(d) (var(d) = 0.032), it reaches the best denoising results at patch size 7. And we can visually see in Fig.16 that the contrast of the denoising images become lower as the patch size increases. For the feeble texture region, for example Fig.10(e) (var(e) = 0.0046), it reaches the best denoising results at the largest patch size. This is because the corrupted feeble texture regions are more like flat regions. From Fig.17, we can clearly see faint details are almost entirely lost when the patch size is small, and under the large patch size, although there is also a part of detail loss, the basic texture can be preserved, which greatly improve the quality of denoising image.

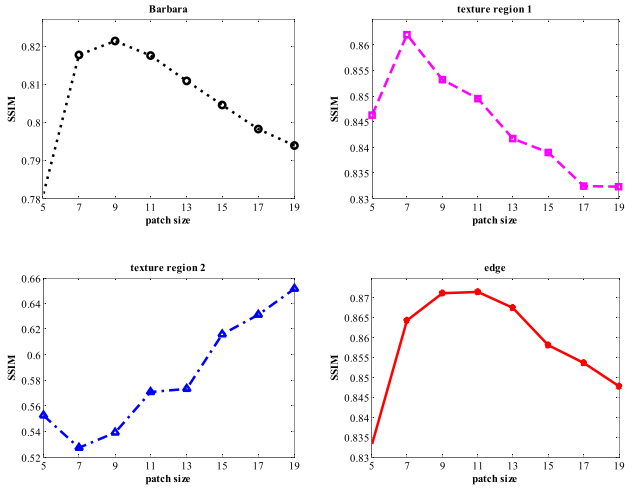



**FIGURE 14.** Comparison of SSIM using different patch sizes on image Barbara corrupted by AWGN of standard deviation 30.

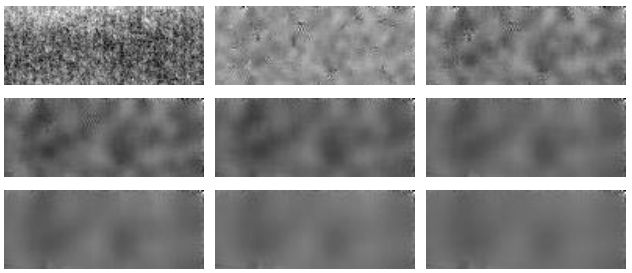


**FIGURE 15.** Comparison of flat region in House corrupted by AWGN of standard deviation 30 using atom size of 5,7,9,11,13,15,17,19.The first one is the clean reference.

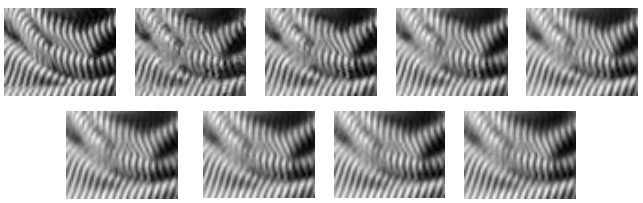


**FIGURE 16.** Comparison of strong texture region in Barbara corrupted by AWGN of standard deviation 30 using atom size of 5,7,9,11,13,15,17,19. The first one is the clean reference.

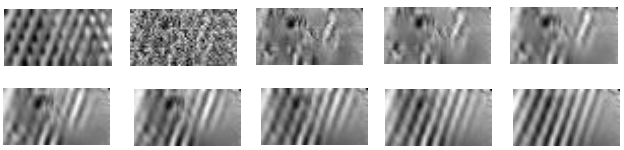


**FIGURE 17.** Comparison of feeble texture region in Barbara corrupted by AWGN of standard deviation 30 using atom size of 5,7,9,11,13,15,17,19 from the third one respectively. The first and second one are the clean and noisy references respectively.

Through the experiments and discussion above, we can see that in the same image,it is propitious to preserve details for texture and edge patches by using a dictionary with small atom size, and for the smooth patches it is better to use a dictionary with large atom size for better removing noises. In this paper, the atom sizes of dictionaries for different

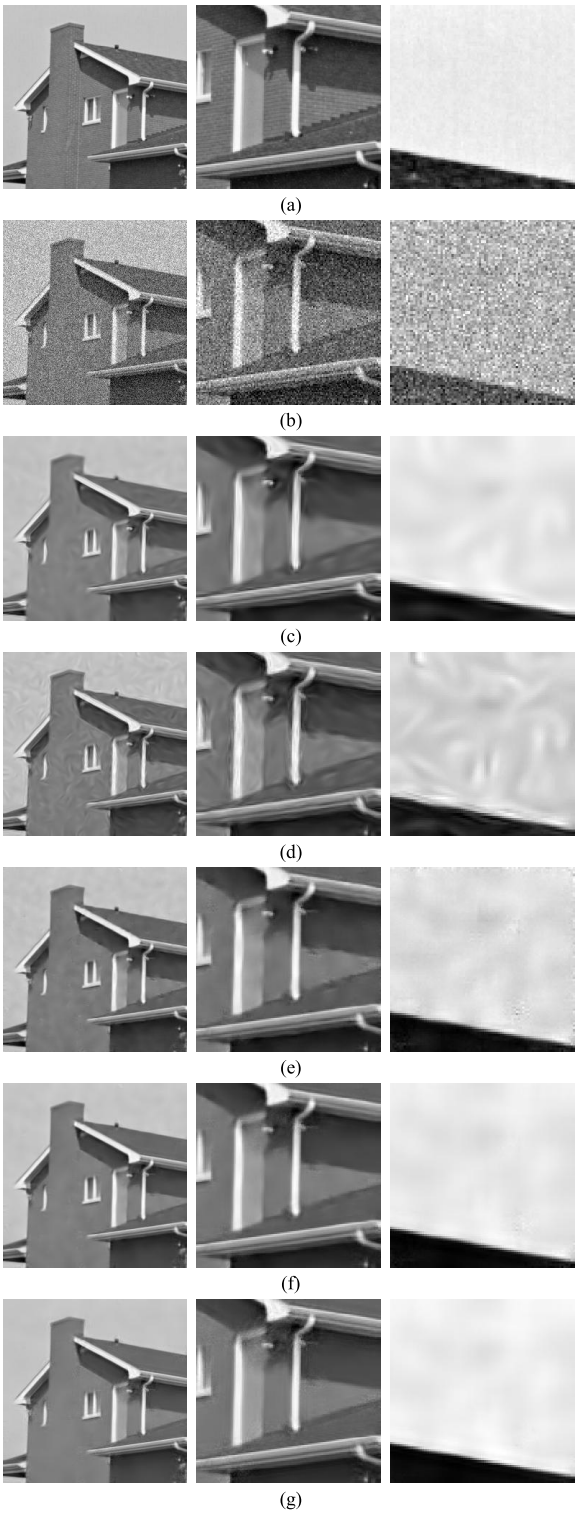


**FIGURE 18.** Comparison of denoising methods on House image. (a) Original image, (b) noisy image corrupted by AWGN of standard deviation 25, (c) ISKR, (d) KLLD, (e) K-SVD, (f) [27], (g) GDwAAS. The second and third columns are partial enlarged details of the first column respectively.

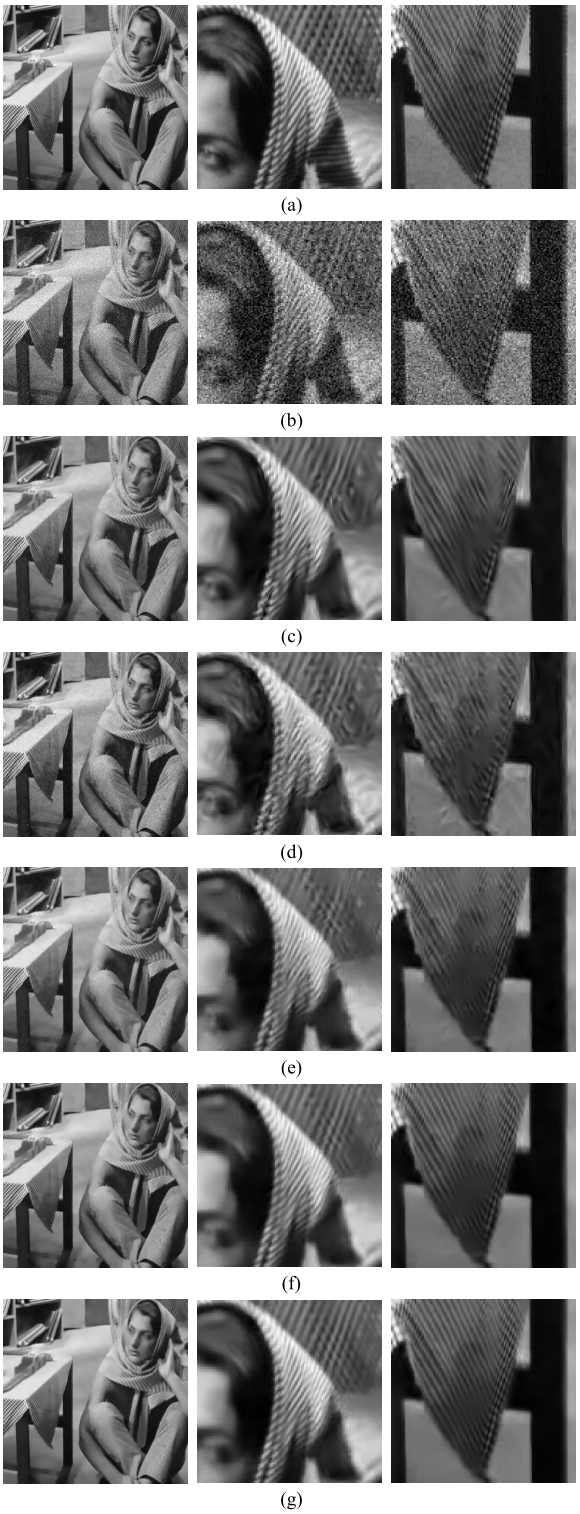


**FIGURE 19.** Comparison of denoising methods on Barbara image. (a) Original image, (b) noisy image corrupted by AWGN of standard deviation 25, (c) ISKR, (d) KLLD, (e) K-SVD, (f) [27], (g) GDwAAS. The second and third columns are partial enlarged details of the first column respectively.

groups are decided as follow:

$$d = \begin{cases} s+2, & \textit{flat} \\ s, & \textit{edge} \\ s-2, & \textit{texture} \end{cases} \tag{20}$$

where $s$ is an empirical variable which varies with the $\sigma$ of noise. In our paper, $s = 9, \sigma < 25, s = 11, 25 \leq \sigma < 40$ and $s = 13, \sigma \geq 40$.

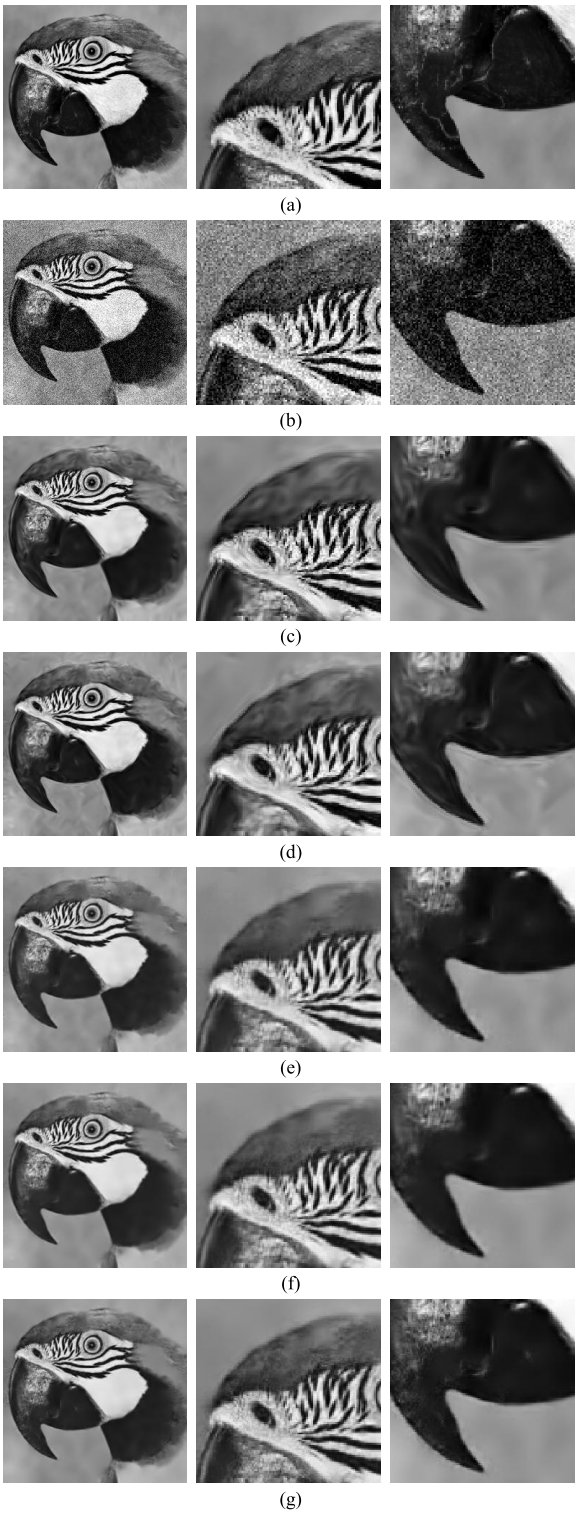



**FIGURE 20.** Comparison of denoising methods on Parrot image. (a) Original image, (b) noisy image corrupted by AWGN of standard deviation 25, (c) ISKR, (d) KLLD, (e) K-SVD, (f) [27], (g) GDwAAS. The second and third columns are partial enlarged details of the first column respectively.

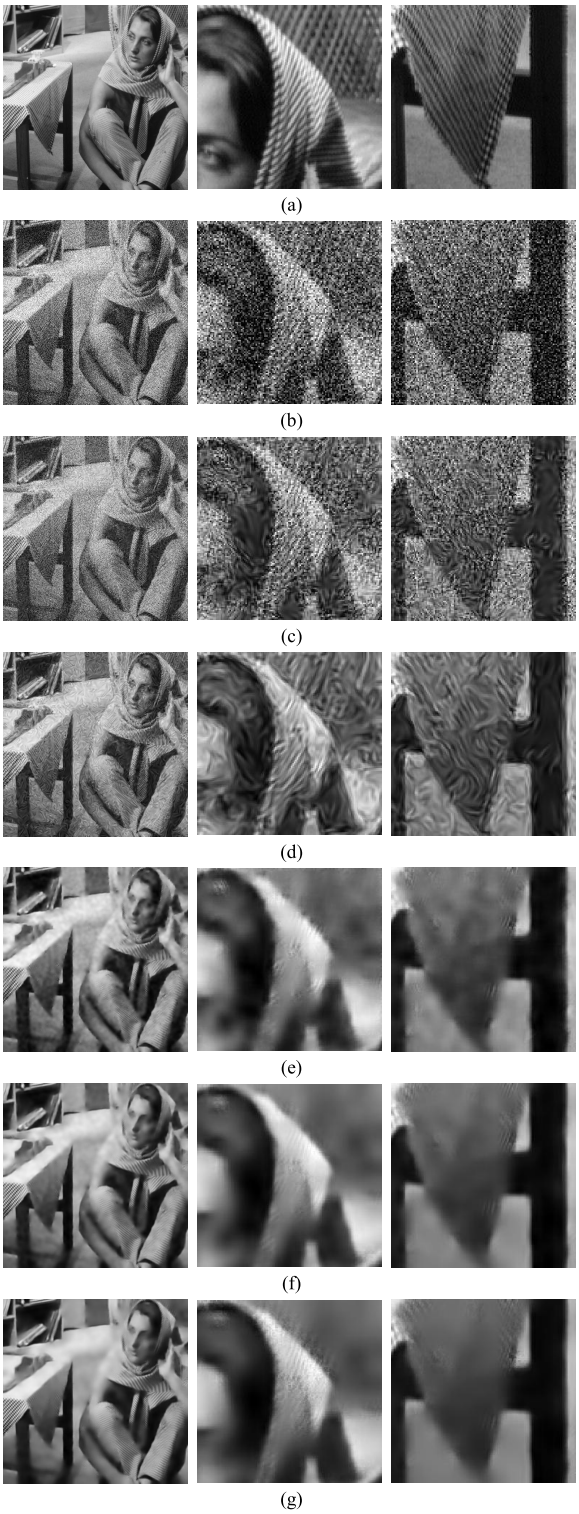



**FIGURE 21.** Comparison of denoising methods on Barbara image. (a) Original image, (b) noisy image corrupted by AWGN of standard deviation 50, (c) ISKR, (d) KLLD, (e) K-SVD, (f) [27], (g) GDwAAS. The second and third columns are partial enlarged details of the first column respectively.

### C. DENOISING

In the above two steps, the image patches are clustered into several groups and the atom size of dictionary for each group is selected adaptively. In (13), we can see that the atom size of the dictionary $\boldsymbol{D}_g$ should be the same as the patch size in $\boldsymbol{Y}_g$. As discussed above, the atom size of a dictionary

**TABLE 4.** Denoising performance of some related methods (ISKR [15], KLLD [22], K-SVD [17] and the algorithm in [27]) under AWGN of different levels, compared with GDwAAS. The results noted are PSNR.

| Methods | ISKR | KLLD | K-SVD | [27] | GDwAAS |
|---|---|---|---|---|---|
| House | | | | | |
| $\sigma$ =15 | 33.7681 | 33.8813 | 34.3646 | 34.5634 | **34.7114** |
| $\sigma$ =20 | 32.6562 | 32.7583 | 33.1866 | 33.402 | **33.5842** |
| $\sigma$ =25 | 31.7634 | 31.5203 | 32.1645 | 32.446 | **32.6237** |
| $\sigma$ =30 | 30.8238 | 30.5142 | 31.1935 | 31.5888 | **31.7214** |
| $\sigma$ =40 | 23.9110 | 27.8469 | 29.3863 | 29.8600 | **30.0495** |
| $\sigma$ =50 | 17.4012 | 24.8562 | 27.1359 | 27.5333 | **28.0038** |
| Barbara | | | | | |
| $\sigma$ =15 | 31.0798 | 30.8066 | 32.3960 | 32.6986 | **32.8573** |
| $\sigma$ =20 | 29.7001 | 29.0690 | 30.7832 | 31.1718 | **31.2987** |
| $\sigma$ =25 | 28.6669 | 27.6646 | 29.5192 | 29.914 | **30.2264** |
| $\sigma$ =30 | 27.5552 | 26.6679 | 28.5014 | 28.8587 | **29.0949** |
| $\sigma$ =40 | 22.6224 | 24.8597 | 26.4019 | 26.7204 | **26.9405** |
| $\sigma$ =50 | 17.2498 | 23.1773 | 24.4360 | 24.7959 | **25.0463** |
| Lena | | | | | |
| $\sigma$ =15 | 29.6429 | 32.2580 | 32.5138 | 32.6087 | **32.7511** |
| $\sigma$ =20 | 29.9759 | 30.8800 | 31.0159 | 31.1211 | **31.2030** |
| $\sigma$ =25 | 29.8361 | 29.7843 | 29.8553 | 29.9574 | **30.0882** |
| $\sigma$ =30 | 28.8232 | 28.7679 | 28.7833 | 28.8344 | **28.9472** |
| $\sigma$ =40 | 22.9215 | 26.6040 | 27.0085 | 26.9495 | **27.1816** |
| $\sigma$ =50 | 17.1635 | 24.3510 | **25.3513** | 25.3078 | 25.3410 |
| Parrot | | | | | |
| $\sigma$ =15 | 30.7538 | 30.8540 | 30.9788 | 30.8668 | **31.1273** |
| $\sigma$ =20 | 29.2133 | 29.5418 | 29.3982 | 29.3187 | **29.5715** |
| $\sigma$ =25 | 28.0206 | 28.4046 | 28.1702 | 28.1057 | **28.3442** |
| $\sigma$ =30 | 26.9544 | 27.4777 | 27.1569 | 27.0546 | **27.4683** |
| $\sigma$ =40 | 22.8974 | 25.4635 | 25.3902 | 25.2027 | **25.4728** |
| $\sigma$ =50 | 17.9110 | 23.3379 | 23.7821 | 23.5465 | **23.8191** |
| average | | | | | |
| $\sigma$ =15 | 31.3112 | 31.9500 | 32.5633 | 32.6844 | **32.8618** |
| $\sigma$ =20 | 30.3864 | 30.5623 | 31.0960 | 31.2534 | **31.4144** |
| $\sigma$ =25 | 29.5718 | 29.3435 | 29.9273 | 30.1058 | **30.3206** |
| $\sigma$ =30 | 28.5392 | 28.3570 | 28.9088 | 29.0841 | **29.3080** |
| $\sigma$ =40 | 23.0881 | 26.1935 | 27.0467 | 27.1832 | **27.4111** |
| $\sigma$ =50 | 17.4314 | 23.9306 | 25.1763 | 25.2959 | **25.5526** |

**TABLE 5.** Denoising performance of some related methods (ISKR [15], KLLD [22], K-SVD [17] and the algorithm in [27]) under AWGN of different levels, compared with GDwAAS. The results noted are SSIM.

| Methods | ISKR | KLLD | K-SVD | [27] | GDwAAS |
|---|---|---|---|---|---|
| House | | | | | |
| =15 | 0.8644 | 0.8649 | 0.8791 | **0.8821** | 0.8817 |
| =20 | 0.8526 | 0.8505 | 0.8617 | 0.8603 | **0.8636** |
| =25 | 0.8420 | 0.8247 | 0.8476 | 0.8529 | **0.8530** |
| =30 | 0.8252 | 0.7954 | 0.8335 | **0.8409** | 0.8388 |
| =40 | 0.5040 | 0.6909 | 0.8024 | 0.8183 | **0.8203** |
| =50 | 0.2078 | 0.5401 | 0.7583 | 0.7777 | **0.7924** |
| Barbara | | | | | |
| =15 | 0.8911 | 0.8826 | 0.9097 | 0.9154 | **0.9174** |
| =20 | 0.8696 | 0.8509 | 0.8785 | 0.8904 | **0.8914** |
| =25 | 0.8474 | 0.8069 | 0.8473 | 0.8609 | **0.8711** |
| =30 | 0.8241 | 0.7709 | 0.8211 | 0.8350 | **0.8424** |
| =40 | 0.6030 | 0.6811 | 0.7540 | 0.7690 | **0.7815** |
| =50 | 0.3155 | 0.5717 | 0.6817 | 0.7002 | **0.7147** |
| Lena | | | | | |
| =15 | 0.8584 | 0.9027 | 0.9070 | **0.9111** | 0.9110 |
| =20 | 0.8669 | 0.8761 | 0.8817 | **0.8862** | 0.8861 |
| =25 | 0.8686 | 0.8510 | 0.8591 | 0.8655 | **0.8663** |
| =30 | 0.8423 | 0.8189 | 0.8328 | **0.8394** | 0.8389 |
| =40 | 0.5479 | 0.7142 | 0.7846 | 0.7898 | **0.7954** |
| =50 | 0.2581 | 0.5920 | 0.7291 | 0.7343 | **0.7400** |
| Parrot | | | | | |
| =15 | 0.8804 | 0.8800 | 0.8888 | 0.8898 | **0.8916** |
| =20 | 0.8561 | 0.8568 | 0.8287 | 0.8572 | **0.8641** |
| =25 | 0.8333 | 0.8261 | 0.8287 | 0.8302 | **0.8319** |
| =30 | 0.8317 | 0.7970 | 0.8070 | 0.8091 | **0.8112** |
| =40 | 0.6043 | 0.7090 | 0.7597 | 0.7595 | **0.7652** |
| =50 | 0.3794 | 0.5972 | 0.7155 | 0.7140 | **0.7172** |
| average | | | | | |
| =15 | 0.8736 | 0.8826 | 0.8962 | 0.8996 | **0.9004** |
| =20 | 0.8613 | 0.8586 | 0.8627 | 0.8735 | **0.8763** |
| =25 | 0.8478 | 0.8272 | 0.8457 | 0.8524 | **0.8556** |
| =30 | 0.8308 | 0.7956 | 0.8236 | 0.8311 | **0.8328** |
| =40 | 0.5648 | 0.6988 | 0.7752 | 0.7842 | **0.7906** |
| =50 | 0.2902 | 0.5756 | 0.7212 | 0.7316 | **0.7411** |

is decided according to (20), so in order to perform the following denoising, here the image patches in a group should be rebuilt according to the atom size of the dictionary for this group. In other word, $n$ is changed with $d$ and $n = d^2$. After this operator, we now perform image denoising under sparse representation framework. For each group, through performing sparse coding and dictionary update alternately several times, the adaptive learning dictionaries and sparse coding coefficients can be obtained. Finally we can get the denoised image patch vectors by

$$\hat{\boldsymbol{x}}_{\boldsymbol{gi}} = \boldsymbol{D}_{\boldsymbol{g}}\boldsymbol{a}_{\boldsymbol{gi}}, \quad g = 1, 2, \cdots G; \ i = 1, 2, \cdots m \qquad (21)$$

where $\hat{\boldsymbol{x}}_{\boldsymbol{gi}}$ is the vector of the $ith$ denoised image patch in the $gth$ group, $\boldsymbol{D}_{\boldsymbol{g}}$ denotes the learning dictionary of the $gth$ group, and $\boldsymbol{a}_{\boldsymbol{gi}}$ is the sparse representation coefficients of $\boldsymbol{x}_{\boldsymbol{gi}}$ over $\boldsymbol{D}_{\boldsymbol{g}}$.

In this paper, we use overlapped patch-based model to perform denoising. So when the vectors are reshaped and replaced back to their original positions, the denoised pixels are overlapped too. In order to get the correct denoising results, the expression in (4) is used to calculate the final results.

## IV. RESULTS

We verify performance of the proposed denoising method through experiments on various images (House, Barbara, Lena, Parrot) at different noise levels ($\sigma$ = 15, 20, 25, 30, 40, 50). The results obtained by GDwAAS are compared to some related methods including the ISKR [15], KLLD [22], K-SVD [17] and the algorithm in [27].

### A. PARAMETER ANALYSIS AND SELECTION

In our experiments, there are several parameters to be tuned. We introduce them in turn according to the steps where these parameters appear in the proposed algorithm framework shown in Fig. 2 in this paper.

In the step 1, one important parameter is the number of clusters $G$ in the grouping step. In general, too few clusters can lead to structurally dissimilar patches being clustered

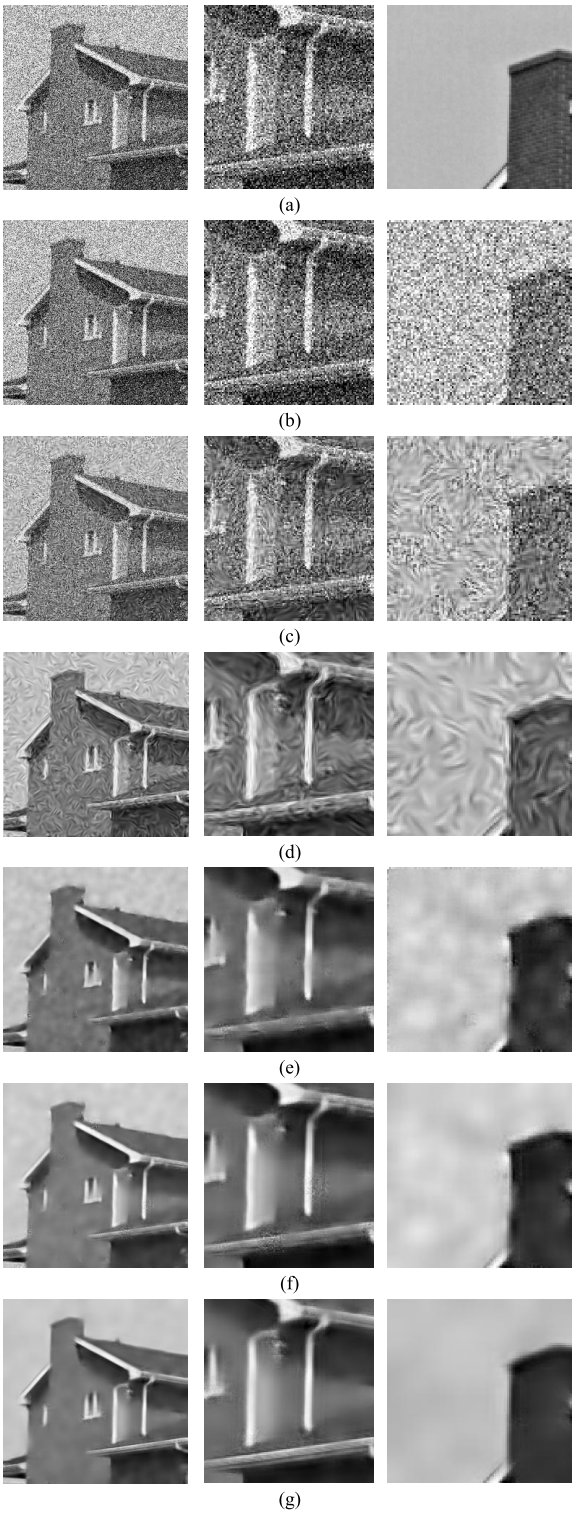



**FIGURE 22.** Comparison of denoising methods on House image. (a) Original image, (b) noisy image corrupted by AWGN of standard deviation 50, (c) ISKR, (d) KLLD, (e) K-SVD, (f) [27], (g) GDwAAS. The second and third columns are partial enlarged details of the first column respectively.

together and cannot reach the purpose of classification in our algorithm. However, too many clusters bring about too few patches in each group, making the dictionary learning impossible or less believable. In addition, similar patches are clustered into different groups constrainedly if $G$ is too large, leading to over grouping. In fact, the optimal number of clusters varies from image to image as it depends on the structural complexity of the image. In this paper, we choose this parameter to be 5 for images having simple structures, such as the House image, and 10 for images having complex structures, such as the Barbara image. The second important parameter is the smoothing parameter $h$ for the kernel formation process. It is worth noting that the ability of extracting geometric structure varies according to $h$. Too big or too small values are both not conducive to obtain the underlying structure. In our experiment we choose it as the same as literature [22]. Another important one is the patch size $\sqrt{n}$. In our experiments $\sqrt{n}$ is set to be 9 in the first step of our algorithm empirically and in the third step $n$ is changed with $d$ and $n = d^2$. The last, but not the least one is the weight factor $t$ used for controlling the weights between the geometric feature and photometric feature. In our experiment it is set to be 0.2.

In the step 2, the atom size $d$ of the dictionary, is set according to (20).

In the step 3, there are also several parameters to be set to get better denoising performance. The first one is the number of K-SVD iterations for the dictionary learning. In our experiments the iterations vary according to the dictionary atom sizes. The same as [27], in this paper the iterations are calculated by $10 \times \lceil d^2/64 \rceil$. The second one is the parameter which is used to control the sparse representation error $||\boldsymbol{Da} - \boldsymbol{x_{ij}}||_2^2$ in solving (3). Generally, $||\boldsymbol{Da} - \boldsymbol{x_{ij}}||_2^2 \leq nC^2\sigma^2$, and $C = 1.15$ [17]. In the proposed algorithm, multiple dictionary atom sizes are designed. So we select $C = sqrt(1 + 2.68/d)$ [27]. The last is the $\lambda$ in (2) and (4). In this paper we set it the same as the reference [17] and [27].

### B. EXPERIMENTAL RESULTS

The results obtained with the House, Barbara, and Parrot images corrupted by AWGN of standard deviation 25 are shown in Figs.18-20, respectively. In order to make a clear observation, the partial enlarged details are shown in the second and third columns of Figs.18-20 respectively. We can subjectively and visually see that GDwAAS is better than the others. ISKR do well in capturing the edges and details, but the protection is so excessive that some distortions and artifacts are introduced. In Fig.19(c), the texture of curtain and tablecloth is clear but distorted seriously. KLLD overcomes the over-protection of ISKR, but the edges and details are lost. For example, in Fig.19(d) the texture of tablecloth is not clear. What is worse, KLLD generates lots of artifacts in the denoisied images especially in the smooth regions, which can be seen in the background areas of house in Fig.18(d). K-SVD is an excellent algorithm in image processing, but it trains one dictionary for the whole image, so the dictionary is lack of robustness and accuracy. In Fig.19(e), the texture of the curtain in the top right of Barbara image is seriously lost and in Fig.18(e), the background is not so smooth as the

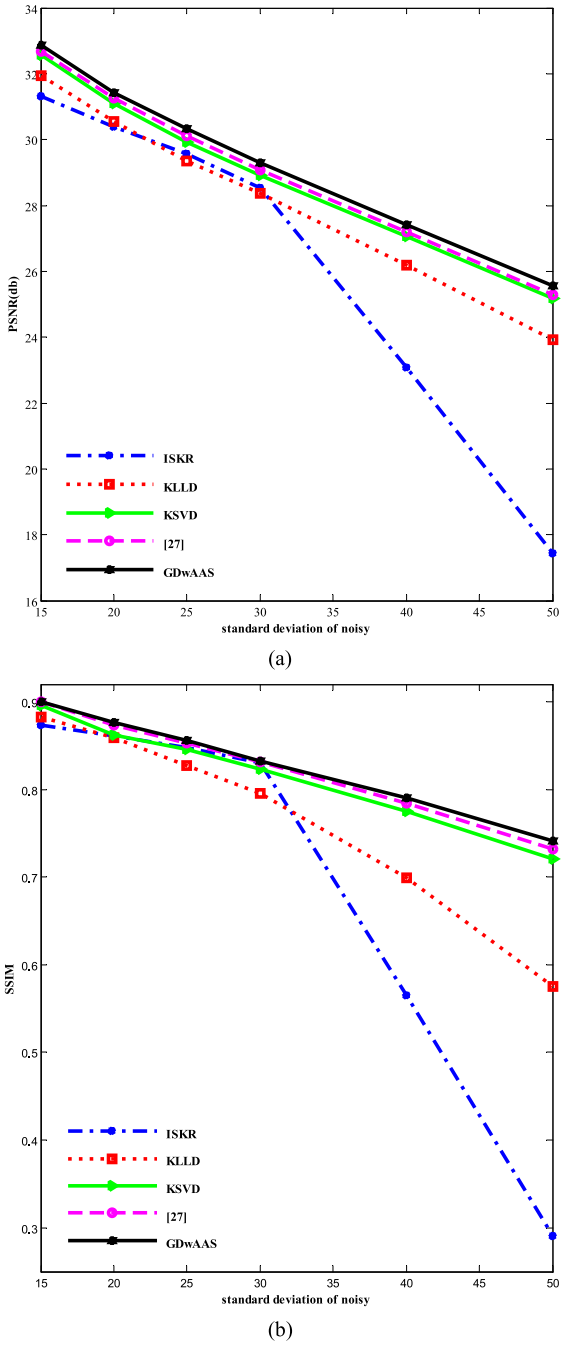



**FIGURE 23.** Comparison of denoising performance on average among ISKR, KLLD, K-SVD, the algorithm in [27] and GDwAAS. (a) PSNR, (b) SSIM.

original one. The method in [27] performs well than ISKR, KLLD and K-SVD, however it is inferior to GDwAAS. In Fig.19(f), the details of the curtain in the top right are a little lost compared with the result in Fig.19(g). In Fig.20(f), the details of the cracks in the parrot's mouth are unclear. The proposed GDwAAS performs better than the others. For example, in the House image, the edges are well protected and the background is smoother; in the Barbara image, the texture of the curtain and tablecloth is more close to the original image and in the Parrot image, the details of the feather on the head is more abundant and the cracks on the Parrot's mouth are more clear than the others.

In the case of high noise level, the superior performance of our algorithm is more obvious compared with the others. The results obtained for the Barbara and House images corrupted by AWGN of standard deviation 50 are shown in Figs. 21-22, respectively. We can see that the denoising performance of ISKR and KLLD drop sharply when the noisy level is high. For ISKR, there was little difference before and after denoising. And for KLLD, artifacts and distortion are more serious. K-SVD shows serious loss of details and unsmooth effects on flat regions, e.g., in Fig. 22(e) the eave of the House is lost seriously and the background of the House is cloudlike and no longer smooth. The method in [27] performs better than K-SVD, however it is still inferior to our proposed GDwAAS. Compared with the results of [27] in Fig. 22(f), the edges of House are more clear, the contrast is more obvious and the background is more smooth in Fig. 22(g). In Fig.21(c)-(f), the stripes on the Barbara's hood are almost invisible, or seriously distorted for the other four algorithms, whereas the proposed GDwAAS preserves the details better in Fig. 21(g) .

Objective comparison results of these methods are shown in Table 4 and Table 5 respectively. From Table 4, we can see that the GDwAAS can improve 0.03-0.47 dB the PSNR values for different images under different noise level ($\sigma$ is from 15 to 50), compared with the method in [27], which is recently published and performs better than the classic K-SVD. From Table 5, the SSIM values of GDwAAS are maximal in most instances. Compared with [27], the maximum of increased SSIM value is 0.015.

For more intuitive observation of the comparison results, two line charts of comparison between the five methods are shown in Fig. 23, which are drawn by the average data of the four images used in this paper. It can be seen from Fig. 23(a) that the sequence of denoising proficiency (measured by PSNR) is ISKR, KLLD, K-SVD, the method in [27] and GDwAAS from low to high. However ISKR is better than KLLD when the standard deviation of noise is between 20 and 30. When the standard deviation of noise is higher than 30, ISKR performs worst. From Fig. 23(b), it is clear that the proposed GDwAAS performs better whether in high or low noisy level. Overall, in both charts, although ISKR, KLLD, K-SVD and the method in [27] have competed when noisy level varies, the proposed GDwAAS algorithm show more proficiency in image denoising.

## V. CONCLUSIONS

In this paper we have proposed an effective image denoising method via sparse representation over grouped dictionaries with adaptive atom sizes. In our work, we first cluster the image into several groups using a feature detector combining geometrical and photometrical features together. The geometrical feature is characterized by SKR coefficients and the photometrical feature is calculated by pixel values. In this step, the approach makes image patches to be classified according

to its underlying structure, which leads to well protection of fine details and structures. Then, groups are classified to three genres, and then the atom size of dictionary is designed for each group respectively. In this step, the grouped dictionaries are further improved from the aspect of their atom sizes. Finally the image is denoised under the sparse representation model. Over all, through efforts in the above three steps, the performance of image denoising is improved. Experimental results presented in this paper demonstrate that the proposed GDwAAS outperforms some related denoising methods. Besides, we can observe that with the increase of noisy level, the qualities of the denoised images are dropped for all methods. Especially for the serious corrupted images ($\sigma \geq 50$), although our algorithm is far superior to the other algorithms mentioned in this paper, there is still a problem of serious loss of details. Thus, researches of dictionary transfer learning [38]–[40], which can use dictionaries from clear image to denoise corrupted images in other domain, will be an important direction in our future work.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments which improved the quality of this paper greatly.

## REFERENCES

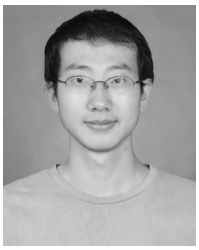
[1] C. A. Micchelli, L. X. Shen, and Y. S. Xu, "Proximity algorithms for image models: Denoising," *Inverse Problems*, vol. 27, no. 4, p. 045009, Apr. 2011.
[2] W. Zeng, X. Lu, and X. Tan, "A local structural adaptive partial differential equation for image denoising," *Multimedia Tools Appl.*, vol. 74, no. 3, pp. 743–757, Feb. 2015.
[3] J. Tian, L. Chen, and L. H. Ma, "A wavelet-domain non-parametric statistical approach for image denoising," *IEICE Electron. Exp.*, vol. 7, no. 18, pp. 1409–1415, Sep. 2010.
[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[5] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Electron. Packag. Manuf.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.
[6] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
[7] R. Eslami and H. Radha, "Translation-invariant contourlet transform and its application to image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3362–3374, Nov. 2006.
[8] S. Mallat and E. LePennec, "Sparse geometric image representations with bandelets," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 423–438, Apr. 2005.
[9] C. Yu and X. Chen, "Remote sensing image denoising application by generalized morphological component analysis," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 33, pp. 83–97, Dec. 2014.
[10] J. Bobin, J.-L. Starck, J. M. Fadili, Y. Moudden, and D. L. Donoho, "Morphological component analysis: An adaptive thresholding strategy," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2675–2681, Nov. 2007.
[11] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Comput. Harmon. Anal.*, vol. 19, no. 3, pp. 340–358, 2005.
[12] T. Dai, W. Lu, W. Wang, J. L. Wang, and S.-T. Xia, "Entropy-based bilateral filtering with a new range kernel," *Signal Process.*, vol. 137, pp. 223–234, Aug. 2017.
[13] S. M. Smith and J. M. Brady, "SUSAN—A new approach to low level image processing," *Int. J. Comput. Vis.*, vol. 23, no. 1, pp. 45–78, May 1997.
[14] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Int. Conf. Comput. Vis.*, Washington, DC, USA, Jan. 1998, pp. 839–846.
[15] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 349–366, Feb. 2007.
[16] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Washington, DC, USA, Jun. 2005, pp. 60–65.
[17] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[18] M. Aharon, M. Elad, and A. Bruckstein, "*rmK*-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[19] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Tokyo, Japan, Sep./Oct. 2009, pp. 2272–2279.
[20] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognit.*, vol. 43, no. 4, pp. 1531–1549, Apr. 2010.
[21] W. Dong, L. Zhang, G. Shi, and X. Li, "nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.
[22] P. Chatterjee and P. Milanfar, "Clustering-based denoising with locally learned dictionaries," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1438–1451, Jul. 2009.
[23] P. Chatterjee and P. Milanfar, "Patch-based near-optimal image denoising," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1635–1649, Apr. 2012.
[24] D. D. Muresan and T. W. Parks, "Adaptive principal components and image denoising," in *Proc. Int. Conf. Image Process.*, vol. 1. Sep. 2003, pp. I101–I104.
[25] V. Katkovnik, K. Egiazarian, and J. Astola, "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule," *J. Math. Imag. Vis.*, vol. 16, no. 3, pp. 223–235, 2002.
[26] S. K. Sahoo and W. Lu, "Image denoising using sparse approximation with adaptive window selection," in *Proc. IEEE 8th Int. Conf. Inf. Commun. Signal Process.*, Dec. 2011, pp. 1–5.
[27] S. K. Sahoo. (Dec. 2016). "Local sparse approximation for image restoration with adaptive block size selection." [Online]. Available: https://arxiv.org/abs/1612.06738
[28] S. K. Sahoo and A. Makur, "Dictionary training for sparse representation as generalization of K-means clustering," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 587–590, Jun. 2013.
[29] S. Sahoo and A. Makur, "Image denoising via sparse representations over sequential generalization of K-means (SGK)," in *Proc. IEEE 9th Int. Conf. Commun. Inf. Signal Process.*, Dec. 2013, vol. 3. no. 3, pp. 1–5.
[30] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 35, no. 1, pp. 111–129, Jan. 2013.
[31] L. Stankovic, I. Orovic, S. Stankovic, and M. Amin, "Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4562–4572, Sep. 2013.
[32] S. Li, Q. Cao, Y. Chen, Y. Hu, L. Luo, and C. Toumoulin, "Dictionary learning based sinogram inpainting for CT sparse reconstruction," *Optik-Int. J. Light Electron Opt.*, vol. 125, no. 12, pp. 2862–2867, 2014.
[33] V. Katkovnik, A. Foi, and K. Egiazarian, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Comput. Vis.*, vol. 86, no. 1, pp. 1–32, Jan. 2010.
[34] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang, "Image and video restorations via nonlocal kernel regression," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 1035–1046, Jun. 2013.
[35] H. Takeda, S. Farsiu, and P. Milanfar, "Deblurring using regularized locally adaptive kernel regression," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 550–563, Apr. 2008.
[36] H. Takeda, S. Farsiu, and P. Milanfar, "Higher order bilateral filters and their properties," in *Proc. SPIE*, vol. 6498, p. 64980S, Feb. 2007.
[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[38] G. Chen, C. Xiong, and J. J. Corso, "Dictionary transfer for image denoising via domain adaptation," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 1189–1192.

[39] L. Cao, C. Wang, and J. Li, "Vehicle detection from highway satellite images via transfer learning," *Inf. Sci.*, vol. 366, pp. 177–187, Oct. 2016.
[40] I. Kuzborskij, F. Orabona, and B. Caputo, "Scalable greedy algorithms for transfer learning," *Comput. Vis. Image Understand.*, vol. 156, pp. 174–185, Mar. 2014.

**LINA JIA** received the M.S. degree in communication and information system from Taiyuan University of Technology, China, in 2010. She is currently pursuing the Ph.D. degree with the North University of China. She joined the Department of Electronic and Information Engineering, Shanxi University, in 2011. Her research interests include image restoration, sparse representation, and pattern recognition.

**SHENGTAO SONG** received the Ph.D. degree in mechanical design and theory from the North University of China in 2013. He is currently an Associate Professor with the North University of China. His research interests include robotics and mechanisms.

**LINHONG YAO** received the Ph.D. degree in fundamental mathematics from Beijing Normal University in 2012. She joined the College of Science, North University of China. Her research interests include partial differential equations, nonlinear analysis, and mathematical method in image processing.



**HANTAO LI** received the Ph.D. degree in theoretical physics from Nanjing University in 2015. He is currently a Teacher with the North University of China. His research interests include therapy and theoretical nuclear physics.

**QUAN ZHANG** received the Ph.D. degree in computer science and technology from Southeast University, Nanjing, China, in 2015. He is currently involved in teaching and research with the North University of China. His research interests include medical image reconstruction and medical image analysis.

**YUNJIAO BAI** received the M.S. degree in signal and information processing from the North University of China, Taiyuan, China, in 2014, where he is currently pursuing the Ph.D. degree. His research interests include image denoising and enhancement.

**ZHIGUO GUI** received the Ph.D. degree in signal and information processing from the North University of China in 2004. He is currently a Professor with the North University of China. His research interests include image processing and image reconstruction.