An Adaptive Contourlet HMM–PCNN Model of Sparse Representation for Image Denoising

GUOAN YANG, (Member, IEEE), ZHENGZHI LU, JUNJIE YANG, AND YUHAO WANG

School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding author: Guoan Yang (gayang@mail.xjtu.edu.cn)

ABSTRACT In this paper, we propose a novel model of sparse representation for image denoising that we call an adaptive contourlet hidden Markov model (HMM)–pulse-coupled neural network (PCNN). In this study, we first adopted a contourlet transform to decompose a noisy image to be some subband coefficients of various directions at various scales. The contourlet emulated extremely well the sparse representation performance of human visual perception, such as its multiscale characteristics, geometric features, and bandpass properties. Second, we used an HMM method to create a statistical model that expressed the coefficient relationships in intrabands, interbands, intrascales, and interscales. Then we used an expectation–maximization training algorithm to obtain the state probability. The result included the state, scale, and direction, the position of the coefficient, the noisy image, and the parameter set of the HMM model. Third, we put the state probability into the PCNN model, which could adaptively optimize the parameters of the HMM model and get better coefficients of clean images. Finally, we transformed the image denoising problem into a Bayesian posterior probability estimation problem. We also reconstructed a denoised image based on the clean coefficients obtained from our proposed method. The experimental results show that the contourlet HMM–PCNN model proposed in this paper is superior to the contourlet with hidden Markov tree model and the wavelet threshold method.

INDEX TERMS Contourlet HMT model, HMM–PCNN model, image denoising, sparse representation, visual perception.

I. INTRODUCTION

Research indicates that most information acquired by humans comes from human vision, which is visually selective; it can perfectly distinguish small amounts of important information from a large amount of visual information. This is called sparse coding [1], [2]. Horace Barlow [3] has proposed an effective coding hypothesis: the human visual perception system can adapt to its environment because the visual perception cells can effectively filter out statistically redundant external signals. The system effectively represents the infinite information of nature by using limited neurons.

Accordingly, researchers hope that computers can be made to simulate the filtering ability of human visual perception. In recent years, computer vision and neural computing, parallel computing, associative memory, and other intelligent information-processing methods have become much more

The associate editor coordinating the review of this manuscript and approving it for publication was Gerardo Di Martino.

closely interconnected. These features have enabled computers to make great progress in simulating human visual perception and information processing [4]. More achievements in the studies of brain science and cognitive science have been applied to the fields of computer vision and image understanding. Two examples are visual attention decision-making and the temporal pattern recognition of the lateral temporal lobe [5], [6]. Neurophysiological studies show that the primary mammalian visual cortex has a sparse coding mechanism that perceives the basic characteristic of a receptive field using a simple cell [2]. This sparse representation model can effectively describe the characteristics of human visual perception [7]. In addition, studies of neurophysiology and psychology show that neurons of the human visual cortex have high selectivity for visual stimuli of various features in static and dynamic states. They are selective for orientation, spatial frequency, speed, color, and other salient features [8].

Olshausen and Field have stated that the natural image is sparse and have described the essential sparsity of image

data [9]. After that, the sparse theory was widely studied and advocated [10]. It is believed that the sparsity property is efficient for representing the rich, varied, and directional information contained in natural images [11]. It has also been proven that the base function in sparse representation has the characteristics of spatial localization, orientation, and bandpass; these are closely related to the properties of the receptive fields of simple cells [2]. Recently, studies further validated the idea that sparse coding performs in a perceptual way that simulates the human visual system (HVS) on natural images [12]. Numerous successful applications have been devised based on the sparse model, including image coding, image retrieval, image denoising, image restoration, and quality assessment.

Image denoising is an important way to improve image quality in electronic imaging systems. Imaging systems have become more easily corrupted by noise with the decrease in size and the increase in channel complexity of CMOS–CCD sensors. Therefore, image denoising is becoming increasingly important for high-resolution imaging.

The most popular image denoising methods are the filtering-based method and the learning-based method [13]. Noisy images include natural images, remote sensing images, infrared images, synthetic aperture radar (SAR) images, and hyperspectral images [14]. There are also several types of noises, such as Gaussian noise, salt and pepper noise, speckle noise, and Poisson noise [15].

In recent years, representative achievements in image denoising include the following research. In [16] Chen et al. proposed a weighted couple sparse representation method for impulse noise removal. The method focused mainly on impulse noise removal and mathematical solutions. In [17], Yuan and Ghanem presented the total variation and the proximal alternating direction method of multipliers, which they called the TV–PADMM method for achieving the impulse noise removal of natural images. In [18], Yin et al. proposed a highly accurate image reconstruction method for multimodal impulse noise suppression on big data. That was a learning-based image denoising method. In [19], Wang et al. proposed a nonconvex weighted ℓ_p norm minimization-based group sparse representation framework for image denoising. That study adopted the generalized soft-thresholding algorithm to solve the nonconvex ℓ_p norm minimization problem. Also, the experiment results were excellent for reducing additive white Gaussian noise (AWGN). In [20], Zha et al. proposed a group sparsity residual constraint for image denoising with an external, nonlocal, self-similarity prior. Thus, they transformed the problem of image denoising into a problem of reducing the group sparsity residual to achieve better image denoising effects for high-density AWGN.

However, the denoising methods of [16]–[18] were suitable only for impulse noise, and in [19] and [20] they did not fully consider the visual perception characteristics.

To make better use of biological visual perception characteristics for image noise removal, we propose an adaptive contourlet-HMM–PCNN model of sparse representation based on a contourlet-HMT model developed by Po and Do [21]. Here, we make full use of the fact that an HMM has a very strong time-warping ability, and a PCNN has a strong classification ability. An organic combination of these two functions may be able to overcome their shortcomings. Also, the proposed approach is superior to the contourlet-HMT method and the wavelet threshold method.

This paper is organized as follows: In Section 2, we propose a sparse image representation method using an HMM–PCNN model in the contourlet domain. In Section 3, we present an adaptive HMM–PCNN modeling implementation in the contourlet domain. In Section 4, we describe experiment results and give an analysis based on an adaptive HMM–PCNN model in the contourlet domain for image denoising applications. In Section 5, we discuss the conclusions and further studies.

II. AN HMM–PCNN MODEL IN THE CONTOURLET DOMAIN

A. A PCNN MODEL

In 1990, Eckhorn proposed the connection model, which reveals pulse synchronization emissions based on the phenomenon of synchronous pulse emission in the cerebral cortex of the cat [22]. In 1996, Olshausen and Field pointed out that human vision has a sparse coding feature [9]. They also successfully illustrated the characteristics of vision nerve cells through a sparse coding maximization hypothesis. This hypothesis was that if the properties of image edges were described accurately by using sparse coding, the sparse method was similar to the dynamic response characteristics of the receptive fields of simple cells [9]. Moreover, the base function for sparse coding successfully simulates the three response properties of the receptive fields of simple cells: localization in a spatial domain, direction characteristics, and direction selectivity in the time and frequency domains. Considering the overcomplete properties of the basis functions, in 1997 Olshausen and Field proposed a sparse coding algorithm for the overcomplete basis and successfully simulated the dynamic response characteristics of the receptive fields of simple cells [23].

In 1999, Johnson proposed the PCNN model by improving the Eckhorn model. PCNN is known as a third-generation artificial neural network (ANN), which implements a function similar to that of the visual cortex in the mammalian brain. A PCNN has the characteristics of both synchronous pulse emission and global coupling. Its mechanisms for representing and processing information are more consistent with the physiological foundation of the human visual system [24]. In the traditional neural network method, the computing cells consist of both perceptron and activation functions. The PCNN model is similar to the real neural network of the mammalian visual cortex in terms of the information processing and signal transmission mechanisms. The model requires no complex training, and the calculating operations of the neural network are relatively simple [25]. However, in recent years, studies have found other neurons with more complex

receptive fields in the mammalian visual cortex, such as the complex cells, the ultracomplex cells, and the grating cells [26]. Among them, the grating cells have a substantial response to a grid-like stimulation, which provides better insight into the reticular texture features. A PCNN not only nearly conforms to the visual perception data measured practically, but also can better describe the basic characteristics of the grating cells.

Research on the human visual perception mechanism shows that the HVS can be regarded as a reasonable and efficient image processing system. The HVS is a series of cells from the retina to the cerebral cortex that can be regarded as a receptive field. The receptive field is the basic structural and functional unit for information processing in the visual system. It is also a region for generating or modulating the cell response on the retina. In addition, neurophysiology studies have shown that the cell receptive field in the primary visual cortex has significant directional sensitivity. A single neuron responds only to some signals of particular frequency bands, such as the edges, contours, and texture features in an image. On the other hand, the spatial receptive field is described as an information coding filter with localized, directional, and bandpass characteristics. Because a PCNN possesses the characteristics of emission and global coupling of synchronous pulses, it can extract useful information from a complex background based on the processing mechanism of the human visual neural system.
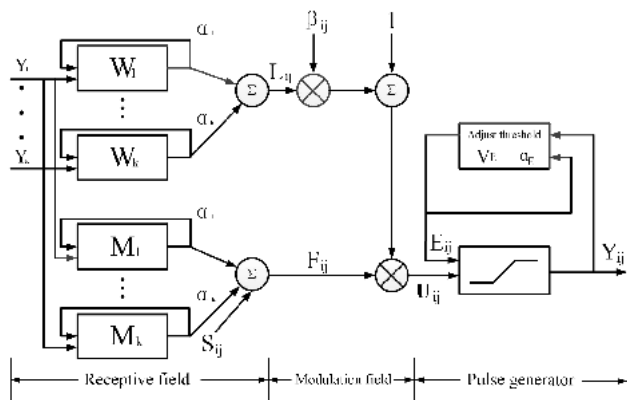

FIGURE 1. Structure chart of a PCNN.

B. IMPLEMENTING A PCNN MODEL

Figure 1 shows that a PCNN is composed of a receiving module, a nonlinear modulating module, and a pulse emission module. Among these, the receiving module provides the information required by the modulating module by integrating the neighborhood pulse signals of the neurons. Those signals include both the link and the feedback input channel accomplished by L_{ij} and F_{ij} respectively. The modulating module implements the nonlinear modulations of the signal transmitted from the receiving module and then obtains the dynamic connection U_{ij}. The pulse emission module carries out the pulse output Y_{ij} by making a comparison between the dynamic connection term U_{ij} and the adjustment

threshold E_{ij}. The model simultaneously adjusts the threshold value of the next iteration according to the pulse output of the neuron. Note that the pulse emission of each neuron has two states: firing and inhibition. In sparse image representation research, the gray value of each pixel is used as the pulse generated by an external stimuli S_{ij}. Following the external stimuli, the feedback signal generated by the pulses of the input of the neighboring neuron into the feed channel, and the link signal generated by the pulses of the neighboring neurons input into the link channel. Here, the PCNN model simulates the process of neuron activity and can be expressed by (1) to (5).

$$F_{ij}(n) = \exp(-\alpha_{kl}^F)F_{ij}(n-1) + S_{ij} + V_F \sum_{kl} M_{ijkl} Y_{kl}(n-1) \tag{1}$$

$$L_{ij}(n) = \exp(-\alpha_{kl}^L)L_{ij}(n-1) + V_L \sum_{kl} W_{ijkl} Y_{kl}(n-1) \tag{2}$$

$$U_{ij}(n) = F_{ij}(n)(1 + \beta_{ij}L_{ij}(n)) \tag{3}$$

$$E_{ij}(n) = \exp(-\alpha_E)E_{ij}(n-1) + V_E Y_{ij}(n) \tag{4}$$

$$Y_{ij}(n) = \begin{cases} 1, & U_{ij}(n) \geq E_{ij}(n) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

From the equations above, we know that the PCNN model consists of six main parameters:

1. The attenuation time constants of α_{kl}^L and α_{kl}^F, which determine the self-attenuation rate in the link and feed input channels.

2. The attenuation time constant α_E of the threshold E_{ij}, where α_E not only regulates the decrement of the threshold E_{ij} after each iteration but also controls the resolution of threshold output adjustments. In other words, the threshold attenuation is faster if E_{ij} is greater and the iteration number of the PCNN model is lower, but the threshold attenuation is slower if E_{ij} is smaller and the iteration number is higher.

3. The amplitude coefficient V_E of the threshold E_{ij}, which determines the overshoot scopes of the threshold E_{ij} after the neurons fire. It also directly influences the pulse emission period of the neurons.

4. The amplitude coefficients of the link and feed channels V_L and V_F, which regulate the amplitude of those channels and can increase or decrease the pulse intensity relayed from the neighboring neurons fired to the central neuron.

5. The link strength β, which controls the contribution of the neighboring neurons to the central neuron for an advance firing. The capturing ability of neurons is stronger with a stronger β; thereby, the scope of pulse synchronization emissions excited by the neuron becomes much greater.

6. The connecting weights $M_{ijk\ell}$ and $W_{ijk\ell}$, which are generally a 3×3 matrix and present the strength of influence through the pulse emission from the neighboring neurons to the central neuron. The strength of

the influence can usually be expressed as the reciprocal of the degree of neuron distances.

From the above analysis, we can see that the values of the parameters in the PCNN mathematical model not only entirely control the operational efficiency but also directly affect the performance of the whole PCNN.

C. PCNN MODEL MECHANISM

The PCNN has many excellent characteristics and marked advantages in the field of image processing and computer vision, including the following:

1. Pulse coupling characteristics, as shown in Figure 1. Here, the information transmission between neurons is implemented by using the pulses; this character can effectively facilitate information transmission in the PCNN model.

2. The nonlinear multiplicative modulation characteristic, where the neuron will realize the above function through connection strength β, and the link and feed input are coupled to be the dynamic connection item U_{ij}. This method can simulate well the internal activities of neurons.

3. The neighbor-capturing characteristics of neurons, where the pulse emission period of the corresponding neurons is shorter if the gray values of the image pixels are greater. Because of the nonlinear multiplicative modulation characteristics of dynamic connection item U_{ij}, each neuron can capture similar neurons in the neighborhood. In Figure 1, we assume that the input signal F_{ij} in the feed channel is a constant, so the dynamic connection item U_{ij} increases when the input signal in the link channel arrives.

4. The threshold mechanism of exponential attenuation, which can simulate the perception mechanism of the mammalian vision system and enable the PCNN to search for a global threshold. In Figure 1, a neuron releases the pulse every time, and E_{ij} generates a step enhancement and exponential attenuation until the next pulse output.

D. A CONTOURLET-HMM–PCNN MODEL

It is known that the progress of image processing algorithms depends largely on the improvement of the image model and the image representation method. The contourlet transform is the most representative algorithm in multiscale geometric analysis theory [27], [28], and it is shown in Figure 2. Compared with the wavelet transform, the contourlet transform has better directional characteristic, which coincides with the main performance of the HVS system, such as the localization feature, the directional characteristic, the bandpass property, and the anisotropic property. Therefore, a contourlet can better simulate the principal characters of the HVS and can also construct a visual model with excellent characteristics. In addition, the PCNN model is consistent with the neural response characteristics of the biological visual system, pos-
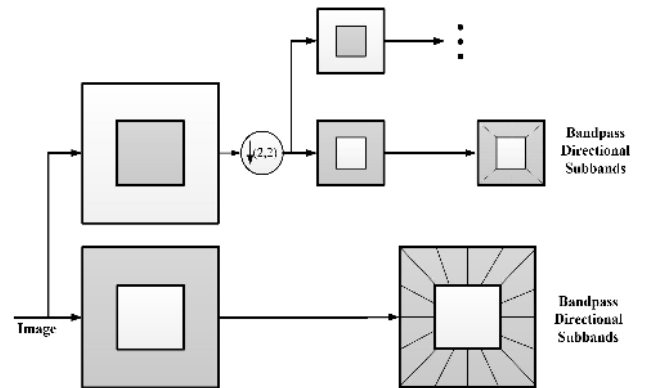

FIGURE 2. Structure chart of a contourlet transform.

sesses a certain adaptivity for data, and has strong properties of pattern classification.

On this basis, we used the HMM with its strong spatiotemporal modeling ability. The hybrid HMM–PCNN model should then better reveal both the superficial geometric features and the deep statistical properties of the visual perception mechanism. Finally, to find a better sparse image representation method, we propose an approach that takes advantage of the contourlet-HMM–PCNN model to simulate the characteristics of the biological visual perception mechanism in the receptive field of the simple cells. This model also builds the essential structures and basic properties of the grating cells and subsequently achieves the basic functions of the visual cortex in the human brain.

The core idea of the contourlet-HMM–PCNN model proposed in this paper is as follows: we first do a contourlet transform on an image to obtain the subband coefficients in various levels and directions. We then use the HMM to describe the statistical properties and correlations of those coefficients and simultaneously train this model using the expectation–maximization (EM) algorithm for solving the parameters of the HMM [29]. Moreover, we enable the state probability of an HMM input into the PCNN models and obtain the final results of the PCNN using nonlinear mapping.

III. AN ADAPTIVE CONTOURLET-HMM–PCNN MODEL

A. AN ADAPTIVE PCNN MODEL

Research indicates that the linking strength β affects the increment of neuron activation, and a large β means neurons tend to fire more easily. In this paper, we propose a contrast operator dev(x, y), which is a measurement of the gray difference between neighborhoods of a target pixel. As the gray difference of a local area in an image increases, the dev(x, y) also increases. Therefore, here we can set an adaptive β through the contrast operator dev(x, y) in the surrounding area of each pixel; accordingly, we can determine the whole neuron activation state. If the contrast of the local area of a pixel is high, we can affirm that the corresponding neuron is in an active state and will easily be fired. Thus, we can normalize the contrast of the local area of the target pixel to obtain an adaptive β as an actual parameter of a PCNN model

as follows:

$$dev(x, y) = \frac{\max f(x, y) - \min f(x, y)}{\max f(x, y)}, \quad \beta = dev(x, y) \quad (6)$$

where $\max f(x, y)$ and $\min f(x, y)$ are the maximum and minimum respectively of the contourlet coefficients in the neighborhood of a target pixel.

In addition, we input the state probability matrix from an HMM training in a PCNN model instead of contourlet coefficients. Then we adjust the PCNN parameters of adaptive V_E and α_E based on the gradient of contourlet coefficients and adaptive β based on the contrast of contourlet coefficients. Among those, V_E controls the increment of thresholds E_{ij}, which increases more when V_E is greater, and α_E controls the decay speed of thresholds E_{ij}, whose decay becomes slower when α_E is greater, and vice versa. The purpose of setting threshold E_{ij} is to amplify the activation difference of corresponding neurons in different areas of an image. Therefore, we accomplished the goals of both reducing the firing thresholds E_{ij} and enhancing the decay speed of the thresholds E_{ij} for the corresponding neuron in the same area of an image. Because a gradient is a measurement of the differences in local areas in an image, we can use the Sobel operator to calculate the gradient of contourlet coefficients; it can be expressed as

$$G_x = [-1\ 0\ 1; -2\ 0\ 2; -1\ 0\ 1] * A,$$
$$G_y = [1\ 2\ 1; 0\ 0\ 0; -1\ -2\ -1] * A,$$
$$G = \sqrt{G_x^2 + G_y^2} \quad (7)$$

where A is a single subband generated by the contourlet decomposition, and G is the gradient of a local image. If the gradient value of the coefficient of the arbitrary position in the subband is smaller, then the difference between neighboring coefficients of the corresponding position is also smaller; this also applies to the pixel domain. Moreover, if the local gradient value of the neuron corresponding to the pixel is smaller, then the firing threshold of this neuron is also smaller. Therefore, in the PCNN model a pixel has a larger gradient where α_E is greater and V_E is smaller, and the following equations are given about V_E and α_E:

$$V_E = 1 - G, \quad \alpha_E = G. \quad (8)$$

Based on the above methods, we amplified the difference between the thresholds of the neurons corresponding to two kinds of pixels in the edge area and the smooth area, simultaneously also obtaining better and more stable results in the image denoising application. In this paper, we set other parameters as follows:

$$\alpha_L = \alpha_F = 0.8, \quad V_L = V_F = 0.5, \ N = 10 \quad (9)$$
$$W = M = [1\ 1\ 1; 1\ 0\ 1; 1\ 1\ 1], \quad (10)$$

where N is the number of iterations, W is the weight matrix of synaptic connections in linking channels, and M is the weight matrix of the synaptic connections in the feeding channel.
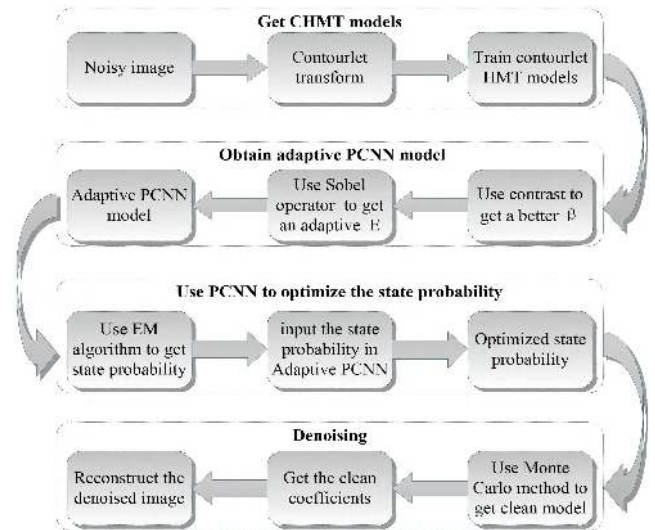

FIGURE 3. Flowchart of adaptive HMM–PCNN model in the contourlet domain for image denoising.

B. FLOWCHART OF THE PROPOSED ALGORITHM FOR IMAGE DENOISING

To show the innovation in this paper, we put the state probability matrices generated through the contourlet-HMT model into a PCNN, and then adopted adaptive parameters V_E, α_E, and β in the PCNN model, thereby showing a better performance in the image denoising application. Figure 3 is a flowchart of image denoising using the HMM–PCNN model in the contourlet domain. Note that CHMT represents an HMM model in the contourlet domain.

The algorithm procedure of an adaptive contourlet-HMM–PCNN model for image denoising is as follows:

Step 1: We used a contourlet to decompose the noisy images and obtain contourlet coefficients. The decomposition levels were [2 2 3 3] in the contourlet domain, and we used a 9/7 filter as a pyramid filter and a Phoong–Kim–Vaidyanathan–Ansari (PKVA) filter as a directional filter to implement the multiscale and directional decomposition for a noisy image.

Step 2: We adopted the HMM to construct a statistical model of contourlet coefficients in various levels and directions. Then we applied the EM algorithm during the training of the HMM model, eventually obtaining the state probability as an output. The state probability can be described as $p(S_{j,k,n} = m | v_{j,k,n}, \theta_z)$, where m, j, k, n, v, θ are state, scale, direction, the position of the coefficient, the noisy image, and the parameter set of the HMM model respectively.

Step 3: We used the contourlet coefficients to adjust the parameters of the PCNN model for obtaining an adaptive PCNN model. We used a contrast to ensure the adaptive β of the linking strength and the Sobel operator to determine both the adaptive V_E of the threshold amplitude and the adaptive α_E of the time attenuation constant. We then input the state probability p from step 2 in the adaptive PCNN model to obtain an optimized one: $p'(S_{j,k,n} = m | v_{j,k,n}, \theta_z)$.

Step 4: We estimated the variance of additive noise using the Monte Carlo method, and then we obtained the variance of the clean image by subtracting the variances of additive noise from the variances of the noisy image:

$$(\sigma_{(j,k,n),m}^{(z)})^2 = (\sigma_{(j,k,n),m}^{(v)})^2 - (\sigma_{(j,k,n)}^{(e)})^2 \qquad (11)$$

where $(\sigma_{(j,k,n),m}^{(z)})^2$, $(\sigma_{(j,k,n),m}^{(v)})^2$ are the variances of the clean image and the noisy image respectively, and $(\sigma_{(j,k,n)}^{(e)})^2$ are the variances of the additive noise. Note that the clean image does not represent the original image; it is true only under the condition of the zero-mean Gaussian noise.

Step 5: We used both the HMM model of the clean image and the optimized one p′ above (as a state probability). Then we easily transformed the image denoising problem into a Bayesian posterior probability estimation problem, where the coefficients of the clean image can be expressed as $E[z_{j,k,n}|v_{j,k,n}, \theta_z]$:

$$E[z_{j,k,n}|v_{j,k,n}, \theta_z] = \sum_m p'(S_{j,k,n} = m|v_{j,k,n}, \theta_z)$$
$$\times \frac{(\sigma_{(j,k,n),m}^{(z)})^2}{(\sigma_{(j,k,n),m}^{(z)})^2 + (\sigma_{j,k,n}^{(e)})^2} v_{j,k,n} \qquad (12)$$

Finally, we could reconstruct a denoised image based on the coefficients of a clean image.

IV. EXPERIMENTS AND ANALYSIS

A. GRAYSCALE IMAGES

To verify the efficiency of an adaptive contourlet-HMM-PCNN model in the field of image processing and computer vision, we did many simulation experiments. The experiments were carried out in MATLAB R2015b on a PC with Intel i7-7700/16G, the size of all gray-scale images was 512 × 512 pixels, the depth was 8 bit, and the format was PNG. To obtain different noisy images, we added zero-mean white Gaussian noise with variances of 30, 40, and 50 to the original images. To evaluate the performance of image denoising using the model proposed in this paper, we compared the results by their peak signal-to-noise ratio (PSNR) and their structural similarity (SSIM) objectively, and by visual quality subjectively.

This paper used 12 standard test images shown in Figure 4. We used an adaptive contourlet-HMM–PCNN model to achieve image denoising. It achieved better image denoising performance through many simulation experiments as below. Here, we show the experiment results of the test images in Table 1 and Table 2. For comparison, we used various noise levels and denoising methods for the same test image, including the wavelet thresholding method with a threshold of $T = 3\sigma$ and the contourlet-HMT model denoising method of Po and Do. We used PSNR and SSIM as evaluation criteria to demonstrate their denoising performance on the test images.

From the experiment results shown in Table 1, Table 2, and Figure 5, we can conclude that an adaptive contourlet-HMM–PCNN model denoised better than the
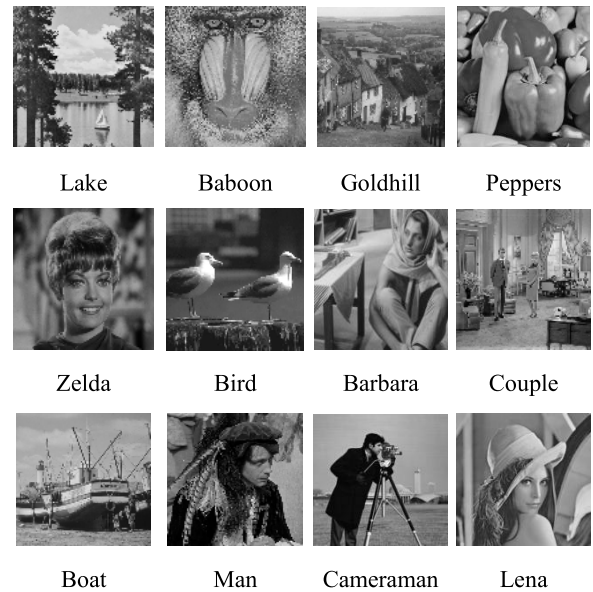

FIGURE 4. Test images.

contourlet-HMT model and the wavelet threshold method in both the PSNR and SSIM estimation criteria, as follows:

- In different noising levels for all the test images, our approach was 0.08 to 3.59 dB higher than that of the wavelet threshold method.
- In different noising levels for the test images of the Lena, Zelda, Cameraman, Baboon, and Couple, our approach was 0.00 to 1.34 dB higher than the contourlet-HMT model of Po and Do, respectively.
- In 50 noising levels for the test image of the Man, in 40 and 50 noising levels for the Boat, in 50 noising levels for the Bird, in 40 and 50 noising levels for the Peppers, in 50 noising levels for the Goldhill, in 50 noising levels for the Lake, and in 40 and 50 noising levels for the Barbara, our approach was 0.06 to 0.34 dB lower than those of the contourlet-HMT model of Po and Do, respectively.
- From an SSIM estimation, at different noising levels, for the test images of the Lena, Zelda, Bird, Peppers, Goldhill, Lake, and Couple, our approach was superior to that of the contourlet-HMT model by 0.0018 to 0.0694.
- From an SSIM estimation, in 50 noising levels for the test image Man, 30 for the Cameraman, in 40 and 50 for the Boat, in 40 and 50 for the Baboon, and in 30, 40, and 50 for the Barbara, our approach was lower than that of the contourlet-HMT model by 0.0011 to 0.0400.
- From the above observations, we can conclude that from a visual estimation of the test images, the proposed method is superior to the contourlet-HMT and wavelet threshold models. However, the proposed method processes images more slowly because of the increased complexity of its algorithm.
- It is well known that the success of sparse representation and image denoising applications is closely dependent on the content and background of the original images. Therefore, for some test images with many textures in the content and in background, there may be a slight performance degradation.

TABLE 1. Comparison of PSNR in denoised images using various algorithms.

Clean image	Noise variance σ	Noisy image (dB)	Wavelet–Thresh. (dB) (Improvement)	Contourlet–HMT (dB) (Improvement)	Adaptive Contourlet-HMM–PCNN (dB)
Lena	30	18.71	26.67 (+1.84)	28.1 (+0.33)	**28.51**
	40	16.34	25.26 (+2.02)	27.0 (+0.28)	**27.28**
	50	14.59	24.20 (+2.07)	26.0 (+0.23)	**26.27**
Zelda	30	18.83	28.24 (+1.97)	30.0 (+0.21)	**30.21**
	40	16.49	27.03 (+1.70)	28.2 (+0.44)	**28.73**
	50	14.73	26.05 (+1.48)	27.0 (+0.46)	**27.53**
Man	30	19.32	24.37 (+0.99)	24.6 (+0.71)	**25.36**
	40	16.99	22.93 (+0.82)	23.7 (+0.03)	**23.75**
	50	15.21	22.27 (+0.08)	**22.69** (−034)	22.35
Cameraman	30	19.04	26.78 (+1.37)	27.4 (+0.67)	**28.15**
	40	16.68	25.43 (+0.98)	26.25 (+0.16)	**26.41**
	50	14.88	23.94 (+1.04)	24.97 (+0.01)	**24.98**
Boat	30	18.78	25.28 (+1.80)	26.68 (+0.40)	**27.08**
	40	16.45	24.12 (+1.70)	**26.00** (−0.18)	25.82
	50	14.74	22.99 (+1.86)	**24.91** (−0.06)	24.85
Bird	30	19.02	27.37 (+1.58)	28.36 (+0.59)	**28.95**
	40	16.82	25.93 (+1.62)	27.42 (+0.13)	**27.55**
	50	15.15	24.92 (+1.31)	**26.32** (−0.09)	26.23
Peppers	30	18.82	26.34 (+1.76)	27.97 (+0.13)	**28.10**
	40	16.48	25.05 (+1.62)	**26.82** (−0.15)	26.67
	50	14.76	23.84 (+1.48)	**25.67** (−0.35)	25.32
Goldhill	30	18.80	25.22 (+1.97)	26.88 (+0.31)	**27.19**
	40	16.45	24.18 (+1.87)	26.03 (+0.02)	**26.05**
	50	14.72	23.19 (+1.95)	**25.21** (−0.07)	25.14
Baboon	30	18.68	21.21 (+3.29)	23.16 (+1.34)	**24.50**
	40	16.31	20.34 (+2.01)	22.35 (+0.00)	**22.35**
	50	15.53	20.00 (+3.59)	21.53 (+2.06)	**23.59**
Lake	30	18.90	24.41 (+1.77)	25.76 (+0.42)	**26.18**
	40	16.66	23.01 (+1.88)	24.76 (+0.13)	**24.89**
	50	15.01	22.09 (+1.75)	**24.09** (−0.25)	23.84
Couple	30	18.74	24.50 (+1.73)	25.67 (+0.56)	**26.23**
	40	16.37	23.25 (+1.82)	24.87 (+0.20)	**25.07**
	50	14.59	22.16 (+1.91)	23.97 (+0.10)	**24.07**
Barbara	30	18.80	23.50 (+2.44)	25.90 (+0.04)	**25.94**
	40	16.48	22.27 (+2.31)	**24.83** (−0.25)	24.58
	50	14.77	21.47 (+2.02)	**23.61** (−0.12)	23.49

B. COLOR IMAGES

We also denoised color images as an extension of the proposed approach in terms of theory and application. The color image denoising procedure was basically the same as that for gray images. The main difference is that we used the contourlet-HMM–PCNN model to decompose the color images with zero-mean white Gaussian noise into RGB three channel images. We then used the same method as that for gray image denoising to denoise the three types of noisy images in the RGB channel. Finally, we fused the denoised images in the RGB three channels into one color image (RGB).

To denoise the color images, we used three color test images—Peppers, Baboon, and Lena. Each was 512 × 512 pixels, 24-bit depth, TIFF format with zero-mean white Gaussian noise variances of 30, 40, and 50 respectively. The color testing images are shown in Figure 6.

To obtain various noisy images, we added zero-mean white Gaussian noise with variances of 30, 40, and 50 to the color testing images. To evaluate the performance of image

TABLE 2. Structural similarity index of denoised images using various algorithms.
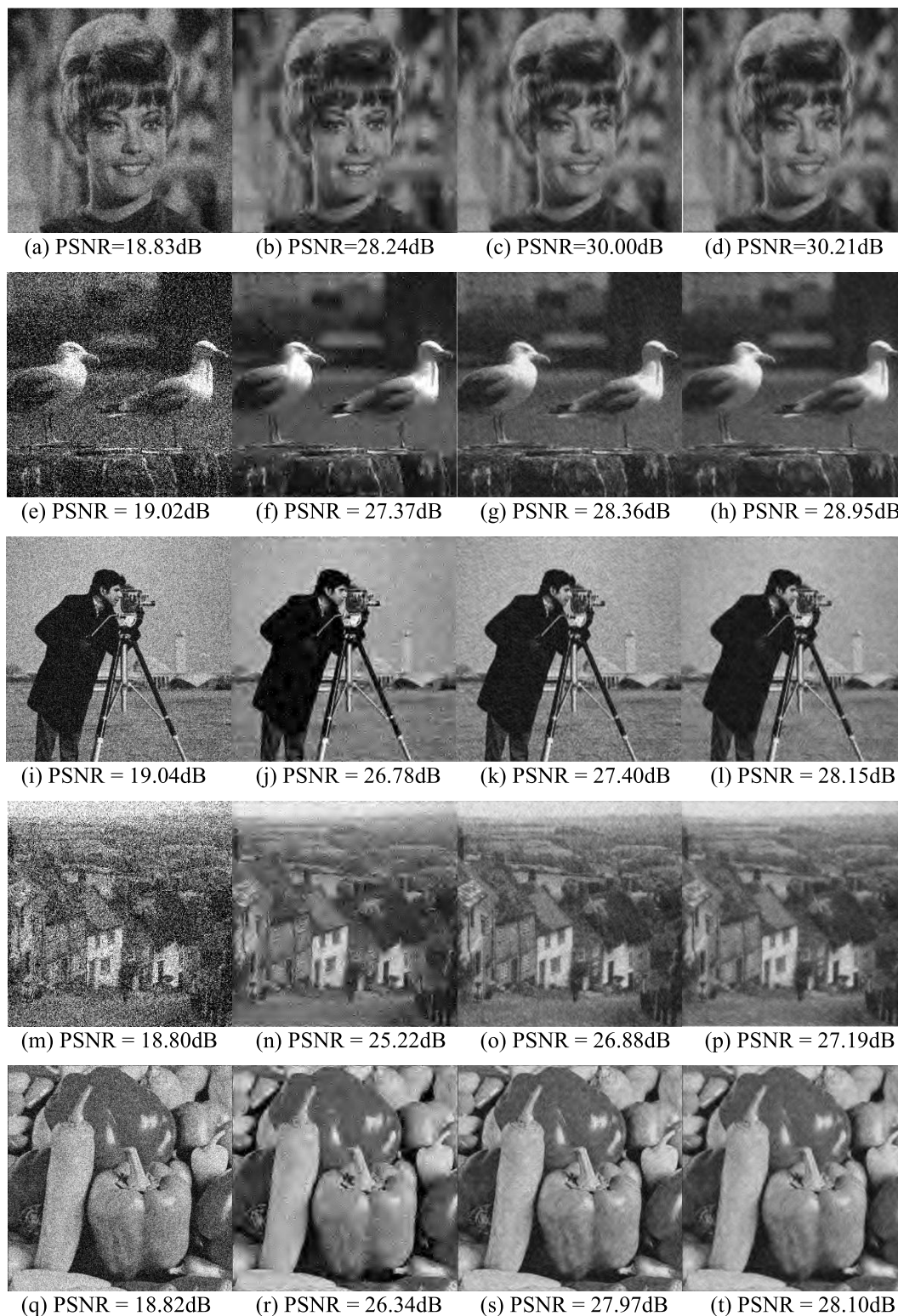
Clean image	Noise variance	Noised image	Contourlet-HMT	Adaptive contourlet-HMM–PCNN
Lena	30	0.2200	0.6891 (+0.0446)	**0.7337**
	40	0.1548	0.6329 (+0.0665)	**0.6994**
	50	0.1162	0.6159 (+0.0707)	**0.6866**
Zelda	30	0.1694	0.7305 (+0.0499)	**0.7804**
	40	0.1109	0.6963 (+0.0561)	**0.7524**
	50	0.0790	0.6559 (+0.0694)	**0.7253**
Man	30	0.3241	0.5903 (+0.0199)	**0.6102**
	40	0.2381	0.5364 (+0.0027)	**0.5391**
	50	0.1824	**0.4849** (−0.0011)	0.4838
Cameraman	30	0.2239	**0.7429** (−0.0028)	0.7401
	40	0.1645	0.6791 (+0.0108)	**0.6899**
	50	0.1273	0.6149 (+0.0464)	**0.6613**
Boat	30	0.2731	**0.7222** (−0.0239)	0.6983
	40	0.2004	**0.6814** (−0.0283)	0.6531
	50	0.1552	0.6089 (+0.00266)	**0.6355**
Bird	30	0.1853	0.7652 (+0.0018)	**0.7670**
	40	0.1303	0.7361 (+0.0309)	**0.7425**
	50	0.0987	0.6984 (+0.0261)	**0.7245**
Peppers	30	0.2267	0.6997 (+0.0185)	**0.7182**
	40	0.1613	0.6394 (+0.0458)	**0.6852**
	50	0.1230	0.5822 (+0.0389)	**0.6211**
Goldhill	30	0.2634	0.6483 (+0.013)	**0.6613**
	40	0.1829	0.6060 (+0.0083)	**0.6143**
	50	0.1356	0.5682 (+0.0143)	**0.5825**
Baboon	30	0.4708	0.6250 (+0.022)	**0.6470**
	40	0.3651	**0.5721** (−0.0127)	0.5594
	50	0.2887	**0.5208** (−0.0345)	0.4863
Lake	30	0.3236	0.6812 (+0.0018)	**0.6830**
	40	0.2454	0.6238 (+0.0229)	**0.6467**
	50	0.1940	0.5924 (+0.0261)	**0.6185**
Couple	30	0.3091	0.6435 (+0.0162)	**0.6597**
	40	0.2261	0.5933 (+0.0149)	**0.6082**
	50	0.1705	0.5395 (+0.0278)	**0.5673**
Barbara	30	0.3442	**0.7372** (−0.0400)	0.6972
	40	0.2607	**0.6740** (−0.0219)	0.6521
	50	0.2050	**0.6106** (−0.0112)	0.5994

denoising using the model proposed in this paper, we compared the results by their PSNR and their SSIM objectively, and by visual quality subjectively. The results are shown in Tables 3 and 4.

Moreover, to show the effects of color image denoising for the zero-mean white Gaussian noise with a variance of 50, we compared visual quality subjectively, as shown in Figure 7.

We found that the effects of the color image denoising were not as good as those of grayscale images in general. This is because of the correlation between the three channels of RGB color space. The higher the correlation, the greater the chromatic aberration in the color images after image denoising. By contrast, the grayscale images had only one channel, so the image denoising performances were not affected by correlation.

FIGURE 5. Denoising results of some test images. The noise variance is 30 and the images, from the left column to the right column, are noisy images (a, e, i, m, q), wavelet thresholding denoised images (b, f, j, n, r), contourlet-HMT denoised images (c, g, k, o, s), and contourlet-HMT–PCNN denoised images (d, h, l, p, t).

C. RUNNING TIME

Most of the computational expense of the proposed approach comes from the training of the HMT model and the calculations of the PCNN model. All the code is implemented in MATLAB with a 3.40-GHz CPU and 16 GB of RAM. The average CPU time of the proposed approach was slightly less

(a) Peppers   (b) Lena   (c) Baboon

FIGURE 6. Three color test images.

TABLE 3. Comparison of PSNR in noised and denoised color images using the proposed algorithm.

Clean image	Noise variance σ	Noised image(dB)	Denoised image(dB)
Peppers	30	20.1644	26.3020
	40	19.9761	25.0197
	50	19.6030	23.9597
Lena	30	20.1463	27.4134
	40	20.0014	26.4662
	50	19.6803	25.2909
Baboon	30	20.0963	22.8431
	40	19.9451	22.4337
	50	19.6239	22.0206

TABLE 4. Structural similarity index of noised and denoised color images using the proposed algorithm.

Clean image	Noise variance σ	Noised image	Denoised image
Peppers	30	0.8161	0.9455
	40	0.8145	0.9389
	50	0.8100	0.9329
Lena	30	0.8038	0.9595
	40	0.8032	0.9573
	50	0.8021	0.9577
Baboon	30	0.7576	0.8180
	40	0.7571	0.8142
	50	0.7556	0.8210

(a) PSNR=23.9597 dB   (b) PSNR=25.2909 dB   (c) PSNR=22.0206 dB

FIGURE 7. Visual quality comparison of denoised images.

than that of the contourlet-HMT model and less than some state-of-the-art methods. For example, for the House testing image, whose size was 256 × 256, the proposed approach took 21 s. By comparison, the contourlet-HMT model took 17 seconds, and the method proposed by [16] took 48 seconds of each outer iteration.

We believe that the proposed approach should be run on a more powerful GPU to further reduce its running time.

V. CONCLUSION

In this paper, the HMM–ANN model, which has been applied successfully in speech processing, is introduced to the field of image processing, computer vision, and pattern recognition. The combined model takes advantage of the pattern classification ability of ANN and the modeling ability of HMM in spatiotemporal use.

In our research, we modulated the image data structures to match the HMM–ANN model and replaced the ANN model with a PCNN to construct a hybrid HMM–PCNN model. In addition, we propose an adaptive contourlet-HMM–PCNN model based on the contourlet-HMT model presented by Po and Do. Finally, we verified the effectiveness of the contourlet-HMM–PCNN model by using an image denoising application.

The performance was better than that of the contourlet-HMT model and the wavelet threshold method for test images. The proposed model had more low-frequency components and was similar to the contourlet-HMT model for test images with more high-frequency components. In future research, we will further improve the theory of multiscale geometric analysis. We will then use human visual attention and saliency features to create a more effective sparse representation theory for texture images.

REFERENCES

[1] T. J. McDermott, A. I. Wiesman, M. S. Mills, R. K. Spooner, N. M. Coolidge, A. L. Proskovec, E. Heinrichs-Graham, and T. W. Wilson, "tDCS modulates behavioral performance and the neural oscillatory dynamics serving visual selective attention," *Hum. Brain Mapping*, vol. 40, no. 3, pp. 729–740, Feb. 2019.

[2] B.-H. Liu, P. Y. Li, Y. J. Sun, Y.-T. Li, L. I. Zhang, and H. W. Tao, "Intervening inhibition underlies simple-cell receptive field structure in visual cortex," *Nature Neurosci.*, vol. 13, pp. 89–96, Nov. 2010.

[3] H. Barlow and A. Gardner-Medwin, "Localist representation can improve efficiency for detection and counting," *Behav. Brain Sci.*, vol. 23, no. 4, pp. 467–468, Aug. 2000.

[4] O. Ben-Shahar and G. Ben-Yosef, "Tangent bundle elastica and computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 1, pp. 161–174, Jan. 2015.

[5] K. D. Harris and T. D. Mrsic-Flogel, "Cortical connectivity and sensory coding," *Nature*, vol. 503, pp. 51–58, Nov. 2013.

[6] Y. Yan, M. J. Rasch, M. Chen, X. Xiang, M. Huang, S. Wu, and W. Li, "Perceptual training continuously refines neuronal population codes in primary visual cortex," *Nature Neurosci.*, vol. 17, no. 10, pp. 1380–1387, Sep. 2014.

[7] Y. Sun, Q. Liu, J. Tang, and D. Tao, "Learning discriminative dictionary for group sparse representation," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3816–3828, Sep. 2014.

[8] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, Mar. 2001.

[9] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, Jun. 1996.

[10] Y. Li, C. Chen, F. Yang, and J. Huang, "Hierarchical sparse representation for robust image registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2151–2164, Sep. 2018.

[11] L. D. Griffin, "The atlas structure of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 234–245, Jan. 2019.

[12] S. J. Fan, T.-T. Ng, B. L. Koenig, J. S. Herberg, M. Jiang, Z. Shen, and Q. Zhao, "Image visual realism: From human perception to machine computation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2180–2193, Sep. 2018.

[13] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, Jun. 2017.

[14] B. Wen, S. Ravishankar, and Y. Bresler, "VIDOSAT: High-dimensional sparsifying transform learning for online video denoising," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1691–1704, Apr. 2019.

[15] R. G. Gavaskar and K. N. Chaudhury, "Fast adaptive bilateral filtering," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 779–790, Feb. 2019.

[16] C. L. P. Chen, L. Liu, L. Chen, Y. Y. Tang, and Y. Zhou, "Weighted couple sparse representation with classified regularization for impulse noise removal," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4014–4026, Nov. 2015.

[17] G. Yuan and B. Ghanem, "l_0TV: A sparse optimization method for impulse noise image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 352–364, Feb. 2019.

[18] J.-L. Yin, B.-H. Chen, and Y. Li, "Highly accurate image reconstruction for multimodal noise suppression using semisupervised learning on big data," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3045–3056, Nov. 2018.

[19] Q. Wang, X. Zhang, Y. Wu, L. Tang, and Z. Zha, "Nonconvex weighted l_p minimization based group sparse representation framework for image denoising," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1686–1690, Nov. 2017.

[20] Z. Zha, X. Zhang, Q. Wang, Y. Bai, Y. Chen, L. Tang, and X. Liu, "Group sparsity residual constraint for image denoising with external nonlocal self-similarity prior," *Neurocomputing*, vol. 275, pp. 2294–2306, Jan. 2018.

[21] D. D.-Y. Po and M. N. Do, "Directional multiscale modeling of images using the contourlet transform," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1610–1620, Jun. 2006.

[22] R. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex," *Neural Comput.*, vol. 2, no. 3, pp. 293–307, 1990.

[23] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, Dec. 1997.

[24] Y. Yang, M. Yang, S. Huang, M. Ding, and J. Sun, "Robust sparse representation combined with adaptive PCNN for multifocus image fusion," *IEEE Access*, vol. 6, pp. 20138–20151, 2018.

[25] B. Cheng, L. Jin, and G. Li, "Infrared and visual image fusion using LNSST and an adaptive dual-channel PCNN with triple-linking strength," *Neurocomputing*, vol. 310, pp. 135–147, Oct. 2018.

[26] J. Poort, A. G. Khan, M. Pachitariu, A. Nemri, I. Orsolic, J. Krupic, M. Bauza, M. Sahani, G. B. Keller, T. D. Mrsic-Flogel, and S. B. Hofer, "Learning enhances sensory and multiple non-sensory representations in primary visual cortex," *Neuron*, vol. 86, no.6, pp. 1478–1490, Jun. 2015.

[27] H. Zhao, X. Zhao, T. Zhang, and Y. Liu, "A new contourlet transform with adaptive directional partitioning," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 843–847, Jun. 2017.

[28] H. Sadreazami and M. A. Amini, "A robust image watermarking scheme using local statistical distribution in the contourlet domain," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 1, pp. 151–155, Jan. 2019.

[29] D. Wu and J. Ma, "A two-layer mixture model of Gaussian process functional regressions and its MCMC EM algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4894–4904, Oct. 2018.

GUOAN YANG received the B.S. degree in industrial automation and control engineering from Jilin University, Changchun, China, in 1986, the M.S. degree in system and automatic control engineering from Tokyo Metropolitan University, Tokyo, Japan, in 1993, and the Ph.D. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2006, respectively. He started his professional career at Hertz Corporation, Tokyo, Japan, as a Research Fellow, and was involved in research on electronics and mobile communication, from 1993 to 2001. He is currently an Associate Professor with the School of Electronic and Information Engineering, Xi'an Jiaotong University, China. Since 2001, he has been with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, where he became affiliated with the image processing, computer vision, and pattern recognition. He has published nearly 40 papers, four patents, and two books. His research interests include image compression, vision computing, wavelet analysis, multiscale geometric analysis, compressed sensing, neural networks, and deep learning. Moreover, he is a Reviewer of several famous journals, such as the IEEE Transactions on Image Processing, the IEEE Transactions on Signal Processing, the IEEE Transactions on Multimedia, and so on.

ZHENGZHI LU received the B.S. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2017. From 2017 to 2019, he was pursuing the M.S. degree in control science and engineering, Xi'an Jiaotong University, where he is currently pursuing the Ph.D. degree. His research interests include image processing, computer vision, wavelet, and multiscale geometric analysis.

JUNJIE YANG received the B.S. degree in control science and engineering from Jilin University, Changchun, China, in 2018. He is currently pursuing the M.S. degree in control science and engineering with Xi'an Jiaotong University. His research interests include image compression, vision computing, multiscale geometric analysis, neural networks, and deep learning.

YUHAO WANG received the B.S. degree in control science and engineering from Southwest Jiaotong University, Chengdu, China, in 2013. From 2013 to 2016, he was a full-time Engineer on autonomous driving research with Xi'an Railway Company. He is currently pursuing the M.S. degree in control science and engineering with Xi'an Jiaotong University. His research interests include image processing, computer vision, road detection, and autonomous driving.

• • •