Received December 11, 2018, accepted January 4, 2019, date of publication January 9, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2891914

CDF Space Covariance Matrix of Gabor Wavelet With Convolutional Neural Network for Texture Recognition

CHAORONG LI[1,2], (Member, IEEE), XIAOYUN GUAN[2], PEIHUI YANG[3], YUANYUAN HUANG[4], WEI HUANG[1], AND HUAFU CHEN[1]

[1]University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Department of Computer Science and Information Engineering, Yibin University, Yibin 644000, China
[3]Department of Computer Science and Technology, Sichuan Vocational College of Post and Telecom, Chengdu 610067, China
[4]Department of Cyberspace Engineering, Chengdu University of Information Technology, Chengdu 610225, China

Corresponding author: Chaorong Li (lichaorong88@163.com)

This work was supported in part by the China Postdoctoral Science Foundation under Grant 2016M602675, in part by the Foundation of the Central Universities in China under Grant ZYGX2016J123, in part by the Project of Education Department of Sichuan Province under Grant 16ZA0328, in part by the Project of Sichuan Science and Technology Program under Grant 2018JY0117, in part by the Key Project of Yibin Science and Technology under Grant 2017ZSF009-9, and in part by the Science and Technology Project of Yibin of China, in 2016.

ABSTRACT A novel method for texture image recognition is proposed in this paper. The aim of the proposed method is to represent texture by combining the Gabor wavelet transform and deep learning which are efficient techniques for image analysis. We developed the cumulative distribution function (cdf) space covariance model of Gabor wavelet (CSCM-GW), which can jointly model multivariate data in cdf space, in the Gabor wavelet domain to represent texture. The images having different sizes will be transformed by CSCM-GW into same size covariance matrices. Because CSCM-GW is based on the covariance matrix which belongs to Riemannian space, it has the high computational cost in the recognition phase. Therefore, we proposed a novel method of texture recognition called CSCM-GWF-CNN which uses CNN to project the fused covariance of CSCM-GW into low-dimensional vector space for reducing the computational cost and improving the recognition performance. The experiments on Brodatz (111) and KTH-TIPS2-b texture databases show that the proposed method is efficient for texture representation and outperforms most of the state-of-the-art recognition methods.

INDEX TERMS Texture recognition, covariance matrix, Gabor wavelet, multidimensional statistical model, convolutional neural network.

I. INTRODUCTION

Texture recognition is a fundamental and promising issue in image processing field, and it involves a lot of research aspects such as content-based information retrieval [1] system, computer vision [2], medical assistant diagnosis [3] and so on. In general, feature extraction is the important step for image classification. Existing methods of image feature extraction include the following classes: local-pattern-based method [4]–[6], Sparse-dictionary-learning-based method [7], [8], deep-learning-based method [9] and the classic wavelet-based method [10]–[12].

The basic idea of the local-pattern-based method is the image is considered being composed of local-patterns which can be encoded with a decimal number. The image features are calculated by counting the number of the different local-patterns in the image. The typical local-pattern methods are Local Binary Pattern (LBP) [4] and its extensions such as Local Ternary Patterns (LTP) [5] and the Local Tetra Patterns (LTrP) [6]. Local-pattern-based method has the quite high computational efficiency for the feature extraction because the generation of local-patterns (features) just needs simple comparing operation between pixels. However, the critical disadvantage of local-pattern-based method is that it is sensitive to image noise due to the local-patterns may be totally different if the intensity of one pixel in a local neighborhood is altered by noise.

The aim of sparse-dictionary-learning-based method is to find a sparse representation of the image by using the atoms, which construct the dictionary, and the linear combination of the atoms [13]. In order to obtain the dictionary and the optimal linear combination, the learning process (called dictionary-learning) is required. The learning algorithm consists of two phases: Dictionary generate and sparse coding with a pre-computed dictionary. K-SVD [14] and SGK [15] are the two state-of-the-arts dictionary-learning algorithms.

Deep-learning-based method has attracted considerable attention in recent years [16]. Various deep-learning-based methods such as sparse autoencoder [17], Convolutional Neural Network [9], [18]–[20] have been designed for image representation or object recognition.

To wavelet-based method, there are two types of features can be extracted from the wavelets coefficients (subbands): one is the wavelet-signature such as the norm-1 and norm-2 energies and standard deviations calculated from the coefficients of each wavelet subband [21]–[23]. The other one is the parameters of probability distribution model which is more efficient than wavelet-signature; the commonly-used probability distribution model includes Generalized Gaussian Model (GGM) [10] and Gaussian Mixture Model (GMM) [24]). Probability distribution model captures the distribution of wavelet coefficients by estimating the model parameter. In fact, probability distribution is the preferred model for modelling the wavelet coefficients and it has a wide range of applications in the fields of image analysis and pattern recognition [10], [25]. Early feature extraction methods in wavelet domain focus on establishing a univariate statistical model for each wavelet subband independently. Researches [26] and [27] have demonstrated that dependence exists in the wavelet transform domain and using the multidimensional distribution in wavelet transform domain can effectively increase discriminative capacity of the wavelet features. Therefore, recent methods employ multivariate distributions such as Multivariate Generalized Gaussian Model (MGGM), Multivariate Laplace Distribution (MLD) [28] and the copula model [29] to join the subbands of the orthogonal wavelet transforms (e.g., the discrete wavelet transform [30] and stationary wavelet transform [31]).

The above-mentioned methods are effective for texture recognition, so our goal is to incorporate two or more than two methods to improve the recognition performance. In this paper, we use both the Gabor wavelet and deep learning to improve the performance of texture recognition methods. Our method is a global texture describing model based on the covariance matrix in Cumulative Distribution Function (CDF) space, called CSCM-GW (CDF Space Covariance Model in Gabor wavelet domain). Namely, we used covariance matrix to model the Gabor wavelet coefficients. Differing from other covariance-matrix-based methods, we first project the coefficients of Gabor wavelet into CDF space to get more robust performance. The challenge lies in that the space of CSCM-GW is not a linear space but a Riemannian manifold which is more difficult than linear space for image analysis. To the best of our knowledge, there is no ideal technique to transform the data from Riemannian manifold into the linear space. To address this issue, we use deep Convolutional Neural Network (CNN) as a transforming approach to project CSCM-GW in to the linear space. There are two contributions in this work:

- We proposed CSCM-GW for texture feature extraction. CSCM-GW can produce robust texture features, and with CSCM-GW the images with different size will be transformed into fixed dimensional covariance matrix.
- We use CNN to project Riemannian manifold into the linear space. Namely, we directly use CNN on the covariance matrix for classifying texture images.

II. RELATED WORKS

The proposed method is built on the statistical model in wavelet domains. Therefore, in this section we mainly introduce the relative work of multidimensional statistical model. Multidimensional statistical model such as Multivariate Gaussian Model (MGM) and its extensions, Multivariate Gaussian Mixture Model (MMGM), MGGM and Gaussian copula, in wavelet domain has shown their excellent ability for texture feature extraction. These models are widely applied in computer vision. In [32] MMGM is used to model over a variety of different color and texture feature spaces, with a view to the retrieval of textured color images from databases. In literature [28] the authors proposed texture retrieval algorithm based on MGGM for the modeling of wavelet subbands. In literature [33] Lasmar and Berthoumieu use Gaussian copula to fit wavelet coefficients and they derived a closed-form Kullback-Leibler Divergence (KLD) as the similarity measure between Gaussian copula. When these statistical models are applied into texture recognition, general we are to compare the dissimilarity between two models, e.g., two MMGMs. Closed-form KLD is a prefer measure between distribution models because it is efficient and is also has the lower computational cost than other measures of the multivariate model. Unfortunately there are no closed-form expressions for the closed-form KLD for most of these models, e.g., copula models excluding Gaussian copula. Generally, multivariate statistical model has a more complex calculation than linear space for texture recognition since the KLD formulas involve the multiplication and inverse of a matrix.

Covariance matrix is also relative to the multivariate model, and it has gained a promising success [34]–[39]. Tuzel *et al.* [34] map each pixel to a 5-dimensional feature space (the five features are image intensities, norm of first and second order derivatives of intensities) and use a covariance matrix to model these features. Both CSCM-GW and covariance descriptor are based on the covariance matrix. Pang *et al.* [35] proposed a method by utilizing Gabor-based region covariance matrices as face descriptors. Both pixel locations and Gabor coefficients are used to form the covariance matrices. Because covariance matrix belongs to Riemannian manifold, Euclidean distance cannot be used. Therefore, Tuzel *et al.* [34] proposed using eigenvalue-based distance Riemannian distance as the measure of the covariance matrix.

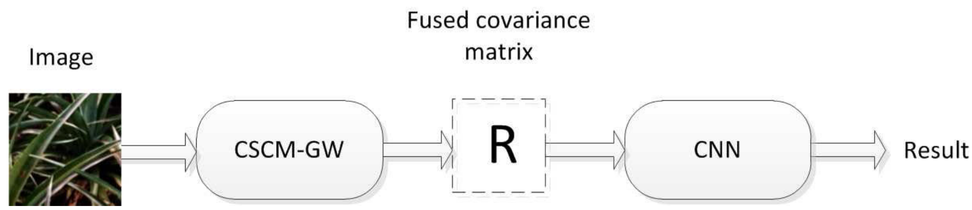

FIGURE 1. The flowchart of the proposed method (CSCM-GWF-CNN) for image classification.

Given two covariance matrices R_1 and R_2, Riemannian distance is defined as:

$$RD\left(R_1, R_2\right) = \sqrt{\sum_{i=1}^{d} ln\lambda_i^2\left(R_1, R_2\right)}, \tag{1}$$

where $\lambda_i^2\left(R_1, R_2\right)_{i=1,\cdots,d}$ are the generalized eigenvalues of R_1 and R_2.

The disadvantage of Riemannian distance lies in the computation cost using formula (1) is quite expensive because it is necessary to calculating the eigenvalues of the two covariance matrices at the image matching step. Recently, a number of researchers employ Log-Euclidean embedding approaches [40]–[43] to transform the covariance matrices into the linear space and use Euclidean distance as the similarity measure of two covariance matrices R_1, R_2. Log-Euclidean between R_1 and R_2 is defined as:

$$LD\left(R_1, R_2\right) = \|log(R_1) - log(R_2)\|_F, \tag{2}$$

where log is the matrix logarithm operator and $\|\cdot\|_F$ denotes the matrix Frobenius norm. Although Log-Euclidean has the low computation cost, much of the spatial information of the covariance matrix is lost. Besides, Minh *et al.* [38] provide a finite-dimensional approximation of the Log-Hilbert-Schmidt (Log-HS) distance, which is the extension of Log-Euclidean, between covariance operators to the large number of images classification.

Multidimensional statistical models including covariance based methods have two folds shortcomings. First, the computational cost of the measures such as Riemannian distance for covariance matrices is more expensive than the measures such as Euclidean distance in linear space. Second, it is difficult to resort to an efficient learning approach to improve the performance. Recently some researchers have proposed some methods to overcome these shortcomings. For example, Harandi *et al.* [44] proposed modeling the mapping from the high-dimensional manifold to the low-dimensional one with an orthonormal projection. Wang *et al.* [39] presented a Discriminative Covariance oriented Representation Learning (DCRL) framework which is a learning based approach to face recognition. Differing from these the above-mentioned methods, in this paper we propose a novel method by using CNN to project the covariance matrix. Our method can efficiently transform the Riemannian manifold into the linear space and meanwhile incorporates the learning ability.

III. THE PROPOSED METHOD FOR TEXTURE CLASSIFICATION

In this section, we describe the framework for the proposed method (see Fig.1). Given an image, we first decompose it by using Gabor wavelet. After the decomposition, there will be generated a number of subbands of Gabor wavelet. These subbands are organized as an observation matrix and CSCM-GW will be used to yield a fused covariance matrices R. In the final step, CNN is used to transform the fused covariance matrix into a feature vector. For simplicity, in final step we directly use the CNN as the classifier. One can just use CNN to transform the R in a vector and then use SVM as the classifier to enhance the performance.

A. CSCM-GW

Gabor wavelet is a quite useful tool of image processing [45]–[47]. It is defined as the convolution on the image with Gabor filters. If the image is decomposed by Gabor wavelet, then there will be $L \times D$ subbands, where L indicates the number of scales and D indicates the number of directions. Specifically, Gabor wavelet subbands can be obtained by convolving the Gabor kernels with the image $I(x, y)$ as follows:

$$g_{l,d}(z) = I(x, y) * \psi_{l,d}(x, y) \tag{3}$$

where $z = (x, y)$. $l = 1, \cdots, L$, and $d = 1, \cdots, D$. The elements of $g_{l,d}(z)$ are complex number, and the magnitudes $M_{l,d}(z)$ and angles $A_{l,d}(z)$ can be respectively computed by using the real part $Re_{l,d}(z)$ and the imaginary part $Im_{l,d}(z)$.

$$M_{l,d}(z) = \sqrt{Re_{l,d}^2(z) + Im_{l,d}^2(z)} \tag{4}$$

$$A_{l,d}(z) = arctan(Re_{l,d}(z)/Im_{l,d}(z)) \tag{5}$$

Since dependence exists between the Gabor subbands, we use a covariance matrix to describe these subbands. Furthermore, in order to obtain more robust performance, we first cast these subbands into their CDF space by using Weibull distribution [46] and then calculate the covariance matrix from these transformed subbands in CDF space, called CDF Space Covariance Matrix of Gabor wavelet (CSCM-GW). At the experiment section, we can observe that CSCM-GW is more efficient for image representation compared to other multivariate distribution models.

Given the 5-scale and 8-direction decomposition of Gabor wavelet, we organize the magnitude and angle subbands into the following observation matrices:

$$Z_m = [M_{00}(x, y), M_{01}(x, y), \cdots, M_{40}(x, y)] \tag{6}$$

$$Z_a = [A_{00}(x, y), A_{01}(x, y), \cdots, A_{40}(x, y)] \tag{7}$$

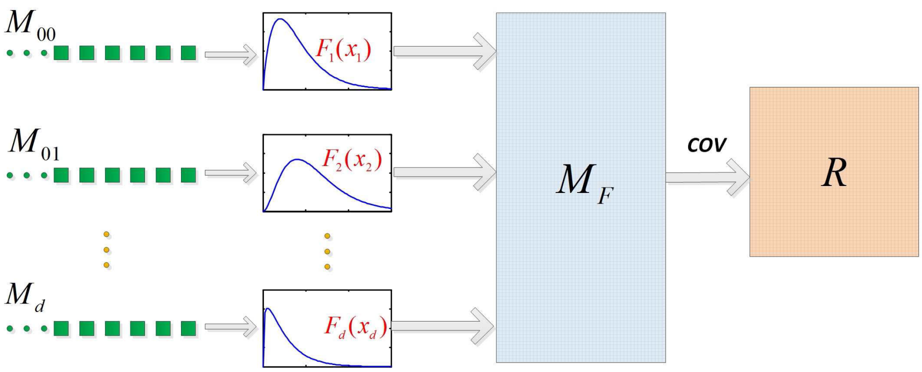

FIGURE 2. CDF Space Covariance Matrix of Gabor wavelet (CSCM-GW). Cov indicates the operation of calculating covariance matrix.

To RGB image, the Gabor wavelet is respectively applied to each of the RGB channels, and then the observation matrices are defined as

$$Z_m^c = [M_{00}^R(x, y), M_{00}^G(x, y), \cdots, M_{40}^B(x, y)] \tag{8}$$

$$Z_a^c = [A_{00}^R(x, y), A_{00}^G(x, y), \cdots, A_{40}^B(x, y)] \tag{9}$$

It can be see that the size of Z_m and Z_a is $N \times 40$, and the size of Z_m^c and Z_a^c is $N \times 120$, where N is the number of pixels in Gabor wavelet subband.

Algorithm 1 CSCM-GW Algorithm

Require: Image $I(x, y)$
Ensure: Fused covariance matrix R

1: Initialize M_F^m
2: Initialize M_F^a
3: Decomposing image $I(x, y)$ using Gabor wavelet according to (3)
4: Organizing observation matrix Z_m and Z_a according to (6)-(7)
5: **for** each column m_i of Z_m **do**
6: Calculating the parameter $\tilde{\alpha}_i$ of Weibull distribution from m_i using ML
7: Calculate CDF $F_i(x|\tilde{\alpha}_i)$ according $\tilde{\alpha}_i$ and c_i
8: Concatenate $F_i(x|\tilde{\alpha}_i)$ into the ith column of M_F^m
9: **end for**
10: Calculating R_m from M_F^m using COV operation using (14)
11: Using the same steps (from step 5-step 9) with normal distribution to calculate M_F^a
12: Calculating R_a from M_F^a using COV operation using (14)
13: Calculating the fused covariance matrix R by using (16) based on R_m and R_a.

The schematic diagram of CSCM-GW is shown in Fig.2, and the procedure of CSCM-GW is listed in algorithm 1 (the algorithm is for gray image). In CSCM-GW, each subband is vectorized as the observations of a variable x_i. It is assumed that the observations of a subband obey a certain probability distribution. Then the observations of variables x_i are used to estimate the distribution $F_i(x|\alpha_i)$ by maximum likelihood (ML), and then the CFD vectors can be calculated according to the estimated parameters $\tilde{\alpha}_i$ and the observations. Because Weibull distribution can well fit the magnitude coefficients of complex wavelets [46], we employ it to model the magnitude subbands of Gabor wavelet. The PDF of Weibull distribution is

$$f_{WBL}(x|\alpha, \beta) = \left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^{\alpha}} \tag{10}$$

where α is the shape parameter, and β is the scale parameter. The CDF of Weibull distribution is:

$$F_{WBL}(x|\alpha, \beta) = 1 - e^{-(x/\beta)^{\alpha}} \tag{11}$$

To the angle subbands of Gabor wavelet, for simplicity, normal distribution is used to model the distribution of the angle subbands.

After all the CDFs corresponding to subbands are determined, these $F_i(x|\tilde{\alpha}_i)$ are concatenated into a matrix M_F (each column of M_F corresponds to a CDF vector). For simplicity, we use F_i to denote the $F_i(x|\alpha_i)$, and then the detailed CDF vector is $F_i = [F_{1,i}, F_{2,i}, \cdots, F_{n,i}]$. To a d-dimensional vector with n observations, M_F has the following express:

$$M_F = [F_1, F_2, \cdots, F_d] = \begin{bmatrix} F_{1,1} & F_{1,2} & \cdots & F_{1,d} \\ F_{2,1} & F_{2,2} & \cdots & F_{2,d} \\ & & \ddots & \\ F_{n,1} & F_{n,2} & \cdots & F_{n,d} \end{bmatrix}. \tag{12}$$

where $0 \le F_{i,j} \le 1$. If the marginal distribution is not given, empirical CDF is used as the estimator of the margin based on Kaplan-Meier algorithm [48].

Finally, based on M_F we can calculate the covariance matrix R by using COV operation. For two distribution vectors F_i and F_j, COV operation is defined as

$$COV(F_i, F_j) = \frac{1}{n-1}\sum_{k=1}^{n}\left(F_{k,i} - \mu_{F_i}\right)\left(F_{k,j} - \mu_{F_j}\right). \tag{13}$$

where $\mu_{F_i} = \frac{1}{n}\sum_{k=1}^{n} F_{k,i}$ is the mean of F_i; $\mu_{F_j} = \frac{1}{n}\sum_{k=1}^{n} F_{k,j}$ is the mean of F_j; n is the number of the observations. Given $F = [F_1, \cdots, F_d]$, covariance matrix R of F is calculated by using $COV(F)$

$$R = COV(F) = (COV([F_1, \cdots, F_d])) = \begin{bmatrix} COV(F_1, F_1) & COV(F_1, F_2) & \cdots & COV(F_1, F_d) \\ COV(F_2, F_1) & COV(F_2, F_2) & \cdots & COV(F_2, F_d) \\ & & \ddots & \\ COV(F_d, F_1) & COV(F_d, F_2) & \cdots & COV(F_d, F_d) \end{bmatrix}. \quad (14)$$

Covariance is statistically expressed as the correlation between two random variables. If variables X and Y are independent, then their covariance is zero. Similarly, if the variables in the random vector are dependent of each other, then every element in the covariance matrix except the main diagonal is equal to zero. In this case, the useful information data are the variances in the main diagonal of the covariance matrix. Thus, there only exist dependencies between the variables, the covariance matrix can show its superior performance. In CSCM-GW, because the covariance matrix is built on Gabor wavelet domain and the dependence exists between the subbands of Gabor wavelet, CSCM-GW has a robust performance for texture recognition.

We use both the magnitude and the angle subbands of Gabor wavelet for texture recognition. In the proposed method, the upper triangular part of the covariance matrix of magnitude subbands (denoted as mag-cov) is concatenated with the lower triangular part of angle subbands (denoted as ang-cov), expressed as:

$$R = triu(R_m) + tril(R_a, -1) \quad (15)$$

where $triu(R_m)$ denotes the operation of cropping the upper triangular matrix R_m including the main diagonal; and $tril(R_a, -1)$ denotes the operation of cropping the lower triangular matrix R_a below the main diagonal. Alternatively, one can fuse the upper triangular matrix of R_a and the lower triangular matrix of R_m, expressed as:

$$R = triu(R_a) + tril(R_m, -1) \quad (16)$$

The scheme of fusing the covariance matrix of magnitude and the covariance matrix of angle subbands is illustrated in Fig.3. Finally, we obtain a fused covariance matrix for an image.

B. DESIGNING THE PROJECTING CNN FOR TEXTURE CLASSIFICATION

In this section, we will design the CNNs to project the fused covariance matrix produced by CSCM-GW. Two types of network structures are implemented: one for the small number of texture classes denoted by CNN#1 [see Fig.4 (a)], one for the large number of texture classes denoted by CNN#2 [see Fig. 4 (b)]. To CNN#1, the CNN has 10 layers; to CNN#2, the CNN has 14 layers. We use BatchNormalization layer (BN) to normalize the activations of the previous layer in each batch, and use dropout layer (drop) to help prevent over-fitting. For CNN#1, one convolution layer and one Full Connection (FC) layer are used; for CNN#2, two convolution layers and two full connection layers are employed. Especially, CNN#1 is tested on KTH-TIPS2-b texture database which is a color texture database and has 11 texture classes, and CNN#2 is tested on Brodatz texture database which is a gray texture database and has 111 texture classes (see experiment section). To the color images, the size of the 2D input matrix (the fused covariance matrix) is 120 × 120; to a gray image, the size of 2D input matrix is 40 × 40. The detailed description of the parameters of the two CNNs is described as follows:

CNN#1:
data: 120 × 120 input matrix;
conv: 25 3×3, stride=1;
pool: maxpooling, 2 × 2, stride=2;
drop: dropout probability is set to 0.6;
FC: 11 nodes for the 11 texture classes.

CNN#2:
data: 40 × 40 input matrix;
conv (first): 25 3×3, stride=1;
conv (second): 12 3×3, stride=1;
drop: dropout probability is 0.6;
FC(first): 1000 nodes;
FC(second): 111 nodes for the 111 texture classes.

FIGURE 3. The fused covariance matrix. Mag-cov indicates the covariance matrix of magnitude subbands; Ang-cov indicates the covariance matrix of angle subbands.

IV. EXPERIMENTS

To evaluate the performance of the proposed methods, we carried out several experiments on the two datasets: Brodatz (111) and KTH-TIPS-2b.

Brodatz(111) dataset [8] is composed of 111 grays-scale images representing a large variety of natural textures and it has been widely used as a validation dataset. The challenge for this database is that there are the relatively large number of classes and the small number of samples per class (see the textures in the first row of Fig.5); and there are also several inhomogeneous textures (see the textures in the second row of Fig.5) which are easily mis-classified. We use the consistent approach as [8], [49], and [50] to create this dataset: Each of the 640 × 640 texture was divided into 9 nonoverlapping subimages (215 × 215), of which 3 subimages were used for training and the remaining 6 for testing. The total number of samples in this dataset is 999.

KTH-TIPS2-b dataset (KTH2(11)) has 11 material classes [51], and each material class has 4 samples and each sample contains 108 color images. These images were generated under the conditions of illumination changes, small rotations, small pose changes and scale changes (see Fig.6). The total number of samples in this dataset is 4752.

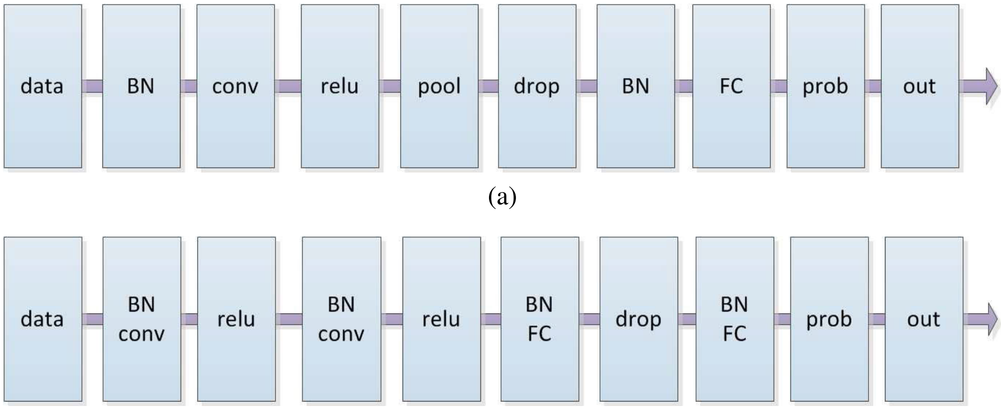

FIGURE 4. The structure of CNN used for projecting the fused covariance matrix. The names of layers are described as follows: data−2D input matrix, conv−Convolution layer, BN− BatchNormalization layer, relu−Relu layer, pool−MaxPooling layer, drop−Dropout layer, FC−Fully connected layer, prob−Softmax layer, out−Classification output layer. (a) For the small number of texture classes CNN#1. (b) For the large number of texture classes CNN#2.

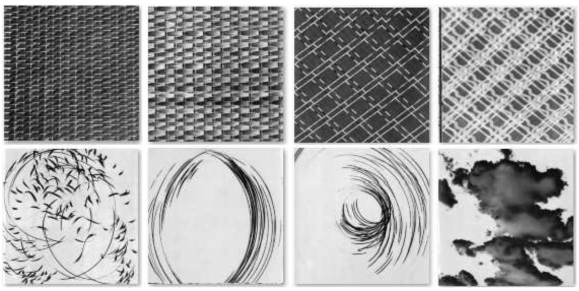

FIGURE 5. Eight example textures from the Brodatz (111) dataset.

FIGURE 6. Two textures with different scales, illumination and sizes from the KTH2 (11) dataset.

We followed the standard evaluation protocol [51]: training on three samples, testing on the remainder.

To better evaluate the proposed methods, we first describe these methods as follows:

- **CSCM-GWM**. Constructing the covariance matrix on the CDF space of magnitude subbands of Gabor wavelet. Riemannian distance of (1) is used as the dissimilarity of two covariance matrices for texture recognition.
- **CSCM-GWA**. Constructing the covariance matrix on the CDF spaces of angle subbands of Gabor wavelet. Riemannian distance is used as the dissimilarity of two covariance matrices.
- **CSCM-GWMA**. The combination of CSCM-GWM and CSCM-GWA. The dissimilarity of two images is the sum of Riemannian distances on CSCM-GWM and CSCM-GWA.
- **CSCM-GWM-CNN**. Constructing the covariance matrix on the CDF spaces of magnitude subbands of Gabor wavelet. CNN is used as the classifier of texture recognition.
- **CSCM-GWA-CNN**. Constructing the covariance matrix on the CDF spaces of angle subbands of Gabor wavelet. CNN is used as the classifier of texture recognition.
- **CSCM-GWF-CNN**. The covariance matrices are constructed on the CDF spaces of magnitude and angle subbands of Gabor wavelet. The fused covariance matrix is used as the input of CNN for texture recognition.

For better evaluating our method we also implemented the covariance matrix in Gabor wavelet domain (Cov-GW) including on the magnitude subbands of Gabor wavelet (called Cov-GWM) (which is similar to the method in [35]) and the angle subbands of Gabor wavelet (Cov-GWA), as well as the combination of Cov-GWM and Cov-GWA (called Cov-GWMA) by using the same combining approach as CSCM-GWMA.

At the beginning, we compared our models CSCM-GWs (including CSCM-GWA, CSCM-GWM and CSCW-GWMA) against covariance matrix models Cov-GWs (including Cov-GWA, Cov-GWM and Cov-GWMA) on the two datasets, and the experimental results of are shown in Table 1. We can observe that the proposed CSCM-GWs always outperform Cov-GWs which directly use the covariance matrix in Gabor domain. Especially, CSCM-GWs show an obvious improvement compared with Cov-GWs on KTH2(11). This experiment validates that the covariance model in CDF space is more effective than the covariance matrix in the original Gabor wavelet domain.

TABLE 1. The comparison of classification accuracies (%) of CSCM-GW and Cov-GW.

Method	Brodatz(111)	KTH2(11)
Cov-GWM	87.69	69.51
CSCM-GWM	**89.64**	**74.79**
Cov-GWA	89.79	71.45
CSCM-GWA	**90.84**	**71.89**
Cov-GWMA	93.09	70.60
CSCM-GWMA	**93.36**	**74.73**
Cov-GWF-CNN	94.29	79.74
CSCM-GWF-CNN	**95.35**	**83.70**

Then we validated the performance of CSCM in Gabor wavelet domain with CNN learning. The recognition accuracies of the CNN based CSCM models (including CSCM-GWA-CNN, CSCM-GWM-CNN and CSCM-GWF-CNN) and the CSCM in Gabor domain without CNN learning (including CSCM-GWA, CSCM-GWM and CSCM-GWMA), are listed in Table 2. In this experiment, the CNN#1 was applied on KTH2 (11), and the CNN#2 is applied on Brodatz (111) since the number of classes of KTH2 (11) is small, while the number of classes of KTH2 (11) is relatively large. It is obvious that the performance of CSCM and covariance model in Gabor domain is significantly improved by using CNN. For example, the recognition accuracy of CSCM-GWF-CNN is improved by about 2% on Brodatz (111) and improved by about 9% on KTH2 (11) than that of CSCM-GWMA.

TABLE 2. The comparison of classification accuracies (%) of CSCM-GW and CSCM-GW-CNN.

Method	Brodatz(111)	KTH2(11)
CSCM-GWM	89.64	74.79
CSCM-GWM-CNN	**89.84**	**81.64**
CSCM-GWA	90.84	71.89
CSCM-GWA-CNN	**92.49**	**79.82**
CSCM-GWMA	93.36	74.73
CSCM-GWF-CNN	**95.35**	**83.70**

From Table 1 and Table 2, we observe that CSCM-GWs has robust performance for texture recognition. CSCM-GWMA has the best classification accuracy compared with CSCM-GWM and CSCM-GWA. Similarly, CSCM-GWF-CNN has the best performance among the CNN-based CSCMs including CSCM-GWM-CNN and CSCM-GWA-CNN. The results in Table 1 and Table 2 demonstrate that our model CSCM has better performance than the covariance matrix in Gabor wavelet domain; CNN significantly improves the performance of the covariance model and CSCM.

About computational performance, the computational time of CSCM-GWF-CNN which uses the fused covariance matrix is equal to CSCM-GWs, and higher than Cov-GWs because CSCM-GWs need to calculate the projection from the subband to its CDF space. However, in the recognition step CSCM-GWF-CNN has a very high efficient. The runtime of CSCM-GWF-CNN for recognizing a sample image on Brodatz (111) is about 0.072ms (Core i7 6700 4GHz CPU, 32GB RAM, Matlab 2017b) and the runtime for recognizing a sample image on KTH2 (11) is about 0.25ms; while the runtime of CSCM-GWMA is about 200ms and 4450ms on the two datasets, respectively (see Table 3).

TABLE 3. The runtime for recognizing a sample image on the two datasets.

Method	Brodatz(111)	KTH2(11)
CSCM-GWM	100ms	2230
CSCM-GWA	100ms	2230
CSCM-GWMA	200ms	4450ms
CSCM-GWF-CNN	**0.072ms**	**0.25ms**

TABLE 4. The comparison of classification accuracies (%) with the sate-of-the-art on the two datasets.

Method	Brodatz(111)	KTH2(11)
MRELBP [52]	93.12	77.91
ScatNet [53]	84.46	68.92
PCANet [54]	90.87	57.70
RandNet [54]	91.14	56.90
FV-AlexNet [18]	98.20	77.90
FV-SIFT [18]	–	81.5
FV-VGGM [55]	**98.70**	79.90
CSCM-GWF-CNN (our)	95.35	**83.70**

We also compared CSCM-GWF-CNN against the state-of-the-art methods. The accuracies of these methods are shown in Table 4. Based on the fused covariance matrix combining with CNN, the proposed method (CSCM-GWF-CNN) obtained 95.35% and 83.70% classification accuracies on the two texture datasets, respectively. On Brodatz (111) database, CSCM-GWF-CNN yielded a promising classification accuracy and slightly lower accuracy than FV-VGGM. However, CSCM-GWF-CNN outperforms all the other methods on KTH2 (11). It should be pointed out that FV-VGGM is a competitive method and is also computationally expensive because its feature dimension is 65 535, making it unfeasible to run in low-power applications. Therefore, overall our method has the robust and excellent recognition ability for texture image classification. Note that we can use FC layer as the features of image and in CNN#1 and CNN#2 the number of FC layer nodes is less than 1000 (the node of FC layer of CNN#1 is 11 and the node of the second FC layer of CNN#2 is 111), so the feature dimensionality produced by CSGM-GWF-CNN is significantly low. It can be seen that the combination of Gabor wavelet and CNN indeed reduces the number of network layers: In CSGM-GWF-CNN, a shadow CNN having few layers can yield promising recognition accuracies; while numerous layers are used in the deep CNNs such as FV-VGGM and RandNet.

V. CONCLUSIONS

We use both the Gabor wavelet and deep learning to implement texture recognition. Our method is a global texture describing the model in CDF space of Gabor wavelet domain based on the covariance matrix, called CSCM-GW (CDF Space Covariance Model). There are two contributions in this work. First, we proposed CSCM-GW for texture feature extraction. CSCM-GW can obtain robust texture feature, and the images of different size will be transformed into fixed-size covariance matrix. The covariance matrix in CDF space is capable of capturing the dependence of the subbands of Gabor wavelet and obtaining significant improvement compared to the covariance matrix which directly employs the subband coefficients. Second, we use CNN to project CSCM-GW, which is based on the fused covariance matrix, into the linear space (CSCM-GWF-CNN). With CNN, the runtime is largely reduced in the recognition phase because the output of the proposed method is a low dimensional vector, and more importantly we also improved the recognition accuracy by incorporating the learning technique.

It should be addressed that in the proposed methods, the structures of the two CNNs are designed based on experiments and maybe not the optimal structures. One can design the more efficient CNN structures for texture recognition. Furthermore, our methods just used the global feature of the image, while the local features are not used. Our methods are similar to covariance descriptor which focuses on extracting the local feature and has promising performance for object recognition and detection [36], [56]. So we believe the performance of our methods will further be improved and can be used to other applications such object recognition and detection if the local features of image are used in our methods.

REFERENCES

[1] M. Mellor, B.-W. Hong, and M. Brady, "Locally rotation, contrast, and scale invariant descriptors for texture analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 52–61, Jan. 2008.

[2] Q. Chen, Z. Song, J. Dong, Z. Huang, Y. Hua, and S. Yan, "Contextualizing object detection and classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 1, pp. 13–27, Jan. 2015.

[3] Y. Zhang, Z. Dong, L. Wu, and S. Wang, "A hybrid method for MRI brain image classification," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10049–10053, 2011.

[4] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[5] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[6] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, May 2012.

[7] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[8] L. Liu and P. W. Fieguth, "Texture classification from random features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 574–586, Mar. 2012.

[9] W. Shao and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Oct. 2016.

[10] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[11] M. Kokare, P. K. Biswas, and B. N. Chatterji, "Rotation-invariant texture image retrieval using rotated complex wavelet filters," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 6, pp. 1273–1282, Dec. 2006.

[12] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[13] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2009, pp. 689–696.

[14] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[15] 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.

[16] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[17] Y. Pu *et al.*, "Variational autoencoder for deep learning of images, labels and captions," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 2352–2360.

[18] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3828–3836.

[19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 91–99.

[20] V. Andrearczyk and P. F. Whelan, "Convolutional neural network on three orthogonal planes for dynamic texture classification," *Pattern Recognit.*, vol. 76, pp. 36–49, Apr. 2017.

[21] C.-M. Pun and M.-C. Lee, "Log-polar wavelet energy signatures for rotation and scale invariant texture classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 590–603, May 2003.

[22] S. Ekici, S. Yildirim, and M. Poyraz, "Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2937–2944, 2008.

[23] P. S. Hiremath and R. A. Bhusnurmath, "Texture classification using partial differential equation approach and wavelet transform," *Pattern Recognit. Image Anal.*, vol. 27, no. 3, pp. 473–479, 2017.

[24] S. C. Kim and T. J. Kang, "Texture classification and segmentation using wavelet packet frame and Gaussian mixture model," *Pattern Recognit.*, vol. 40, no. 4, pp. 1207–1221, Apr. 2007.

[25] M. S. Allili, N. Baaziz, and M. Mejri, "Texture modeling using contourlets and finite mixtures of generalized Gaussian distributions and applications," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 772–784, Apr. 2014.

[26] J. Portilla and E. P. Simoncelli, "Texture modeling and synthesis using joint statistics of complex wavelet coefficients," in *Proc. IEEE Workshop Stat. Comput. Theories Vis.*, 1999, pp. 1–32.

[27] 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.

[28] L. Bombrun, Y. Berthoumieu, N.-E. Lasmar, and G. Verdoolaege, "Multivariate texture retrieval using the geodesic distance between elliptically distributed random variables," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 3637–3640.

[29] S. Sakji-Nsibi and A. Benazza-Benyahia, "Copula-based statistical models for multicomponent image retrieval in thewavelet transform domain," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 253–256.

[30] G. Verdoolaege, S. De Backer, and P. Scheunders, "Multiscale colour texture retrieval using the geodesic distance between multivariate generalized Gaussian models," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2008, pp. 169–172.

[31] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized Gaussian distributions," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5960–5971, Dec. 2013.

[32] H. Permuter, J. Francos, and I. Jermyn, "A study of Gaussian mixture models of color and texture features for image classification and segmentation," *Pattern Recognit.*, vol. 39, no. 4, pp. 695–706, 2006.

[33] N.-E. Lasmar and Y. Berthoumieu, "Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2246–2261, May 2014.

[34] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2006, pp. 589–600.

[35] Y. Pang, Y. Yuan, and X. Li, "Gabor-based region covariance matrices for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 989–993, Jul. 2008.

[36] R. Wang, H. Guo, L. S. Davis, and Q. Dai, "Covariance discriminative learning: A natural and efficient approach to image set classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2496–2503.

[37] Y. Chen, P. Lin, Y. He, and Z. Xu, "Classification of broadleaf weed images using Gabor wavelets and lie group structure of region covariance on Riemannian manifolds," *Biosyst. Eng.*, vol. 109, no. 3, pp. 220–227, 2011.

[38] H. Q. Minh, M. S. Biagio, L. Bazzani, and V. Murino. (2016). "Kernel methods on approximate infinite-dimensional covariance operators for image classification." [Online]. Available: https://arxiv.org/abs/1609.09251

[39] W. Wang, R. Wang, S. Shan, and X. Chen, "Discriminative covariance oriented representation learning for face recognition with image sets," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5749–5758.

[40] Z. Huang, R. Wang, S. Shan, X. Li, and X. Chen, "Log-Euclidean metric learning on symmetric positive definite manifold with application to image set classification," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 720–729.

[41] Q. Liu, G. Shao, Y. Wang, J. Gao, and H. Leung, "Log-Euclidean metrics for contrast preserving decolorization," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5772–5783, Dec. 2017.

[42] E. Ghorbel, J. Boonaert, R. Boutteau, S. Lecoeuche, and X. Savatier, "An extension of kernel learning methods using a modified Log-Euclidean distance for fast and accurate skeleton-based human action recognition," *Comput. Vis. Image Understand.*, vol. 175, pp. 32–43, Oct. 2018.

[43] M.-T. Pham. (2018). "Efficient texture retrieval using multiscale local extrema descriptors and covariance embedding." [Online]. Available: https://arxiv.org/abs/1808.01124

[44] M. Harandi, M. Salzmann, and R. Hartley, "Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 48–62, Jan. 2018.

[45] F. Mirzapour and H. Ghassemian, "Improving hyperspectral image classification by combining spectral, texture, and shape features," *Int. J. Remote Sens.*, vol. 36, no. 4, pp. 1070–1096, 2015.

[46] C. Li, Y. Huang, and L. Zhu, "Color texture image retrieval based on Gaussian copula models of Gabor wavelets," *Pattern Recognit.*, vol. 64, pp. 118–129, Apr. 2017.

[47] H. R. Shahdoosti and N. Javaheri, "Pansharpening of clustered MS and Pan images considering mixed pixels," *IEEE Trans. Geosci. Remote. Lett.*, vol. 14, no. 6, pp. 826–830, Jun. 2017.

[48] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, 2nd ed. Hoboken, NJ, USA: Wiley, 2002.

[49] X. Qi, R. Xiao, C.-C. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise rotation invariant co-occurrence local binary pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2199–2213, Nov. 2014.

[50] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognit.*, vol. 62, pp. 135–160, Feb. 2017.

[51] B. Caputo, E. Hayman, and P. Mallikarjuna, "Class-specific material categorisation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2005, pp. 1597–1604.

[52] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Median robust extended local binary pattern for texture classification," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1368–1381, Mar. 2016.

[53] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.

[54] T.-H. Chan, K. Jia, S. Gao, J. Lu, and Z. Zeng, Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[55] T.-Y. Lin and S. Maji, "Visualizing and understanding deep texture representations," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2791–2799.

[56] P. Li, Q. Wang, H. Zeng, and L. Zhang, "Local log-Euclidean multivariate Gaussian descriptor and its application to image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 803–817, Apr. 2017.

CHAORONG LI received the Ph.D. degree from the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, in 2013, and the M.S. degree in the computer science and engineering from the Western Normal University of China, Nanchong, China, in 2007. He is currently an Assistant Professor with the College of Computer and Information Engineering, Yibin University, Yibin, China. His research interests include wavelet analysis, pattern recognition, computer vision, and information security.

XIAOYUN GUAN received the B.S. degree in the computer science and engineering from Southwest Normal University, in 2005, and the M.S. degree in the computer science and engineering from Sichuan Normal University, in 2013. He is currently an Associate Professor with the College of Computer and Information Engineering, Yibin University, Yibin, China. His research interests include image and pattern recognition.

PEIHUI YANG received the B.S. degree in computer science and engineering from Southwest Normal University, in 2005, and the M.S. degree in computer science and engineering from Sichuan Normal University, in 2013. She is currently a Lecturer with the Department of Computer Science and Technology, Sichuan Vocational College of Post and Telecom, Chengdu, China. Her research interests include image and pattern recognition.

YUANYUAN HUANG received the B.S., M.S., and Ph.D. degrees in computer science from the University of Electronic Science and Technology of China, in 2004, 2007, and 2013, respectively. From 2009 to 2011, he was a Visiting Scientist with the University of Washington, Seattle, WA, USA. He has been an Assistant Professor with the Chengdu University of Information Technology, since 2013. His research interests include image/video processing and cyberspace engineering.

WEI HUANG received the M.S. degree in biomedical engineering from the University of Electronic Science and Technology of China, in 2017, where he is currently pursuing the Ph.D. degree in biomedical engineering. His research interests include visual cognitive decoding, deep learning, computer vision, and functional magnetic resonance imaging.

HUAFU CHEN received the Ph.D. degree in biomedical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2004. He was a Visiting Scholar with The University of Texas at San Antonio, USA, from 2005 to 2006. He is currently a Professor with the University of Electronic Science and Technology of China. He has published 140 SCI index papers, including *Brain*, *NeruoImage*, *Human Brian Mapping*, the IEEE Transactions on Biomedical Engineering, the IEEE Transactions on Medical Imaging, and *Physical Review E*. His current research interests include magnetic resonance imaging pattern recognition, brain functional and structure network analysis method, and their applications in cognitive science and neurological problems.