IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 14, 2019, accepted January 30, 2019, date of publication February 18, 2019, date of current version March 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2899024

Facial Expression Recognition Using Fusion Features Based on Center-Symmetric Local Octonary Pattern

MIN HU[1], YAQIN ZHENG[1], CHUNJIAN YANG[1], XIAOHUA WANG[1], LEI HE[2], AND FUJI REN[1,3]

[1]School of Computer and Information, Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine, Hefei University of Technology, Hefei 230602, China
[2]School of Mathematics, Hefei University of Technology, Hefei 230602, China
[3]Graduate School of Advanced Technology and Science, University of Tokushima, Tokushima 7708502, Japan

Corresponding author: Xiaohua Wang (xh_wang@hfut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672202 and Grant 61673156, in part by the State Key Program of the NSFC-Shenzhen Joint Foundation under Grant U1613217, and in part by the State Key Program of the National Natural Science of China under Grant 61432004.

ABSTRACT A local feature descriptor has gained a lot of interest in many applications, such as image retrieval, texture classification, and face recognition. This paper proposes a novel local feature descriptor: center-symmetric local octonary pattern (CS-LOP) for facial expression recognition. A CS-LOP operator not only considers the difference of the gray value between central pixels and neighboring pixels in all eight directions but also compares the gray value of four pairs of center-symmetric pixels. Besides, this paper used the CS-LOP to extract diverse features from the preprocessed facial image, the feature map of gradient magnitude, and the feature map of Gabor, and also to make extracted features more abundant and detailed. To evaluate the performance of the proposed method, experiments on JAFFE and CK facial expression datasets demonstrate that the proposed method outperforms the method using the individual descriptor. Compared with other state-of-the-art methods, our approach improves the overall recognition accuracy.

INDEX TERMS Facial expression recognition, feature extraction, center-symmetric local octonary pattern, feature fusion.

I. INTRODUCTION

Since facial expressions can be represented by appearance changes on the face and they are one of the most potent, natural and immediate means for a human being to communicate their emotions and intentions. Facial expression recognition as a branch of affective computing promotes the development of various fields, such as security technology, human-computer interaction, educational supervision, medical testing, entertainment, etc. [1]– [3]. Owing to its practical applications and bright prospects, facial expression recognition became a research focus for many scholars and they also have made significant progress in recent years [4]– [6], [27].

The process of expression recognition can be roughly divided into three parts: image acquisition and preprocessing, image feature extraction and selection, classification and recognition [7], [8]. The feature extraction for facial images is one key part that determines the accuracy of facial expression recognition. At present, there are two main approaches to describe facial images: geometric-feature-based and texture-feature-based. The first approach based on geometric feature presents facial images by encoding region of interest (ROI), such as mouth, eyes, eyebrows, nose, etc. ROI can describe facial image efficiently using a few features, but its performance relies on the locations of key facial components. In this category, AAM and ASM are two common methods. ASM proposed by Cootes *et al.* [11] is inefficient because of the uncertainty of convergence and iterations. AAM [12] is invariant to scale and rotation. However, the texture-feature-based methods can avoid those problems innately. In this category, there are some widely used methods: LBP [9], [13], [9], LTP [14], HOG [15], GLCM [16], SIFT [17] etc. To make full use of different methods, many researchers began to fuse those

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

features [10], [18], [22], [51], [52] and achieved better results than using an individual feature. Recently, deep learning, which integrates both feature extraction and learning procedure within deep networks, is being widely used for facial expression recognition [19]– [21]. However, some drawbacks of deep learning have emerged gradually: a large number of training samples are required; the neural network can easily converge to a local minimum; the generalization of models are poor etc.

The feature extraction algorithm used in this paper is texture-feature-based. LBP operator proposed by Ojala *et al.* [13] is an effective descriptor for extracting texture features and classification. LBP has some attractive properties such as rotation invariance, robustness against monotonic gray level transformation and easy to encode [23]. However, LBP only compares the gray value between central pixels and neighboring pixels. It does not consider the gray value of four pairs of center-symmetric pixels, and also ignore part of the structural information of the original image. After that, many variations about LBP have been proposed: CS-LBP [24], ULBP [35], ELBP [36], RILBP [37], LTP [14] etc. For example, CS-LBP reduces the feature dimension from LBP, but it also has some problems: ignore central pixels, the optimal threshold is hard to determine and must be chosen from experiments.

To solve the above problems of LBP and CS-LBP, this paper proposes a novel texture feature descriptor: Center-Symmetric Local Octonary Pattern (CS-LOP). It not only compares the gray value between central pixels and neighboring pixels in all eight directions but also compares the gray value of four pairs of center-symmetric pixels. CS-LOP reduces feature vector dimension based on LBP and avoids selecting the optimal threshold. Furthermore, this paper apply CS-LOP operator to extracts Gabor and gradient magnitude features from facial images, then obtain the other two features: CS-LOP based on feature maps of Gabor and CS-LOP based on feature map of gradient magnitude. After that, the three features as mentioned above are fused to achieve feature-level fusion for facial expression recognition.

The rest of this paper is organized as follows. Section II introduces the related works about LBP and CS-LBP briefly; section III describes a facial feature extraction method based on CS-LOP and then introduces the feature fusion; section IV presents experimental results and analysis. This paper is concluded in section V.

II. RELATED WORK

A. LOCAL BINARY PATTERN (LBP)

Texture information is very important for pattern analysis of images, and local binary pattern (LBP) were considered to gain texture information from images as a simple but very efficient texture descriptor. LBP was first proposed by Ojala *et al.* [13] and from then on LBP has been found to be a significant feature for texture representation. LBP obtain a binary pattern of the differences between the central pixels and neighboring pixels. It labels the pixels of an image by thresholding the 3*3 neighborhood of each pixel with the central value and considering the result as a binary number. Moreover, LBP code is computed in formula (1) and (2):

$$LBP_{N,R}(C) = \sum_{i=0}^{N-1} S(P_i - P_c)2^i \tag{1}$$

$$S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2}$$

where P_c denotes the gray value of the central pixel, P_i (i = 0,1,…, N-1) is the gray value of neighboring pixel centered on c, N is the total number of involved neighbors, R is the radius of the neighborhood.

B. CENTER-SYMMETRIC LOCAL BINARY PATTERN (CS-LBP)

The CS-LBP operator [24] is one of the variants of the LBP and SIFT descriptor, inheriting the desirable properties of both texture features and gradient-based features. LBP produces a rather long histogram and has 256 different binary patterns, whereas CS-LBP only produces 16 binary patterns. Instead of comparing each pixel with center one, CS-LBP utilizes center-symmetric to encode an image by thresholding the difference of gray value of four pairs center-symmetric pixels as a binary number. So CS-LBP encoding method is defined in formula (3) and (4):

$$CS-LBP_{N,R}(C) = \sum_{i=0}^{(N/2)-1} S\left(P_i - P_{i+(N/2)}\right)2^i \tag{3}$$

$$S(x) = \begin{cases} 1, & x \geq T \\ 0, & x < T \end{cases} \tag{4}$$

where P_i (i = 0, 1,…, (N/2)-1) denotes the gray value of neighboring pixel centered on c, and N is the total number of involved neighbors. R is the radius of the neighborhood, T is a threshold that is chosen in experiments.

This paper take N = 9, R = 1 and Figure 1 show the calculation process of LBP operator and CS-LBP operator respectively.

III. APPROACH

A. CENTER-SYMMETRIC LOCAL OCTONARY PATTERN (CS-LOP)

Considering LBP operator that mentioned in section II, it only considers the difference of gray value between the central pixels and neighboring pixels but ignores to compare the difference of center-symmetric pairs of pixels. Besides, CS-LBP operator reduces the feature vector dimension and simplifies the computational complexity on the basis of LBP. However, CS-LBP does not consider the central pixels and the threshold (T) has a significant influence on experimental results. Inspired by the two feature descriptors as mentioned earlier, this paper proposes a novel descriptor: Center-Symmetric Local Octonary Pattern (CS-LOP). CS-LOP not only compares the gray value between the central pixels and neighboring pixels but also take the gray value of center-symmetric

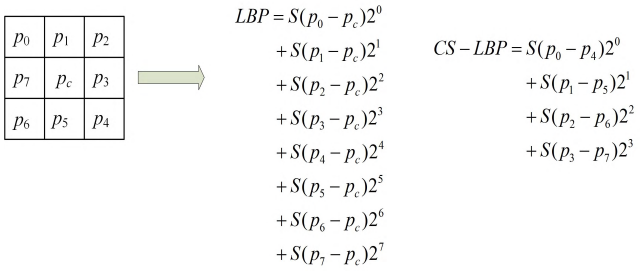


FIGURE 1. LBP and CS-LBP operators for a neighbor with 8 pixels.

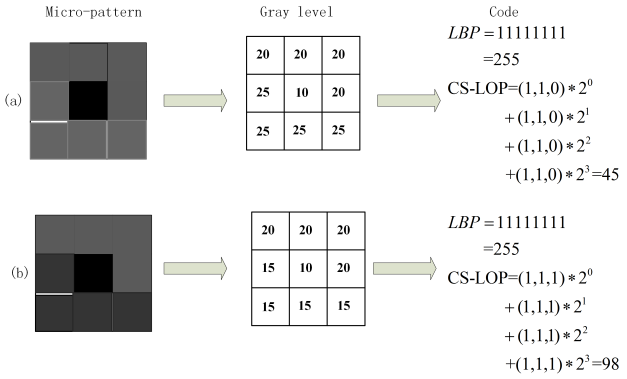


FIGURE 2. Different micro-patterns with LBP code and CS-LOP code.

neighboring pixels into account so that the extracted texture features are more complete and detailed. The encoding about CS-LOP is shown in formula (5), (6) and (7).

$$CS-LOP_{N,R}(C) = \sum_{i=0}^{(N/2)-1} \mu(s(p_i - p_c), \quad s\left(p_{i+(\frac{N}{2})} - p_c\right), s(p_i - p_{i+(\frac{N}{2})}))2^i \tag{5}$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{6}$$

$$\mu(a, b, c) = a * 2^0 + b * 2^1 + c * 2^2 \tag{7}$$

where N denotes the total number of involved neighbors, p_c is the gray value of the central pixel, p_i (i = 0, 1,. . .,((N/2)-1)) corresponds to the gray value of neighboring pixel. R is the radius of the neighborhood. This paper employs μ(a, b, c) for the first time here. Different from LBP and CS-LBP operators encoding the result as a binary number and mapping the value into four-digit decimal numbers from 0 to 3, CS-LOP adds one binary code that maps the value into [0, 7] interval, which makes the coded values more diverse and captures the delicate difference of features between pixels so that the extracted features more precise and accurate. Figure 2 shows code results for LBP and CS-LOP.

From a human visual perspective, the two micro-patterns represented in the image of Figure 2 (a) and (b) are very similar. Therefore, the generated LBP codes are the same. However, applying the CS-LOP operator generates different code values because CS-LOP operator compares the center-symmetric pairs of pixels from four directions. It is shown that local texture feature descriptor: CS-LOP can capture the delicate differences and represent in the code values from unobvious regions.

B. CS-LOP BASED ON FEATURE MAP OF GRADIENT MAGNITUDE

The Feature Map of Gradient Magnitude (FMGM) is formed by extracting texture features utilizing gradient magnitude on an image sample. The idea of gradient magnitude derives from the histogram of oriented gradient, which refers to four neighboring pixels in the horizontal and vertical directions [25] and produces the corresponding gradient value by their magnitudes. It is described in formula (8) and (9).

$$\begin{cases} G_x(c) = p_3 - p_7 \\ G_y(c) = p_1 - p_5 \end{cases} \tag{8}$$

$$M(c) = \sqrt{G_x(c)^2 + G_y(c)^2} \tag{9}$$

As can be seen in Figure 1, c is the central pixel of 3*3 rectangle region, p_1, p_3, p_5, p_7 are four neighboring pixels centered on c. $G_x(c)$ denotes the horizontal gradient at c and $G_y(c)$ is the vertical gradient. M(c) is the gradient magnitude on pixel c.

After generating feature maps of gradient magnitude by above approach, then CS-LOP operator is used to extracting texture features from the feature map of gradient magnitude and obtain the feature: CS-LOP$_{\mathrm{FMGM}}$. With the introduction of gradient magnitude, the extracted features are complete.

C. CS-LOP BASED ON FEATURE MAP OF GABOR

Gabor wavelet [26] is a well-known descriptor representing texture information and very similar to the stimulus-response of simple cells in the human visual system. It provides excellent characteristic selection about orientation and scale. Furthermore, Gabor wavelet is robust and adapt to the changes of illumination well. The kernel function of Gabor wavelet has the same characteristics as two-dimensional reflection region of simple cells in the cerebral cortex. It captures different spatial information of frequency, position, orientation from an image and extract subtle local transformation effectively. The kernel function is defined in formula (10).

$$\psi_{\mu,v}(z) = \frac{\|k_{\mu,v}\|}{\delta^2} e^{-(\frac{\|k_{\mu,v}\|^2 \|z\|^2}{2\delta^2})} [e^{ik_{\mu,v}z} - e^{-\frac{\delta^2}{2}}] \tag{10}$$

where μ, v indicate the kernel orientation and scale of the Gabor filter; z represents the spatial coordinate of pixel; $\|\cdot\|$ is a norm descriptor; the parameter δ determines the radius of the Gaussian function; $k_{\mu,v}$ is a vector and $k_{\mu,v}, = k_v \exp(i\varphi_\mu)$, $k_v = k_{max}/\lambda^v$, $\varphi_\mu = \pi\mu/8$, k_{max} denotes the maximum of sampling frequency; λ is spatial factor among filters in the frequency domain. The Gabor feature of facial image is obtained by convolving facial image with filters. $f(z)$ denotes facial image, and then convolve with filter $\psi_{\mu,v}(z)$ is shown in formula (11).

$$G_{\mu,v}(z) = f(z) * \psi_{\mu,v}(z) \tag{11}$$

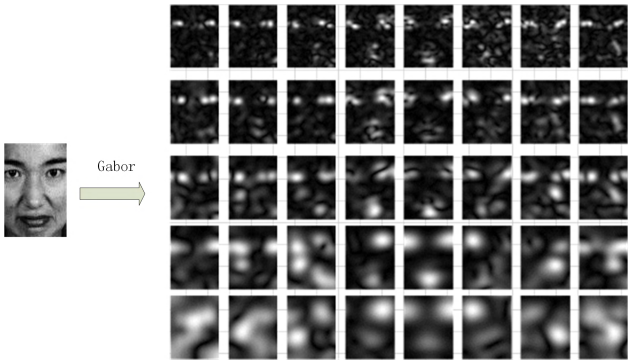


FIGURE 3. Feature maps of Gabor for a face image.

where * represents convolution operation, for a filter with five scales and eight orientations, $v \in \{0, 1, 2, 3, 4\}$, $\mu \in \{0, 1, 2, \ldots, 7\}$ and $G_{\mu,v}(z)$ is a norm of Gabor features. Taking one sample image as an example, its corresponding Gabor feature is shown in Figure 3 below.

As shown in Figure 3, a transformation of each sample image produces 40 corresponding images, and Gabor wavelet expands extracted features from different scales and orientations. We conduct CS-LOP operator for extracting features again based on feature map of Gabor to obtain the feature: $\text{CS-LOP}_{\text{Gabor}}$.

D. FEATURE FUSION

At present, the single feature extraction method cannot meet the requirements for final recognition rate. Considering different features have different capabilities to represent images, and also in order to take advantages of various feature extraction method, more and more scholars have begun to research fusing different features to achieve better results. Widely used methods are feature-level fusion and decision-level fusion. This paper adopts the former fusion scheme. The experimental results indicate that CS-LOP cascades $\text{CS-LOP}_{\text{FMGM}}$ with weight and then fuses with $\text{CS-LOP}_{\text{Gabor}}$can achieve the best result. The detailed explanation will be discussed in section IV.

IV. EXPERIMENTS AND DISCUSSION

In this section, experiments are conducted for facial expression recognition to validate the efficiency of the proposed methods and compared with some state-of-the-art methods. All experiments were implemented by using Visual Studio 2013 and OpenCV 2.4.9.

A. DATASETS CONSTRUCTION

In order to evaluate the proposed algorithm, experiments used two famous facial image datasets: The Japanese Female Facial Expression (JAFFE) [29] and Cohn-Kanade (CK) [30]. The JAFFE database contains 213 images of 10 Japanese females. The CK database consists of over 2000 facial images of 210 subjects whose age are between 18 and 50. Women accounted for 69% and men accounted for 31%. One example of expressions about the two data sets is shown in Figure 4 (from left to right, the label of expressions are: anger, disgust, fear, happy, neutral, sad, surprise).

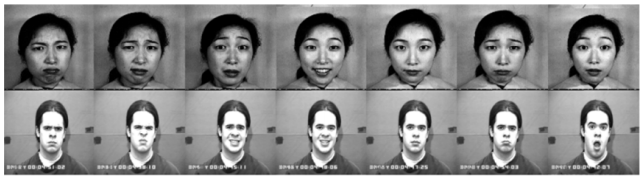


FIGURE 4. Facial expression examples in JAFFE and CK.

Our experiments select three images of each expression in JAFFE with seven basic emotions and four images in CK with six basic emotions (lack of neutral). This paper executed experiments by using person-dependent and person-independent cross-validation testing schemes [31], herein called N-fold and N-person cross-validation, respectively, to evaluate the performance of facial expression recognition accurately. In the N-fold cross-validation, all facial images are divided into N groups randomly and selected N-1 groups as training set and the rest as the test set. This scheme was repeated N times and take average results as the final recognition accuracy. In the N-person cross-validation, all facial images are partitioned into N groups according to person and exclude one person out of training set as the test set and N-1 person as the training set, so it always ensures the person-independence itself.

B. STEPS OF EXPERIMENT

1) PREPROCESS

Image preprocessing can be divided into three steps: (a) detecting the position of human eyes from image with Haar-like and AdaBoost [32] algorithm, and using the coordinates of two eyes for geometric transformation to eliminate the effect of posture. (b) detecting and cropping the region of interest and normalize the image size (JAFFE:64 $\times$ 96, CK: 96 $\times$ 96); (c) processing image with Gaussian filter [33] to eliminate the effect of noise and improve the recognition accuracy.

2) FEATURE EXTRACTION AND FUSION

This paper adopts CS-LOP descriptor to extract texture features from the preprocessed facial image, feature maps of gradient magnitude, feature maps of Gabor respectively, and then get three features: CS-LOP, $\text{CS-LOP}_{\text{FMGM}}$, $\text{CS-LOP}_{\text{Gabor}}$. Then, the first two feature histograms are fused with weighted ratio and cascades the third features to obtain the final features.

3) CLASSIFICATION

Recognizing the category of facial expression by carrying out SVM [34] with Polynomial Function kernel and using automatic training function to obtain the optimal parameters.

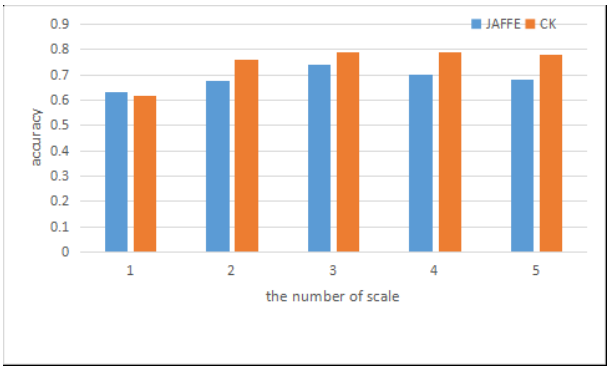


FIGURE 5. The average accuracy of different scale (N-person).

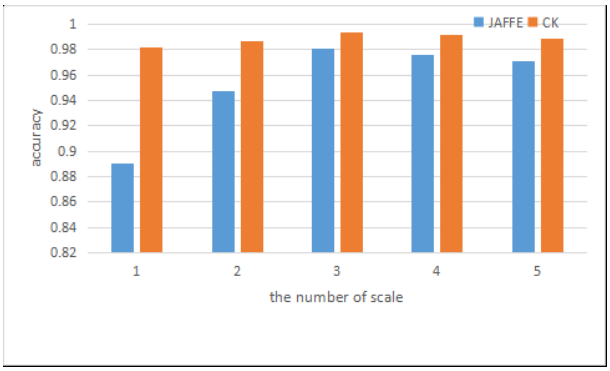


FIGURE 6. The average accuracy of different scale (N-fold).

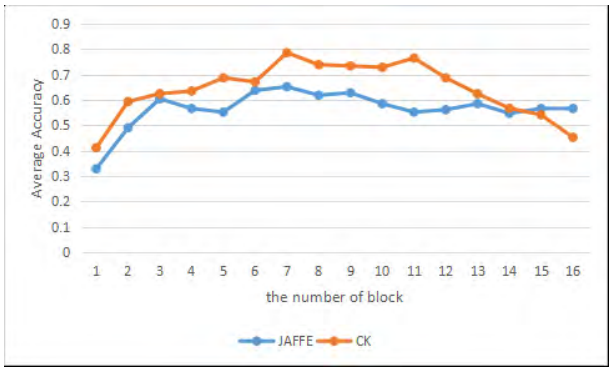


FIGURE 7. The accuracy of CS-LOP for the number of block in JAFFE and CK (N-person).

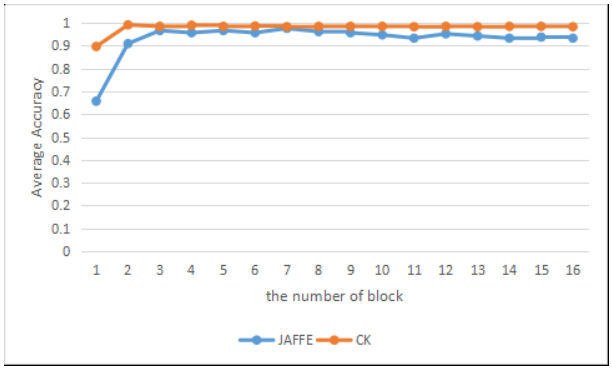


FIGURE 8. The accuracy of CS-LOP for the number of block in JAFFE and CK (N-fold).

C. EXPERIMENTAL RESULTS AND ANALYSIS

1) THE SELECTION OF GABOR FEATURE MAPS

As shown in Figure 3, the Gabor filter of five scales and eight orientations is used to transform images and each image generates 40 Gabor feature maps. The transformed Gabor feature maps increase the number of training samples and the training time significantly. In order to reduce time complexity, this paper explores the number of scale channels (μ) used by Gabor filters. We select the first $8^*\mu$ sheets ($\mu = 1, 2, 3, 4, 5$) from Gabor feature maps for feature extraction. The experimental results are shown in Figure 5 and Figure 6.

It can be seen from Figure 5 and Figure 6 that different scale of Gabor filter influence the final recognition result. Compared with other numbers of scale, scale = 3 achieves better recognition performance on JAFEE and CK. In person-independent strategy, scale = 3 or 4 shows almost the same accuracy on CK. However, the feature dimension of scale = 3 is smaller than scale = 4. Compared with previous papers exploiting Gabor filter of five scales, setting scale = 3 not only improves the overall recognition rate but also removes redundant information and reduces computation complexity.

2) THE SELECTION OF BLOCK

Before fusing the three features as mentioned earlier, it is necessary to determine the optimal number of blocks. In the process of feature extraction, the number of block has a great influence on the recognition result. Too few blocks make extracted local features insufficient. Too many blocks increase time complexity and feature dimension. Therefore, experiments are conducted under three feature extraction methods to select appropriate block number. The experimental results of CS-LOP are shown in Figure 7 and Figure 8.

It can be seen from Figure 7 that in the person-independent strategy, For JAFFE, CS-LOP yields the highest recognition rate (65.24%) when the number of blocks = 7; For CK, the corresponding accuracy reaches a maximum of 78.65% then begin to decrease obviously. Therefore, the number of blocks for CS-LOP method is set to 7 in N-person. Besides, it can be seen from Figure 8 that in the person-dependent strategy, when the number of blocks is taken as 1, (i.e. feature extracted from the whole image) the recognition rate is not ideal; but recognition rate improves greatly when the number of blocks increases to 2 and then begins to stabilize. In JAFFE, the average accuracy can up to 97.62% when the number of blocks = 7; In CK, there is no significant difference in recognition rate and it remains basically stable with the increase of number of blocks. To keep the consistency and facilitate experiments, the number of blocks is set to 7 on JAFFE and CK in N-fold strategy.

This paper also tests the optimal blocks for CS-LOP$_{\text{FMGM}}$ and CS-LOP$_{\text{Gabor}}$ in N-person and N-fold. The experimental results are shown in Figure 9 (a)-(d). For CS-LOP$_{\text{FMGM}}$

TABLE 1. The average accuracy(%) of fused feature for each expression on JAFFE and CK.

Expression	N-person		N-fold	
	JAFFE	CK	JAFFE	CK
anger	80	90.63	96.67	96.67
disgust	76.67	84.38	96.67	100
fear	63.33	75	100	100
happy	80	100	100	100
neutral	63.33	----	100	----
sad	66.67	87.5	100	100
surprise	96.67	100	100	100
average	76.67	89.58	99.05	99.34

TABLE 2. The recognition rate (%) for different feature descriptors in JAFFE and CK.

Descriptors	N-person		N-fold	
	JAFFE	CK	JAFFE	CK
LBP	66.19	77.08	96.19	98.34
CS-LBP	62.38	81.25	95.71	98.51
ULBP[35]	62.38	71.35	95.71	98.34
ELBP[36]	66.67	79.17	97.14	98.34
RILBP[37]	64.29	78.13	96.19	98.34
HOG	68.10	75.76	95.71	98.10
CS-LOP	65.24	78.65	97.62	99.00
CS-LOP$_{\text{FMGM}}$	62.39	78.65	95.71	99.00
CS-LOP$_{\text{Gabor}}$	74.29	79.17	96.67	99.17
Proposed	76.67	89.58	99.05	99.34

operator, we set block = 7 on JAFFE and CK both in N-person and N-fold. For CS-LOP$_{\text{Gabor}}$, we set block=6 on JAFFE and block = 4 on CK both in N-person and N-fold strategies.

3) FEATURE FUSION

Section III proposes three different features: CS-LOP, CS-LOP$_{\text{FMGM}}$ and CS-LOP$_{\text{Gabor}}$. There are also various fusion methods for these three approaches. As shown in Figure 10 (a) and (b), the experimental results indicate that CS-LOP and CS-LOP$_{\text{FMGM}}$ are complementary for different facial expressions in JAFFE.

As can be seen from Figure 10 (a) and (b), CS-LOP operator performs better than CS-LOP$_{\text{FMGM}}$ in some facial expressions: happy, neutral, surprise. However, CS-LOP$_{\text{FMGM}}$ achieves better results in other expressions (N-person: anger, disgust, fear; N-fold: anger, sad). For integrating the advantages of above methods, feature-level-fusion is adopted in this paper. The results of fusion indicate that overall recognition rate is higher than the individual method. Similarly, fusion result on CK is same as JAFFE. In this paper, the weight ratio between CS-LOP and CS-LOP$_{\text{FMGM}}$ is set to 3:1. After that,

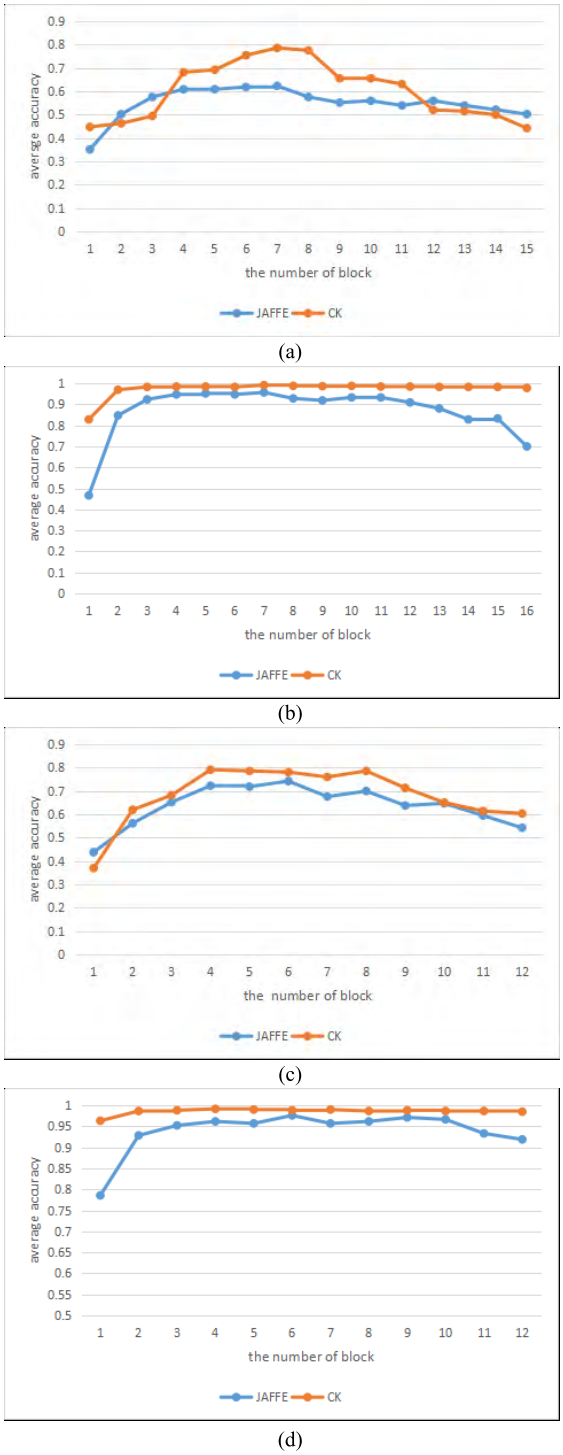


FIGURE 9. (a) The accuracy of CS-LOP$_{\text{FMGM}}$ for the number of block (N-person). (b) The accuracy of CS-LOP$_{\text{FMGM}}$ for the number of block (N-fold). (c) The accuracy of CS-LOP$_{\text{Gabor}}$ for the number of block (N-person). (d) The accuracy of CS-LOP$_{\text{Gabor}}$ for the number of block (N-fold).

we concatenate fused features with CS-LOP$_{\text{Gabor}}$ to obtain the final features and then feed into SVM.

D. COMPARISON OF EXPERIMENTAL RESULTS

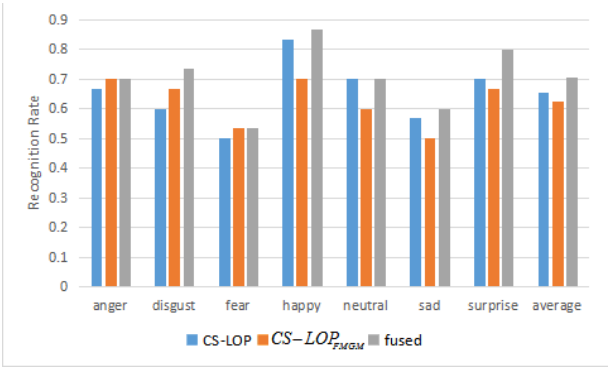
To illustrate the performance of proposed methods, feature fusion is executed on JAFFE and CK to obtain the recognition

TABLE 3. Accuracy (%) comparison of different facial expression recognition methods (N-person).

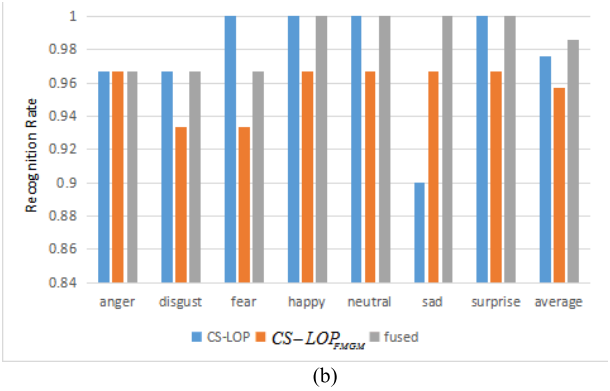
Method	JAFFE	Method	CK
NEDP_s[38]	67.97	SLFDA[39]	77.92
LDTP[31]	67.61	Hybrid(CLM based)[40]	78.00
Collaborative Representation [42]	71.58	CS-LSBP+HOAG[43]	81.43
Image pyramid + decision tree[41]	74.76	Image pyramid + decision tree[41]	83.10
Proposed method	76.67	Proposed method	89.58

TABLE 4. Accuracy (%) comparison of different facial expression recognition methods (N-fold).

Method	JAFFE	Method	CK
Image pyramid + decision tree[41]	91.43	Image pyramid + decision tree[41]	97.84
LDTP[31]	93.2	K-ELBP[49]	92.3
Local curvelet transform[44]	94.65	Local curvelet transform[44]	95.17
Feature co-clustering[45]	96.25	Haar + LR[46]	98.73
Deep CNN[46]	97.71	Pairwise feature selection[48]	98.1
LBP + LPQ + Gabor[47]	98.57	Gabor+ PCA+ SRC[49]	99.12
Proposed method	99.05	Proposed method	99.34



(a)



(b)

FIGURE 10. (a) The recognition rate of different expressions in JAFFE (N-person). (b) The recognition rate of different expressions in JAFFE (N-fold).

rate of each expression in N-person and N-fold schemes. Table 1 shows the average accuracy for each expression.

From the overall recognition results in Table 1, since the subject in the training set was excluded out of testing set in the person-independent, which contribute to the recognition rate is relatively low compared to person-dependent. Whereas, some expressions such as surprise, happy always be recognized well.

In order to demonstrate the effectiveness of the proposed feature extraction method for facial expression recognition in this paper, we compare the proposed method with several common local feature extraction algorithms. The comparison results are shown in Table 2.

As can be seen from Table 2, the proposed CS-LOP$_{\text{Gabor}}$ in this paper improves recognition rate nearly 9% than other descriptors in person-independent on JAFFE, which indicates the availability of proposed CS-LOP$_{\text{Gabor}}$ feature. In addition, the recognition rates of almost all descriptors have no significant difference in person-dependent scheme, which indicates the reasonability of those feature descriptors proposed by previous scholars. Most important of all, experimental results show that CS-LOP in this paper performs better than present descriptors on CK and our fused method outperforms all other methods both in person-independent and person-dependent schemes.

Additionally, to further illustrate the reasonability of overall algorithm, our proposed method is compared with other literatures in both person-independent and person-dependent strategies and comparison results are shown in Table 3 and Table 4. As can be seen from following tables, proposed method is not only performs better than those traditional feature descriptors but also superior to some methods using deep neural networks. In conclusion, the proposed CS-LOP operator and its fused methods show their superiority and excellence for facial expression recognition.

V. CONCLUSION

In this paper, we propose a new local feature descriptor: CS-LOP for facial expression recognition. It not only

compares the gray value between central pixels and neighboring pixels in all eight directions but also compares the gray value of four pairs of center-symmetric pixels. CS-LOP reduces feature vector dimension on the basis of LBP and avoids selecting the optimal threshold. Compared with previous descriptors, such as: LBP and CS-LBP using two-bit coding, one more bit code value added by CS-LOP operator enables more accurate and detailed features among pixels from unobvious regions can be captured and represented. Afterwards, we use CS-LOP operator to extract features from the preprocessed facial image, feature map of gradient magnitude and feature map of Gabor to obtain the corresponding CS-LOP, CS-LOP$_{\text{FMGM}}$ and CS-LOP$_{\text{Gabor}}$ features, and then fuse the three features for facial expression recognition. Additionally, this paper also analyzes the performance of the proposed algorithm under person-dependent and person-independent strategies, and experimental results on JAFFE and CK datasets show that the CS-LOP achieves better performance than the operators proposed by other literatures and improves the overall recognition accuracy compared with some state-of-the-art methods. In the future, we will study on cross-datasets experiments and how to improve the recognition rate of micro-expressions.

REFERENCES

[1] D. Kim, *Automated Face Analysis: Emerging Technologies and Research*. Pennsylvania, PA, USA: IGI Global, 2009.

[2] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, and G. Fortino, "Facial expression recognition utilizing local direction-based robust features and deep belief network," *IEEE Access*, vol. 5, pp. 4525–4536, 2017.

[3] F. Ren and Z. Huang, "Automatic facial expression learning method based on humanoid robot XIN-REN," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 6, pp. 810–821, Dec. 2016.

[4] P. Ekman and W. V. Friesen, *Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions*. California, CA, USA: Ishk, 2003.

[5] F. De la Torre and J. F. Cohn, "Facial expression analysis," *Visual Analysis of Humans*. London, U.K.: Springer, 2011, pp. 377–409.

[6] F. Ren and Y. Wu, "Predicting user-topic opinions in twitter with social and topical context," *IEEE Trans. Affective Comput.*, vol. 4, no. 4, pp. 412–424, Oct./Dec. 2013.

[7] R. Gross and V. Brajovic, "An image preprocessing algorithm for illumination invariant face recognition," in *Proc. Int. Conf. Audio-Video-Based Biometric Person Authentication*. Berlin, Germany: Springer, 2003, pp. 10–18.

[8] S. Abe, "Feature selection and extraction," in *Support Vector Machines for Pattern Classification*. London, U.K.: Springer, 2010, pp. 331–341.

[9] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, 2009.

[10] B. Fernando, E. Fromont, D. Muselet, and M. Sebban, "Discriminative feature fusion for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 3434–3441.

[11] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.

[12] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.

[13] T. Ojala, M. Pietikäinen, and T. Maenpaa, "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.

[14] 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.

[15] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based HoG features," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. Workshops (FG)*, Mar. 2011, pp. 884–888.

[16] R. F. Walker, P. T. Jackway, and D. Longstaff, "Genetic algorithm optimization of adaptive multi-scale GLCM features," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 17, no. 1, pp. 17–39, 2003.

[17] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Comput. Vis. Image Understand.*, vol. 113, no. 3, pp. 345–352, Mar. 2009.

[18] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 699–714, May 2005.

[19] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1805–1812.

[20] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proc. ACM Int. Conf. Multimodal Interact.*, 2015, pp. 435–442.

[21] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–10.

[22] Z. Huang and F. Ren, "Facial expression recognition based on multi-regional D–S evidences theory fusion," *IEEJ Trans. Elect. Electron. Eng.*, vol. 12, no. 2, pp. 251–261, 2017.

[23] Y. Guo, G. Zhao, and M. Pietikäinen, "Discriminative features for texture description," *Pattern Recognit.*, vol. 45, no. 10, pp. 3834–3843, Oct. 2012.

[24] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognit.*, vol. 42, no. 3, pp. 425–436, 2009.

[25] N.-S. Vu and A. Caplier, "Face recognition with patterns of oriented edge magnitudes," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 313–326.

[26] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 200–205.

[27] G. Muhammad, M. Alsulaiman, S. U. Amin, A. Ghoneim, and M. F. Alhamid, "A facial-expression monitoring system for improved healthcare in smart cities," *IEEE Access*, vol. 5, pp. 10871–10881, 2017.

[28] A. H. Gunatilaka and B. A. Baertlein, "Feature-level and decision-level fusion of noncoincidently sampled sensors for land mine detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 577–589, Jun. 2001.

[29] F. Cheng, J. Yu, and H. Xiong, "Facial expression recognition in JAFFE dataset based on Gaussian process classification," *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1685–1690, Oct. 2010.

[30] T. Kanade, Y. Tian, and J. F. Cohn, "Comprehensive database for facial expression analysis," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Mar. 2000, pp. 46–53.

[31] B. Ryu, A. R. Rivera, J. Kim, and O. Chae, "Local directional ternary pattern for facial expression recognition," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 6006–6018, Dec. 2017.

[32] G. Rätsch, T. Onoda, and K.-R. Müller, "Soft margins for AdaBoost," *Mach. Learn.*, vol. 42, no. 3, pp. 287–320, 2001.

[33] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, no. 6, pp. 1233–1258, Dec. 1987.

[34] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local SVM approach," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3, Aug. 2004, pp. 32–36.

[35] A. Rölle *et al.*, "Effects of human cytomegalovirus infection on ligands for the activating NKG2D receptor of NK cells: Up-regulation of UL16-binding protein (ULBP)1 and ULBP2 is counteracted by the viral UL16 protein," *J. Immunol.*, vol. 171, no. 2, pp. 902–908, 2003.

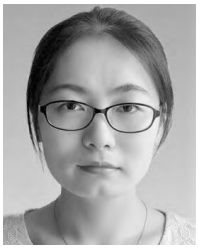
[36] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "On matching sketches with digital face images," in *Proc. 4th IEEE Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2010, pp. 1–7.

[37] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognit.*, vol. 43, no. 3, pp. 706–719, 2010.

[38] M. T. B. Iqbal, M. A.-A. Wadud, B. Ryu, F. Makhmudkhujaev, and O. Chae, "Facial expression recognition with neighborhood-aware edge directional pattern (NEDP)," *IEEE Trans. Affective Comput.*, to be published.

[39] Z. Wang, Q. Ruan, and G. An, "Facial expression recognition using sparse local Fisher discriminant analysis," *Neurocomputing*, vol. 174, pp. 756–766, Jan. 2016.
[40] S. W. Chew, P. J. Lucey, S. Lucey, J. Saragih, J. Cohn, and S. Sridharan, "Person-independent facial expression detection using constrained local models," in *Proc. FG Facial Expression Recognit. Anal. Challenge*, 2011, pp. 915–920.
[41] A. M. Ashir and A. Eleyan, "Facial expression recognition based on image pyramid and single-branch decision tree," *Signal, Image Video Process.*, vol. 11, no. 6, pp. 1017–1024, 2017.
[42] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2950–2959.
[43] H. U. Min, T. Wendi, W. Xiaohua, X. Liangfeng, and Y. Juan, "Facial expression recognition based on local texture and shape features," *J. Electron. Inf. Technol.*, vol. 40, no. 6, pp. 1338–1344, 2018.
[44] A. Uçar, Y. Demir, and C. Güzeliş, "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering," *Neural Comput. Appl.*, vol. 27, no. 1, pp. 131–142, 2016.
[45] S. Khan, L. Chen, X. Zhe, and H. Yan, "Feature selection based on co-clustering for effective facial expression recognition," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, vol. 1, Jun. 2016, pp. 48–53.
[46] L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang, and T. Zhang, "Deep convolutional neural network for facial expression recognition using facial parts," in *Proc. IEEE 15th Intl Conf Dependable, Autonomic Secure Comput., 15th Intl Conf Pervasive Intell. Comput., 3rd Intl Conf Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech)*, Nov. 2017, pp. 1318–1321.
[47] B. Zhang, G. Liu, and G. Xie, "Facial expression recognition using LBP and LPQ based on Gabor wavelet transform," in *Proc. 2nd IEEE Int. Conf. Comput. Commun. (ICCC)*, Oct. 2016, pp. 365–369.
[48] M. J. Cossetin, J. C. Nievola, and A. L. Koerich, "Facial expression recognition using a pairwise feature selection and classification approach," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 5149–5155.
[49] X. Lu L. Kong, M. Liu, and X. Zhang, "Facial expression recognition based on Gabor feature and SRC," in *Chinese Conference on Biometric Recognition*. Cham, Switzerland: Springer, 2015, pp. 416–422.
[50] M. Guo, X. Hou, Y. Ma, and X. Wu, "Facial expression recognition using ELBP based on covariance matrix transform in KLT," *Multimedia Tools Appl.*, vol. 76, no. 2, pp. 2995–3010, 2017.
[51] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.
[52] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, 2015.

MIN HU received the M.S. degree in industrial automation from Anhui University, China, in 1994, and the Ph.D. degree in computer science from the Hefei University of Technology, Hefei, China, in 2004, where she is currently a Professor with the School of Computer and Information. Her research interests include digital image processing, artificial intelligence, and data mining.



YAQIN ZHENG is currently pursuing the master's degree with the Hefei University of Technology. Her research interests are image processing and facial expression recognition.

CHUNJIAN YANG is currently pursuing the master's degree with the Hefei University of Technology. His research interests are image processing and pattern recognition.

XIAOHUA WANG received the Ph.D. degree in computer science from the Hefei Institute of Physical Science, Chinese Academy of Sciences, China, in 2005. She is currently an Associate Professor with the School of Computer and Information, Hefei University of Technology. Her research interests include affective computing, artificial intelligence, and visual pattern recognition.

LEI HE received the Ph.D. degree, in 2005. She is currently an Associate Professor with the Hefei University of Technology. Her research interest is computer application.

FUJI REN received the Ph.D. degree from the Faculty of Engineering, Hokkaido University, Sapporo, Japan, in 1991. He is currently a Professor with the Department of Information Science and Intelligent Systems, University of Tokushima, Tokushima, Japan. His current research interests include natural language processing, machine translation, artificial intelligence, language understanding and communication, robust methods for dialogue understanding, and affective information processing and knowledge engineering.