IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 29, 2019, accepted August 14, 2019, date of publication August 22, 2019, date of current version September 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936976

Facial Expression Recognition Based on Fusion Features of Center-Symmetric Local Signal Magnitude Pattern

MIN HU[1,2], CHUNJIAN YANG[1,2], YAQIN ZHENG[1,2], XIAOHUA WANG[1,2], LEI HE[3], AND FUJI REN[2,4], (Senior Member, IEEE)

[1]Key Laboratory of Knowledge Engineering with Big Data, Ministry of Education, Hefei University of Technology, Hefei 230009, China
[2]School of Computer and Information, Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine, Hefei University of Technology, Hefei 230009, China
[3]School of Mathematics, Hefei University of Technology, Hefei 230009, China
[4]Graduate School of Advanced Technology and Science, University of Tokushima, Tokushima 7708502, Japan

Corresponding authors: Xiaohua Wang (xh_wang@hfut.edu.cn) and Lei He (lei_he80@163.com)

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 NSFC-Shenzhen Joint Foundation under Grant U1613217, and in part by the Fundamental Research Funds for the Central Universities of China under Grant ACAIM190101.

ABSTRACT Local feature descriptors play a fundamental and important role in facial expression recognition. This paper presents a new descriptor, Center-Symmetric Local Signal Magnitude Pattern (CS-LSMP), which is used for extracting texture features from facial images. CS-LSMP operator takes signal and magnitude information of local regions into account compared to conventional LBP-based operators. Additionally, due to the limitation of single feature extraction method and in order to make full advantages of different features, this paper employs CS-LSMP operator to extract features from Orientational Magnitude Feature Maps (OMFMs), Positive-and-Negative Magnitude Feature Maps (PNMFMs), Gabor Feature Maps (GFMs) and facial patches (eyebrows-eyes, mouths) for obtaining fused features. Unlike HOG, which only retains horizontal and vertical magnitudes, our work generates Orientational Magnitude Feature Maps (OMFMs) by expanding multi-orientations. This paper build two distinct feature maps by dividing local magnitudes into two groups, i.e., positive and negative magnitude feature maps. The generated Gabor Feature Maps (GFMs) are also grouped to reduce the computational complexity. Experiments on the JAFFE and CK+ facial expression datasets showed that the proposed framework achieved significant improvement and outperformed some state-of-the-art methods.

INDEX TERMS Facial expression recognition, center-symmetric local signal magnitude pattern, local representation, feature fusion.

I. INTRODUCTION

Facial expressions are important aspect of behavior and nonverbal communication for people to express their inner feelings. Indeed, expression can be represented by appearance changes on the face and plays an essential role in human interactions. Recently, as one of the most representative applications of affective computing, facial expression recognition (FER) [1], [2], [48] has received considerable attention in various fields, such as human-computer interaction, medical treatment and intelligent control [3], [4].

In terms of features, FER system used to be divided into two categories: geometric-feature-based [5] and appearance-feature-based methods [6]. Geometric-feature-based method extracts shape information and locations of main facial components and encodes region of interest (ROI), like eyes, mouths, noses, etc. Geometric-feature-based describes facial images using a few features but the recognition result is not satisfying since it is difficult to detect under appearance changes. Whereas, appearance-feature-based method extracts facial texture caused by expression changes and represents facial images by using image filters which are applied on the holistic or local regions. In this category, there are some holistic methods: Principal Component Analysis (PCA) [7], Information Discriminant Analysis (IDA) [8], Linear Discriminant Analysis (LDA) [9], and local approaches, such as Scale-Invariant Feature Transform (SIFT) [10], Local Binary Pattern (LBP) [11] and its variants: Center-Symmetric Local Binary Pattern (CS-LBP) [12], Local Ternary Pattern (LTP) [13], Local Directional Ternary

The associate editor coordinating the review of this article and approving it for publication was Gustavo Olague.

Pattern (LDTP) [14], Local Phase Quantization (LPQ) [15], Local Gabor Binary Pattern (LGBP) [16], Completed Local Mapped Pattern (CLMP) [40], Completed Local Binary Pattern (CLBP) [41] and so on. However, in many current pattern recognition systems and some complex tasks [38], [39], the drawbacks of single feature has revealed gradually. It is often the case that no single class of feature is rich enough to capture all of the dominant information. Finding and combining complementary feature sets has thus become an active research topic and also makes great achievements [17], [18], [38]. In recent years, with the wide spread of neural networks, deep learning as an end-to-end tools has been applied in FER [19]–[21], but it comes with some evident drawbacks: large demand for training samples, easy to overfit for neural networks, the poor generalization ability of models, high requirements of hardware etc. Reference [48] introduces almost relevant work of FER from all aspect, including facial expression datasets, conventional feature extraction methods and deep neural networks etc.

Ojala *et al.* [11] first introduced LBP method in 2002 and it is a valid descriptor for extracting texture features and classification, and also has some excellent properties such as rotation invariance, robustness against monotonic gray level transformation and easy to encode. Whereas LBP simply compares the numerical values between peripheral pixels and central pixel, and does not focus on the specific difference of peripheral pixels, which result in losing some details in the image. Since Ojala's work, a vast number of LBP variants has been proposed and continued to be developed. CS-LBP, as one of the representative variants of LBP and SIFT descriptor, reduces feature dimensionality and time complexity greatly. But CS-LBP neglects central pixels and must choose the appropriate threshold from experiments. In addition, HOG [22] maintains an attractive invariance in both geometric and optical deformation of images. It was used in pedestrian detection widely. However, HOG also fails to take the central points into account in the calculation process of requiring the horizontal gradient and vertical gradient, and all the cells are the same size, which makes HOG lacking variability and flexibility. On the other hand, for the two algorithms of CLMP [40] and CLBP [41], the local grayscale differences are decomposed into three complementary components: the signal (S), the magnitude (M) and the center (C), which ignores the internal relationship among the three components. The threshold C_I in CLBP, the threshold τ_I and curve slope β in CLMP are hard to determine.

In order to solve the aforementioned problems, this paper proposes modified versions, Center-Symmetric Local Signal Magnitude Pattern (CS-LSMP), Orientational Magnitude Feature Maps (OMFMs) and Positive-and-Negative Magnitude Feature Maps (PNMFMs). The CS-LSMP represents the local magnitude differences between the neighboring pixels and central pixel, and also combines their signal information. Then the local differences are compared with the average magnitude in order to reflect the distinction among pixels. The OMFMs are collection of feature maps from different orientations, increasing gradient magnitude of diagonal orientations and also pay attention to the central pixels. The PNMFMs are collection of positive magnitude feature maps and negative magnitude feature maps, which reserve more local magnitude information. Moreover, this paper also employs CS-LSMP operator to extract texture features from dominant local facial patches (eyebrow-eye, mouth). After that, all above features are used to achieve feature-level fusion and histograms are cascaded to form ultimate features.

The main contributions of our work are summarized as follows:

(1) We present a new feature map: Orientational Magnitude Feature Maps (OMFMs). Multi-Orientational Magnitude Pattern (OMP) with different local radius consist of various OMFMs, which increases the flexibility of feature scale.
(2) We introduce a formal definition of Positive-and-Negative Magnitude Feature Maps (PNMFMs), which reserve more local magnitude information in images.
(3) We construct a grouping method of GFMs. It not only reduces the feature dimension and time complexity effectively but also improves the recognition rate.
(4) CS-LSMP operator records the gray value difference between neighboring and central pixels, and the difference takes signal and magnitude information into account. Four-bit encoded CS-LSMP makes the code values more diverse and distinguishable.

The reminder of this paper is structured as follows: in section II we introduce the related work about LBP, CS-LBP, HOG and give a brief introduction about LBP-based texture descriptors. Then, section III describes the feature extraction method based on OMFMs, PNMFMs, GFMs, CS-LSMP operator and feature fusion. We carry out experiments of the proposed method and discuss results in section 4. Lastly, we make a conclusion in section 5.

II. RELATED WORK

A. LOCAL BINARY PATTERN

Ojala *et al.* [11] introduced Local Binary Pattern (LBP) as a means of summarizing local gray-level structures. The LBP operator was proposed to encode the pixel-wise information. LBP takes a local neighborhood around each pixel and then thresholds the pixels of the neighborhood at the value of the central pixel. Then LBP uses the resulting binary-valued image patch as a local image descriptor. It was originally defined for 3×3 neighborhoods, giving 8-bit integer LBP codes based on the eight pixels around the central one. And the LBP code result is given in Eq. (1) and (2).

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

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

where p_c is the gray value of the central pixel, $p_i (i = 0, 1, \ldots, N-1)$ denotes 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

The Center-Symmetric Local Binary Pattern (CS-LBP), as a variant of LBP and SIFT descriptor, inheriting the desirable properties of both texture features and gradient-based features. A comparison of LBP yields a longer histogram and has 256 different binary patterns, CS-LBP only generates 16 binary patterns. Different from comparing each peripheral pixels with the central one, CS-LBP operator encodes a facial image as a binary number by thresholding the gray-level differences of pairs of opposite pixels with respect to the central pixel. Therefore, the definition of CS-LBP is given in Eq. (3) and (4).

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

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

where $p_i (i = 0, 1, \ldots, (N/2)-1)$ and $p_{i+(N/2)}$ correspond to the gray level values of peripheral pixels. T is a threshold that needs to be specified by experiments. It should be noted that CS-LBP is related to gradient operator closely as it considers gray level differences between pairs of opposite pixels in a local neighborhood.

CS-LBP descriptor not only lower feature dimensionality but also has a better robustness. However, the CS-LBP descriptor only compares the gray value of center symmetric pixels but ignores the central pixel and the texture information is not represented entirely.

The two operators: LBP and CS-LBP describe each pixel by the relative gray values of its neighboring pixels, see Figure 1 for an illustration with $N = 8$ and $R = 1$.

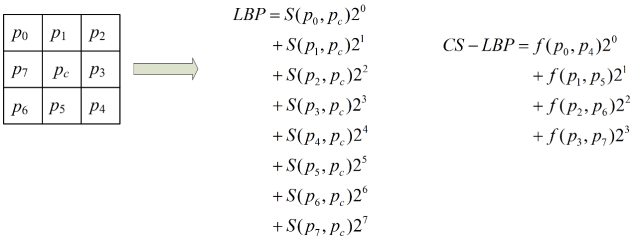

FIGURE 1. Calculation of LBP and CS-LBP operators for a neighborhood of 8 pixels.

C. HISTOGRAM OF ORIENTED GRADIENT

Histogram of Oriented Gradient (HOG) was first developed by Dalal and Triggs in 2005 as a robust descriptor for pedestrian detection. HOG extracts features related to the distribution of target edges in local regions as a means of representing target shape. The gist of the HOG algorithm is to describe the local shape information of images by analyzing the gradient magnitude distribution from different gradient orientations statistically. HOG consists of an intelligent grouping of gradient information (cells and blocks), as well as well-engineered histograms of gradient orientations. HOG features are descriptions of orientations histograms. Taking Figure 1 as an example, convolving the first-order differential template $[-1, 0, 1]$ with image, the horizontal gradient $G_x(c)$ and the vertical gradient $G_y(c)$ of the central pixel c can be calculated by the Eq. (5).

$$\begin{cases} G_x(c) = p_3 - p_7 \\ G_y(c) = p_5 - p_1 \end{cases} \tag{5}$$

And the gradient magnitude M(c) and gradient orientation $\theta(c)$ of the central pixel can be obtained from Eq. (6) and (7).

$$M(c) = \sqrt{G_x(c)^2 + G_y(c)^2} \tag{6}$$

$$\theta(c) = arctan(G_y(c)/G_x(c)) \tag{7}$$

where $p_1, p_3\ p_5, p_7$ are four peripheral pixels. The gradient magnitudes of the pixels with the same gradient orientation are accumulated to form a gradient histogram.

D. BRIEF REVIEW OF LBP VARIANTS

Among the local approaches, LBP-based methods emerged as one of the most prominent texture descriptors, because of their outstanding performance. The pioneering LBP and its success in various computer vision applications has inspired the development of effective LBP variants. Due to its flexibility, the LBP method can be easily modified to make it more suitable to meet the requirements of different applications. Since Ojala's work, a vast number of LBP variants has been proposed and continue to be developed. Guo and Zhang [41] developed a completed modeling of LBP to produce CLBP which combines magnitudes of local differences as well as their signs. The method consists in converting the gray level of the central pixel into a binary code, namely CLBP-Center (CLBP_C), using a global threshold. The image local differences yield two complementary components: sign and magnitude. Two operators, namely CLBP-Sign (CLBP_S which is therefore the same as the original LBP) and CLBP-Magnitude (CLBP_M), are then defined to code the sign and magnitude of the image local differences, respectively. Vieira *et al.* [40] proposed CLMP for improving the classification of rotated images. The local grayscale differences in the CLMP formulation are decomposed into two complementary components: the signal (S) and the magnitude (M), generating two operators CLMP_S and CLMP_M. The introduction of curve slope increase the uncertainty. Huang *et al.* [13] proposed LTP for face recognition. LTP, which extends original LBP to 3-valued codes using a threshold, is introduced to reduce noise sensitivity of the original LBP. El merabet and Ruichek proposed Local Concave-and-Convex Micro-structure Pattern (LCCMSP) [42] which is based on concave-and-convex characteristics of 3×3 grayscale image patches. Ryu *et al.* [14] proposed LDTP for texture classification. The LDTP operator, basically, encodes at the same time the

information related to image contrast and the directional pattern features based on local derivative variations, etc.

III. PROPOSED APPROACH

In this section, our approach will be described in detail, including features extraction and features fusion.

A. ORIENTATIONAL MAGNITUDE FEATURE MAPS

Orientational Magnitude Feature Maps (OMFMs) are collection of Orientational Magnitude Pattern (OMP) from different orientations. The HOG in section II generates the gradient magnitude but neglects central pixels, and simply calculates the magnitude of horizontal and vertical orientations in a cell of the same size. The proposed OMP not only focuses on the gray value of central point but also increases the multi-orientation magnitudes according to the radius of neighborhood. As the radius of the neighborhood increases, the number of involved neighboring pixels increases, so does the number of OMP. All of the OMP from various orientations are cascaded to form the OMFMs. With the introduction of OMFMs, the local information of images can be better represented. And the OMP can be defined in Eq. (8).

$$M_{i,N,R}(c) = \sqrt{\frac{(p_i - p_c)^2 + (p_{i+(N/2)} - p_c)^2}{2}} \tag{8}$$

where $M_{i,N,R}(c)$ is defined as the i_{th} orientation magnitude of the given N and R of the central point c. Figure 2 shows OMPs feature extracted by one facial image with $R = 1$ of four different orientations. (from (a) to (d) are visualizations of the OMP in four orientations: horizontal, vertical and two diagonal orientations). As shown in Figure 2, OMP depicts the region of interest of the face. According to Eq. (8), $(N/2)$ OMPs are obtained and cascaded to form the final OMFMs.

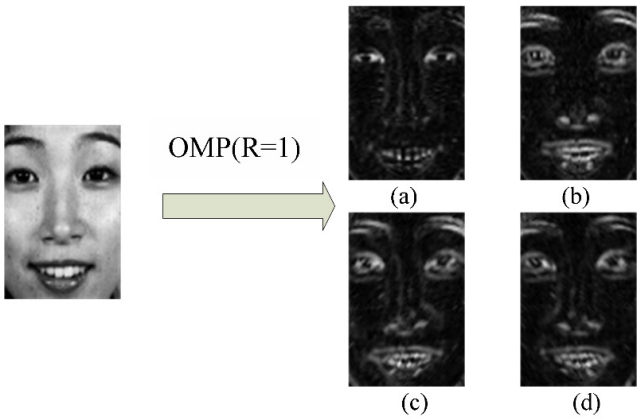

FIGURE 2. The schematic diagram of OMP (R = 1).

B. GROUPING OF GABOR FEATURE MAPS

Gabor wavelet [23] is a well-known descriptor representing texture information of an image and is very similar to the stimulus-response of simple cells in the human visual system. Gabor feature is highly capable of describing textures used in different research, such as identity recognition, image classification and facial expression recognition. Gabor wavelet has better robustness and adapt to the changes of illumination well. It also provides better selectivity for multi-orientation and multi-scale. Considering the advantages of the Gabor filters, we employ the Gabor filters to extract features from the input face images. The kernel function captures various spatial information of frequency, position, orientation from an image and represent subtle local transformation effectively. For a Gabor filter with five scales and eight orientations, $\nu \in \{0, 1, 2, 3, 4\}$, $\mu \in \{0, 1, 2, \ldots, 7\}$, each facial image generates 40 Gabor feature maps (GFMS). In this paper, we separated four consecutive sheets in order, and obtained separated ten groups of feature maps, as shown in Figure 3, and then a group of feature maps was added for each experiment.

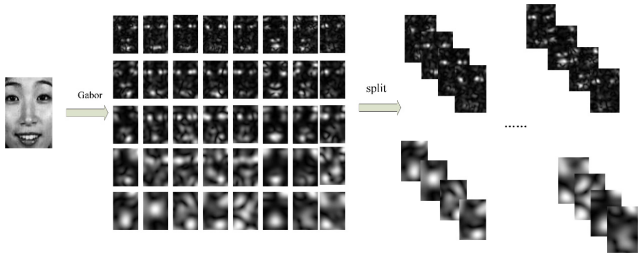

FIGURE 3. The operation of GFMs for one face image.

C. POSITIVE- AND-NEGATIVE MAGNITUDE FEATURE MAPS

Referring to Figure 1, given a central pixel p_c and its neighboring pixels $p_i (i = 0, 1, \ldots, N-1)$, we can simply calculate the local magnitude $\boldsymbol{m}_i$ between p_c and p_i in Eq. (9).

$$m_i = p_i - p_c \tag{9}$$

The local magnitude vector $[m_0, m_1, \ldots, m_{N-1}]$ characterizes the image local structure at p_c. Moreover, m_i can be further divided into two parts, greater than 0 (positive) and less than 0 (negative). This paper introduces two weight matrices. When the m_i is positive, we set the weight value of corresponding position to 1 in the weight matrix and the left is set to 0. When the m_i is negative, we set the weight value to -1 and the left is set to 0 in Eq. (10)-(12), where W is weight matrix.

$$w = \begin{bmatrix} a_0 & a_1 & a_2 \\ a_7 & a_c & a_3 \\ a_6 & a_5 & a_4 \end{bmatrix} \tag{10}$$

$$a_{i,positve} = \begin{cases} 1, & m_i \geq 0 \\ 0, & m_i < 0 \end{cases} \tag{11}$$

$$a_{i,negatve} = \begin{cases} -1, & m_i \leq 0 \\ 0, & m_i > 0 \end{cases} \tag{12}$$

Then we obtain positive weight matrix and negative weight matrix. Figure 4 shows an example. Figure. 4(a) is the original 3*3 local structure with central pixel being 25.

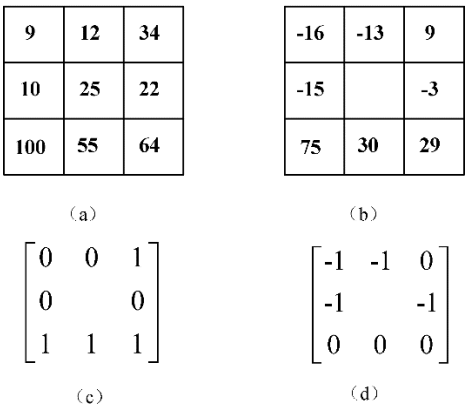

FIGURE 4. (a) 3 × 3 sample block; (b) the local magnitude; (c) positive weight matrix; (d) negative weight matrix.

The magnitude vector (see Figure. 4(b)) is [−16, −13, 9, −3, 29, 30, 75, −15]. After Eq. (11) and (12), the positive weight matrix (see Figure.4(c)) and negative weight matrix (see Figure. 4(d)) are obtained. Then two weight matrices are convolved with magnitude of local neighborhood respectively in order to obtain Positive Magnitude Pattern (PMP) and Negative Magnitude Pattern (NMP) according to Eq. (13) and (14), where * represents convolution operation. PMP and NMP together form positive-and-negative magnitude feature maps (PNMFMs).

$$\varphi_{positive}(c) = \frac{m_i * W_{positive}}{\sum a_{i,positive}} \tag{13}$$

$$\delta_{negative}(c) = -\frac{m_i * W_{negative}}{\sum a_{i,negative}} \tag{14}$$

D. CENTER-SYMMETRIC LOCAL SIGNAL MAGNITUDE PATTERN

The Center-Symmetric Local Signal Magnitude Pattern (CS-LSMP) descriptor is proposed on the basis of LBP and CS-LBP for extracting image texture features. LBP only compares the gray values between the peripheral pixels and central pixels: the former is smaller than the latter one, then code value is assigned to 0; otherwise code value is 1. However, the LBP operator ignores the gray value differences of local peripheral pixels, resulting in losing local texture details. In comparison with the LBP, CS-LBP halves the number of pixel pairs for the same number of neighbors, thus it significantly reduces the dimensionality from 256 to 16. However, in CS-LBP mechanism, it is quite difficult to choose an appropriate threshold (T) in general cases and it does not take central points into consideration.

The binary expression of the LBP makes this method very popular because it is very simple. On the other hand, the binary formulation, although simple, also discard information on local differences in the presence of noise. To deal with these drawbacks and extract robust features, the CS-LSMP operator not only captures the grayscale value differences between the neighboring and central pixels but also adds the signal information, and then compares the magnitude differences of opposite pixels with average magnitude, which makes the extracted texture features among pixels are more meticulous and easy to distinguish. Thus, the resulting value of CS-LSMP are shown in Eq. (15) and (16).

$$CS - LSMP_{N,R}(c) = \sum_{i=0}^{(N/2)-1} \mu(S(m_i, m_c), S(m_{i+(N/2)}, m_c))2^i \tag{15}$$

$$\mu(a, b) = a * 2^0 + b * 2^1 \tag{16}$$

where N, R, m_i and $S(x, y)$ are mentioned before. This paper employs $\mu(a, b)$ for the first time here. Different from LBP and CS-LBP operators encoding the result as 0 and 1, $\mu(a, b)$ maps the coded value into the [0, 3] interval, which makes the results more diverse and distinguishable. And the m_c is the average of m_i, that is $m_c = \frac{1}{N}\sum_{i=0}^{N-1} m_i$. It is known that the average gray level is widely accepted statistical parameter. Taking $N = 8$ as an example, the calculation process of CS-LSMP is shown in Figure 5.

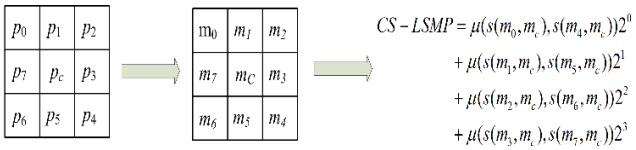

FIGURE 5. The CS-LSMP operator with 8 neighbor pixels.

IV. EXPERIMENTS AND DISCUSSION

In this section, we carried out experiments for facial expression recognition to validate the efficiency of the proposed frameworks and compared with different 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 are conducted on two public available famous facial expression datasets: The Japanese Female Facial Expression (JAFFE) and the Extended Cohn-Kanade (CK+) [24]. The statistics of two datasets are summarized. JAFFE dataset consists of 213 images from 10 Japanese female subjects. Every subject has 3 or 4 examples of all the six basic expressions and also has a sample of neutral expression. The CK+ dataset comprises 593 image sequences (from neutral to apex) of 123 subjects who were instructed to perform a series of 23 facial displays but only 327 sequences are assigned to seven labels. In these sequences, the expression label of many images are distributed unevenly and some labels are not be given. In order to compare with other methods, all of the contempt expressions were removed. One example of expressions about the two datasets is shown in Figure 6 (from left to right, the label of expressions are: anger, disgust, fear, happy, neutral, sad, surprise).

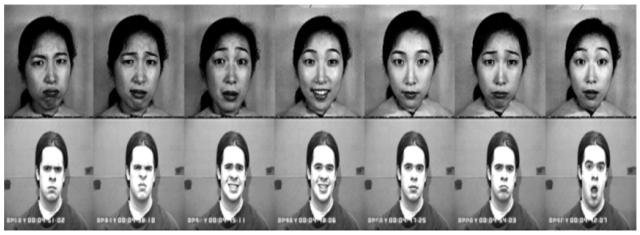

FIGURE 6. Facial expression samples in JAFFE and CK+ datasets.

Our experiments selected three images of each subject in JAFFE dataset with seven basic expressions and three peak frames of each subject in CK+ dataset with six basic expressions (excluding neutral expression). This paper used person-independent cross-validation scheme, which is N-person to evaluate the performance of proposed algorithm [14].

In the N-person cross-validation, all image samples were partitioned into N groups according to subject and excluded one subject out of training set as the testing set, which ensured the person-independence itself. Then this scheme was repeated N times and average results were taken as the final average recognition. Recognition rate is calculated by the Eq. (17).

$$Recognition\ rate = \frac{true\ samples}{all\ samples} \tag{17}$$

B. STEPS OF EXPERIMENT

1) PREPROCESS

In this experiments, the image samples were preprocessed through the following steps: ① detecting the position of human eyes in the image with Haar-like and AdaBoost algorithm, and using the coordinates of two eyes for geometric transformation to eliminate the effect of posture. ② detecting and cropping the region of human faces and normalizing the image size to 64 × 96. ③ processing images with Gaussian filter to eliminate the effect of noise and improving the quality of images; ④ detecting and cropping the region of human eyes-eyebrows and mouths, then normalizing these regions to 64 × 32 size.

2) FEATURE EXTRACTION AND FUSION

In this work, image features are extracted and fused through the following steps (see in Figure 7): ① utilizing OMP, PMP and NMP for preprocessed whole facial image, mouth and eyebrow- eye respectively to generate OMFMs and PNMFMs; ② employing CS-LSMP operator extracts texture features from OMFMs and PNMFMs, and then cascading histograms to obtain Feature 1; ③ obtaining GFMs from the preprocessed face image and then employing CS-LSMP operator to obtain Feature 2; ④ Cascading Feature 1 and Feature 2 to form ultimate feature histograms.

3) CLASSIFICATION

Recognizing the category of facial expressions by carrying out SVM with Polynomial Function kernel and using automatic training function to acquire the optimal parameters.

C. EXPERIMENTAL RESULTS AND ANALYSIS

1) THE SELECTION OF NUMBERS AND BLOCKS IN GFMS

Gabor are commonly adopted as a feature as they are robust to misalignment. As we all know, adopting Gabor filters has a high computational cost, and the dimensionality of the output is very large, especially if they are applied holistically with a wide range of frequencies, scales and orientations. In order to reduce time complexity, our work examined the recognition rate of GFMs with different groups by adding four feature maps at a time.

Before fusing aforementioned features, it is essential to confirm the reasonable number of blocks for each feature extraction method. As we all know, the number of blocks has a vital effect on the final recognition result. A small value of blocks will make the extracted image features insufficient, whereas a large value of blocks will also result in higher feature dimensionality and time complexity. Therefore, experiments are executed to determine appropriate block number. So the experimental results on JAFFE dataset are described in Figure 8.

Figure 8 shows the recognition rate under different number of blocks and GFMs. When the number of block is set to

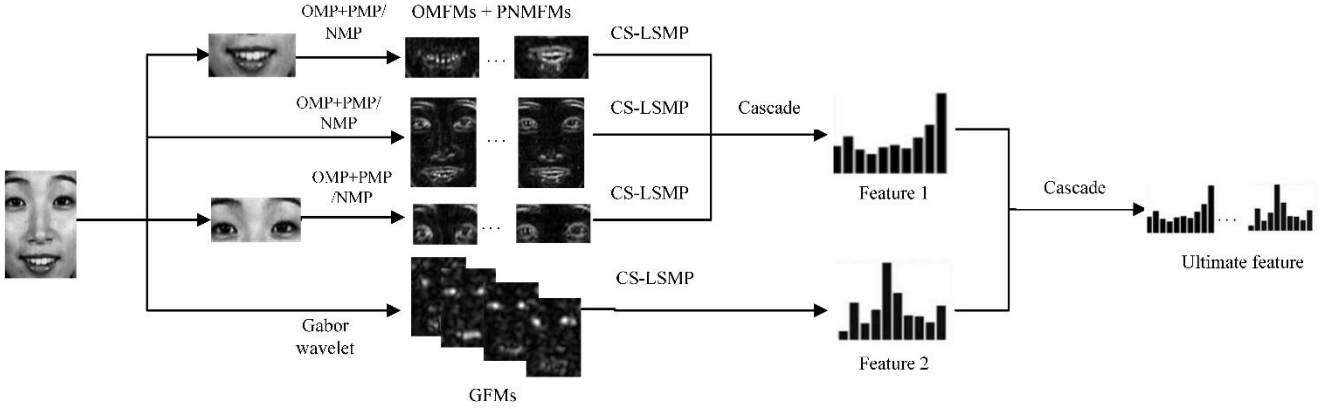

FIGURE 7. The overall diagram of feature extraction and fusion.

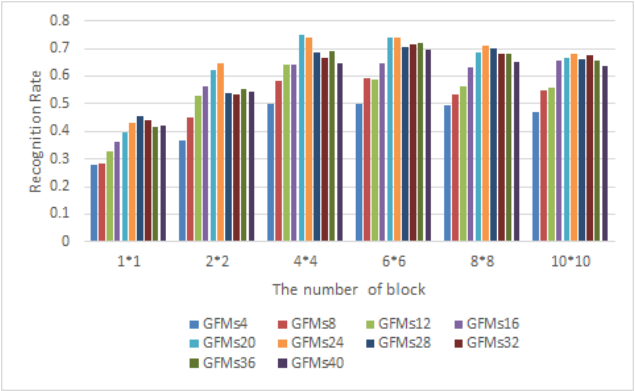

FIGURE 8. The recognition rate by varying the number of GFMs and blocks on JAFFE dataset.

1*1 or 2*2, the experimental performance using GFMs is not ideal at all. With the increase of local block, the recognition rate increases gradually and reaches the best result when the number of block = 4*4 then begin to decline. On the other hand, as compared with other sheets of GFMs, GFMs20 reaches a maximum of 74.76%. In addition, our selection of GFMs reduces feature dimensionality greatly rather than previous methods using all 40 feature maps. Similarly, we also get the same result on CK+ dataset. GFMs20 also performs favorably than other groups of GFMs and then we test the optimal number of block as shown in Figure 9 on CK+ dataset. It is evident that the number of blocks is also set to 4*4.

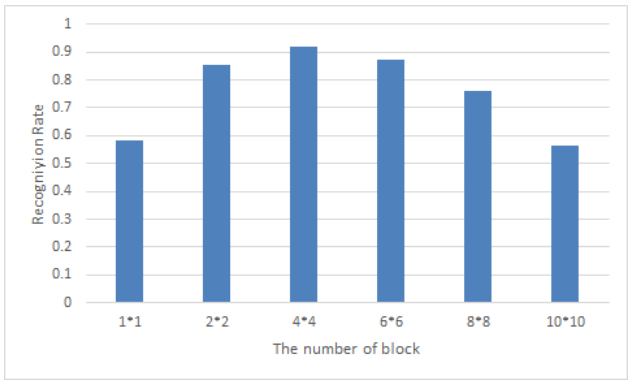

FIGURE 9. The recognition rate of GFMs for the number of block on CK+ dataset.

2) THE SELECTION OF BLOCK AND RADIUS IN OMFMS AND PNMFMS

OMFMs is a combination of multi-orientation OMP under different local radius. With the increase of the local radius, the number of involved neighboring points around the central pixel also increases, making the number of OMP increases (see in Figure 10). Similar to OMP, the local radius is also involved in PMP and NMP.

Figure 10 takes R = 1 and R = 2 as examples to show that the number of orientation increases with the increase of R.

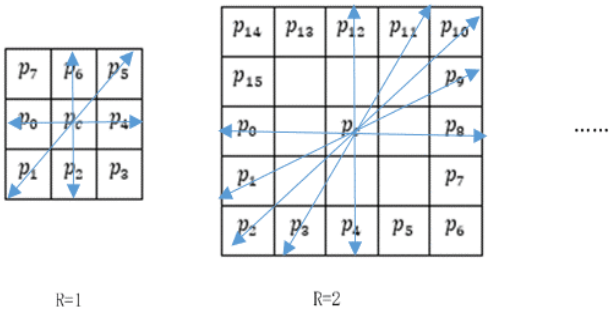

FIGURE 10. The description of OMP (R = 1 and R = 2).

Experimental results in Figure 11 delicate that the OMFMs achieve the better recognition rate on JAFFE when R = 3 and block = 4*4.

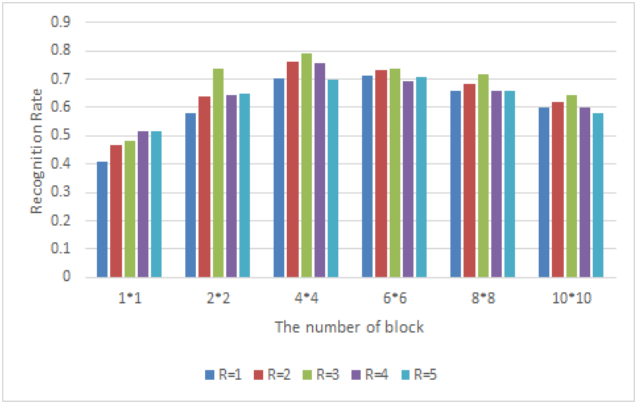

FIGURE 11. The recognition rate by varying the Radius of OMFMs and blocks on JAFFE dataset.

After that, our work assumed the R to be 3 and continued to complete the effect of the number of blocks on recognition rate on CK+ dataset. As shown in Figure 12, when the number of blocks = 4*4, the average accuracy reaches 95.31%. As we except, the pervious results and analysis on JAFFE and CK+ datasets validated that the optimal number of blocks

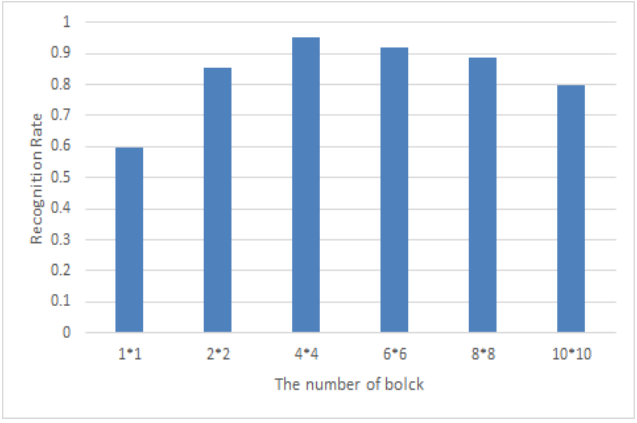

FIGURE 12. The recognition rate of OMFMs for the number of blocks on CK+ dataset.

was set to 4*4 and the average accuracy achieved the higher accuracy. Results demonstrated that the proposed method has uniformity on different datasets and also proved the general applicability of overall algorithms.

Similar to the experiment of OMFMs, we tested the optimal radius and the number of blocks for PNMFMs. Figure 13 shows that the PNMFMs achieve better recognition rate on JAFFE dataset when R = 2 and block = 6*6. On the premise that the radius is set to 2, we verify that the experimental results on CK+ dataset perform better when the number of blocks = 6*6 (See in Figure 14).

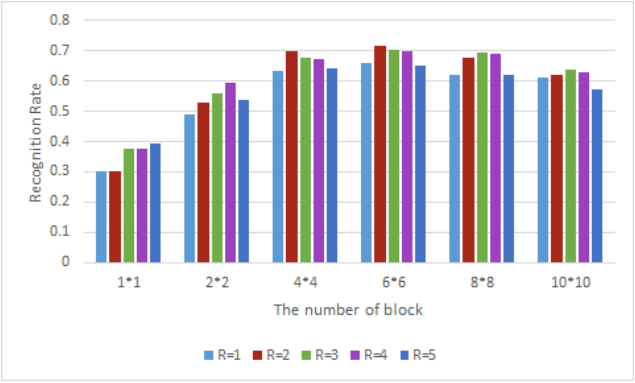

FIGURE 13. The recognition rate by varying the Radius of PNMFMs and blocks on JAFFE dataset.

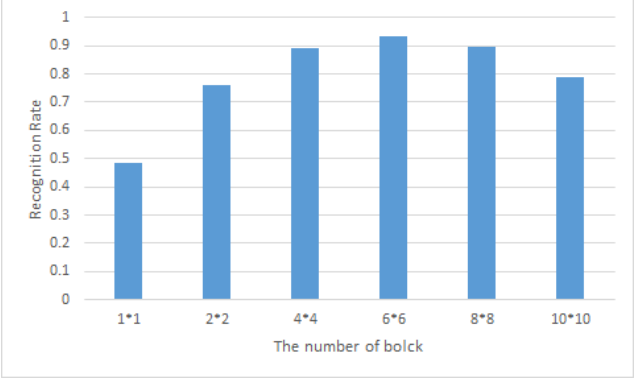

FIGURE 14. The recognition rate of PNMFMs for the number of blocks on CK+ dataset.

3) THE SELECTION OF THE NUMBER OF BLOCKS IN FACIAL PATCHES

Facial patches contain a lot of discriminant information and are conducive to expression recognition. In this paper, these facial patches are added to extract CS-LSMP features. When we express emotions, the nose patch contains less useful characteristics than eyebrow, eye and mouth. Meanwhile, in order to reduce the redundancy and training time, the nose patch is discarded. We selected eyebrow-eye and mouth patches. Figure 15 also shows the recognition rate for different block numbers of the added facial patches.

In order to make the parameters consistent, the number of blocks of the eyebrows and the mouth is the same in the experiment. Compared previous experimental results, these added facial patches further improves the average accuracy. And when the number of blocks of eyebrow-eye, mouth is 2*3 or 2*4, experiments achieve the best accuracy of 80% and 97.67% on JAFFE and CK+ datasets, respectively. In order to lower the computational complexity, the number of blocks is set to 2*3.

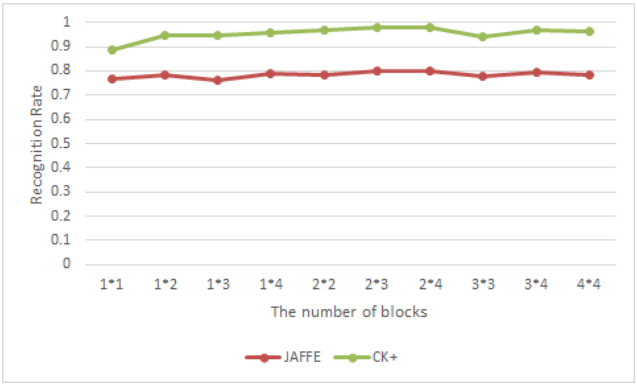

FIGURE 15. The recognition rate for different blocks of added facial patches on JAFFE and CK+ dataset.

D. COMPARISON OF EXPERIMENTAL RESULTS

In order to illustrate the effectiveness of the proposed feature extraction and fusion methods for facial expression recognition, this paper compares the recognition rate of proposed algorithm with some conventional feature extraction algorithms and CNN, and ensures that the recognition rate of each method is obtained under its optimal number of blocks. The comparison results are shown in Table 1.

TABLE 1. The recognition rate (%) for different methods on JAFFE and CK+ datasets.

Methods	JAFFE	CK+
LBP	66.19	85.87
CS-LBP	62.38	92.29
HOG	65.41	89.03
ULBP[33]	62.38	89.9
ELBP[34]	66.67	92.53
RILBP[35]	64.29	88.19
CS-LMP[36]	67.62	91.91
CLBP[41]	64.29	89.02
CLMP[40]	60.00	85.94
LCCMSP[42]	60.48	83.39
RAL-BGC[43]	61.90	91.19
LDENP[44]	62.38	90.04
ARCS-LBP[45]	60.48	87.15
CNN (data augmentation)	69.99	94.28
OMFMs + CS-LSMP	79.05	95.31
GFMs + CS-LSMP	74.76	91.88
PNMFMs + CS-LSMP	71.43	93.13
Our fused method	82.86	98.15

It can be seen from Table 1, extracting the CS-LSMP features from OMFMs, PNMFMs and GFMs improves the recognition rate greatly. Compared with some conventional feature descriptors, the best accuracy on JAFFE dataset only achieves 67.62% (CS-LMP [36]), and OMFMs + CS-LSMP yields average accuracy of 79.05%. The better performance of ELBP [34] achieves 92.53% on CK+ dataset and this paper is 95.31%. Our experiments also used neural networks for facial expression recognition. Due to the different scale of two datasets, we adopted and data augmentation (image rotation, random crop) and fine-tine to train respective best networks. So the parameter of network from two dataset are different. CNN with data augmentation yields recognition rate of 69.99% on JAFFE and 94.28% on CK+, which also indicates the validity of the CS-LSMP descriptor proposed in our work. To further improve the average accuracy, fusion strategy is adopted to obtain hybrid features. Experimental results illustrate that fused features are further enhanced based on the previously mentioned individual feature, and yield the final recognition accuracy 82.86% on JAFFE and 98.15% on CK+, respectively. Our proposed method performs better than other well-known descriptors, which demonstrates its feasibility and advancement.

Moreover, this paper also examined the runtime of multiple feature descriptors to recognize one image under the same experimental configuration. Taking JAFFE dataset as an example, we rank the processing time (in milliseconds) of various descriptors, including time of feature extraction and SVM classification (see in Figure 16). One can see that the proposed PNMFM have comparable times to the traditional LBP. Compared to original Gabor features, our fused method shorten the runtime and improve the overall recognition results greatly.

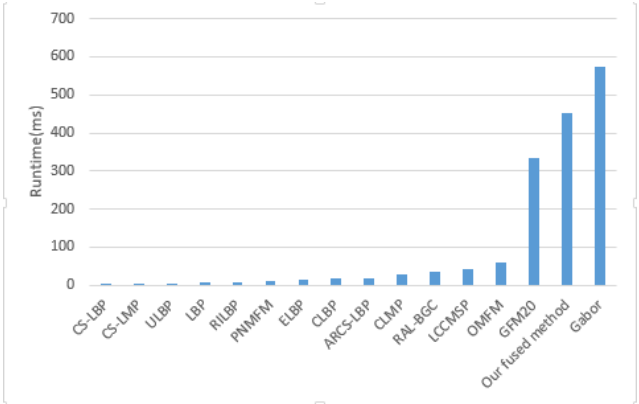

FIGURE 16. Runtime (in ms) of multiple descriptors in JAFFE dataset.

In addition, to further demonstrate the reasonability of overall algorithm, the proposed method in this paper was compared with several methods of facial expression recognition in other papers and ensured the same experimental strategy, then the comparison results on JAFFE and CK+ datasets are shown in Table 2 and 3, respectively. As Table 2 reveals, the performance of our proposed method on JAFFE dataset is superior to other excellent methods in recent years, and average accuracy is improved by 4.29% on the basis of [28] with 78.57%, which demonstrates the effectiveness of our approach. Moreover, [28] uses all Gabor features and the time complexity is much higher than our proposed method. Table 3 indicates that the proposed method on CK+ is not only superior to conventional feature extraction methods but also better than some methods using neural networks (In [29], the feature extracted from the appearance feature-based network is fused with the geometric feature in a hierarchical CNN, and also resulted in high feature dimensionality). In summary, fusion features based on CS-LSMP operator show their superiority and excellence in facial expression recognition.

TABLE 2. Accuracy (%) comparison of different methods on JAFFE dataset.

Methodology	Classifier	Accuracy
Local directional ternary pattern [14]	SVM	67.61
Special designed CNN [47]	SRC	68.08
Collaborative Representation [26]	ProCRC	71.58
Image pyramid + decision tree [25]	SVM	74.76
CS-LOP + Gabor [27]	SVM	76.67
LPQ+ Gabor+ CA-LDA [28]	SVM	78.57
Our proposed	SVM	82.86

TABLE 3. Accuracy (%) comparison of different methods on CK+ dataset.

Methodology	Classifier	Accuracy
Deep neural network [32]	Softmax	93.2
Local directional ternary pattern [14]	SVM	94.2
Spatio-Temporal RBM-based [46]	Softmax	95.66
HOG-TOP + hybrid features [30]	SVM	95.7
Spatial-Temporal LBP + Gabor [31]	SVM	95.8
CNN+ appearance / geometric features [29]	Softmax	96.46
Our proposed	SVM	98.15

V. CONCLUSION

This paper presents a new descriptor, CS-LSMP for feature extraction of facial expression recognition. The CS-LSMP operator obtains the corresponding magnitude by comparing the gray value between the central pixel and peripheral pixels, while retains signal of magnitude and then compares with the average magnitude. CS-LSMP reduces feature dimensionality and avoids the selection of threshold. Compared with LBP

and CS-LBP using a two-bit binary code of 0 or 1, CS-LSMP first extends the coding range to [0, 3] interval so that some delicate and detailed features between pixels can be captured and represented. In addition, we performed CS-LSMP feature extraction on OMFMs, PNMFMs, group of GFMs and facial patches such as eyebrows-eyes and mouths. The subject-independent experiments on JAFFE and CK+ datasets indicated that the method proposed in this paper improved the overall accuracy greatly and the consistency of experimental parameters also demonstrated the universality of our proposed algorithms. Since two public expression datasets were built in a limited laboratory environment, we will continue to study the expression recognition under uncontrollable and wild environment. In the future, we will try to explore more challenging tasks, such as micro-expression recognition and expression recognition across data sets.

REFERENCES

[1] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," in *Proc. 9th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Santa Barbara, CA, USA, Mar. 2011, pp. 921–926. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/FG.2011.5771374

[2] T. Wu, S. Fu, and G. Yang, "Survey of the facial expression recognition research," in *Advances in Brain Inspired Cognitive Systems* (Lecture Notes in Computer Science), H. Zhang, A. Hussain, D. Liu, and Z. Wang, Eds. Berlin, Germany: Springer, Jul. 2012, pp. 392–402. doi: 10.1007/978-3-642-31561-9_44.

[3] M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. F. Cohn, "FERA 2015—Second facial expression recognition and analysis challenge," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, Ljubljana, Slovenia, May 2015, pp. 1–8. doi: 10.1109/FG.2015.7284874.

[4] A. Wood, M. Rychlowska, S. Korb, and P. Niedenthal, "Fashioning the face: Sensorimotor simulation contributes to facial expression recognition," *Trends Cognit. Sci.*, vol. 20, no. 3, pp. 227–240, 2016.

[5] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines," *Sensors*, vol. 13, no. 6, pp. 7714–7734, 2013.

[6] J. Chen, D. Chen, Y. Gong, M. Yu, K. Zhang, and L. Wang, "Facial expression recognition using geometric and appearance features," in *Proc. ACM 4th Int. Conf. Internet Multimedia Comput. Service*, 2012, pp. 29–33.

[7] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[8] Z. Nenadic, "Information discriminant analysis: Feature extraction with an information-theoretic objective," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1394–1407, Aug. 2007.

[9] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 14, no. 8, pp. 1724–1733, 1997.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[11] 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.

[12] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with center-symmetric local binary patterns," in *Computer Vision, Graphics and Image Processing*. Berlin, Germany: Springer, 2006, pp. 58–69.

[13] M. Huang, Z. Mu, H. Zeng, and S. Huang, "Local image region description using orthogonal symmetric local ternary pattern," *Pattern Recognit. Lett.*, vol. 54, no. 4, pp. 56–62, 2015.

[14] 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.

[15] S. Yang and B. Bhanu, "Facial expression recognition using emotion avatar image," in *Proc. Face Gesture*, Mar. 2011, pp. 866–871.

[16] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Oct. 2005, pp. 786–791.

[17] 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.

[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] A. Ruiz-Garcia, M. Elshaw, A. Altahhan, and V. Palade, "A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots," *Neural Comput. Appl.*, vol. 29, no. 7, pp. 359–373, 2018.

[20] V. Mavani, S. Raman, and K. P. Miyapuram, "Facial expression recognition using visual saliency and deep learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2783–2788.

[21] B.-F. Wu and C.-H. Lin, "Adaptive feature mapping for customizing deep learning based facial expression recognition model," *IEEE Access*, vol. 6, pp. 12451–12461, 2018.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2005, pp. 886–893.

[23] 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.

[24] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn–Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops (CVPR)*, Jun. 2010, pp. 94–101.

[25] 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.

[26] 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. (CVPR)*, Jun. 2016, pp. 2950–2959.

[27] M. Hu, Y. Zheng, C. Yang, X. Wang, L. He, and F. Ren, "Facial expression recognition using fusion features based on center-symmetric local octonary pattern," *IEEE Access*, vol. 7, pp. 29882–29890, 2019.

[28] 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.

[29] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure," *IEEE Access*, vol. 7, pp. 41273–41285, 2019.

[30] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial expression recognition in video with multiple feature fusion," *IEEE Trans. Affect. Comput.*, vol. 9, no. 1, pp. 38–50, Jul. 2016.

[31] L. Zhao, Z. Wang, and G. Zhang, "Facial expression recognition from video sequences based on spatial-temporal motion local binary pattern and Gabor multiorientation fusion histogram," *Math. Problems Eng.*, vol. 2017, Jan. 2017, Art. no. 7206041.

[32] 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.

[33] A. Rölle, M. Mousavi-Jazi, M. Eriksson, J. Odeberg, C. Söderberg-Nauclér, D. Cosman, K. Kärre, and C. Cerboni, "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. Immunology*, vol. 171, no. 2, pp. 902–908, 2003.

[34] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "3-D face recognition using eLBP-based facial description and local feature hybrid matching," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1551–1565, Oct. 2012.

[35] 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.

[36] C. T. Ferraz, O. Pereira, Jr., and A. Gonzaga, "Feature description based on center-symmetric local mapped patterns," in *Proc. 29th Annu. ACM Symp. Appl. Comput.*, 2014, pp. 39–44.
[37] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification," *IEEE Access*, vol. 7, pp. 18940–18950, Jan. 2019.
[38] X. Lu, Y. Yang, W. Zhang, Q. Wang, and Y. Wang, "Face verification with multi-task and multi-scale feature fusion," *Entropy*, vol. 19, no. 5, p. 228, 2017.
[39] Z. Yang, H. Wang, Y. Han, and X. Zhu, "Discriminative multi-task multi-view feature selection and fusion for multimedia analysis," *Multimedia Tools Appl.*, vol. 77, no. 3, pp. 3431–3453, 2018.
[40] R. T. Vieira, T. T. Negri, and A. Gonzaga, "Improving the classification of rotated images by adding the signal and magnitude information to a local texture descriptor," *Multimedia Tools Appl.*, vol. 77, no. 23, pp. 31041–31066, 2018.
[41] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jan. 2010.
[42] Y. El Merabet and Y. Ruichek, "Local concave-and-convex micro-structure patterns for texture classification," *Pattern Recognit.*, vol. 76, pp. 303–322, Apr. 2018.
[43] I. E. Khadiri, M. Kas, Y. El Merabet, Y. Ruichek, and R. Touahni, "Repulsive-and-attractive local binary gradient contours: New and efficient feature descriptors for texture classification," *Inf. Sci.*, vol. 467, pp. 634–653, Oct. 2018.
[44] A. Pillai, R. Soundrapandiyan, S. Satapathy, S. C. Satapathy, K.-H. Jung, and R. Krishnan, "Local diagonal extrema number pattern: A new feature descriptor for face recognition," *Future Gener. Comput. Syst.*, vol. 81, pp. 297–306, Apr. 2018.
[45] Y. E. Merabet, Y. Ruichek, and A. E. Idrissi, "Attractive-and-repulsive center-symmetric local binary patterns for texture classification," *Eng. Appl. Artif. Intell.*, vol. 78, pp. 158–172, Feb. 2019.
[46] S. Elaiwat, M. Bennamoun, and F. Boussaid, "A spatio-temporal RBM-based model for facial expression recognition," *Pattern Recognit.*, vol. 49, pp. 152–161, Jan. 2016.
[47] E. B. Sonmez, "A study on facial expression recognition," *Gazi Univ. J. Sci.*, vol. 30, pp. 19–27, 2017.
[48] S. Li and W. Deng, "Deep facial expression recognition: A survey," 2018, *arXiv:1804.08348*. [Online]. Available: https://arxiv.org/abs/1804.08348

MIN HU received the M.S. degree in industrial automation and the Ph.D. degree in computer science from the Hefei University of Technology, Hefei, China, in 1994 and 2004, respectively, 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.

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

YAQIN ZHENG is currently pursuing the master's degree with the Hefei University of Technology. Her research interests include image processing and facial expression 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 M.S. and Ph.D. degrees with the Hefei University of Technology, in 2007 and 2015, respectively, where she is currently an Associate Professor. Her research interests include computer application, data structure, and software engineering.

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.