IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON HEALTHCARE INFORMATION TECHNOLOGY FOR THE EXTREME AND REMOTE ENVIRONMENTS

Received April 11, 2019, accepted May 12, 2019, date of publication May 16, 2019, date of current version May 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917230

Classroom Micro-Expression Recognition Algorithms Based on Multi-Feature Fusion

LIDAN MAO[1], NING WANG[1], LEI WANG[2], AND YU CHEN[3]

[1]College of Foreign Languages, Hebei Agricultural University, Baoding 071001, China
[2]Physical Education Department, Hebei Agricultural University, Baoding 071001, China
[3]Hebei Software Institute, Baoding 071000, China

Corresponding author: Lidan Mao (maolidan@yeah.net)

ABSTRACT The extraction of facial features is the key to micro-expression recognition. This paper puts forward a micro-expression recognition algorithm through multi-feature fusion. In this algorithm, the change of local binary pattern (LBP) feature distribution is correlated with projection error. For fast and accurate detection, the research data were all extracted from professional facial expression databases, the images had basically the same positions in each expression library, and the pure face image was acquired through manual segmentation from the selected expression library. Through comparison and its application in an intelligent classroom environment, it is proved that the proposed algorithm clearly outperforms the original LBP algorithm. The proposed method can be extended to other image recognition and classification problems.

INDEX TERMS Classroom micro-expression recognition, projection error, multi-feature fusion, local binary pattern (LBP).

I. INTRODUCTION

Micro-expression recognition is generally achieved in four steps, namely, face detection, image preprocessing, feature extraction and feature classification. The face detection is generally implemented using knowledge model, template matching or statistical model [1]–[3]. Whichever the basis, a face template should be prepared through sample training before the detection process, such as to eliminate the background and other noises from the target image, highlighting the face area. The image preprocessing is a process of dimension reduction [4]. With a huge amount of data, the original image belongs to the high-dimensional space. The dimension of the image needs to be reduced through transformation or mapping. The most popular ways of preprocessing include gray processing, histogram equalization and scale normalization. The feature extraction refers to the extraction of global and local features of facial expressions, including but not limited to appearance, geometry, deformation and motion [5]. The different expressions can be recognized by observing the relative positions of feature points [6]. The feature classification relies on the selection of a classifier out of the numerous available classifiers. It is impossible to recognize micro-expressions accurately without a suitable classifier.

Inspired by deep learning, this paper puts forward a micro-expression recognition algorithm through multi-feature fusion. In this algorithm, the change of local binary pattern (LBP) feature distribution is correlated with projection error. Moreover, the statistical model-based face detection method was adopted due to good performance in complex environment. For fast, accurate detection, the research data were all extracted from professional facial expression databases, the images had basically the same positions in each expression library, and the pure face image was acquired through manual segmentation from the selected expression library.

II. LITERATURE REVIEW

A. APPEARANCE FEATURE-BASED MICRO-EXPRESSION RECOGNITION

The existing studies on micro-expression recognition mainly tackle face detection, feature extraction and classification [3], with special emphasis on the extraction of appearance features. The appearance features, ranging from density, edge to texture, are commonly extracted by local binary pattern (LBP)‘[7], Gabor filter [8] and linear discriminant analysis (LDA) [9]. Below are some representative studies on micro-expression recognition using appearance feature.

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

Wang [10] introduced differential active appearance model and manifold learning to recognize micro-expressions. Li *et al.* [11] combined Gabor wavelet and kernel principal component analysis (PCA) into a micro-expression recognition method, and proved the robustness of the method through experiments. Lee *et al.* [12] improved the independent PCA, and classified expressions based on the resolution ability of principal components. Shojaeilangari *et al.* [13] extracted appearance features through Gabor transform and reduced the image dimension by PCA. Zhong *et al.* [14] employed Gabor transform to extract appearance features from the upper half, the lower half and the entire area of faces. Zhang *et al.* [15] proposed a maximum stable extremum region algorithm with excellent affine invariance.

Face is a flexible body. Facial expressions are mostly the result of facial muscles. Every expression is a gradual process of change. It is difficult for existing algorithms to distinguish different expressions accurately. At present, there are many facial expression databases for facial expression recognition, but there is no universal standard in the world. It will lead to errors in the accuracy of recognition. In addition, in the process of facial expression recognition, the existing algorithms usually separate face recognition from facial expression recognition. In this paper, the existing algorithms are improved from the aspect of feature extraction. It improves the disadvantage that a single feature cannot accurately represent all the information of the image, and the recognition rate of expression is improved.

B. DEEP LEARNING-BASED MICRO-EXPRESSION RECOGNITION

Deep learning is an important field of computer vision. The deep learning network boasts multiple intermediate layers between the input layer and the classifier, and learns high-level semantic features from abundant training data. This technology has been introduced to micro-expression recognition.

For example, Zhao *et al.* [16] designs a micro-expression recognition method based on convolution neural network (CNN), which extracts nonlinear facial features through multi-layer convolution and integration, and verifies the accuracy of the method against several facial expression benchmarks. Uddin *et al.* [17], Sun and Bin [18] successfully apply neural network models like deep belief network (DBN) and deep Boltzmann machine (DBM) to micro-expression recognition.

III. PRELIMINARIES

The image preprocessing directly bears on the recognition accuracy of micro-expressions. This step is of great necessity because the original image often has poor evenness, contrast and noisiness, owing to the limitation of image acquisition device and the unstable light condition. As mentioned before, micro-expression images are usually preprocessed gray processing, histogram equalization or scale normalization.

A. GRAY PROCESSING

Gray processing means converting the original color image to a grayscale image. After all, the color image is too sensitive to the light condition. In computer system, the colors in the original image can be represented by many color models, such as RGB (red, green and blue), HSI (hue, saturation, intensity) and YCbCr. All the other models can be derived from the most basic color model RGB.

In the RGB system, the three primary colors, i.e. red, green and blue, correspond to the three axes, pure black is placed at (0, 0, 0) and pure white is positioned at (255, 255, 255). The pixel values of an image can be expressed as coordinates in the system. When R = G = B, the pixel shows a gray color. Thus, the grayscale range for each pixel is 0∼255. Since our eyes are more sensitive to green than blue, the three components should be property weighed:

$$Gray = 0.299 \times R + 0.587 \times G + 0.114 \times B \tag{1}$$

FIGURE 1. Grey processing of an image on angry face.

B. HISTOGRAM EQUALIZATION

Histogram equalization eliminates the illumination impact and skin color difference in the original image. The grayscale distribution of the target image can be illustrated by the statistical graph called grayscale histogram. In fact, this distribution is about the correlation between the grayscale of a pixel r and its visualization probability. For each pixel in the image, the grayscale r is a random value in the interval [0, 1], that is, $f(x)$ is a random variable. Then, the grayscale distribution of the original image can be described by the probability density function $p_r(r)$:

$$p_r(r_k) = \frac{n_k}{n}, \quad k = 0, 1, \ldots, L-1 \tag{2}$$

where n is the total number of pixels in the image; n_k is the number of pixels with the grayscale r_k L is the total number of grayscale levels in the image. The curve of the $p_r(r_k)$ function is the grayscale image of the original image. Then, the grayscale image should be normalized by the following transform and inverse transform functions:

$$\begin{aligned} S_k &= T(r_k) = \sum_{j=1}^{k} p_k(r_k) \\ &= \sum_{j=1}^{k} \frac{n_j}{n}, \quad k=0, 1, \ldots, L-1 0 < r < 1 \end{aligned} \tag{3}$$

$$r_k = T^{-1}(s_k), \quad k = 0, 1, \ldots, L-1 \tag{4}$$

Under the same light condition, the faces with similar shapes should have similar grayscale distributions, i.e. similar

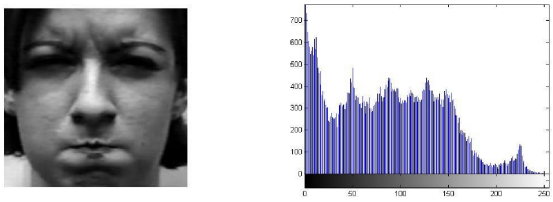


FIGURE 2. Micro-expression image and its histogram before equalization.

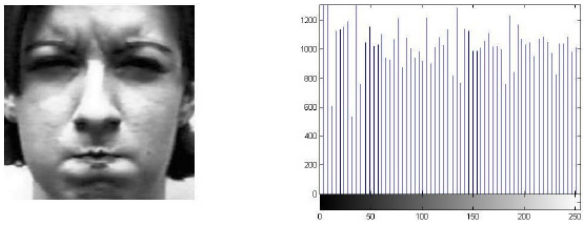


FIGURE 3. Micro-expression image and its histogram after equalization.

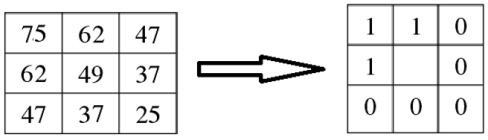


FIGURE 4. Example of LBP algorithm.

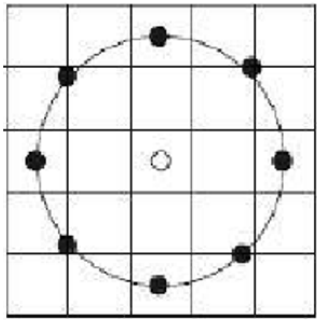


FIGURE 5. Extended LBP of a circular neighborhood.

means and variances of grayscale. Figures 2 and 3 compare a micro-expression image before and after histogram equalization.

C. SCALE NORMALIZATION

Scale normalization is necessary when the shooting angle and distance of the camera change during image acquisition. The key steps of scale normalization are translation, rotation and scaling. First, the eyes on the face are moved to the same horizontal line; then, the image is cropped to place the face at the center; finally, the cropped image is scaled so that the entire face has the same pixel size. The specific steps of scale normalization are as follows:

Step 1: The positions of both eyes are determined by face detection, and denoted as E_l and E_r, respectively.

Step 2: The image is rotated such that line E_lE_r is parallel to the horizontal direction.

Step 3: The image is cropped to the size of $2d * 2d$ ($d = |\overline{E_lE_r}|$), with O, the midpoint of E_lE_r, at the center.

Step 4: The cropped image is scaled to a unified pixel size.

IV. ALGORITHM DESIGN

The neutral expression is the starting point of every micro-expression. The subblocks in each micro-expression image have different LBP feature distribution from their counterparts in the neutral expression image. If the micro-expression image is projected to the neutral expression image, the information richness of the micro-expression increases with the projection error. Thus, the LBP feature was integrated with projection error to extend the interval of normalized LBP feature and enhance feature differentiation.

The LBP is a good descriptor of the grayscale and rotation invariance of textures. Focusing on surface features, the LBP algorithm [19] can extract local invariants from the entire face or a specific point on the face, which is sensitive to face changes like illumination and occlusion. By the algorithm, a point in the grayscale image is selected as the center pixel, and its grayscale is taken as the threshold. Then, the grayscale of each adjacent pixel is compared against the threshold. The resulting set of binary sequences is converted into decimal numbers, and taken as the eigenvalues of the point.

The original LBP algorithm can be expressed as a nine-square grid. As shown in Figure 4, the grayscales of the eight adjacent pixels are compared with the grayscale of the center pixel. If the adjacent pixel has a greater grayscale than the center pixel, the corresponding square will be marked as 1; otherwise, the corresponding square will be marked as 0.

Following the clockwise direction, a 8-bit binary number can be obtained from the nine-square grid: 11000001. The decimal number converted from this binary number is the LBP value of the pixel at that point. Thus, the LBP value can be calculated as:

$$\mathrm{LBP}(x_c, y_c) = \sum_{k=0}^{7} Z(I_k, I_c)2^k \tag{5}$$

where I_c is the grayscale of the center point; I_k is the grayscale of an adjacent point; $Z(I_k, I_c)$ can be obtained by:

$$Z(I_k, I_c) = \begin{cases} 1, & if\ I_k - I_c \geq 0 \\ 0, & if\ I_k - I_c < 0 \end{cases}$$
$$k = 0, \quad 1, 2, \ldots, 7 \tag{6}$$

Through the pairwise comparison, a monotonous grayscale curve can be obtained for the local area, which agrees with the grayscale variation in micro-expression images under changing light conditions. Therefore, the extraction of local features may help achieve automatic micro-expression recognition.

The LBP algorithm has a limited coverage (9 pixels in the nine-square grid), which constrains its expression ability. In this paper, the original LBP algorithm in extended to cover a wider range [20]. The extended LBP algorithm can be denoted as (P, R), where P is the number of pixels adjacent to the central pixel and R is the radius of the circular neighborhood (Figure 5). The LBP features are denoted as LBP_{PR}.

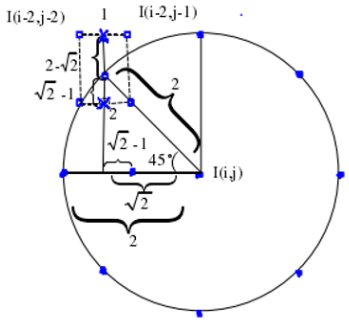


FIGURE 6. The plaintext histogram.

As shown in Figure 5, the circular neighborhood has a radius of 2 and eight adjacent pixels. Obviously, half of the adjacent pixels are located at the center of square, and another half are not located at square centers. The grayscale of each center adjacent pixel equals that of the local square, and that of each non-center adjacent pixel was determined by bilinear interpolation.

Let $I(i, j)$ be the grayscale of the center point in row j, column i of the neighborhood. Then, the eight adjacent pixels were arranged on the circle (Figure 6). For the non-center pixel in the upper left corner, its grayscale can be derived from that of its four adjacent pixels.

The first step of bilinear interpolation is the horizontal interpolation of pixels 1 and 2. The grayscale of pixel 1 can be obtained by linear interpolation of $I(i-2, j-2)$ and $I(i-2, j-1)$ on the same line:

$$\text{Value}(1) = I(i-2, j-2) + (2-\sqrt{2}) \times (I(i-2, j-1) - I(i-2, j-2)) \quad (7)$$

Similarly, the grayscale of pixel 2 can be calculated as:

$$\text{Value}(2) = I(i-1, j-2) + (2-\sqrt{2}) \times (I(i-1, j-1) - I(i-1, j-2)) \quad (8)$$

Then, the vertical interpolation of the two pixels can be carried out. The grayscale of each non-center pixel can be calculated as:

$$\text{Value} = \text{Value}(1) + (2-\sqrt{2}) \times (\text{Value}(2) - \text{Value}(1)) \quad (9)$$

Then, the grayscale was compared to that of the center pixel. The pixel position was marked 1 if the grayscale is greater than that of the center pixel, and 0 if otherwise. After each pixel in the neighborhood was assigned a weight, the value of the local texture feature of the central pixel $\text{LBP}_{8,2}$ can be determined by adding up the weighted grayscales of all pixels.

A. PROJECTION ERROR

Figure 7 illustrates a 2D vector projection, in which vector X is being projected on vector Y. As shown in the figure, the project vector is αY and the projection error is $\|X - \alpha Y\|_2$. To minimize the projection error, the most

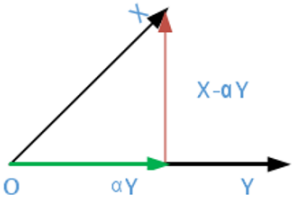


FIGURE 7. 2D vector projection.

suitable α needs to be selected through continuous adjustment. The error vector can be minimized when it is perpendicular to the data vector, that is, the inner product of the error vector and the data vector is 0.

Next, a 2D vector projection analysis was performed to find the projection error between each subblock in the micro-expression image and its counterpart in the neutral expression image. Let $A = \{Y_i | i = 1, 2, \ldots, N\}$ be a set of neutral expression images, and $B = \{X_i^k | i = 1, 2, \ldots, \text{N}; k = 1, 2, \ldots, 7\}$ be a set of micro-expression images. Each image in the expression library was divided into λ subblocks:

$$Number\ of\ subblock(\lambda) = \frac{Image\ size}{Block\ size} \quad (10)$$

This paper considers the k-th micro-expression image and the k-th neutral expression image. Let x_m^k and y_m be the m-th subblock of the k-th micro-expression image and the m-th subblock of the k-th neutral expression image ($m \epsilon \{1, 2, \ldots, \lambda\}$). As mentioned before, the subblocks in each micro-expression image have different LBP feature distribution from their counterparts in the neutral expression image. Thus, the projection error between the m-th subblock of the micro-expression image and the m-th subblock of the neutral expression image can be expressed as:

$$e_m^k = \left\| proj_{y_m} x_m^k \right\|_2 = \left\| x_m^k - \frac{\langle x_m^k \cdot y_m \rangle}{\langle y_m \cdot y_m \rangle} y_m \right\|_2 \quad (11)$$

The projection errors of other subblocks can be determined similarly. The magnitude of the projection error directly reflects the information richness of each subblock.

V. RESULTS AND DISCUSSION

To verify the effectiveness of the extended LBP algorithm, the LBP features extracted by this algorithm were compared with those extracted by the original LBP algorithm based on a student micro-expression library. The support vector machine (SVM) classifier was used to eliminate any other interference.

The student micro-expression library contains 617 images from 19 students, covering 12 expressions. Each student has 2 to 3 pictures of each expression. All the pictures in the database are 720 × 576 in size. And the objects in the library have changes in lighting and posture, which is more in line with the reality of classroom teaching scene.

The LBP and extended LBP algorithms were adopted to analyze the same micro-expression images. The tested micro-expressions include anger, contempt, disgust, fear,

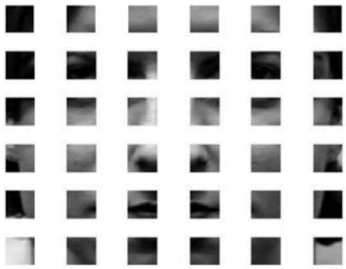


FIGURE 8. Neutral expression image.

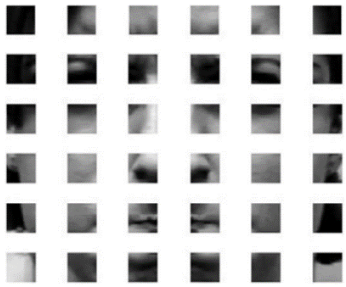


FIGURE 9. Anger image.

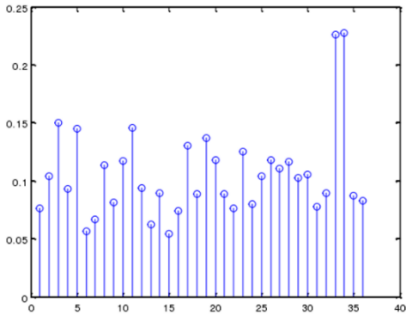


FIGURE 10. Subblock projection error.

happiness, sadness, neutrality and surprised. For simplicity, a neutral expression image and an anger image were cited as an example. Both images were divided into 36 blocks (Figures 8 and 9). The differences of anger from neutrality can be clearly seen by comparing the subblocks of the two images.

The projection error of each subblock is shown in Figure 10, where the abscissa is the serial number of subblocks and the ordinate is the projection error.

The projection errors were relatively large in subblocks 33 and 34, revealing obvious differences between anger and neutral expression. The two subblocks correspond to the mouth. By contrast, subblocks 6, 13 and 7 had small projection errors, an evidence of small differences. The three subblocks correspond to cheeks and hair.

Besides, the obvious and slight differences appeared alternatively from subblocks 31∼36. As a result, the LBP eigenvalue of the six sub-blocks were divided by the projection error of each subblock. The normalized LBP eigenvalue of each subblock is shown in Figure 11(a) and the ratio of LBP

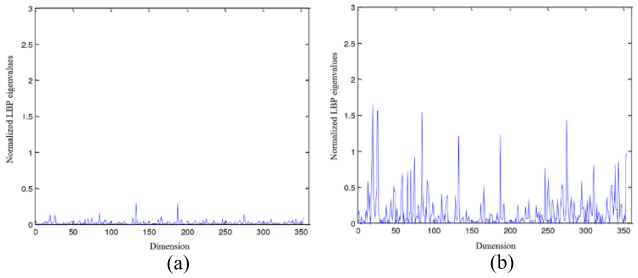


FIGURE 11. Line chart of LBP features in subblocks 31∼36. (a) LBP features between subblocks 31∼36. (b) LBP features between subblocks 31∼36 divided by projection errors.

eigenvalue to projection error of each subblock is presented in Figure 11(b).

It can be seen that the intervals in Figure 11(b) were larger than those in Figure 11(a). The eigenvalues were stretched and the slope was increased, making the LBP features more distinguishable.

The extended LBP and original LBP algorithms were further contrasted through cross-validation. Each algorithm was applied to micro-expression recognition for 10, 20, 30 and 40 times. The mean recognition rate of each algorithm in each application is recorded in Table 1 below.

TABLE 1. Comparison of mean recognition rates.

Method	Recognition rate of 10 times	Recognition rate of 20 times	Recognition rate of 30 times	Recognition rate of 30 times
Original LBP algorithm	93.68%	95.17%	95.89%	96.89%
Extended LBP algorithm	93.89%	95.56%	97.31%	98.16%

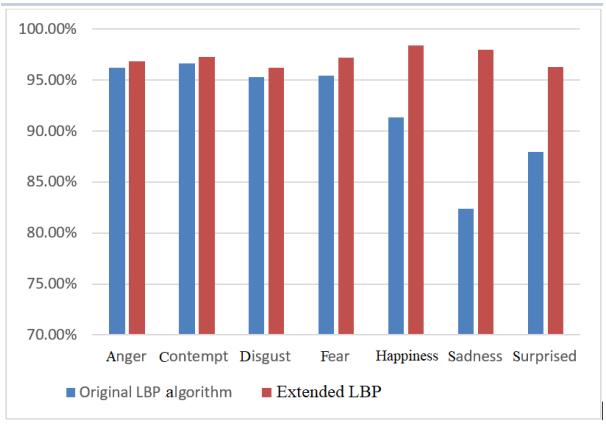


FIGURE 12. Mean recognition rate of each micro-expression.

Figure 12 shows the recognition rates of the two algorithms for each micro-expression. The extended LBP algorithm clearly outperformed the original algorithm in students' dataset. Therefore, the proposed method can eliminate the

negative impact of the difference between individual expressions, and achieve a good recognition rate at a fast speed.

VI. CONCLUSIONS

Considering the dense distribution of normalized LBP features of each micro-expression, this paper puts forward a micro-expression recognition method that divides the LBP feature extracted from each sub-block of the original image. The original LBP algorithm was improved with the projection error between the corresponding subblocks in the micro-expression image and the neutral expression image. Through comparison, it is proved that the extended LBP algorithm clearly outperforms the original algorithm. Micro-expression occurs very quickly and attempt to hide real emotions and potentials. The advancement and development of micro-expression recognition technology and algorithm make it possible to realize effective teaching, and help students to understand their inner state, understand their real ideas, and promote meaningful learning. The proposed method can also be extended to other image recognition and classification problems. The future research will further enhance the robustness of our algorithm under various real-world interferences.

REFERENCES

[1] G. Hermosilla, J. R.-del-Solar, R. Verschae, and M. Correa, "A comparative study of thermal face recognition methods in unconstrained environments," *Pattern Recognit.*, vol. 45, no. 7, pp. 2445–2459, Jul. 2012.

[2] G. Sun and S. Bin, "A new opinion leaders detecting algorithm in multi-relationship online social networks," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4295–4307, 2018.

[3] X. R. Song, S. Gao, and C. B. Chen, "A novel vehicle feature extraction algorithm based on wavelet moment," *Traitement Signal*, vol. 35, nos. 3–4, pp. 223–242, 2018.

[4] I. Li, J.-L. Wu, and C.-H. Yeh, "A fast classification strategy for SVM on the large-scale high-dimensional datasets," *Pattern Anal. Appl.*, vol. 21, no. 4, pp. 1023–1038, Nov. 2017.

[5] S. M. Lajevardi and Z. M. Hussain, "Automatic facial expression recognition: Feature extraction and selection," *Signal, Image Video Process.*, vol. 6, no. 1, pp. 159–169, 2012.

[6] A. Vinciarelli *et al.*, "Open challenges in modelling, analysis and synthesis of human behaviour in human–human and human–machine interactions," *Cognit. Comput.*, vol. 7, no. 4, pp. 397–413, Aug. 2015.

[7] 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.

[8] P. Li, L. Zhang, and Y. Zhao, "Feature fusion with neighborhood-oscillating tabu search for oriented texture classification," *IFIP-The International Federation for Information Processing*, vol. 187, 2012, pp. 671–680.

[9] M. H. Siddiqi, R. Ali, A. M. Khan, Y. T. Park, and S. Lee, "Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1386–1398, Apr. 2015.

[10] X. K. Wang, X. Mao, and I. Mitsuru, "Human face analysis with non-linear manifold learning," *J. Electron. Inf. Technol.*, vol. 33, no. 10, pp. 2531–2535, Oct. 2011.

[11] J.-B. Li, J.-S. Pan, and Z.-M. Lu, "Face recognition using Gabor-based complete Kernel Fisher Discriminant analysis with fractional power polynomial models," *Neural Comput. Appl.*, vol. 18, no. 6, pp. 613–621, Sep. 2009.

[12] S. H. Lee, K. N. K. Plataniotis, and Y. M. Ro, "Intra-class variation reduction using training expression images for sparse representation based facial expression recognition," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 340–351, Jul./Sep. 2014.

[13] S. Shojaeilangari, W. Y. Yau, J. Li, and E. K. Teoh, "Dynamic facial expression analysis based on extended spatio-temporal histogram of oriented gradients," *Int. J. Biometrics*, vol. 6, no. 1, pp. 33–52, 2014.

[14] L. Zhong, Q. Liu, P. Yang, J. Huang, and D. N. Metaxas, "Learning multiscale active facial patches for expression analysis," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1499–1510, Aug. 2015.

[15] G. Zhang H. Kai, B. Zhang, H. Fu, and J. Zhao, "A natural scene text extraction method based on the maximum stable extremal region and stroke width transform," *J. Xi'an Jiaotong Univ.*, vol. 51, no. 1, pp. 135–140, 2017.

[16] X. Zhao, X. Shi, and S. Zhang, "Facial expression recognition via deep learning," *IETE Tech. Rev.*, vol. 32, no. 5, pp. 347–355, Sep. 2015.

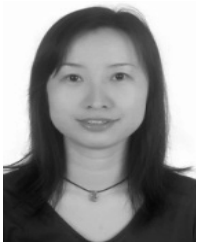
[17] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, and G. Fortino, "Facial expression recognition utilizing local direction-based robust features and deep belief network," *IEEE Access*, vol. 5, pp. 4525–4536, 2017.

[18] G. Sun and S. Bin, "Router-level Internet topology evolution model based on multi-subnet composited complex network model," *J. Internet Technol.*, vol. 18, no. 6, pp. 1275–1283, 2017.

[19] J. Lu, G. Wang, and Z. Pan, "Nonlocal active contour model for texture segmentation," *Multimedia Tools Appl.*, vol. 76, no. 8, pp. 10991–11001, Apr. 2017.

[20] H. Tang, B. Yin, Y. Sun, and Y. Hu, "3D face recognition using local binary patterns," *Signal Process.*, vol. 93, no. 8, pp. 2190–2198, Aug. 2013.

LIDAN MAO is currently a Lecturer with the College of Foreign Languages, Hebei Agricultural University. Her major research directions are ESL, language education, and emerging technologies in learning.



NING WANG is currently a Lecturer with the College of Foreign Languages, Hebei Agricultural University. Her major research direction is English education.

LEI WANG is currently a Lecturer with the Physical Education Department, Hebei Agricultural University. His major research direction is physical education and training.

YU CHEN is currently a Lecturer with the Hebei Software Institute. His major research direction is computer application.