A Method of Facial Expression Recognition Based on Gabor and NMF[1]

Jun Zhou, Sue Zhang, Hongyan Mei, and Dawei Wang

School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou, China

e-mail: bishuilinglan@126.com

Abstract—The technology of facial expression recognition is a challenging problem in the field of intelligent human-computer interaction. An algorithm based on the Gabor wavelet transformation and non-negative matrix factorization (G-NMF) is presented. The main process includes image preprocessing, feature extraction and classification. At first, the face region containing emotional information is obtained and normalized. Then, expressional features are extracted by Gabor wavelet transformation and the high-dimensional data are reduced by non-negative matrix factorization (NMF). Finally, two-layer classifier (TLC) is designed for expression recognition. Experiments are done on JAFFE facial expressions database. The results show that the method proposed has a better performance.

Keywords: facial expression recognition, Gabor wavelet transform, non-negative matrix factorization.

DOI: 10.1134/S1054661815040070

1. INTRODUCTION

The study of psychologist A. Mehrabian shows that information is transmitted by language, voice and facial expressions in our daily communication. In these three ways, the information transmitted through facial expressions accounts for 55% of the amount of information [1]. Therefore, a lot of valuable information can be got by the classification of facial expressions. With the development of computer science and technology, the intelligent human-computer interaction is becoming a hot research in the area of artificial intelligence and the emotional communication between human and computer has been paid more attention in recent years. Consequently, the research on facial expression is of great significance.

Feature extraction is a key step for facial expression recognition. The information from different orientations and scales can be derived using Gabor wavelet transformation, so it has been widely applied. Liu accomplished expression recognition by integrating Gabor filter and sparse representation [2]. Li completed it with local Gabor filters to extract features [3]. However, Gabor wavelet transform algorithm produces high-dimensional data. It is difficult to deal with these data and the process increases time and space complexity, so the data dimensionality needs to be reduced. Principal Component Analysis (PCA) and Linear Discrimination Analysis (LDA) are commonly used, such as [4, 5]. In facial expression recognition, the grey value of input image is non-negative and the common characteristic of these methods is that the reduction of the rank can not guarantee the non-negativity of original data. This characteristic leads to the decrease of recognition rate.

This paper presents facial expression recognition based on Gabor wavelet transform and non-negative matrix factorization. Features of preprocessed image are extracted using Gabor wavelet transform. High-dimensional data generated from Gabor wavelet transform is reduced by the application of non-negative matrix factorization. Nearest neighbor classifier is trained to achieve the first layer classification and the final result is obtained by recording the votes of different classes in the first layer result.

2. THE GENERAL PROCESS OF FACIAL EXPRESSION RECOGNITION

The process of facial expression recognition mainly includes feature extraction and expression classification. The general procedure of facial expression recognition is shown by Fig. 1. The preprocessing is to eliminate some interference factors which have influences on the feature extraction and classification. The purpose of feature extraction is to obtain the attribute information which can characterize the class of input image. Classification is a differential process according to the characteristics of input image. Commonly used methods of feature extraction are based on image or gray values. [7] is an application of method based on image, which regards the scales of eyes, eyebrows and mouth as the attribute characteristics for expression classification. Gabor wavelet transform and Local Binary Pattern (LBP) [8] are the ideas based on gray values. Gabor wavelet transform reflects the image

[1] The article is published in the original.

Received May 10, 2014

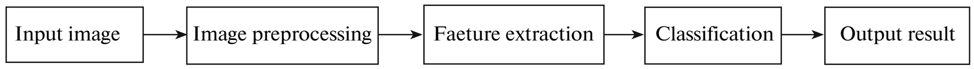


Fig. 1. Diagram of facial expression recognition process.

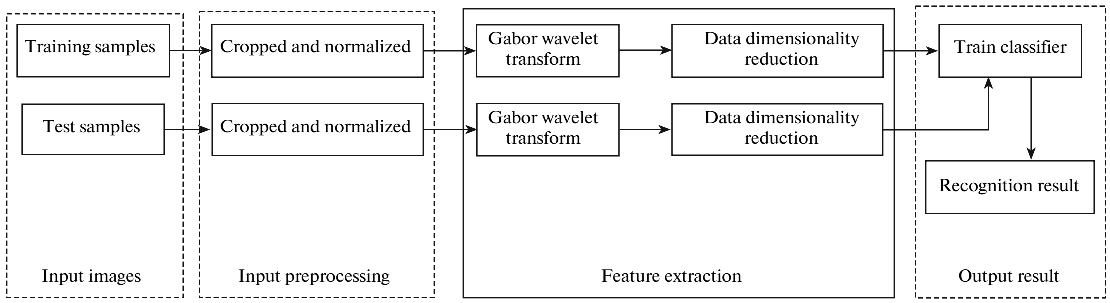


Fig. 2. Facial expression recognition process proposed.

features through the gray information in different scales and orientation. While LBP responses it with the changes of gray in local area. Common methods of classification are based on the distance, such as support vector machine [9] and nearest neighbor classifier. The support vector machine must search for optimal classification surface. Nearest neighbor classifier (KNN) just needs the calculation of distance between different elements.

In the image preprocessing, the area containing expression is gained and normalization of scale and gray is conducted. In the feature extraction, various expression features are extracted using measures based on images or gray values. In the classification and identify, classifier is trained to finish the classification and the results are outputted.

3. FACIAL EXPRESSION RECOGNITION BASED ON GABOR AND NMF

Facial expression is recognized by combining the Gabor wavelet transform with non-negative matrix factorization in this paper. Nearest neighbor classifier and Voting Statistics is adopted to complete the two-layer classification. By doing experiments on JAFFE database, the recognition rate of anger, happiness, sadness and surprise reach 100%. The process of facial expression recognition presented in this paper is shown by Fig. 2.

3.1. Image Preprocessing

Feature information for expression recognition focus on facial area. Hair, accessories and other background factors have adverse effects on expression recognition. For this reason, facial area containing all face organs should be cropped. The manual way is adopted in the paper. But, the sizes of the images after cropping have some differences, which make the dimensionalities are inconsistent. The processes of feature extraction and image recognition will be influenced by above reason. Hence, it is necessary to normalize the images and all of them are normalized to 64×64.

Under natural condition, the problem of uneven illumination is inevitable, which makes blurred situation appeared in a part of images. In order to reduce the impact from environmental factors, histogram equalization is used to enhance the contrast and sharpness of images so that more effective information can be extracted. The preprocessing of image is show by Fig. 3.

3.2. Expression Feature Extraction

In feature extraction based on image, different expressions are reflected through the changes of face organ scale. It is needed to locate the feature region accurately. Whereas, there are some difficulties in positioning exactly. Therefore, feature extraction based on gray-scale is widely used.

3.2.1. Gabor wavelet transform. In this paper, Gabor wavelet transform is used to complete the feature extraction. It is a method based on gray and belongs to the Fourier transform with a Gaussian window [10]. Gabor wavelet transform denotes the image features by gray information in different scales and orientation. The information carried by each pixel is described by a group of data. Compared with other methods, Gabor wavelet transform can reflect the local imperceptible changes of the gray more effective. As Gabor transform has the ability to describe the local gray changes, it is not sensitive to the changes of illumination and position. It also has good robustness.

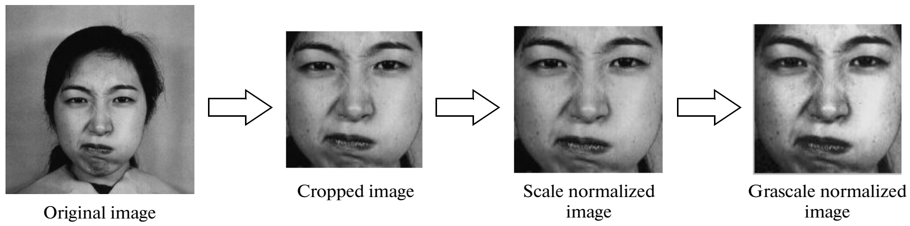


Fig. 3. Image preprocessing.

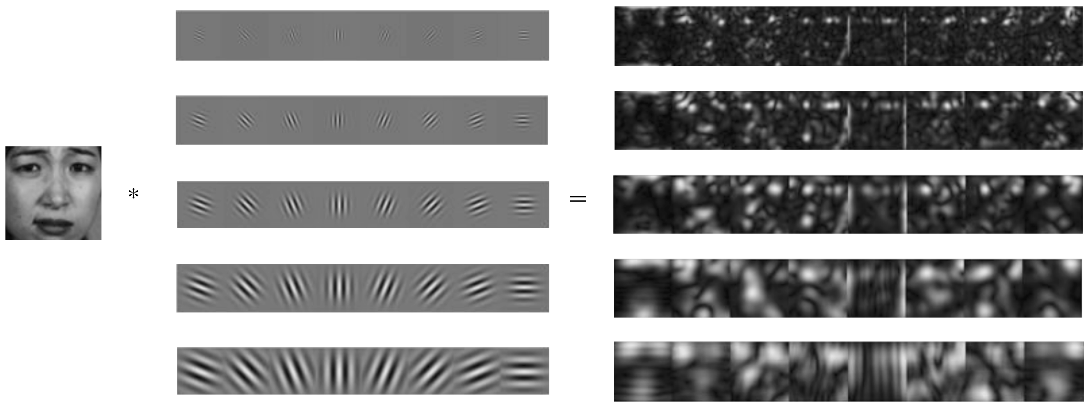


Fig. 4. The filtering process of expression image with 5-scale and 8-orientation Gabor filter bank.

Gabor wavelet kernel function is defined as follows:

$$\varphi^{j}_{v,\mu}(z) = \frac{\|k^{j}_{v,\mu}\|^2}{\sigma^2}\exp\left(-\frac{\|k^{j}_{v,\mu}\|^2\|z\|^2}{2\sigma^2}\right) \times\left[\exp(ik^{j}_{v,\mu}z) - \exp\left(-\frac{\sigma^2}{2}\right)\right]. \quad (1)$$

Among them, i is the complex operator v and μ determine scales and orientations of the filter. Different filters consist of different v and μ. $z = (x, y)$ expresses the coordinates of image. $k_{v,\mu} = k_v e^{i\phi_\mu}$ is center frequency, where $k_v = \frac{k_{\max}}{f^v}$, $\phi_\mu = \mu\frac{\pi}{8}$. $k_{\max} = \frac{\pi}{2}$ is the maximum frequency, and $f = \sqrt{2}$ is the scale factor. σ decides the bandwidth of filters, and it can be defined as $\sigma = 2\pi$.

The Gabor feature value of the image at the $z = (x, y)$ is the convolution of the gray value and Gabor filter $\varphi_{v,\mu}(x, y)$:

$$G(x, y) = I(x, y)*\varphi_{v,\mu}(x, y). \quad (2)$$

Among (2), "*" is the convolution operator. The process of filtering is shown in Fig. 4.

In the image feature extraction, in order to obtain a wealth of expression information, $v = 0, 1, 2, 3, 4$ and $\mu = 0, 1, 2, 3, 4, 5, 6, 7$ are selected in this paper.

3.2.2. Data dimensionality reduction. Non-negative matrix factorization theory is presented by Lee and Seung in Nature in 1999. Under the condition that all the elements in the matrix are non-negative and the sum of any row is not zero, the matrix can be decomposed without negative values appearance. All gray values in the image matrix are non-negative and the gray value in the subsequent image gained after decomposition should also meet the non-negativity. In the result of matrix factorization by principal component analysis, some elements in the basis matrix are negative. It doesn't conform to the characteristic that the gray value of image is non-negative. Compared with principal component analysis, non-negative matrix factorization makes the image decomposition more visualization and makes it easier to interpret and understand the physical significance. In addition, the basis matrix and decomposition coefficient are calculated by multiple iterations which can not only verify the convergence of data but also take up less storage space. The principal component analysis needs to calculate the covariance matrix of the image matrix, which requires a lot of memory space and equipments with high standard. In a word, non-negative matrix factorization has the advantages of simple realization and less storage space [11].

NMF is a multivariate analysis method. Non-negative matrix $V_{m\times n}$ is decomposed:

$$V_{m\times n} = W_{m\times r}H_{r\times n}, \quad (3)$$

where $V_{m\times n}$ is training matrix, $W_{m\times r}$ is basis image matrix, and $H_{r\times n}$ is coefficient matrix. Each image is a column vector of $V_{m\times n}$, which consists of n m-dimen-

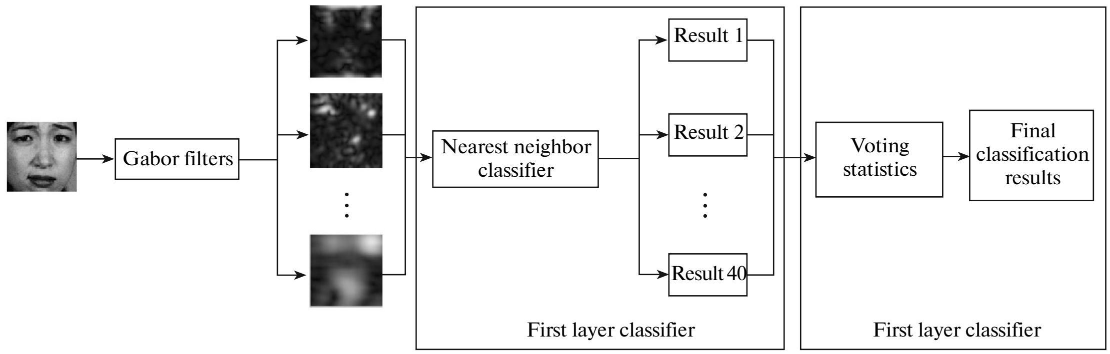


Fig. 5. Two-layer classifier.

sional images. Thus, each image is a linear combination of basis images.

The number of feature vectors must satisfy the following conditions [12]:

$$r < \frac{mn}{m+n}. \tag{4}$$

NMF iteration rules are as follows:

$$W_{ik} \longleftarrow W_{ik} \sum_{j=1}^{m} \frac{V_{ij}}{(WH)_{ij}} H_{ij}, \tag{5}$$

$$W_{ik} \longleftarrow \frac{W_{ik}}{\sum_{p=1}^{n} W_{pk}}, \tag{6}$$

$$H_{kj} \longleftarrow H_{kj} \sum_{i=1}^{n} W_{ik} \frac{V_{ij}}{(WH)_{ij}}. \tag{7}$$

By the iterative calculation, feature subspace composed by the basis images can be achieved. $H = (W^TW)^{-1}W^TV$ can be obtained from $V = WH$. Then, the sample images are projected into feature space, finishing the reduction of data dimension.

In the feature extraction, $v = 0, 1, 2, 3, 4$ and $\mu = 0, 1, 2, 3, 4, 5, 6, 7$ are selected, constituting 40 different filters with 5-scale and 8-orientation. Each image becomes 40 images after filtering by Gabor filters. If the 40 images as a column vector are put into the training matrix, the dimensionalities of the matrix is $64 \times 64 \times 40$. The exorbitant dimensionality increases the complexity of iterative computation. In this paper, the images filtered by one filter are decomposed respectively. The dimensionality is reduced and the operation efficiency is improved.

3.3. Classifier Design

In the recognition process, the nearest neighbor classifier is trained for classification firstly. The filtered images are classified respectively with the nearest neighbor classifier. One image is judged for many times so as to reduce the probability of erroneous judgment. Since one image becomes 40 images after filtering by Gabor filter bank, each image will has 40 test results. Further judgment is taken for the 40 test results by counting the number of every class. The class with the largest numbers is the final results. This is a two-layer classification. Through the two-layer classification, arithmetic robustness can be improved to a certain extent. Classifier design is shown in Fig. 5.

Filtered images as input images of the first layer classifier input into classifier, obtaining the results of the first layer classification. Then, the result as the input data of the second layer classification input to second classifier, achieving the final results.

4. ALGORITHM DESCRIPTIONS

Assuming $G(v, u)$ is the filter bank, where $v = 0, 1, 2, 3, 4$ and $\mu = 0, 1, 2, 3, 4, 5, 6, 7$. $G_i(j)$ represents the ith image filtered by the jth filter. $i = 1, 2, ..., 213$ is the number of the sample sets and $j = 1, 2, ..., 40$ is the number of the filters. So $g(j) = \{G_1(j), G_2(j), ..., G_{213}(j)\}$ expresses the image set filtered by the jth filter. $V(j)$ is the training matrix composed by $g(j)$. $r(j)$ is the classification result of this image set. The procedure of facial expression recognition based on Gabor wavelet transform and non-negative matrix factorization is described as follows:

Input: Images of JAFFE facial expression database, and a total of 213.

Output: Classification results of each image.

Begin

Step 1: 213 images are filtered respectively with one filter, obtaining a filtered image set $g(1) = \{G_1(1), G_2(1), ..., G_{213}(1)\}$.

Step 2: The images in $g(1)$ as column vectors are successively put into the training matrix $V(1)$, and the matrix $V(1)$ is decomposed. The feature vector of each image is gained by the feature subspace projection.

Step 3: Nearest neighbor classifier is trained for classification, achieving classification results $r(1)$.

Step 4: While $j \leq 40$, step 1, step 2, and step 3 are looped.

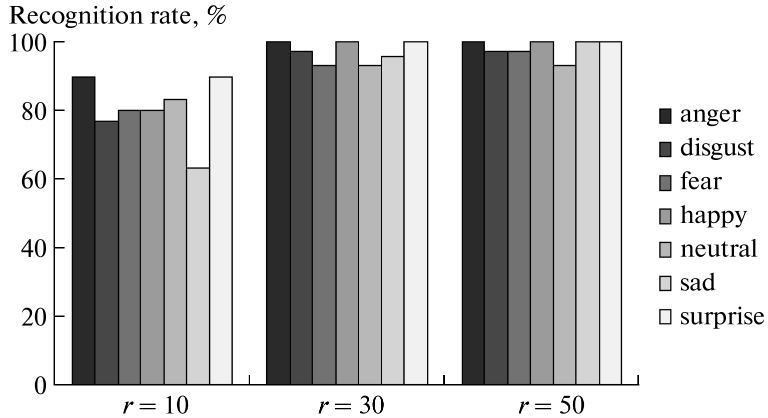


Fig. 6. Test result with different r.

Step 5: Votes for 40 results of each image are counted. The class with the largest number of votes is the final test result.

End

5. EXPERIMENTAL RESULTS AND ANALYSIS

Experiment is implemented at MATLAB7.8.0 (R2009a) programming environment. Japanese female facial expression database (JAFFE) is tested. There are 213 images of 10 female. The size of each image is 256 × 256. It contains seven kinds of basic expressions (anger, disgust, fear, happy, neutral, surprise and sad). Each expression has 2–4 images.

The experimental data is divided by two steps, including vertical and transverse division. Firstly the vertical division is conducted. In the feature extraction, the images filtered by one filter are divided into the same group, and there are 40 groups. The vertical division of data has come true. Each group of vertical division is further divided into 3 groups so as to finish the transverse division. In the first group, the first image of each expression of each person is selected as a test image, and the other images are regard as the training images and so on. In the third group, if the expression only has two images, the first one is chose to test and the rest are the training images. So that each test set has 70 images and the training set has 143 images. Cyclic testing is conducted, implementing the cross validation.

During the decomposition, the number of feature cardinality decides the dimension of the sample images after projecting into the feature subspace, which directly affects the image classification. Thus, different r is selected for testing. When r equals 10, 30 and 50, the recognition rates of various expressions are shown in Fig. 6.

Figure 6 shows that the recognition rate increases gradually with the increasing of r. When r increases from 10 to 30, the recognition rates raise obviously. However, when r changes from 30 to 50, it is not significantly improved. This shows that if a smaller r is chose, the feature vector only reflects the common characteristics of the image and ignores some important information reflecting the class. If a larger r is selected, the feature vector has a large number of redundant information and noise information, which have little contribution but increase the amount of computation. Therefore, the choice of r plays a vital role in the final judgment.

In order to verify the effectiveness of the method, experiments on different methods were carried out as following. The images are directly decomposed by NMF (Image + NMF). Images filtered by a group of filters are directly decomposed by NMF (Gabor + NMF). Images filtered by one filter are decomposed with PCA and classified with two-layer classifier (Gabor + PCA + TLC). Images filtered by one filter are decomposed by NMF and classified with two-layer classifier (Gabor + NMF + TLC). The results are shown in Table 1.

The average recognition rate in the Table 1 is the mean value of various expressions. Table 1 shows that the method proposed in this paper can effectively improve the recognition rate.

CONCLUSIONS

Images in JAFFE facial expression database are few. The number is increased after Gabor filtering, which

Table 1. Result of different algorithms

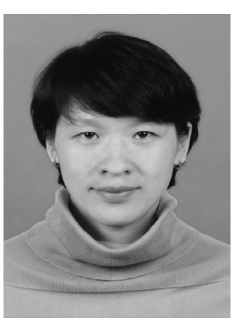
Methods	Image + NMF + KNN	Gabor + NMF + KNN	Gabor + PCA + TLC	Gabor + NMF + TLC
Recognition rate, %	83.7	93.7	96.1	98.1

eases the insufficiency of samples. Image matrix decomposing with NMF algorithm is more interpretable, and each image can be seen as a set of linear superposition of the basis images. In the process of non-negative matrix factorization, appropriate choice of r is very important, which not only extracts the characteristics responding classes information but also reduces the interference caused by redundant information and noise information. Testing images filtered by one filter respectively and accomplishing the second classification by recording the votes are accordant with the idea of probability and statistics. To a certain extent, it improves the robustness of the algorithm. All images used in the experiments are static images, but the facial expression is always changing at every moment. So it still has limitations to track and identify dynamic expressions, which is also the next research work.

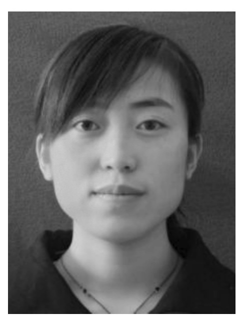
REFERENCES

1. Bin Jiang, Kebin Jia, and Guosheng Yang, "Research advance of facial expression recognition," Comput. Sci. **38** (4), 25–31 (2011).
2. Weifeng Liu, Caifeng Song, Yanjiang Wang, and Lu Jia, "Facial expression recognition based on gabor features and sparse representation," in *Proc. 12th IEEE Int. Conf. on Control Automation Robotics and Vision (ICARCV)* (Guangzhou, 2012), pp. 1402–1406.
3. Xiaoli Li, Qiuqi Ruan, and Chengxiong Ruan, "Facial expression recognition with local gabor filter, in *Proc. 10th IEEE Int. Conf. on Signal Processing (ICSP)* (Beijing, 2010), pp. 1013–1016.
4. Shi Dongcheng and Jiang Jieqing, "The methos of facial expression recognition based on DWT-PCA/LDA," in *Proc. 3rd IEEE Int. Congress on Image and Signal Processing (CISP)* (Yantai, 2010), pp. 1970–1974.
5. Zhiguo Niu and Xuehong Qiu, "Facial expression recognition based on weighted principal component analysis and support vector machines," in *Proc. 3rd IEEE Int. Conf. on Advance Computer Theory and Engineering (ICACTE)* (Berlin, 2010), pp. V3-174–V3-178.
6. Guoyi Zhang and Zilu Ying, "Facial expression recognition based on non-negative matrix factorization," J. Wuyi Univ. (Nat. Sci. Ed.) **23** (3), 28–32 (2009).
7. Nazil Perveen, Shubhrata Gupta, and Kesari Verma, "Facial expression recognition using facial characteristic points and Gini index," in *Proc. IEEE Students Conf. on Engineering and System (SCES)* (Allahabad, 2012), pp. 1–6.
8. Wencheng Wang, Faliang Chang, Jianguo Zhao, and Zhenxue Chen, "Automatic facial expression recognition using local binary pattern," in *Proc. 8th IEEE Word Congress on Intelligent Control and Automation(WCICA)* (Jinan, 2010), pp. 6375–6378.
9. Zilu Ying, Jinghai Tang, Jingwen Li, and Youwei Zhang, "Support vector discriminant analysis and its application to facial expression recognition," Acta Electron. Sinica **36** (4), 725–730 (2008).
10. Xiaomin Liu and Yujin Zhang, "Facial expression recognition based on Gabor histogram feature and MVBoost," J. Comput. Res. Develop. **44** (7), 1089–1096 (2007).
11. Weixiang Liu, Nanning Zheng, and Qubo You, "The application of non-negative matrix factorization in pattern recognition," Chinese Sci. Bull. **51** (3), 241–250 (2006).
12. Yuan-Hsiang Chang and Gong-Yun Jheng, "Customizable facial expression recognition using non-negative matrix factorization," in *Proc. IEEE Int. Conf. on Electrical and Control Engineering (ICECE)* (Yichang, 2011), pp. 2088–2091.

Jun Zhou, professor. Born in 1966. Received M. Sc. Degrees from the Northeast Normal University and received Ph.D. degrees from the Northeastern University. Presently, she is a professor in Liaoning University of Technology. Scientific interest: artificial intelligence, data mining, and image processing.



Hongyan Mei. Born in 1978. Received M. Sc. degrees from the Liaoning University of Technology and received Ph. D. degrees from the Beijing University of Posts and Telecommunications. Scientific interest: data mining and big data analysis.



Sue Zhang. Born in 1987. Graduate from Xinzhou Teacher University, in 2007. Presently, she is a postgraduate at School of Electronics and Information Engineering, Liaoning University of Technology. Scientific interest: image processing and pattern recognition.

Dawei Wang. Born in 1988. Graduated from Shan Dong Jiaotong University. Presently, he is a postgraduate at School of Electronics and Information Engineering, Liaoning University of Technology. Scientific interest: image processing, data mining and big data analysis.