

# Facial Feature Recognition based on ASNMF Method

Jing Zhou<sup>1\*</sup>, Tianjiang Wang<sup>2</sup>

<sup>1</sup> School of Mathematics and Computer Science, Jiangnan University  
Wuhan, 430056 China  
[e-mail: zhj131@jhun.edu.cn]

<sup>2</sup> School of Computer Science and Technology, HuaZhong University of Science and Technology, Wuhan Hubei  
430074, China

[e-mail: tjwang@hust.edu.cn]

\*Corresponding author: Jing Zhou

*Received February 3, 2019; revised May 24, 2019; accepted June 23, 2019;  
published December 31, 2019*

---

## Abstract

Since Sparse Nonnegative Matrix Factorization (SNMF) method can control the sparsity of the decomposed matrix, and then it can be adopted to control the sparsity of facial feature extraction and recognition. In order to improve the accuracy of SNMF method for facial feature recognition, new additive iterative rules based on the improved iterative step sizes are proposed to improve the SNMF method, and then the traditional multiplicative iterative rules of SNMF are transformed to additive iterative rules. Meanwhile, to further increase the sparsity of the basis matrix decomposed by the improved SNMF method, a threshold-sparse constraint is adopted to make the basis matrix to a zero-one matrix, which can further improve the accuracy of facial feature recognition. The improved SNMF method based on the additive iterative rules and threshold-sparse constraint is abbreviated as ASNMF, which is adopted to recognize the ORL and CK+ facial datasets, and achieved recognition rate of 96% and 100%, respectively. Meanwhile, from the results of the contrast experiments, it can be found that the recognition rate achieved by the ASNMF method is obviously higher than the basic NMF, traditional SNMF, convex nonnegative matrix factorization (CNMF) and Deep NMF.

---

**Keywords:** scale normalization, improved iterative step sizes, improved iterative rules, threshold sparse, facial feature extraction, ASNMF method

## 1. Introduction

**F**acial feature recognition is a complex task due to the variation of human appearance with age, race, gender and obstacles (such as glasses and masks), which affects the accuracy of feature extraction, thus reduces the recognition rate. And the more accurate the extracted facial features are, the higher the recognition accuracy will be. Therefore, the research of feature extraction methods is critical for the facial recognition. At present, facial feature extraction methods mainly include the appearance and geometry feature extraction, deformation and motion feature extraction, global and local feature extraction. Appearance features include image density, edge, texture and some more distinctive features, and the common appearance-based feature extraction methods include Local Binary Pattern (LBP) method [1], Linear Discriminant Analysis (LDA) [2] and directional gradient histogram method [3]. The appearance feature extraction methods are usually simple and direct, and can be fused with other methods to improve the performance. Such as in literature [1], the LBP-based feature is fused with the color information feature, and can measure the similarity of color images with rich color information. Literature [2] fused the LDA and principal component analysis methods at rank level using borda count method, thus the recognition accuracy over individual face representations is significant improved. Geometric method of feature extraction can recognize facial expressions by measuring facial geometric features, such as facial reference points, distance and curvature of variable significant areas, etc. The recognition rate of geometric method alone will decrease when the number of samples increases, thus the improvements are proposed recently. Such as in literature [4], based on the geometric feature extraction method, a simple and novel feature descriptor for face recognition called local diagonal extreme number pattern is proposed. With the new geometric feature descriptor, the recognition rate can be improved.

Besides the above methods for facial feature extraction, the deformation feature extraction methods such as the point distributed model [5] and motion module method [6], are also proposed to extract features based on the deformation of facial sense organs under different expressions. Meanwhile, the feature extraction method based on motion is also proposed to judge facial expression according to the movement trend and direction of facial feature points. The representative method is the optical flow method [7]. These methods are widely adopted to extract the features of video image sequence to track the facial expression region. Since the computational complexity of these methods is large, the local feature extraction methods have attracted much attention in recent years. For local feature extraction method [8], the five senses are separated and treated differently according to their relative importance. For example, the changes of eyes and mouth are more obvious than the nasal region when facial expression changes, so the proportion of the nose can be reduced in the analysis. Typical methods of local feature extraction are the local principal component analysis [9] and Gabor wavelet transform [10]. However, local feature extraction methods can only extract local features, which results in inaccurate accuracy. Therefore, global feature extraction method is introduced. For global feature extraction method, the scope of global extraction is all the pixels in the image, and then the whole feature of the image can be extracted. Usually, the global feature of an image is the reflection of the macro information of the image, which is advantageous in the representation of the approximate contour of the facial image, thus it achieved significant performance for facial recognition. Principal component analysis [11], independent component analysis [12] and nonnegative matrix factorization (NMF) [13] are the most widely adopted global methods at present,

among which the NMF method is more explanatory and then suitable for facial feature extraction than others [14][15].

Nonnegative matrix factorization is realized under the condition that all elements of the matrix are nonnegative, which is always adopted for dimension reduction and feature extraction [16][17]. And the improved NMF algorithms have been proposed, such as the SNMF method [18], CNMF method [19] and Deep NMF method [20]. The SNMF method considers the redundant information hidden in the complex data by adding the L-norm sparse constraints in the iteration rules. CNMF method is derived from Semi-NMF [21], and it replaces the basis matrix with the non-negative convex combination of the original matrix. And the Deep NMF method is to stack one-layer NMF into multiple layers to learn the hierarchical relationships among features or hierarchical projections. Among these NMFs, the SNMF can control the sparsity [22] of the decomposed matrix, thus it can be adopted to control the sparsity in facial feature extraction to improve the recognition rate. But the traditional SNMF method is iterated by the multiplicative iteration rules, and the improvements for the iteration step sizes are not considered. Actually, it is difficult to improve the recognition accuracy when the iterative step sizes are not appropriate. Therefore, the new iterative step sizes are proposed in this paper to improve the iterative rules, and then a new additive sparse nonnegative matrix factorization method can be derived, whose recognition rate is obviously higher than basic NMF method, SNMF method, CNMF method and Deep NMF method.

## 2. Image Preprocessing

The captured facial image will be affected by the performance of the image acquisition equipment or the change of illumination, which causes the light and shade difference and noise, and then the target image is different from the actual image data. Therefore, the acquired facial images should be pre-processed to improve image quality, remove noise and normalize image size, and are prepared for later feature extraction and classification.

### 2.1 Conversion of color face to gray image

Since the color image is sensitive to light source, it is easy to be affected by illumination. Direct use of color image will have negative impact on facial expression recognition. Thus, the color face will be converted to gray image firstly. The color images in CK+ facial expression dataset are represented by RGB color model, which can represent pixel value of an image by three color components, that is, red(R), green(G) and blue(B). The range of values for each color component is 0-255, and then the gray value can be calculated by the weighted average value of three color components. Since the human eye is more sensitive to green and less sensitive to blue, we can give the following weighting formula as Eq. (1) to get the converted gray value, and the effect is shown in Fig. 1.

$$Gray = 0.299 \times R + 0.578 \times G + 0.114 \times B \quad (1)$$



(a) original color facial image (b) converted gray facial image

**Fig. 1.** Conversion from facial color image to gray image

## 2.2 Face detection and scale normalization

The images in CK+ dataset are captured from videos, which carry lots of interference information adverse to recognition, such as hair, clothes, subtitle of video, etc. Meanwhile, uncertainty of face location will also reduce the accuracy of facial feature recognition. Thus, the face region will be detected first. The adaboost cascade classifier [23] is adopted in this paper to detect the face of the original face image, get the face part of the image and remove the complex background, text and other interference information. However, the size inconsistency of the detected face will also affect feature extraction and classification, thus it is necessary to normalize the detected faces with the operations of translation, scaling and rotation. Then the size of the resulting face image is  $276 \times 200$ , and the face is adjusted to the same angle to achieve face alignment. The results of face detection and scale normalization are shown in Fig. 2.

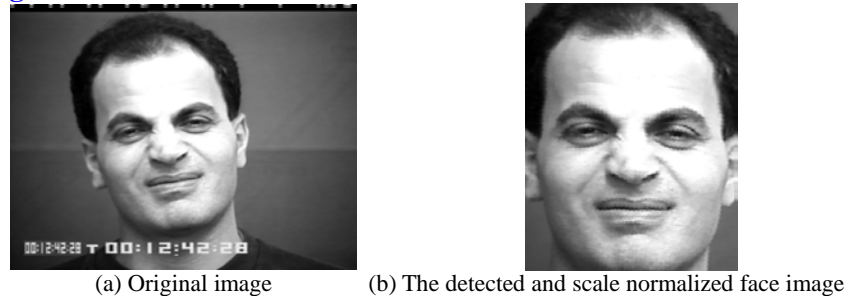


Fig. 2. The results of face detection and scale normalization

## 2.3 Histogram equalization

The uneven illumination causes difficult extracting for the texture features of detected face and then the accuracy of facial feature recognition will decrease. Therefore, it is necessary to adopt the histogram equalization method to equalize the illumination, and make pixel values of image uniformly distributed at each gray scale. Histogram equalization makes the image gray histogram evenly distributed by stretching the gray scale contrast of samples, and then the detailed contrast of the image can be enhanced.

The implementation steps of histogram specification are as follows:

- (1) Calculating the numbers of pixels of image with different gray values  $v_q$ ,  $q=0,1,\dots,L-1$
- (2) Computing the gray histogram  $p_q=v_q/v$ , where  $v$  is the total number of pixels for image to be processed

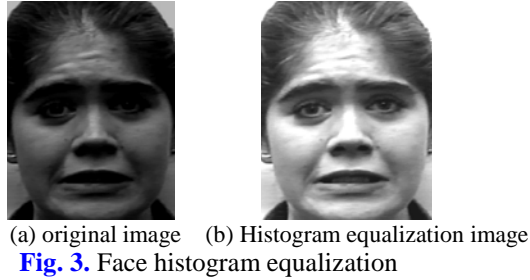
- (3) Calculating cumulative distribution function  $s_q = \sum_{l=0}^q p_l$

- (4) Calculating the transformed gray scale  $u = \text{int} \left[ (L-1)s_q + 0.5 \right]$

- (5) Calculating the numbers of pixels of the transformed image  $n_u$ ,  $u=0,1,\dots,L-1$

- (6) Calculating the gray histogram of the transformed image  $p_u=v_u/v$ ,  $u=0,1,\dots,L-1$ .

Histogram equalization was applied to a face image of the CK+ face expression dataset, and the result is shown in Fig. 3.



## 2.4 Low-frequency information extraction by wavelet transform

The wavelet transform can be adopted to extract the low-frequency feature from an image. Two-dimensional mother wavelet function  $\varphi(x, y)$  can be constructed by Eq. (2), and a set of functions depending on parameter  $a, b, c$  is the wavelet basis function.

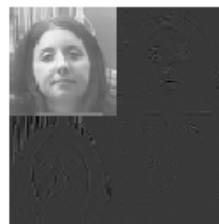
$$\varphi_{a,b,c}(x, y)dt = |a|^{-1} \varphi\left(\frac{x-b}{a}, \frac{y-c}{a}\right), b, c \in \mathbb{R}, a \in \mathbb{R} - \{0\} \quad (2)$$

The coiflets basis wavelet is a family of orthogonal wavelets based on daubechies wavelet, whose scale function  $\varphi(x)$  and wavelet function  $\psi(x)$  with  $N$ -order vanishing distance satisfy the following basic properties shown as Eq.(3) and Eq.(4). Meanwhile, coiflets wavelet not only retains the two-scale relation, orthogonality and smoothness of daubechies wavelet, but also has a generalized vanishing distance of its scale function  $\varphi(x)$ , as shown in Eq. (4).

$$\begin{cases} \varphi(x) = \sqrt{2} \sum_{r=0}^{3N-1} p_r \varphi(2x-r) \\ \psi(x) = \sqrt{2} \sum_{r=0}^{3N-1} (-1)^r p_{3N-1-r} \varphi(2x-r) \end{cases} \quad (3)$$

$$M_n = \int (x-M)^n \varphi(x) dx = \delta_{0,n}, n = 0, 1, \dots, N-1 \quad (4)$$

Here,  $M_n$  is the vanishing distance of the scale function  $\varphi(x)$  and  $p_k$  is the filter coefficient. To satisfy the Eq.(3), the filter coefficient  $p_r$  increases from  $2N$  to  $3N$ , and the compact support of the corresponding scaling function and wavelet function also increases from  $[0, 2N-1]$  to  $[0, 3N-1]$ . In fact, the larger the vanishing moment, the support length is longer, and the corresponding filter is flatter. Thus, the coiflets wavelet has better compact support and vanishing moments in both time domain and frequency domain, and the ability to approximate and reconstruct smooth signal is stronger. Meanwhile, coiflets wavelet has other good properties, such as orthogonality, time-frequency compact support, smoothness, fast algorithm of wavelet and so on. Moreover, the scale function of coiflets wavelet has quasi-interpolation property, so its scale expansion coefficient can be obtained directly by sampling the signal at a single point, thus avoiding the numerical integration operation needed in daubechies wavelet analysis. So, the coiflets wavelet transform is chosen to extract the low-frequent feature, and the extracted low-frequency feature is in the upper left corner of **Fig. 4**.



**Fig. 4.** The extracted low-frequency of coiflets wavelet transform

### 3. ASNMF with Improved Additive Iterative Rules

The traditional SNMF method decomposes a nonnegative matrix  $V$  of high dimension into the product of two low-rank nonnegative matrices  $W$  and  $H$ , and the L1-norm is adopted to sparse the matrices  $W$  and  $H$ . The matrix  $W$  is the basis matrix and  $H$  is the coefficient matrix. To improve the accuracy of the SNMF method, the new iteration step sizes will be proposed.

#### 3.1 The improved additive iteration rules for SNMF method

The nonnegative matrix  $V$  is a linear mixture model with additive noise, and the objective function of SNMF is defined as Eq. (5).

$$J(W, H) = \frac{1}{2} \sum_{ij} \|V - WH\|^2 + \alpha \|W\| + \beta \|H\| \quad (5)$$

The iterative rules described as Eq.(6) and Eq.(7) can be obtained by using gradient descent method.

$$W'_k = W_k - \mu \frac{\partial J}{\partial W_k} \quad (6)$$

$$H'_{kj} = H_{kj} - \lambda \frac{\partial J}{\partial H_{kj}} \quad (7)$$

The  $\frac{\partial J}{\partial W_k}$  and  $\frac{\partial J}{\partial H_{kj}}$  can be solved by computing the partial derivatives of Eq. (5). Substituting them into the Eq. (6) and Eq.(7), the iteration rules described as Eq. (8) and Eq. (9) can be obtained.

$$W'_k = W_k + \mu \left[ (VH^T)_{ik} - (WHH^T)_{ik} - \alpha I_{ik} \right] \quad (8)$$

$$H'_{kj} = H_{kj} + \lambda \left[ (W^T V)_{kj} - (W^T WH)_{kj} - \beta I_{kj} \right] \quad (9)$$

Set the improved iteration step sizes  $\mu$  and  $\lambda$  as Eq. (10) and Eq. (11), however, the original iteration step sizes  $\mu'$  and  $\lambda'$  of traditional SNMF method are described as Eq. (12) and Eq.(13).

$$\mu = \frac{W_k}{(WHH^T)_{ik}} \frac{(VH^T)_{ik}}{(VH^T)_{ik} + (WHH^T)_{ik}} \quad (10)$$

$$\lambda = \frac{H_{kj}}{(W^T WH)_{kj}} \frac{(W^T V)_{kj}}{(W^T V)_{kj} + (W^T WH)_{kj}} \quad (11)$$

$$\mu' = \frac{W_k}{(WHH^T)_{ik}} \quad (12)$$

$$\lambda' = \frac{H_{kj}}{(W^T WH)_{kj}} \quad (13)$$

The  $\mu$  and  $\lambda$  also satisfy the gradient iteration descending criterion, which can definitely decrease the target function described as Eq. (5), and the convergence proof will be given later.

Substituting the new iteration step sizes described as Eq.(10) and Eq.(11) to Eq.(8) and Eq.(9), respectively, the new additive iteration rules of  $W$  and  $H$  can be derived as shown in Eq. (14) and Eq. (15). The new iteration step sizes described as Eq.(10) and Eq.(11) of the new additive rules as Eq.(14) and Eq.(15) are obviously smaller than the original iteration step sizes described as Eq.(12) and Eq.(13), thus the searching accuracy during the iteration process based on Eq.(14) and Eq.(15) is higher than that of the original SNMF method, and the accuracy can be improved.

$$W_k' = W_k + W_k \frac{(VH^T)_{ik}((VH^T)_{ik} - (VH^T)_{ik} - \alpha l_{ik})}{(VH^T)_{ik}((VH^T)_{ik} + (VH^T)_{ik})} \quad (14)$$

$$H_{kj}' = H_{kj} + H_{kj} \frac{(WV)_{kj}((WV)_{kj} - (WV)_{kj} - \beta l_{kj})}{(WV)_{kj}((WV)_{kj} + (WV)_{kj})} \quad (15)$$

When the target function as Eq. (5) reaches the minimum value, or the number of iterations reaches the upper limit, the algorithm is convergent and terminated, and the optimal  $W$  and  $H$  can be obtained.

The computational complexity for the new iteration rule as Eq.(14) is  $O(N \times (mnk + (mkn + mnk) + mk))$ , that is  $O(N \times (mnk + mk))$ , where  $N$  is the number of iterations to convergence,  $m$  is the number of rows of data  $V$ ,  $n$  is the number of columns of data  $V$  and  $k$  is the rank of decomposition. And the computational complexity for the new iteration rule as Eq.(15) is  $O(N \times (kmn + mk^2 + nk^2 + kn))$ . The original iteration rules of traditional SNMF are described as Eq. (16) and Eq.(17), and the computational complexities of  $W$  and  $H$  are  $O(N \times (mnk + mk))$  and  $O(N \times (kmn + mk^2 + nk^2 + kn))$ , respectively. Therefore, the computational complexities are unchanged when the traditional multiplicative iteration rules of the SNMF model is substituted to the new additive iteration rules as Eq.(14) and Eq.(15).

$$W_k = W_k \frac{(VH^T)_{ik} - \alpha l_{ik}}{(VH^T)_{ik}} \quad (16)$$

$$H_{kj} = H_{kj} \frac{(WV)_{kj} - \beta l_{kj}}{(WV)_{kj}} \quad (17)$$

The proposed new SNMF method based on the new additive iterative rules described as Eq.(14) and Eq.(15) is more precise than the traditional SNMF method due to the smaller iterative step sizes described as Eq.(10) and Eq.(11), which will result in the higher searching accuracy and more precise feature.

### 3.2 Convergence proof

Similar to the NMF method, the objective function  $J(H)$  as Eq.(5) of the SNMF can be described as Eq. (18). Since the second derivative of Eq.(5) for traditional SNMF method is equal to that of the NMF method, the auxiliary function  $G(H, H^t)$  for convergence proof of the traditional SNMF can refer to that of NMF described in literature [13], which is described as Eq. (19) with  $\mathcal{K}(H) = (WVH)_{kj} / H_{kj}$ .

$$J(H) = J(H) + (H - H^t)^T \nabla J(H) + \frac{1}{2} (H - H^t)^T W^T W (H - H^t) \quad (18)$$

$$\mathcal{Q}(H, H^t) = J(H) + (H - H^t)^T \nabla J(H) + \frac{1}{2} (H - H^t)^T \mathcal{K}(H) (H - H^t) \quad (19)$$

Because that  $(H_{kj} - H_{kj}^t)_{kj} [\mathcal{K}(H_{kj}) - W^T W] (H_{kj} - H_{kj}^t) \geq 0$  has been proved in literature [13], the inequality of  $\mathcal{Q}(H, H^t) \geq J(H)$  is proved, and then  $J(H^{t+1}) \leq \mathcal{Q}(H^{t+1}, H^t) \leq \mathcal{Q}(H^t, H^t) = J(H^t)$  can be obtained for the traditional SNMF method, where  $H^{t+1}$  is the minimum of auxiliary function  $G(H, H^t)$ . Thus the  $J(H)$  described as Eq.(5) for the traditional SNMF method is proved to be non-increased. Meanwhile, by iterating the update of  $H^{t+1} = \arg \min_H \mathcal{Q}(H, H^t)$ , a sequence of estimates that converge to a local minimum  $H_{\min} = \arg \min_H J(H)$  of the objective function can be obtained, and then the convergence of traditional SNMF is proved.

To further prove the convergence behavior of the improved SNMF method with new additive rules as Eq.(14) and Eq.(15) proposed in this paper, an auxiliary function of  $Q(H, H^t)$  should be constructed for the objective function  $J(H)$ , which satisfies the conditions of Eq. (20).

$$Q(H, H^t) \geq J(H), Q(H, H) = J(H) \quad (20)$$

Meanwhile,  $H^{t+1}$  is also the minimum of auxiliary function  $Q(H, H^t)$  and can be described as Eq. (21).

$$H^{t+1} = \arg \min_H Q(H, H) \quad (21)$$

From the Eq. (20) and Eq. (21), we can get the Eq. (22).

$$J(H^{t+1}) \leq Q(H^{t+1}, H) \leq Q(H, H) = J(H) \quad (22)$$

Then the objective function  $J(H)$  can be proved to be non-increasing and convergent based on the new additive rules as Eq.(14) and Eq. (15). Thus, the construction of the auxiliary function satisfied Eq. (20) is important.

Refer to Eq.(19), an auxiliary function of  $Q(H, H^t)$  can be constructed for  $J(H)$  as Eq.(23).

$$Q(H, H) = J(H) + (H - H)^T \nabla J(H) + \frac{1}{2} (H - H)^T \Xi(H) (H - H) \quad (23)$$

$$\text{Where } \Xi(H) = ((\bar{W}\bar{W}H)_{kj} / H_{kj}) ((\bar{W}\bar{V} + \bar{W}\bar{W}H)_{kj} / (\bar{W}\bar{V})_{kj}) \quad (24)$$

Since the inequality of  $Q(H, H) \geq J(H)$  has been proved, we only need prove  $Q(H, H) \geq \mathcal{Q}(H, H)$ , and then  $Q(H, H) \geq J(H)$  can be proved.

Based on Eq. (23) and Eq. (19), we only need prove  $\Xi(H) \geq \mathcal{K}(H) = (\bar{W}\bar{W}H)_{kj} / H_{kj}$ , then  $Q(H, H) \geq \mathcal{Q}(H, H)$  can be proved.

From Eq. (24), we can derive as following,

$$\begin{aligned} \Xi(H) &= ((\bar{W}\bar{W}H)_{kj} / H_{kj}) ((\bar{W}\bar{V} + \bar{W}\bar{W}H)_{kj} / (\bar{W}\bar{V})_{kj}) \\ &= ((\bar{W}\bar{W}H)_{kj} / H_{kj}) (1 + (\bar{W}\bar{W}H)_{kj} / (\bar{W}\bar{V})_{kj}) \geq (\bar{W}\bar{W}H)_{kj} / H_{kj} = \mathcal{K}(H) \end{aligned} \quad (25)$$

From the Eq. (25), we can get  $Q(H, H) \geq \mathcal{Q}(H, H) \geq J(H)$ , and then  $J(H)$  with the new iteration rules is proved to be non-increasing and convergent based on the updating rule of Eq. (21), that is, the convergence of the new additive iteration rules for new SNMF method can be proved.

Based on the Eq. (21), the iteration rule of  $H$  can be derived as Eq. (26), that means Eq.(15) can be obtained.

$$H_{kj}^{t+1} = \arg \min_H Q(H, H) = H_{kj}^t + H_{kj}^t \frac{(\bar{W}\bar{V})_{kj} ((\bar{W}\bar{V})_{kj} - (\bar{W}\bar{W}H)_{kj} - \beta I_{kj})}{(\bar{W}\bar{W}H)_{kj} ((\bar{W}\bar{W}H)_{kj} + (\bar{W}\bar{V})_{kj})} \quad (26)$$

Similarly, the iteration of  $W$  described as Eq. (27) can be derived, that means Eq. (14) can be obtained.

$$W_{ik}^{t+1} = \arg \min_H Q(W, W) = W_{ik}^t + W_{ik}^t \frac{(VH^T)_{ik} ((VH^T)_{ik} - (WHH^T)_{ik} - \alpha I_{ik})}{(WHH^T)_{ik} ((WHH^T)_{ik} + (VH^T)_{ik})} \quad (27)$$

### 3.3 Threshold sparsity for the basis matrix

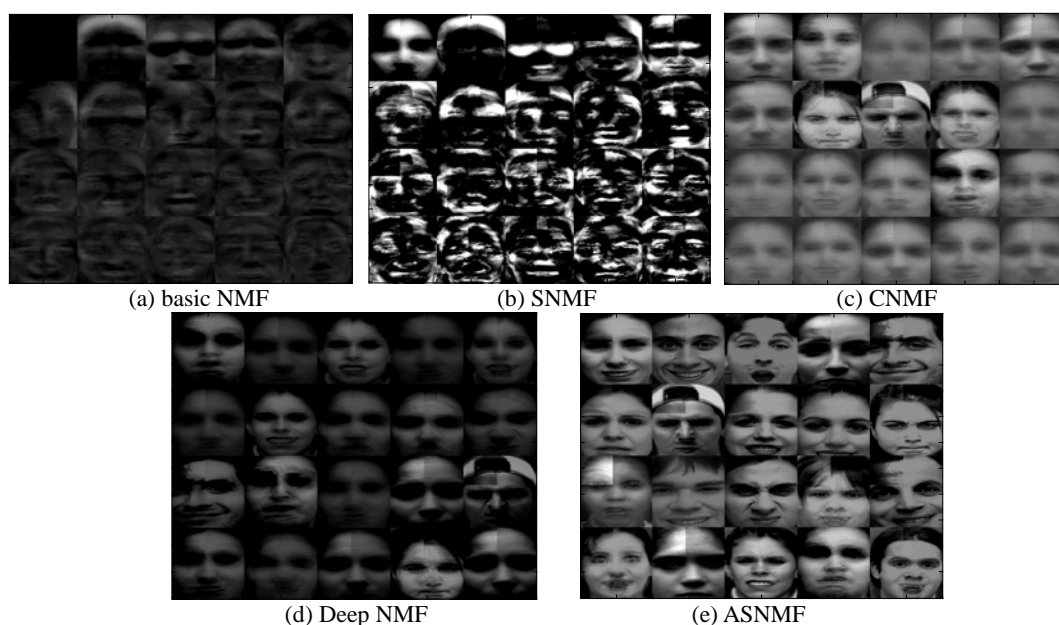
To further improve the recognition rate, the redundant information in the feature data represented by the basis matrix  $W$  can be further reduced. Thus the threshold judgment strategy is adopted to increase the sparsity of the basis matrix  $W$  of the improved additive SNMF method, and then the recognition accuracy can be further improved. The threshold judgment is applied to the basis matrix  $W$  in additive iteration process, that is, a threshold is selected to reset the value of the basis matrix  $W$  while using the Eq. (14) and Eq. (15) for updating the  $W$  and  $H$ . When the  $W$  is produced each iteration, each data of column in the basis matrix  $W$  is reset according to the threshold. If the data is greater than the threshold, it is set to value of one, otherwise the data is set to zero. The basis matrix  $W$  is sparsed to 0 and 1 matrix, which can make the weight of the important feature larger, and the weight of the



less important feature smaller, and then the facial features become more concentrated and easy to be extracted, thus, the accuracy can be improved.

Threshold judgment is introduced from the initial iteration, and the initial values of  $W$  matrix are random values between 0 and 1, thus the threshold can only be set between 0 and 1. From the initial iteration, if the threshold is a large number between 0 and 1, only a small number of values in  $W$  matrix will exceed the threshold, and most of the other values of  $W$  matrix will be set to 0, resulting in most of the original useful feature information in  $W$  matrix will be set to zero. After several iterations, most of the values in  $W$  matrix are 0, resulting in serious loss of features, and then the recognition accuracy will be greatly reduced. Thus, the threshold can't be large number between 0 and 1. A small threshold can ensure that redundant information is eliminated, while retaining most of the useful feature information. And the threshold of 0.01 can get better results in this paper.

The above-mentioned improved additive SNMF method based on the new additive iterative rules and threshold judgment is named ASNMF method. To verify the effectiveness of the new method, the ASNMF, basic NMF, SNMF, CNMF and Deep NMF are applied on the ORL and CK+ face databases respectively, and then the basis matrix images extracted from the training samples, called "feature faces," can be obtained and shown in Fig. 5 (a)-(e), respectively. As can be seen from Fig. 5, the expression contours of "feature faces" of  $W$  matrix obtained by ASNMF method based on the new additive iterative rules proposed in this paper are clearer than other four NMF methods, that means the ASNMF method can express the expression features more clearly.



**Fig. 5.** Feature faces of the basis matrices obtained by the five NMF methods

#### 4. Classification Based On SVM

The classifier should be selected after facial feature extraction. Facial feature classifiers include hidden Markov model (HMM), neural network, support vector machine (SVM), adaboost algorithm and K nearest neighbor algorithm (KNN), etc. SVM classifier has good generalization ability, and by which the strict bound of generalization ability of the model

can be given, which is not available in other learning machines. Meanwhile, compared with other methods, the establishment of SVM classifier requires less prior intervention. In fact, the less human intervention, the more objective the classification results are. Meanwhile, even if the sample is insufficient, SVM also can achieve accurate recognition. Therefore, SVM classifier is adopted to recognize the facial features in this paper.

SVM can classify the facial features of multiple categories with the "one-to-one" and "one-to-many" strategies. Since the results of "one-to-one" strategy are more accurate, it is adopted in SVM in this paper. The  $M$  classes of the samples can be divided into two categories respectively, and then  $M(M-1)/2$  classifiers were constructed. The categories of facial expressions in the CK+ dataset is 8, meanwhile, the categories of faces in the ORL database is 40, and then the "one to one" strategy can be adopted to construct 28 classifiers for CK+ dataset and 780 classifiers for ORL dataset, respectively. And from the following experimental results, the "one-to-one" SVM classifier can correctly classify and recognize facial feature extracted by the ASNMF method for ORL dataset and CK+ dataset, respectively, that is, it can correctly recognize the same expression of different people, and also can recognize the same people of different facial expression.

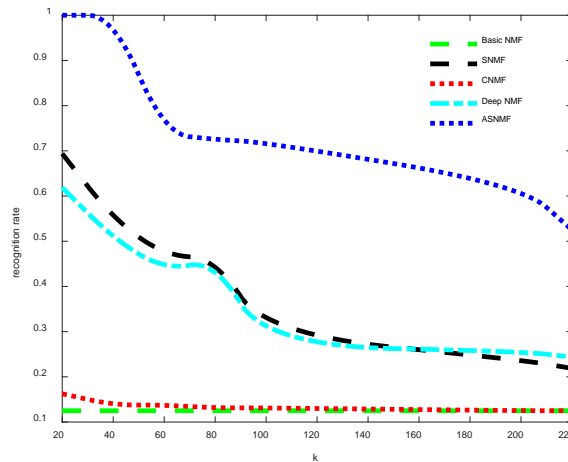
## 5. Experimental Results and Analysis

To illustrate the validity of the proposed ASNMF method, two datasets are chosen for experiments respectively, which are CK+ dataset and the ORL dataset. The CK+ dataset is the extended Cohn-Kanade data provided by the Carnegie Mellon University, which is composed by the image sequences with the emotion labels. The eight emotion labels of CK+ dataset are anger, contempt, disgust, fear, happiness, neutrality, sadness and surprise. In experiments, each expression category contains 40 images of different people, and then there are 320 images of the total eight expressions. The images in CK+ dataset were digitized into  $640 \times 480$  pixel arrays with 8-bit gray-scale or 24-color values. And there are 20 training samples for each expression, that is, there are 160 training samples for the eight expressions, meanwhile, there are 20 test samples for each expression and 160 test samples for eight expressions. The facial expressions and details of each person in the dataset vary in varying degrees. The facial expression posture is also different, and variation of the depth rotation angle and plane rotation angle can reach 20 degrees and the scale of expression reaches 10%. Another dataset in experiments is the ORL dataset provided by Cambridge University. There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). The size of each image is  $92 \times 112$  pixels with 8-bit gray-scale values.

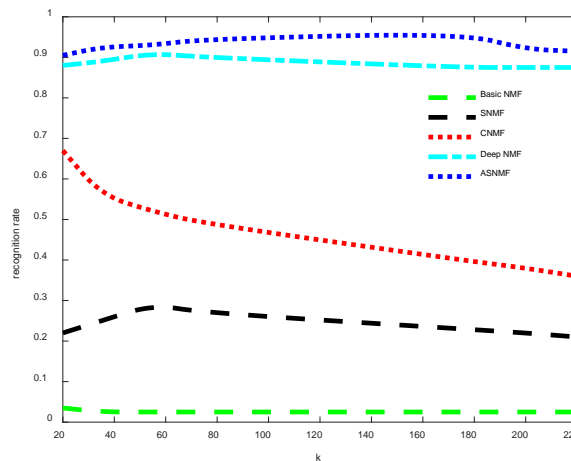
To verify the effectiveness of the ASNMF method by comparison experiments, several NMF methods such as basic NMF, traditional SNMF, CNMF, Deep NMF and the ASNMF method, are all adopted to extract and recognize facial features for CK+ dataset and ORL dataset, respectively. And the results of recognition rates for five NMF methods with different decomposition rank  $k$  for CK+ dataset and ORL dataset are shown in [Fig. 6](#) and [Fig. 7](#), respectively.

As can be seen from [Fig. 6](#), for CK+ dataset, the highest recognition rate can reach 100% by adopting the ASNMF method proposed in this paper when  $k$  is less than 40. Secondly, the recognition rate of the traditional SNMF method reaches 69.375%, which adopts the L1-norm to sparse the basis matrix and coefficient matrix, and the recognition rate decreases with the increase of  $k$  value. Next, the recognition rate of Deep NMF method reaches

61.875% and the CNMF is adopted to extract features, whose recognition rate only reaches 16.25%. Finally the recognition rate with the basic NMF features only reaches 12.5%, and the recognition rate can't be effectively improved by changing the value of rank  $k$ . Therefore, the highest recognition rate of ASNMF method is 30.625% higher than that of SNMF method, 38.125% higher than Deep NMF, 83.75% higher than CNMF and 87.5% higher than basic NMF.



**Fig. 6.** Comparison of the recognition rate of the five NMF methods for CK+ dataset



**Fig. 7.** Comparison of the recognition rate of the five NMF methods for ORL dataset

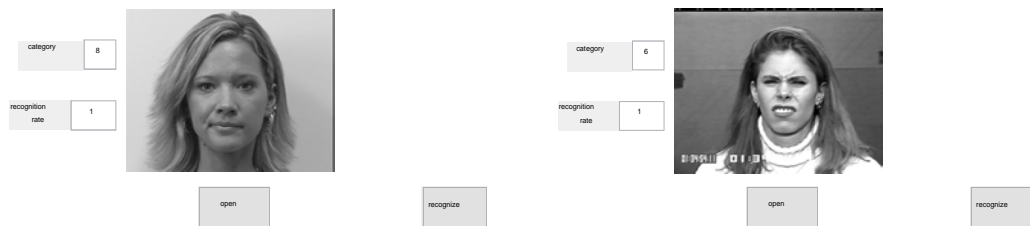
As can be seen from **Fig. 7**, for the ORL dataset, the recognition rate reaches 96% by adopting the ASNMF method proposed in this paper. Meanwhile, the recognition rate of Deep NMF reaches 91%. The recognition rates of CNMF reaches 67% and the recognition rate of SNMF reaches 30%. However, the recognition rate of basic NMF only reaches 3.5%. Therefore, the highest recognition rate of the ASNMF method is 5% higher than Deep NMF, 29% higher than CNMF, 66% higher than SNMF and 92.5% higher than basic NMF.

The recognition rate of ASNMF method based on new additive and sparse iteration rules is higher than other NMF methods because that the search accuracy of the iteration process based on the new smaller iteration step sizes is higher, which results in higher recognition accuracy. Moreover, the threshold judgment is adopted, which make the feature data

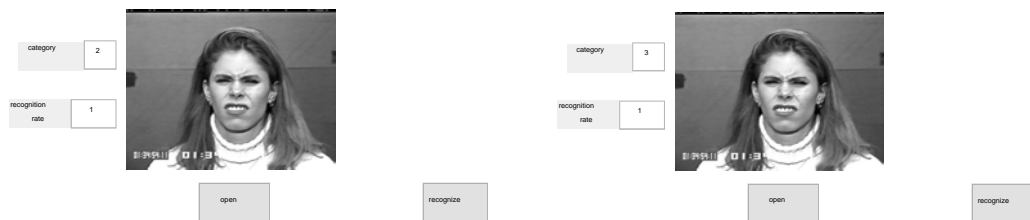
obtained by the new additive iterative rules sparser and more concentrated than the feature data obtained by the traditional multiplicative iterative rules in other four NMF methods. Since the extracted facial feature is more concentrated and easier to recognize, a higher recognition rate based on the new additive and sparse iterative rules can be achieved. High recognition rate means accurate feature extraction, so the feature extracted by ASNMF method is more accurate than other methods, and then the error of the reconstruction based on the feature basis matrix of ASNMF is smaller.

Based on ASNMF method, one facial feature recognition software is realized, which can recognize the facial features of CK+ dataset and ORL dataset, respectively. To achieve the highest recognition rate 100% for CK+ dataset, the value of decomposition rank  $k$  adopts 20, and for ORL dataset, the value of decomposition rank  $k$  changes to 175, and then the highest recognition rate 96% can be achieved. Meanwhile, the categories of eight facial expressions of CK+ dataset are numbered from 1 to 8, for instance, the contempt expression is numbered as 8, the disgust expression is numbered as 6, the happy expression is numbered as 2 and the surprise expression is numbered as 3, and so on. Similarly, the categories of 40 persons for ORL dataset are numbered from 1 to 40.

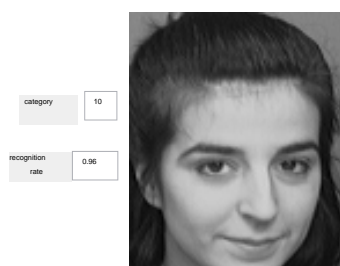
The experimental results for CK+ dataset based on ASNMF method are shown in **Fig. 8 - Fig. 11**. It can be seen that the various expressions can be correctly recognized respectively, for example, contempt expression is correctly recognized as category 8 shown in **Fig. 8**, disgust expression is correctly recognized as category 6 shown in **Fig. 9**, happy expression is correctly recognized as category 2 shown in **Fig. 10**, and surprise expression is correctly recognized as category 3 shown in **Fig. 11**, and so on. Meanwhile, from the **Fig. 8** to **Fig. 11**, it can be found that the different people with the same expression can also be correctly recognized. And the experimental results for ORL dataset are shown in **Fig. 12-Fig. 15**. From **Fig. 12** and **Fig. 13**, it can be seen that ASNMF method can recognize the different expressions belonging to the same person. Meanwhile, ASNMF method can recognize the person wearing glasses or not as shown in **Fig. 14**, and also can recognize the person whose eyes open or closed as shown in **Fig. 15**.



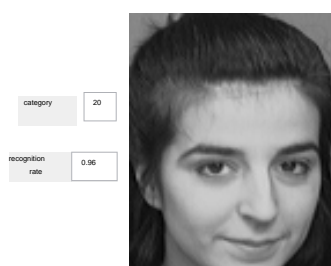
**Fig. 8.** Contempt expression correctly recognized. **Fig. 9.** Disgust expression correctly recognized.



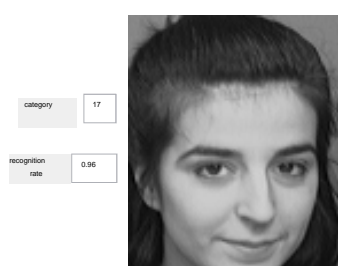
**Fig. 10.** Happy expression correctly recognized. **Fig. 11.** Surprise expression correctly recognized.



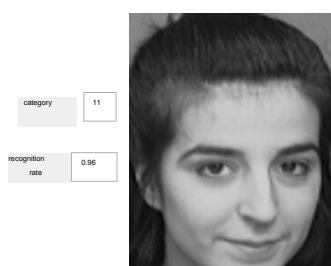
**Fig. 12.** Different facial expressions of same person



**Fig. 13.** Different facial expressions of a person



**Fig. 14.** Person wearing glasses or not.



**Fig. 15.** Person whose eyes open or closed.

## 6. Conclusion

In this paper, a new SNMF method with improved additive and threshold-sparse iterative rules, named ASNMF method, is proposed to improve the accuracy of facial feature extraction and recognition. Firstly, the color face images in dataset are converted to gray images to avoid the influence of illumination. Then the adaboost cascade classifier and scale normalization are adopted to detect the face region and align faces, respectively. Meanwhile, histogram specification is adopted to enhance the detailed contrast of the expression image, and the low-frequency information of expression image is extracted to reduce noises of face images by coiflets wavelet transform. Then the new iterative step sizes are proposed to improve the traditional iterative rules for SNMF method, which are smaller than the traditional iterative step sizes, and can improve the recognition accuracy. By analyzing the time complexity, we find that the improved iterative rules based on the new step sizes will not change the computational complexities of the SNMF. Then we prove the convergence of the improved SNMF method with the new iterative rules. To further reduce the redundant information in the feature data and achieve higher recognition accuracy, the threshold-sparse strategy is adopted to sparse the basis matrix  $W$ . Finally, the SVM classifier is chosen to identify the extracted facial features.

To verify the effectiveness, the ASNMF method is adopted to recognize the CK+ expression dataset and ORL face dataset respectively, and from the results of the comparative experiments with different  $k$  on the feature recognition rate, it can be shown that ASNMF method proposed in this paper can significantly improve the recognition rate for the two datasets, and achieve 100% and 96% rate for CK+ and ORL, respectively. For CK+ dataset, the highest recognition rate of ASNMF method is 30.625%, 38.125%, 83.75% and 87.5% higher than that of SNMF method, Deep NMF, CNMF and basic NMF, respectively. And for ORL dataset, the highest recognition rate of ASNMF method is 5%, 29%, 66% and 92.5% higher than that of Deep NMF, CNMF, SNMF and basic NMF, respectively.

Meanwhile, from the simulation results of the facial feature recognition software based on ASNMF, it can be found that the ASNMF method can recognize the different people with the same expression and the different expressions belonging to the same person.

## Acknowledgments

This paper is sponsored by the research project of provincial teaching reform in Hubei Province (No. 2017301).

## References

- [1] P. Liu, J. M. Guo, K. Chamnongthai and H. Prasetyo, "Fusion of color histogram and LBP-based features for texture image retrieval and classification," *Information Sciences*, vol. 390, no. 1, pp. 95-111, June, 2017. [Article\(CrossRefLink\)](#)
- [2] S. N. Borade, R. R. Deshmukh and S. Ramu, "Face recognition using fusion of PCA and LDA: Borda count approach," in *Proc. of Mediterranean Conference on Control and Automation, IEEE press*, pp.21-24, June 21-24, 2016. [Article\(CrossRefLink\)](#)
- [3] Y. Pang, Y. He, Y. Yuan and K. Wang, "Robust CoHOG Feature Extraction in Human-Centered Image/Video Management System," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 458-468, April, 2012. [Article\(CrossRefLink\)](#)
- [4] A. Pillai, R. Soundrapandiyan, S. Satapathy, S. C. Satapathy, K. H. Jun and R. Krishnan, "Local diagonal extrema number pattern: A new feature descriptor for face recognition," *Future Generation Computer Systems*, vol. 81, pp. 297-306, April, 2018. [Article\(CrossRefLink\)](#)
- [5] K. Petersen, M. Nielsen and S. S. Brandt, "Conditional Point Distribution Models," in *Proc. of International MICCAI Workshop on Medical Computer Vision*, pp. 1-10, September, 2011. [Article\(CrossRefLink\)](#)
- [6] A. Dewan, T. Caselitz, G. D. Tipaldi and W. Burgard, "Motion-based detection and tracking in 3D LiDAR scans," in *Proc. of IEEE International Conference on Robotics and Automation(ICRA)*, pp. 4508-4513, May 16-21, 2016. [Article\(CrossRefLink\)](#)
- [7] A. Lowhur and M. C. Chuah, "Dense Optical Flow Based Emotion Recognition Classifier," in *Proc. of IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems*, pp. 573-578, October 19-22, 2015. [Article\(CrossRefLink\)](#)
- [8] F. Liu, J. Tang, Y. Song, Y. Bi, and S. Yang, "Local structure based multi-phase collaborative representation for face recognition with single sample per person," *Information Sciences*, vol. 346, pp. 198-215, June, 2016. [Article\(CrossRefLink\)](#)
- [9] M. S. S. Kumar, R. Swami and M. Karuppiah, "An Improved Face Recognition Technique Based on Modular LPCA Approach," *Journal of Computer Science*, vol. 7, no. 12, pp. 1900-1907, December, 2011. [Article\(CrossRefLink\)](#)
- [10] L. A. Camentab, F. J. Galdames, K. W. Bowyer and C. A. Perez, "Face recognition under pose variation with local Gabor features enhanced by Active Shape and Statistical Models," *Pattern Recognition*, vol. 48, no. 11, pp. 3371-3384, November, 2015. [Article\(CrossRefLink\)](#)
- [11] F. Artoni, A. Delorme and S. Makeig, "Applying dimension reduction to EEG data by Principal Component Analysis reduces the quality of its subsequent Independent Component decomposition," *NeuroImage*, vol. 175, pp. 176-187, July, 2018. [Article\(CrossRefLink\)](#)
- [12] F. Artoni, A. Delorme, S. Makeig, "A visual working memory dataset collection with bootstrap Independent Component Analysis for comparison of electroencephalographic preprocessing pipelines," *Data in Brief*, vol. 22, pp. 787-793, February, 2019. [Article\(CrossRefLink\)](#)
- [13] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems (NIPS)*, vol. 1, pp. 556-562, 2001. [Article\(CrossRefLink\)](#)
- [14] Z. Li, J. Tang and X. He, "Robust Structured Nonnegative Matrix Factorization for Image Representation," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1947-1960, May, 2018. [Article\(CrossRefLink\)](#)

- [15] Y. Lu, C. Yuan, W. Zhu and X. Li, "Structurally incoherent low-rank nonnegative matrix factorization for image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5248-5260, November, 2018. [Article\(CrossRefLink\)](#)
- [16] Y. Chen, T. Z. Huang, X. L. Zhao and L. J. Deng, "Hyperspectral image restoration using framelet-regularized low-rank nonnegative matrix factorization," *Applied Mathematical Modelling*, vol. 63, pp. 128-147, November, 2018. [Article\(CrossRefLink\)](#)
- [17] X. Li, L. Wang, Q. Cheng, P. Wu, W. Gan, L. Fang, "Cloud removal in remote sensing images using nonnegative matrix factorization and error correction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 148, pp. 103-113, February, 2019. [Article\(CrossRefLink\)](#)
- [18] S. Z. Lai, "Research on several non-negative matrix factorization methods and applications," A *Master Thesis of University of Electronic Science and Technology*, pp. 17-29, June, 2014. [Article\(CrossRefLink\)](#)
- [19] G. S. Cui, X. L. Li, Y. S. Dong, "Subspace clustering guided convex nonnegative matrix factorization," *Neurocomputing*, vol. 292, pp. 38-48, May, 2018. [Article\(CrossRefLink\)](#)
- [20] Z. X. Guo, S. H. Zhang, "Sparse Deep Nonnegative Matrix Factorization," *arXiv:1707.09316v1[cs.CV]*, pp. 1-13, July, 2017. [Article\(CrossRefLink\)](#)
- [21] Y. Wang, "Proximal gradient method for convex and Semi-Nonnegative matrix factorization," A *Dissertation of Northeast Normal University for Master Degree*, pp. 5-6, June, 2015. [Article\(CrossRefLink\)](#)
- [22] Q. L. Ye, J. Yang, F. Liu, C. X. Zhao, N. Ye and T. M. Yin, "L1-norm distance linear discriminant analysis based on an effective iterative algorithm," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 114-129, August, 2018. [Article\(CrossRefLink\)](#)
- [23] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *Computer Vision and Pattern Recognition (CVPR)*, vol.1, pp. 511-518, December 8-14, 2001. [Article\(CrossRefLink\)](#)



**Jing Zhou** received the B.E. from Wuhan University of Technology in department of Control Technology in 2002, the M.E. degree in pattern recognition and intelligent systems in 2005 and Ph.D. degree in control science and engineering in 2009 from Huazhong University of Science and Technology (HUST), Wuhan, China. She has published 24 journal articles and conference papers. Her current research interests include artificial intelligent and image processing.



**Tianjiang Wang** Ph.D., Professor, School of Computer Science, Huazhong University of Science and Technology. He received the B.E. from Mathematics Department of Huazhong University of Technology in 1982, the M.E. degree from Information Management Department of Wuhan Institute of Technology in 1988 and Ph.D. degree from Computer College of Huazhong University of Science and Technology (HUST) in 2002, Wuhan, China. And he is the Part-time Academic Visitor to the University of Illinois at Champaign (UIUC) in 2008, and the Part-time Professor, School of Computer Science, Charles Sturt University, Australia. He has presided over (or participated in) more than 10 scientific research projects, such as the National Tenth Five-Year Plan of Science and Technology, the National 863 Plan, the National Natural Science Foundation, the key projects of science and technology in Hubei Province, the CIMS demonstration project in Hubei Province and the major projects entrusted by large enterprises in China. He has published 65 journal articles and conference papers. His current research interests include artificial intelligent and image processing.