www.ietdl.org

Published in IET Computer Vision
Received on 11th July 2013
Revised on 19th November 2013
Accepted on 16th December 2013
doi: 10.1049/iet-cvi.2013.0171

ISSN 1751-9632

Facial expression recognition considering individual differences in facial structure and texture

Jizheng Yi[1], *Xia Mao*[1], *Lijiang Chen*[1], *Yuli Xue*[1], *Angelo Compare*[2]

[1]*School of electronic and information engineering, Beihang University, Beijing, People's Republic of China*
[2]*Department of human sciences, University of Bergamo, Bergamo, Italy*
E-mail: moukyoucn@yahoo.com.cn

Abstract: Facial expression recognition (FER) plays an important role in human–computer interaction. The recent years have witnessed an increasing trend of various approaches for the FER, but these approaches usually do not consider the effect of individual differences to the recognition result. When the face images change from neutral to a certain expression, the changing information constituted of the structural characteristics and the texture information can provide rich important clues not seen in either face image. Therefore it is believed to be of great importance for machine vision. This study proposes a novel FER algorithm by exploiting the structural characteristics and the texture information hiding in the image space. Firstly, the feature points are marked by an active appearance model. Secondly, three facial features, which are feature point distance ratio coefficient, connection angle ratio coefficient and skin deformation energy parameter, are proposed to eliminate the differences among the individuals. Finally, a radial basis function neural network is utilised as the classifier for the FER. Extensive experimental results on the Cohn–Kanade database and the Beihang University (BHU) facial expression database show the significant advantages of the proposed method over the existing ones.

1 Introduction

Emotion recognition has been an important direction in the field of human–computer interaction (HCI). To establish a friendly and harmonious HCI model, a large number of researchers start from voice, facial expression or text and expect to achieve better HCI effectiveness through single-mode or multi-mode. As a meaningful field of the emotion recognition, facial emotion recognition has been a strong impetus to the development of HCI. Numerous outstanding research achievements have emerged [1–12] in recent years. Paul and Friesen [13, 14] proposed six basic expressions (joy, sadness, surprise, anger, disgust and fear) in 1971, and established the facial action coding system in 1978. However, automatic facial expression recognition (FER) in the real sense began in the 1990s, and many research achievements based on Paul and Friesen's theory sprouted during this period. Besides, some researchers [15–17] proposed other facial expressions in addition to the six basic ones, which greatly enriched the contents of the FER. With the rapid development of computer technology, automatic FER has achieved remarkable results. Generally speaking, different approaches of the FER can roughly be divided into two main categories: one is the holistic approach, and the other is the feature-based approach. The main difference between the holistic approach and the feature-based approach is that the former takes all the pixels of the face image as the input data, but the latter only need the local shape or the texture information.

The holistic approach takes the entire face region as the input data for the FER system. Thus, the capturing process of the local texture information is very important. To obtain the ideal texture information, it is necessary to extract the original expression features by taking advantage of the information which consists of the facial shape, the local texture, the optical flow and so on. However, these original features usually have the problem of information redundancy and a high dimension. Therefore the original features must be preprocessed before classification. Achieving this goal mainly depends on feature extraction and dimensionality reduction. Three of the most popular techniques for feature extraction and dimensionality reduction are principal component analysis (PCA) [18], linear discriminant analysis (LDA) [19] and independent component analysis (ICA) [20]. To effectively eliminate the interfering factors, the PCA reduces the dimension of the data space and extracts the most essential features, but does not consider the distinction among the different categories of data. LDA is a supervised learning algorithm and selects a suitable projection direction by maximising the ratio of the between-class distance to the within-class distance. The intrinsic limitation of the classical LDA algorithm is that the objective function requires the non-singularity of one scatter matrix. The ICA is a computational method for separating a multivariate signal into additive subcomponents, and also a special case of blind source separation. However, for the shortcoming of the components being treated independently, the ICA is not widely used. The PCA, the LDA and the ICA are only good at extracting information with the Euclidean structure. However, the expression features usually lie on the non-linear submanifold hiding in the face image space.

Therefore the study of the dimensionality reduction goes deep into the field of the non-linear manifold. Thus, some non-linear dimensionality reductions have been proposed to solve this problem, for example, multidimensional scaling, isometric mapping, locally linear embedding, Laplacian eigenmap and stochastic neighbour embedding (SNE) [21–25].

The feature-based approach for the FER can also gain a satisfactory recognition effect. Kotsia *et al.* [26] used the shape and texture information for the FER and obtained very good results. Mahdi and Hamed [27] proposed a fuzzy FER system based on the colour images. They also gave the feature extraction of the facial organs and the fuzzy classification in detail, and obtained a successful experimental result. Shi *et al.* [28] verified the effect of the marked points related to the facial geometry information in face recognition (FR) and FER, and pointed out that the geometric coordinates of the marked points could not be directly used to train the classifier, because different people might have different facial features. Therefore the relative distance became the primary means to measure the geometric distance. Park *et al.* [29] used the active appearance model (AAM) to extract the facial geometrical features, thus realised the purpose of the FER. Valstar and Pantic [30] proposed a method for face geometry feature extraction and motor unit identification, and predicted the facial expression. However, for some of the different facial expressions, the geometrical features might be similar although the local texture features were different. For this reason, Valstar and Pantic's research became more difficult. Song *et al.* [31] used the active shape model (ASM) to detect the feature points of the face images preprocessed by the graphic processing unit, then combined the image ratio features with the facial animation parameters as the input of the support vector machine (SVM), and ultimately obtained good recognition results. All the methods discussed above have the same problem that they do not consider the effect of the individual differences on the recognition results. To solve this problem, this paper proposes a novel FER algorithm by exploiting the structural characteristics and the texture information hiding in the image space, and gives three facial features: feature point distance ratio coefficient (FPDRC), connection angle ratio coefficient (CARC) and skin deformation energy parameter (SDEP).

This paper is organised as follows: Section 2 describes an overview of our FER system and the AAM. Section 3 demonstrates the proposed facial feature extraction and gives three concepts: FPDRC, CARC and SDEP. Section 4 is the experimental results and analysis. The conclusions and future work are given in Section 5.

2 Overview of the proposed FER system

Fig. 1 shows the overview of the proposed FER system. The basic idea of the system is to merge the structural characteristics and the texture information, and then classify the extracted features by using a radial basis function (RBF) neural network. The system consists of three parts: feature point detection, feature extraction and classification.

For the feature point detection, we select two face images with neutral and certain expressions, respectively, from the facial expression sequence, as illustrated in Fig. 2. The classic feature point localisation algorithm AAM [32] is utilised to locate the feature points of the face images. Different from the ASM [33], the AAM uses not only the information of the face shape, but also the texture information. The AAM's shape model is constituted of the two-dimensional (2D) coordinates of the vertices that make up the mesh. As shown in (1), S denotes the set of vertices, and v is the number of vertices

$$S = \left(x_1, y_1, x_2, y_2, \ldots, x_v, y_v\right)^{\mathrm{T}} \tag{1}$$

After the alignment, the statistical estimation and the PCA, the final shape model S is expressed by a linear combination of a mean shape S_0 and the orthogonal basis $\{S_i\}(i = 1, 2, \ldots, n)$

$$S = S_0 + \sum_{i=1}^{n} \rho_i S_i \tag{2}$$

where $\{\rho_i\}$ denotes a group of shape parameters, and S_i is the ith eigenvector that corresponds to the ith largest eigenvalue.

Based on S_0, the texture model of the AAM is established. After the alignment, the statistical estimation and the PCA,

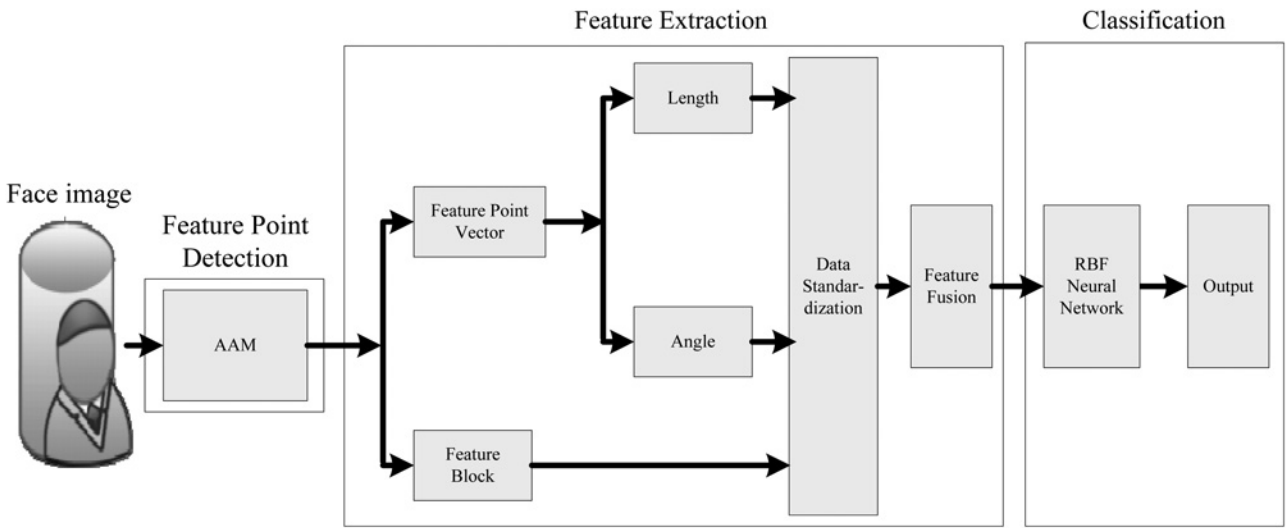

Fig. 1 *Diagram of the proposed FER system*

The proposed system includes three parts: feature point detection, feature extraction and classification

IET Comput. Vis., 2014, Vol. 8, Iss. 5, pp. 429–440
doi: 10.1049/iet-cvi.2013.0171

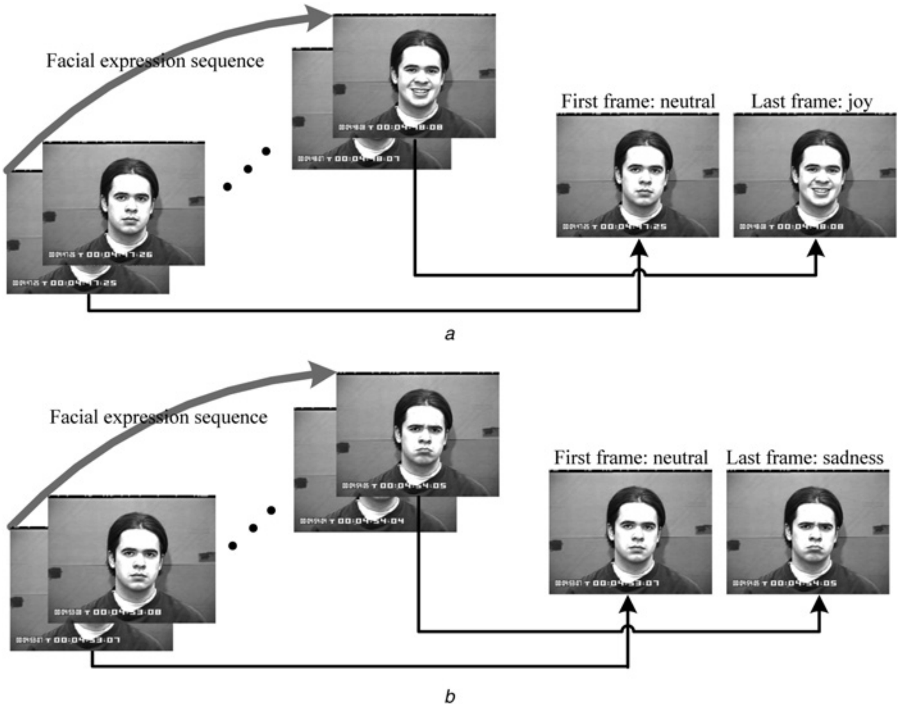

Fig. 2 *Selected two face images from the facial expression sequence (the Cohn–Kanade database)*

Here, the database gives two selecting processes for the joy and the sadness emotions, and the processes are the same for the other four expressions
a Facial expression sequence with the expression of joy
b Facial expression sequence with the expression of sadness

the appearance variable T can be obtained by the following

$$T = T_0 + \sum_{i=1}^{m} \lambda_i \boldsymbol{T}_i \tag{3}$$

where T_0 is the mean texture, $\{\boldsymbol{T}_i\}$ $(i = 1, 2, \ldots, m)$ are the m texture basis vectors and $\{\lambda_i\}$ denote a group of texture parameters.

For the feature extraction, the characteristic parameters of the FER are divided into feature point vectors containing the structural characteristics and feature blocks containing the texture information. As we all know, the shapes of the facial organs differ from man to man, which leads to differences of magnitude and direction of facial muscle movement while people are giving a certain expression. In addition, the different sizes of the face images could directly make the positions of the feature points different, and make the FER lack uniform conditions. Therefore, to achieve satisfactory FER results, we should not only apply the position information of the feature points, which ignores the differences in the facial structure. Based on this consideration, we perform feature point detection on two face images which are selected from the facial expression sequence and show neutral and certain expression, respectively. Then, according to the movements of the feature points and the dynamic texture variations of the feature blocks, we give the concept of a ratio coefficient and propose a feature point distance ratio coefficient (FPDRC), a connection angle ratio coefficient (CARC) and a skin deformation energy parameter (SDEP), which are expected to eliminate the negative effects of the individual differences. After obtaining the three characteristics (FPDRC, CARC and SDEP), a problem of big data has emerged, which seriously affects the efficiency of the FER. However, it is found that not all the features are helpful for the final recognition results in the experiments. Those FPDRCs and CARCs always fluctuating near 1 under all the expressions are eventually removed, because the movements of the corresponding feature points are not significant. For the SDEP, we extract its most essential features by the PCA. In our experiments, the three characteristics mentioned above prove to be effective in reducing the data dimension, and make the training and the classification of the RBF neural network be more effective and faster. Details of this part have been described in Section 3.

The third part of our system is classification. As shown in Fig. 1, we utilise the RBF neural network for data training and classification after obtaining the fused features. The artificial neural network (ANN) is a bionic neural network, and a mathematical model of distributed and parallel information processing. As an important member of the ANN, the RBF neural network has the following advantages: the nature of a global approximation and no problem of local minima; strong input and output mapping functionality; a linear relationship between the connection weights and the output; a powerful classification ability and fast convergence speed. Fig. 3 represents the structure of the RBF neural network.

The RBF neural network consists of three layers: the input layer, the hidden layer and the output layer. Each hidden layer neuron corresponds to one centre. The RBF takes the Gaussian function. Upon receiving the input of the expression vectors, the RBF neural network executes a non-linear transformation based on the Euclidean distance between the vectors and the centres, and then performs a linear operation to the output of the hidden layer by the connection weights. The final results are stored in the output layer. This paper designs a radial basis network by the MATLAB toolbox function and combines the three

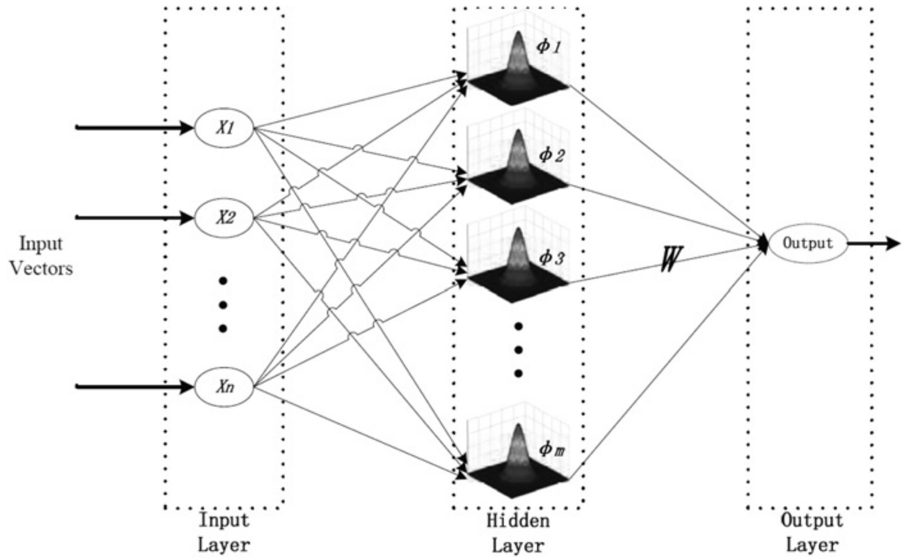

Fig. 3 *Structure of the RBF neural network*

In the proposed algorithm, the RBF takes the Gaussian function and the final outputs of the RBF neural network can be set as 1, 2, 3, 4, 5 and 6, which correspond to joy, sadness, surprise, anger, disgust and fear, respectively

proposed characteristics (FPDRC, CARC and SDEP) as the inputs of the RBF neural network. The final outputs of the RBF neural network can be set as 1, 2, 3, 4, 5 and 6, which correspond to joy, sadness, surprise, anger, disgust and fear, respectively.

3 Facial feature extraction

By considering the individual differences, this section gives the facial feature extraction process in detail. After marking the feature points, we present two innovative concepts based on the feature point vectors: FPDRC and CARC. We also propose the SDEP. Finally, we mix the three features to form a new facial feature vector for the FER.

3.1 Feature point vector

There are many differences between individuals, especially the facial appearance. Different persons' facial organs cause a lot of differences in the same expression. For example, people usually keep their eyes widely open when they feel surprised. The eyes-open amplitude (distance between the upper eyelid and the lower eyelid) of small-eyed people with the surprised expression may be equal to or even smaller than the eyes-open amplitude of large-eyed ones with the neutral expression. Therefore we should consider the great effect of the individual differences on the FER.

In our approach, a total of 106 points are located in a face image by the AAM. 26 of 106 points are selected as the feature points and shown in Fig. 4, where '+' indicates the location of each feature point, and the arrow represents the corresponding number of the feature point. The feature point numbers and the corresponding names are described in Table 1.

Here, $p_i (i = 1, 2, 3, \ldots, 26)$ denote the selected feature points. The feature points of the neutral and the other expressions are denoted as p_i^{neutral} and $p_i^{\text{expression}}$, respectively. Fig. 5 shows the results of some samples' feature point detections. The feature point vector is determined by two different points. Fig. 6 shows the feature point vector of the feature point p_{18}. Firstly, we calculate

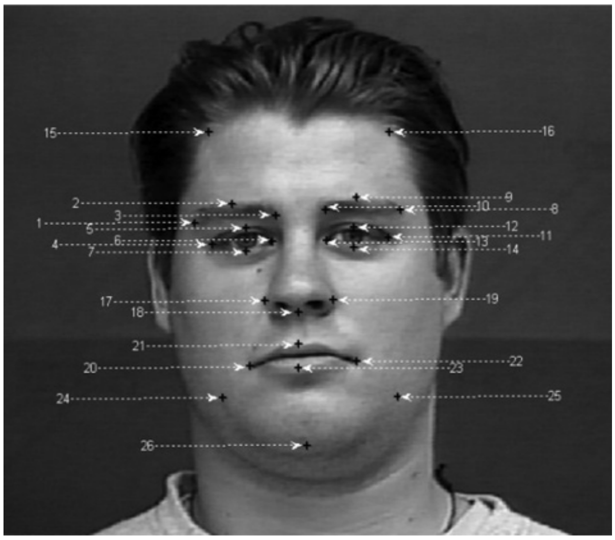

Fig. 4 *Feature points for the FER 26 of 106 points are selected as the feature points and are thought to be the most representative in the proposed algorithm*

Table 1 Feature point numbers and corresponding names

Number	Name	Number	Name
1	right eyebrow right	14	left eyelid down
2	right eyebrow middle	15	right forehead
3	right eyebrow left	16	left forehead
4	right eyelid outer	17	right ala nasi
5	right eyelid up	18	nasal columella
6	right eyelid inner	19	left ala nasi
7	right eyelid down	20	right lip corner
8	left eyebrow left	21	up lip middle
9	left eyebrow middle	22	left lip corner
10	left eyebrow right	23	down lip middle
11	left eyelid outer	24	right cheek
12	left eyelid up	25	left cheek
13	left eyelid inner	26	chin

IET Comput. Vis., 2014, Vol. 8, Iss. 5, pp. 429–440
doi: 10.1049/iet-cvi.2013.0171

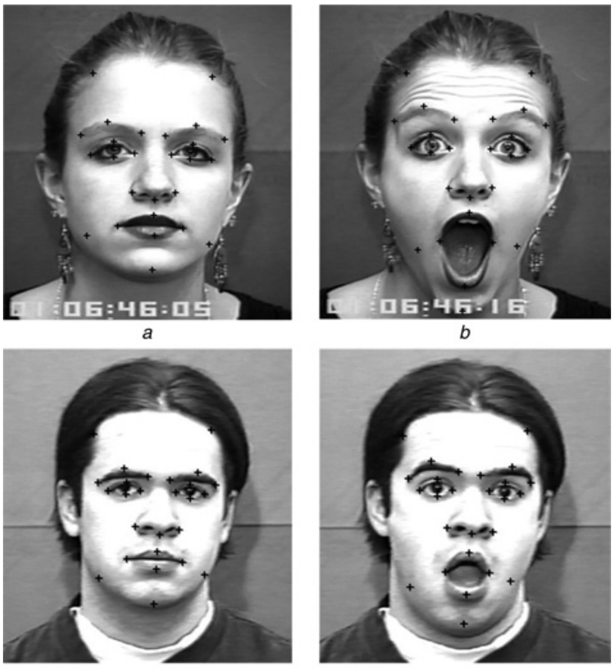

Fig. 5 *Feature point detection results of some examples*
a and *c* Exhibit neutral emotion
b and *d* Exhibit surprise emotion

the Euclidean distance $d_{ij}(i, j = 1, 2, \ldots, 26\ i \neq j)$ between two different feature points

$$d_{ij} = \left\| p_i - p_j \right\| = \sqrt{\left(x_i - x_j\right)^2 - \left(y_i - y_j\right)^2} \tag{4}$$

Secondly, we represent the line between p_i and p_j and the line between p_i and p_k by l_{ij} and l_{ik}, respectively

$$l_{ij} = p_j - p_i = (x_j - x_i, y_j - y_i) \tag{5}$$

$$l_{ik} = p_k - p_i = \left(x_k - x_i, y_k - y_i\right) \tag{6}$$

where $i, j, k = 1, 2, \ldots, 26$, and $i \neq j \neq k$. Finally, we define the angle between l_{ij} and l_{ik} as follows

$$\alpha_{l_{ij}l_{ik}} = \arccos\left\{\left(l_{ij} \bullet l_{ik}\right) / \left(\left|l_{ij}\right| \times \left|l_{ik}\right|\right)\right\} \tag{7}$$

3.2 FPDRC and CARC

After knowing the coordinate values of p_i^{neutral} and $p_i^{\text{expression}}$, we firstly calculate the Euclidean distances d_{ij}^{neutral} and $d_{ij}^{\text{expression}}$ by (4). Relative to d_{ij}^{neutral}, the changing of $d_{ij}^{\text{expression}}$ has three situations, as illustrated in Fig. 7*a*. Then, we standardise them and obtain the initial FPDRC, as

$$k_d = d_{ij}^{\text{expression}} / d_{ij}^{\text{neutral}} \tag{8}$$

In Fig. 8, we give the FPDRCs of two samples whose expressions are surprise. The total number of the feature points is 26. According to $C_{26}^2 = 325$, the maximum value of the abscissa of Fig. 8 is 325.

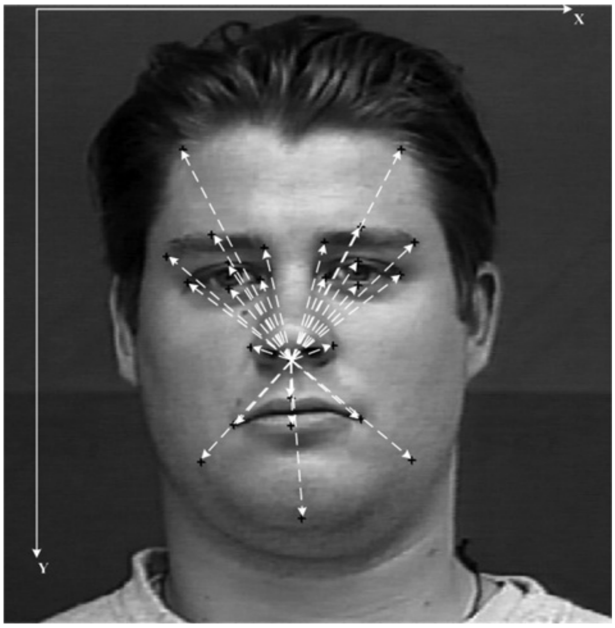

Fig. 6 *Feature point vectors of p_{18}.*
With the expression changing from neutral to certain expressions, the corresponding feature point vectors change

We have performed a large number of experiments and found that some FPDRCs in the histogram always fluctuate near 1 under all expressions. These FPDRCs are not helpful and even negatively affect the following training and the classification of the RBF neural network. Hence, we remove the FPDRCs which meet the condition as

$$\left|k_d - 1\right| \leq K \tag{9}$$

where K is a constant and its value is discussed in Section 4. The rest are included in the final inputs of the FER experiment as $k_{d\text{-final}}$.

We calculate the angles $\alpha_{l_{ij}l_{ik}}^{\text{neutral}}$ and $\alpha_{l_{ij}l_{ik}}^{\text{expression}}$ by (7). Relative to $\alpha_{l_{ij}l_{ik}}^{\text{neutral}}$, the changing of $\alpha_{l_{ij}l_{ik}}^{\text{expression}}$ has three situations, as illustrated in Fig. 7*b*. Then, we standardise them and obtain the initial CARC, as

$$k_\alpha = \alpha_{l_{ij}l_{ik}}^{\text{expression}} / \alpha_{l_{ij}l_{ik}}^{\text{neutral}} \tag{10}$$

Fig. 9 presents the CARCs of performer 1 whose expression is surprise. The total number of the feature points is 26. According to $C_{26}^3 = 2600$, the maximum value of the abscissa of Fig. 9 is $2600 \times 3 = 7800$.

Similarly, we remove those CARCs which meet the condition as

$$\left|k_\alpha - 1\right| \leq K \tag{11}$$

The remaining ones are included in the final inputs of the FER experiment as $k_{\alpha\text{-final}}$.

3.3 Skin deformation energy parameter

Although showing certain expressions, people's faces often have the expression lines which perform on the face images as texture variations. These changes make the original

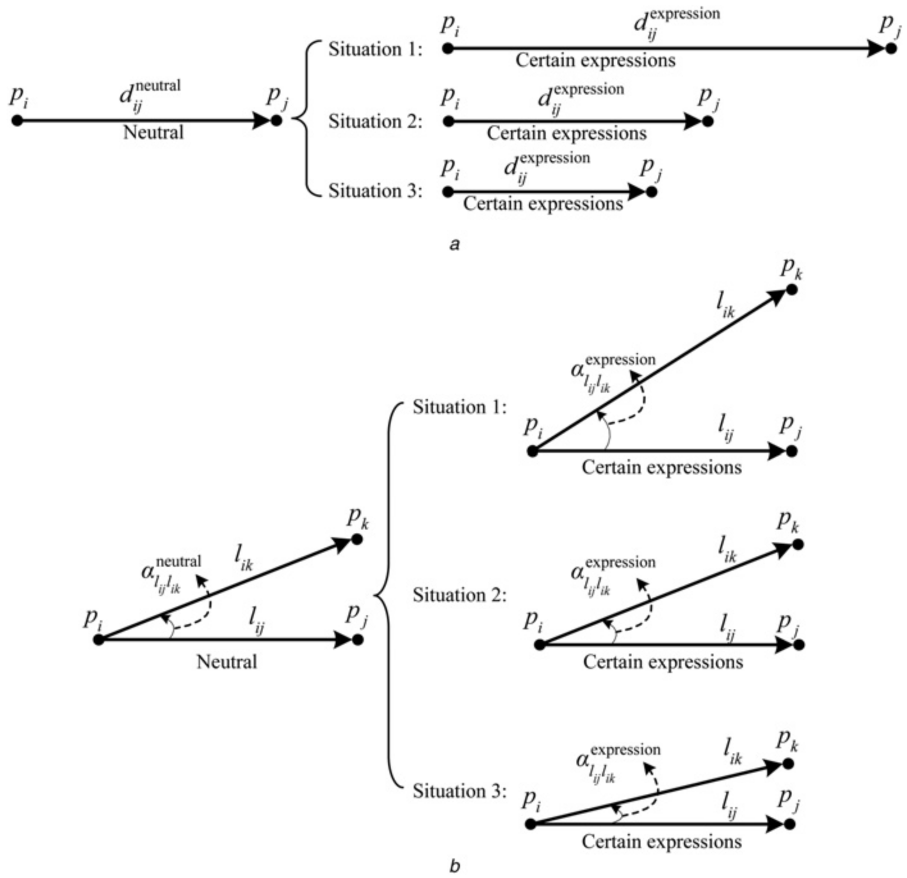

Fig. 7 *Distance and connection angle of the feature points*

The three changing situations of the distance and the connection angle are bigger, equal and smaller
a Distance
b Connection angle

smooth surface of the facial skin complicated and the energy distribution of the image changed. Using texture changes for the FER has been a research focus. Song *et al.* [31] selected the characteristic block with the texture changes for wrinkles detection and proposed the concept of the skin deformation parameters to achieve better recognition effect.

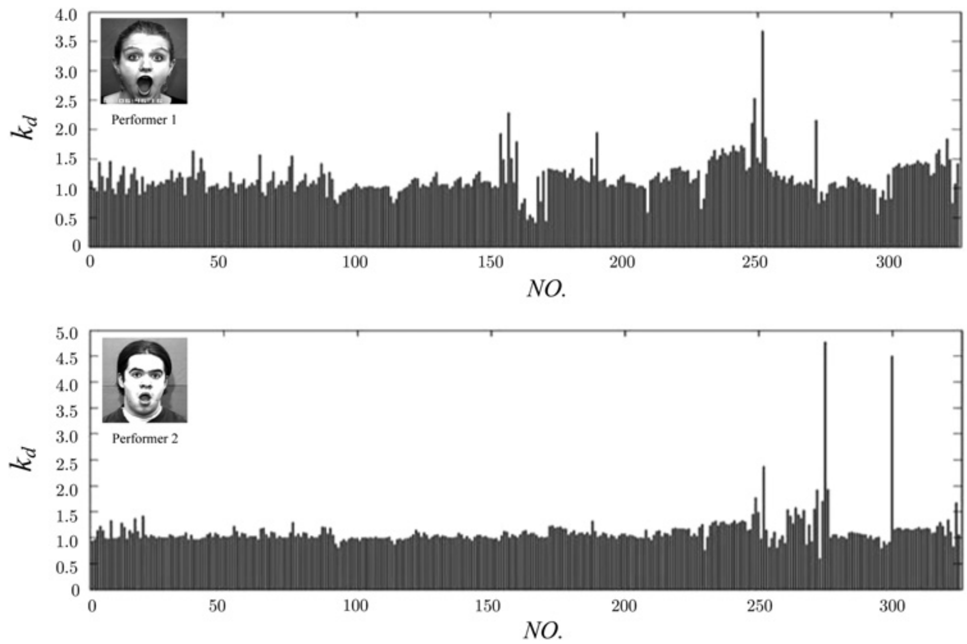

Fig. 8 *FPDRCs of the two samples*

The maximum value of the abscissa is 325

doi: 10.1049/iet-cvi.2013.0171

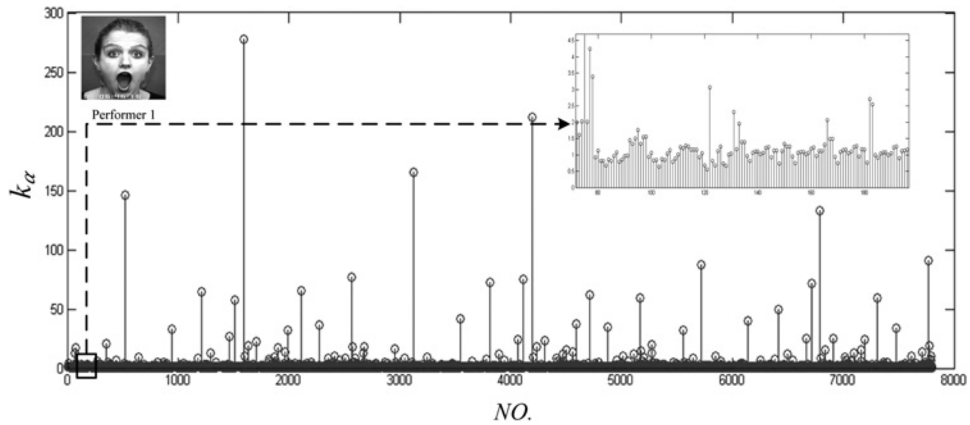

Fig. 9 *CARCs of performer 1*
The maximum value of the abscissa is 7800

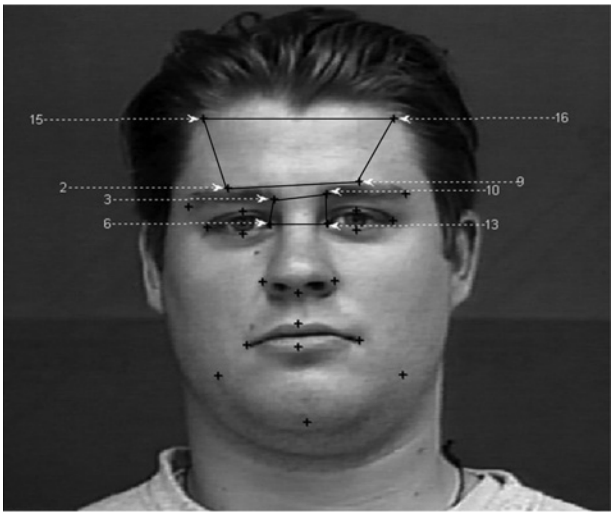

Fig. 10 *Selected feature blocks for the FER*
Here, the two selected blocks will be adjusted to 155 × 65 and 35 × 25, respectively

Liu *et al.* [34] pointed out that the expression ratio image can capture the facial local motion information, just as the expression lines. By considering that the expression lines in the forehead and the glabella are the most obvious ones, we select the two feature blocks for the FER, which are shown in Fig. 10.

The two feature blocks are referred to as $F_{15,16,9,2}$ and $F_{3,10,13,6}$, where the subscript represents the serial number of the feature point. For the case of $F_{3,10,13,6}$, we note that its textures change with the occurrence of the expression and the movement of the corresponding facial muscles. These changes consist of a grey-value variation and an edge appearance. The more complex the image is, the more component of high frequency it contains. According to this fact, we give the new concept of the SDEP, whose calculation process is shown in Fig. 11.

Owing to the changes of the feature points' positions, the size of the corresponding feature block will change. Therefore we make the size of the selected feature blocks uniform by the median interpolation method before the Fourier transform. $\boldsymbol{F}_{15,16,9,2}$ and $\boldsymbol{F}_{3,10,13,6}$ are adjusted to 155 × 65 and 35 × 25, respectively. The Fourier transform of a feature block is shown in Fig. 12.

The matrices of $\boldsymbol{F}_{15,16,9,2}$ and $\boldsymbol{F}_{3,10,13,6}$ after the Fourier transform are depicted in (12) and (13), respectively

$$B1 = \begin{pmatrix} b1_{11} & \cdots & b1_{1N} \\ \vdots & \ddots & \vdots \\ b1_{M1} & \cdots & b1_{MN} \end{pmatrix} \tag{12}$$

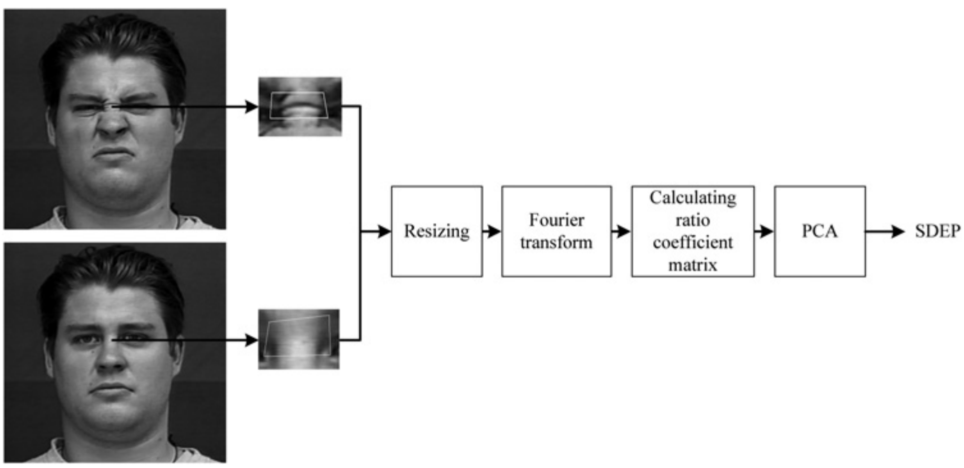

Fig. 11 *Calculation process of the SDEP*

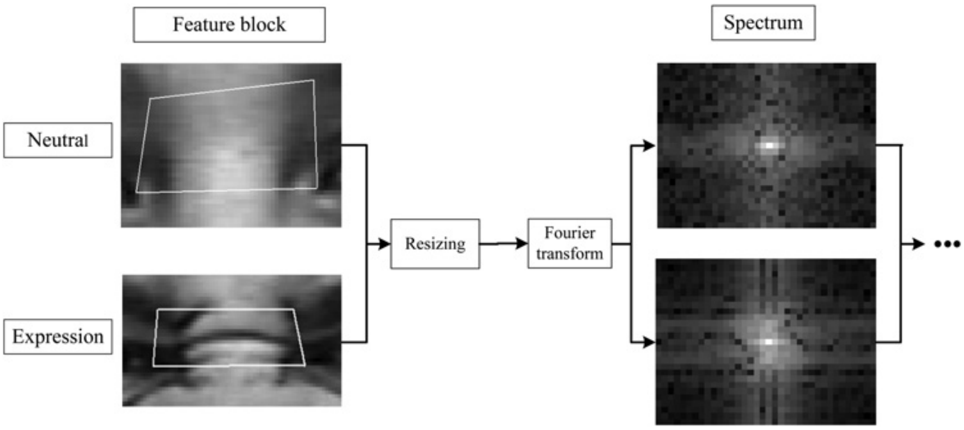

Fig. 12 *Fourier transform of a feature block*

$$B2 = \begin{pmatrix} b2_{11} & \cdots & b2_{1N} \\ \vdots & \ddots & \vdots \\ b2_{M1} & \cdots & b2_{MN} \end{pmatrix} \tag{13}$$

where $N \times M$ can be 155×65 or 35×25. Thereby, the ratio coefficient matrix $\boldsymbol{B}$ is calculated by

$$\begin{aligned} \boldsymbol{B} = B1/B2 &= \begin{pmatrix} b1_{11}/b2_{11} & \cdots & b1_{1N}/b2_{1N} \\ \vdots & \ddots & \vdots \\ b1_{M1}/b2_{M1} & \cdots & b1_{MN}/b2_{MN} \end{pmatrix} \\ &= \begin{pmatrix} b_{11} & \cdots & b_{1N} \\ \vdots & \ddots & \vdots \\ b_{M1} & \cdots & b_{MN} \end{pmatrix} \end{aligned} \tag{14}$$

Finally, after the processing of the PCA, we obtain the final SDEP which is denoted as $k_{s\text{-final}}$.

By merging the FPDRC, the CARC and the SDEP, we obtain the final input feature for the FER, as

$$k_{\text{final}} = \left(k_{d-\text{final}}, k_{\alpha-\text{final}}, k_{s-\text{final}}\right) \tag{15}$$

4 Experimental results and analysis

We have performed some experiments that demonstrate the effect of the proposed FER method. All the experiments are implemented in a Matlab2011 environment on a computer with an Intel (R) Core (TM)2 Duo CPU with a clock speed of 2.2 GHz, a 2 GB RAM and Windows XP professional.

4.1 Database

Two facial expression databases are used in our experiment: the first is the famous Cohn-Kanade database [35, 36], and the second is the BHU (Beihang University) facial expression database which is built by our laboratory [15]. The Cohn-Kanade database includes more than 100 performers from different regions (15% are African-American and 3% are Asian or Latino), with different colours, ages (ranging from 18 to 43) and genders (65% female). Besides, the image resolution of the Cohn-Kanade database is 640×490. For this database, 100 performers who have more than 500 frontal face image pairs are selected to perform our FER experiments. In the training process, 60 performers are employed to train the classifier. The BHU facial expression database consists of three parts: 18 kinds of single expressions, 3 kinds of mixed expressions and 4 kinds of complex expressions. Thirty-two Chinese college students from different ages (ranging from 21 to 25) and genders (14 males and 18 females) have participated in the data collection [15]. We choose 20 performers for the training and the rest for classifying. Each subject has at least six basic expressions. Hence, there are a total of 192 frontal face image pairs. Some samples of the BHU facial expression database are shown in Fig. 13.

4.2 Results and analysis

To evaluate the performance of the proposed algorithm, we implement our FER experiment in three stages. The first stage is to search the best value of K which is given in (9) and (11). The second stage is to evaluate the final FER rates of our proposed algorithm under the best value of K. The final stage is to compare our proposed algorithm with the other methods.

We compared the FER performance by using the RBF neural network with different K. Figs. 14 and 15 show the FER rates for different K. The two figures show that the average rate becomes smaller after a peak value and obtains the highest point when K is 0.15. It should be noted that it is necessary to determine a suitable value of K for a new given database.

Fig. 13 *Some samples of the BHU facial expression database*

IET Comput. Vis., 2014, Vol. 8, Iss. 5, pp. 429–440
doi: 10.1049/iet-cvi.2013.0171

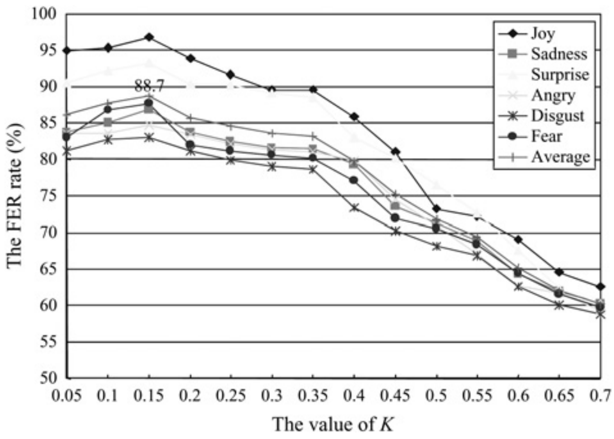

Fig. 14 *Comparison of the FER rates of the Cohn–Kanade database*

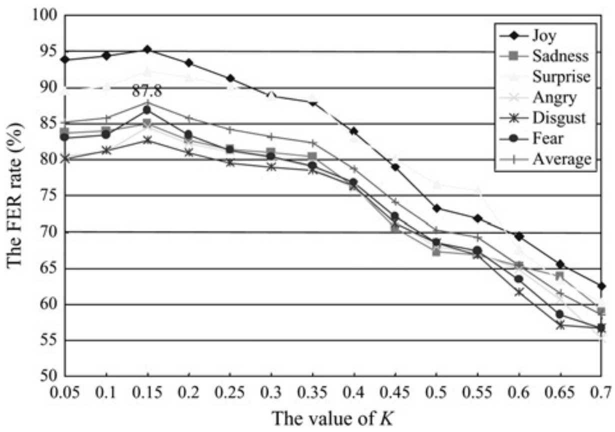

Fig. 15 *Comparison of the FER rates of the BHU facial expression database*

For the best value of K, Tables 2 and 3 show the FER rates of the different expressions of the Cohn-Kanade database and the BHU facial expression database, respectively. As shown in Table 2, joy and surprise hold the top two of the FER rates: 96.7 and 93.4%, respectively. For the other four expressions, the FER rates are distributed between 80 and 90%, respectively, and the lowest recognition rate is 82.9%. Tables 2 and 3 also give the error recognition rates of the six basic expressions. The probability is 8.8% for angry being recognised as sadness, 6.3% for disgust being recognised as angry and 5.6% for sadness being recognised as disgust. For the Cohn-Kanade database, the final average recognition rate is 88.7%.

For the BHU facial expression database, the recognition results are shown in Table 3. The distribution of the experimental results of the BHU facial expression database is similar to the Cohn-Kanade database, and the overall recognition performances are lower than the latter. The final average recognition rate is 87.8%.

Fig. 16 compares the FER performances by using different features of input: PCA, LDA, SNE, FPDRC, CARC and a group of FPDRC, CARC and SDEP, whereas the classifier is the RBF neural network, and also makes the comparison of the FER rates with different methods given by the authors [11, 37, 38]. It presents that the proposed algorithm outperforms the others. For the Cohn-Kanade database, the recognition rate increases by 20.2, 14.1, 6.2, 10.8, 6.7, 10.6 and 1.9%, respectively. For the BHU facial expression database, the recognition rate increases by 23.5, 16.1, 6, 12.6, 7.5, 11 and 3.2%, respectively. Fig. 16 also shows the FER performances by using the FPDRC only, the CARC only and a group of FPDRC, CARC and SDEP. Compared with the recognition rates 84.1 and 82.6%, the results increase by 4.6 and 5.2%, respectively, which indicate that the method merging the three proposed facial features is more effective. In Fig. 16, we know that our proposed FER algorithm which merges the structural characteristics and

Table 2 FER rates of the Cohn–Kanade database (%)

Input	Output					
	Joy	Sadness	Surprise	Angry	Disgust	Fear
joy	**96.7**	0	0.5	0.6	0	2.2
sadness	1.4	86.8	0.4	5.6	4.6	1.2
surprise	1.2	0.3	93.4	0.5	0.6	4
angry	0	8.8	0	84.7	1.4	5.1
disgust	0	5.4	0.8	6.3	82.9	4.6
fear	0.2	0.6	3.5	4.4	3.6	87.7
the average FER rate: 88.7						

Table 3 FER rates of the BHU facial expression database (%)

Input	Output					
	Joy	Sadness	Surprise	Angry	Disgust	Fear
joy	95.3	0.1	0.7	0.3	0	3.6
sadness	0.7	85	0.4	5.1	5.3	3.5
surprise	1.6	0	92.3	0.9	0.8	4.4
angry	0	9.4	0.6	84.6	2.2	3.2
disgust	0	5.2	0.5	6.8	82.7	4.8
fear	0.4	3.4	0.3	4.3	4.7	86.9
the average FER rate: 87.8						

doi: 10.1049/iet-cvi.2013.0171

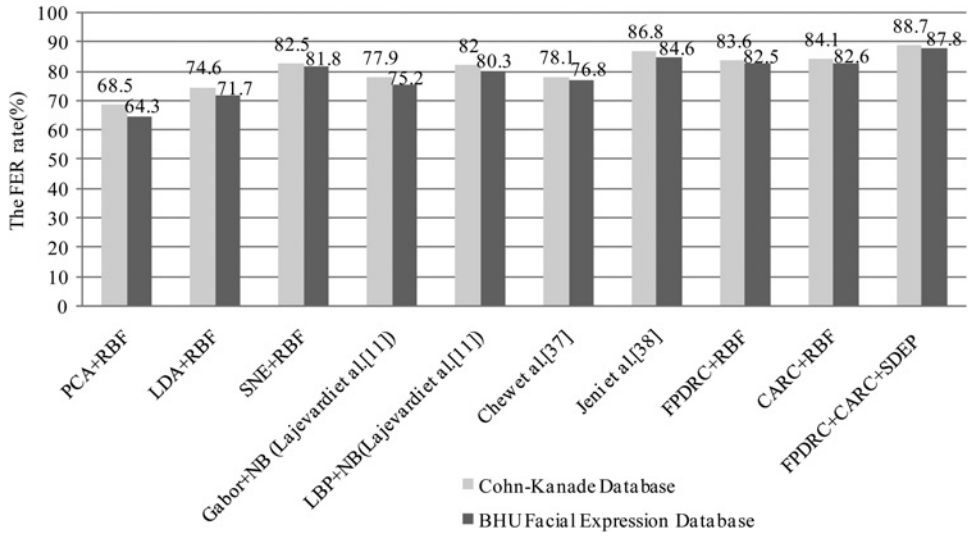

Fig. 16 *Comparison of the FER performances with different feature inputs*
Here, the NB represents a Naive Bayesian

the texture information is more robust and works better than the existing methods.

Fig. 17 gives the FER rates by using different types of classifiers: the SVM, the ADABOOST and the RBF neural network. To ensure a fair comparison, the SVM, the ADABOOST and the RBF are given the same training and testing data in our experiments. It should be noted that the training and the testing data are non-linear. As shown in Fig. 17, with the same inputs of the FPDRC, the CARC and the SDEP, the proposed algorithm using the RBF neural network has better recognition effect. Through a series of ratio normalisation approaches, the feature points

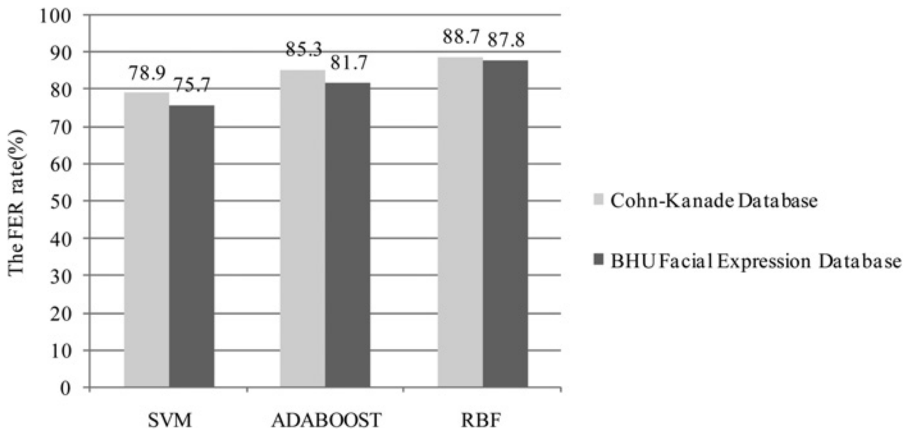

Fig. 17 *Comparison of the FER performances with different classifiers*

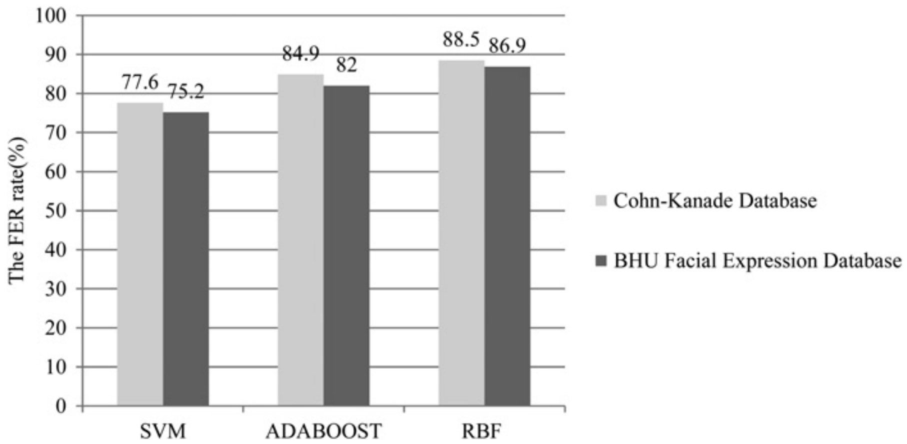

Fig. 18 *Comparison results of the verified experiment*

doi: 10.1049/iet-cvi.2013.0171

have different contributions to different expressions and the trained RBF classifier gives suitable weights to different feature points. That is the reason why the RBF performs better than the other classifiers.

To test and verify the resistance to noise, we perform one more experiment in our paper. For the testing data of this experiment, the locations of the feature points 11, 12, 13 and 14 in the expression images are manually set as the same values with the corresponding feature points' locations in the neutral images. Similar to Fig. 17, the experimental results shown in Fig. 18 do not change much. Although these feature points' locations are fixed, the locations of the feature points 4, 5, 6 and 7 are changing, and these changes are still able to make our algorithm robust.

5 Conclusion and future work

In this paper, we have proposed a new algorithm for the FER which merges three novel features: FPDRC, CARC and SDEP. In comparison with the other FER methods, it demonstrates better recognition performance on the Cohn–Kanade database and the BHU facial expression database. Based on the idea of the feature point vector, the FPDRC and the CARC are defined as the distance ratio coefficient and the connection angle ratio coefficient of the feature points marked by the AAM, respectively. To calculate the final FPDRC and CARC, we have eliminated the interference of the excess coefficients by the value of K. To take advantage of the facial texture information, we also have proposed the SDEP which represents the changing of the skin deformation energy. The RBF neural network is used as the classifier. The results of our experiments prove that the utilisation of the structural characteristics and the texture information of the face image is robust and better for the FER.

We frankly recognise that our algorithm has two limitations. Firstly, in spite of having good experimental results, the method requires both a neutral and an expressive image, which is quite restrictive. Secondly, this work does not test the robustness of the proposed method under several environmental conditions (noisy, illumination variations, head pose variations etc.). Therefore the authors will mainly devote their energy and enthusiasm to the two previously mentioned limitations in the future. Furthermore, the authors plan to discover how to make better use of the space–time information hiding in the facial expression sequences. The separation of the identity information and the motion information will be our next research focus.

6 Acknowledgment

This work was supported in part by the Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 30400002012102002), and the National Natural Science Foundation of China (Grant no. 61103097). The authors would like to thank the associate editor and the anonymous reviewers for their valuable suggestions. This paper also utilises the Cohn–Kanade database and the BHU facial expression database. The authors would like to thank the providers of these databases.

7 References

1 Wang, X., Mao, X., Mitsuru, I.: 'Human ace analysis with nonlinear manifold learning', *J. Electron. Inf. Technol.*, 2011, **33**, (10), pp. 2531–2535
2 Kotsia, I., Pitas, I.: 'Facial expression recognition in image sequences using geometric deformation features and support vector machines', *IEEE Trans. Image Process.*, 2007, **16**, (1), pp. 172–187
3 Mao, X., Xue, Y.: 'Human-computer affective interaction' (Science Press, 2011)
4 Li, Z., Mao, X.: 'EEMML: the emotional eye movement animation toolkit', *Multimedia Tools Appl.*, 2012, **60**, (1), pp. 181–201
5 Xue, Y., Mao, X., Caleanu, C.D., Lv, S.: 'Layered fuzzy facial expression generation of virtual agent', *Chin. J. Electron.*, 2010, **19**, (1), pp. 69–74
6 Mao, X., Li, Z.: 'Generating and describing affective eye behaviors', *IEICE Trans. Inf. Syst.*, 2010, **E93-D**, (5), pp. 1282–1290
7 Wang, X., Mao, X., Caleanu, C.D.: 'Nonlinear shape-texture manifold learning', *IEICE Trans. Inf. Syst.*, 2010, **93**, (7), pp. 2016–2019
8 Moore, S., Bowden, R.: 'Local binary patterns for multi-view facial expression recognition', *Comput. Vis. Image Underst.*, 2011, **115**, (4), pp. 541–558
9 Choi, J.Y., Ro, Y.M., Plataniotis, K.N.: 'Color local texture features for color face recognition', *IEEE Trans. Image Process.*, 2012, **21**, (3), pp. 1366–1380
10 Lajevardi, S.M., Wu, H.: 'Facial expression recognition in perceptual color space', *IEEE Trans. Image Process.*, 2012, **21**, (8), pp. 3721–3733
11 Lajevardi, S.M., Hussain, Z.M.: 'Automatic facial expression recognition: feature extraction and selection', *Signal Image Video Process.*, 2012, **6**, (1), pp. 159–169
12 Zhao, G., Huang, X., Taini, M., Li, S.Z., Pietikälnen, M.: 'Facial expression recognition from near-infrared videos', *Image Vis. Comput.*, 2011, **29**, (9), pp. 607–619
13 Paul, E., Friesen, W.V.: 'Constants across cultures in the face and emotion', *J. Personality Soc. Psychol.*, 1971, **17**, (2), pp. 124–129
14 Paul, E., Friesen, W.V.: 'Investigator's guide to the facial action coding system, part two' (CA: Consulting Psychologists Press, Palo Alto, 1978)
15 Xue, Y., Mao, X., Zhang, F.: 'Design and realization of BHU facial expression database', *J. Beijing Univ. Aeronaut. Astronaut.*, 2007, **2**, pp. 224–228
16 Sim, T., Baker, S., Bsat, M.: 'The CMU pose, illumination, and expression (PIE) database'. Proc. Fifth IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2002, pp. 46–51
17 Martinez, A.R., Aleix, M.: 'The AR face database', CVC Technical Report, 1998, p. 24
18 Jolliffe, I.T.: 'Principal component analysis' (Springer-Verlag New York, 1986)
19 Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: 'Eigenfaces vs. fisherfaces: recognition using class specific linear projection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 711–720
20 Comon, P.: 'Independent component analysis'. Higher-Order Statistics, 1992, pp. 29–38
21 Borg, I., Groenen, P.: 'Modern multidimensional scaling: theory and applications' (Springer Verlag, 2005)
22 Tenenbaum, J.B., Silva, V.D., Langford, J.C.: 'A global geometric framework for nonlinear dimensionality reduction', *Science*, 2000, **290**, (5500), pp. 2319–2323
23 Roweis, S.T., Saul, L.K.: 'Nonlinear dimensionality reduction by locally linear embedding', *Science*, 2000, **290**, (5500), pp. 2323–2326
24 Belkin, M., Niyogi, P.: 'Laplacian eigenmaps for dimensionality reduction and data representation', *Neural Comput.*, 2003, **15**, (6), pp. 1373–1396
25 Hinton, G., Roweis, S.: 'Stochastic neighbor embedding', *Adv. Neural Inf. Process. Syst.*, 2002, **15**, pp. 833–840
26 Kotsia, I., Zafeiriou, S., Pitas, I.: 'Texture and shape information fusion for facial expression and facial action unit recognition', *Pattern Recognit.*, 2008, **41**, (3), pp. 833–851
27 Mahdi, I., Hamed, S.H.: 'A novel fuzzy facial expression recognition system based on facial feature extraction from color face images', *Eng. Appl. Artif. Intell.*, 2012, **25**, (1), pp. 130–146
28 Shi, J., Samal, A., Marx, D.: 'How effective are landmarks and their geometry for face recognition', *Comput. Vis. Image Underst.*, 2006, **102**, (2), pp. 117–133
29 Park, S., Shin, J., Kim, D.: 'Facial expression analysis with facial expression deformation'. Proc. 19th Int. Conf. on Pattern Recognition, 2008, pp. 1–4
30 Valstar, M.F., Pantic, M.: 'Biologically vs. logic inspired encoding of facial actions and emotions in video'. Proc. IEEE Int. Conf. on Multimedia and Expo, 2006, pp. 325–328
31 Song, M., Tao, D., Liu, Z., Li, X., Zhou, M.: 'Image ratio features for facial expression recognition application', *IEEE Trans. Syst. Man Cybern. B, Cybern.*, 2010, **40**, (3), pp. 779–788
32 Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: 'Active shape models-their training and application', *Comput. Vis. Image Underst.*, 1995, **61**, (1), pp. 38–59

33 Cootes, T.F., Edwards, G.J., Taylor, C.J.: ‘Active appearance models’. Computer Vision-ECCV’98, 1998, pp. 484–498
34 Liu, Z., Shan, Y., Zhang, Z.: ‘Expressive expression mapping with ratio images’. Proc. 28th Annual Conf. on Computer Graphics and Interactive Techniques, 2001, pp. 271–276
35 Kanade, T., Cohn, J.F., Tian, Y.: ‘Comprehensive database for facial expression analysis’. Proc. Fourth IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2000, pp. 46–53
36 Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: ‘The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression’. Proc. 2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, 2010, pp. 94–101
37 Chew, S., Lucey, P., Lucey, S., Saragih, J., Cohn, J.F., Sridharan, S.: ‘Person-independent facial expression detection using constrained local models’. Proc. 2011 IEEE Int. Conf. on Automatic Face & Gesture Recognition and Workshops, 2011, pp. 915–920
38 Jeni, L.A., Lörincz, A., Nagy, T., *et al.*: ‘3D shape estimation in video sequences provides high precision evaluation of facial expressions’, *Image Vis. Comput.*, 2012, **30**, (10), pp. 785–795

IET Comput. Vis., 2014, Vol. 8, Iss. 5, pp. 429–440
doi: 10.1049/iet-cvi.2013.0171