International Journal of Automation and Computing 11(5), October 2014, 459-468
DOI: 10.1007/s11633-014-0835-0

Feature Representation for Facial Expression Recognition Based on FACS and LBP

Li Wang Rui-Feng Li Ke Wang Jian Chen

State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, Harbin 150080, China

Abstract: In expression recognition, feature representation is critical for successful recognition since it contains distinctive information of expressions. In this paper, a new approach for representing facial expression features is proposed with its objective to describe features in an effective and efficient way in order to improve the recognition performance. The method combines the facial action coding system (FACS) and "uniform" local binary patterns (LBP) to represent facial expression features from coarse to fine. The facial feature regions are extracted by active shape models (ASM) based on FACS to obtain the gray-level texture. Then, LBP is used to represent expression features for enhancing the discriminant. A facial expression recognition system is developed based on this feature extraction method by using K nearest neighborhood (K-NN) classifier to recognize facial expressions. Finally, experiments are carried out to evaluate this feature extraction method. The significance of removing the unrelated facial regions and enhancing the discrimination ability of expression features in the recognition process is indicated by the results, in addition to its convenience.

Keywords: Local binary patterns (LBP), facial expression recognition, active shape models (ASM), facial action coding system (FACS), feature representation.

1 Introduction

Feature representation is an active topic in the field of computer vision. Since 1990s, facial expression recognition (FER) has started to become a focus in the field of computer vision and machine learning. For recognizing facial expressions, feature representation or extraction plays an important role. Thus, an effective representation method is needed. Generally, good extraction methods own strong and robust discrimination capability to improve recognition performance effectively[1]. The descriptors of facial feature represent spatial, temporal or both of them in order to make a distinction among different expressions.

There are many methods for feature extraction or representation in the work of facial expression recognition. Active shape model was used to extract shape features[2], and active appearance model (AAM)[3, 4] was used to extract gray level features for FER. Some manifold methods were applied to FER by extracting the intrinsic essences of the image space[5−7], such as locality preserving projection, fuzzy nearest feature line-based manifold embedding, and linear discriminant analysis. But there are limitations on the small size database for its dimensional reduction property. Early in the last century, a method for automatically classifying facial images based on labeled elastic graph matching with 2D Gabor wavelet representation and linear discriminant analysis was proposed in [8]. Then, Gabor wavelet or log-Gabor features were used to recognize facial expressions by many researchers[9−11]. There are many feature extraction methods, such as Gabor filters, the scale invariant feature transform (SIFT), local binary patterns (LBP)[1, 12]. On account of computing amount and memory, Gabor feature has less effect on the recognition efficiency. And LBP, as a simple, efficient and robust local descriptor, has been employed in many domains, including face recognition[13, 14], facial expression recognition[15−17], and texture analysis[18−20]. Many extended researches on LBP have been done well, which motivated their applications and development for its advantages. Since feature images are downsampled before recognition for FER, feature extraction after detecting faces in images is pivotal in the recognizing course without considering the dimensions of face regions in the images.

From the previous studies on FER, there are face detection, feature extraction and expression classification on the still 2D images, image sequences, videos or 3D face structures by modeling 3D face on a single image or multiple-view images. In this paper, we focus on feature extraction or representation issue, which is determinant for facial expression recognition or analysis. As known to us, local feature descriptors adopted in the whole face region will inevitably generate more unrelated information. However, facial expressions are mainly generated from regions around eyebrows, mouth and cheek. Therefore, it is necessary to remove the useless regions and strengthen the expression regions. The main contribution of this work is to propose an approach to extract the effective facial expression regions and enhance the feature representation by using LBP. The main steps are as follows: Firstly, the face feature regions are extracted by active shape model (ASM) to remove the redundancy facial regions for expression recognition. This model is used to detect face and reduce the computation amount. Then, the facial expression features are represented by LBP to enrich expression information accurately in the main face regions extracted by ASM based on facial action coding system (FACS). The whole feature is simple, useful and utilizable. The K nearest neighborhood (K-NN)

Regular paper
Special Issue on Massive Visual Computing
Manuscript received January 17, 2014; accepted June 20, 2014
This work was supported by National Natural Science Foundation of China (No. 61273339)

classifier is used for FER based on the properties of the proposed feature representation method.

The organization of this paper is as follows. In Section 2, the facial feature extraction method on LBP and its derivative operator are presented. The FER system composed of feature extraction and classification is introduced in Section 3. The verification and assessment on the approach by experiments and result analysis based on confusion matrices of each facial expression recognition are shown in Section 4. Finally, conclusions are drawn in Section 5.

2 Local binary patterns

2.1 The original operator

LBP is a non-parametric operator, which is computed by comparing the gray-level values of the center pixel and pixels in its local neighborhood. And then the information of spatial structure is obtained according to the gray-level differences. Given an image with G gray levels, the joint probability distribution of gray levels for successive pixels according to the image coordinate is denoted by $p(g_0,\ g_1)$, $g_0,\ g_1 = 0,\ 1,\ \cdots,\ G-1$[21]. The distribution can be represented as follows after subtracting g_0 from g_1, because the average error is rather small in proportion to the average distribution. The joint probability distribution of gray-level image is defined as

$$p(g_0, g_1 - g_0) = p(g_0)p(g_1 - g_0). \tag{1}$$

LBP is computed to reduce the computation expense and statistical unreliability. Given a pixel in the image, the operator of LBP for this pixel is defined as

$$LBP_{(P,R)} = \sum_{i=1}^{P} s(g_0,g_i)2^{i-1}, s(g_0,g_i) = \begin{cases} 1, & \text{if } g_i \geqslant g_0 \\ 0, & \text{if } g_i \leqslant g_0 \end{cases} \tag{2}$$

where $(P,\ R)$ represents the number of pixels and the radius in the local neighborhood, g_0 is the gray level value of the center pixel, and g_i is the value of the local neighborhood. The original operators of LBP and computation process are shown in Fig. 1.

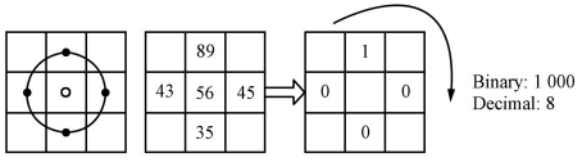

Fig. 1 The original operator of LBP

2.2 Uniform patterns

Since the original operator only computes LBP in the 4-neighborhood, bilinear interpolation is used to calculate the pixels not falling among the integer image coordinates to extend LBP application in large scale. The different sacles and size neighborhoods of LBP are shown in Fig. 2. The parameters $(P,\ R)$ of LBP with different radii and numbers of pixles in the neighborhood are (8, 2), (16, 3), and (8, 3), respectively.

The holistic method of feature representation is sensitive to illumination changes, while it is robust to rotation invariance. To make LBP more robust against these changes, more extension LBP operators have been proposed. Many attentions are paid to forming a more efficient descriptor with properties of rotation invariance and temporal transition. The "uniform" LBP is much more significant and fundamental than other LBP patterns, which is derived from the original LBP descriptor[22].

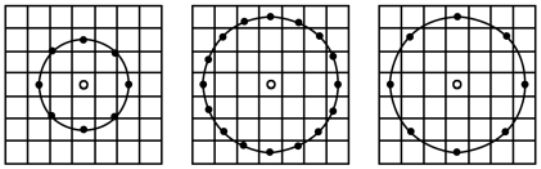

Fig. 2 LBPs with different radii and neighborhoods

The definition of the "uniform" LBP is that the U value is not more than two bitwise transitions, which is to say that the number of changes from 1 to 0 or from 0 to 1 is 0, 1, or 2. For example, 00110000 is a "uniform" pattern, but 00101000 is a nonuniform pattern. These patterns may describe the underlying texture properties and reduce noisy patterns and computation amount. The dimensionality 2^p of the original LBP reduces to $P^{P-1}+3$ for a "uniform" LBP. It is defined as

$$LBP^{u2}_{P,R} = \begin{cases} \sum\limits_{p=0}^{P-1} s(g_p - g_0), & \text{if } U(LBP_{P,R} \leqslant 2) \\ P\,(P-1) + 2, & \text{otherwise} \end{cases} \tag{3}$$

and their U values are computed as

$$U(LBP_{(P,R)}) = |s\,(g_{P-1} - g_c) - s\,(g_0 - g_c)| + \sum_{p=1}^{P-1} |s\,(g_p - g_c) - s\,(g_{p-1} - g_c)|. \tag{4}$$

3 The method for FER

The method proposed for FER in this paper consists of three parts: facial region extraction, feature representation, and expression recognition. The framework is shown in Fig. 3. After facial region extraction by ASM on the basis of FACS, facial features are extracted by LBP, and feature histograms are constructed to form feature vectors. These facial regions are composed of action unites (AUs), which produce expressions when human beings communicate with each other to realize non-language communications. At last, the K-NN classifier is used to recognize expressions based on the dissimilarity between testing samples and the existing images. The distances between them are measured by χ^2 statistic.

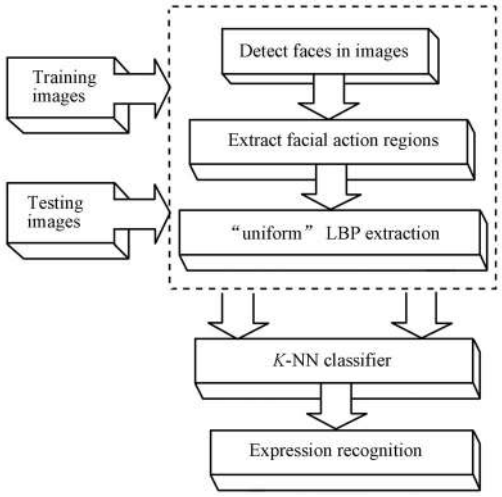

Fig. 3 The framework for FER

3.1 Active shape models

ASM, as a parameterized linear statistical model, was proposed[23]. It has been applied to the domains of medical image analysis, FER and tracking problems[24], and handled changes of face pose and occlusion effectively. Here, ASM is used to detect faces and extract facial regions by landmarks, which label the contours of objects, such as shapes of eyes and mouth.

Sample images with landmarks for ASM are shown in Fig. 4. Given one landmark labeled on the face features represented by image coordinate value (x_i, y_i), all the landmarks described the face region are formed as a vector:

$$s = (x_1, y_1, \cdots, x_n, y_n)^{\mathrm{T}}. \tag{5}$$

Fig. 4 Sample images with landmarks

It contains 68 vertices in each sample, which are used in this paper to construct the facial region distributing around eyes, eyebrows, nose, and mouth. After transformation, scale and rotation of the coordinates based on landmarks, ASM is constructed after calculating principle component analysis on the matrix constructed by the landmark vectors. It is described as

$$s = s_0 + \sum_{i=1}^{m} s_i p_i \tag{6}$$

where s_0 is the mean shape, p_i and s_i are the shape parameter and variation selected to characterize the main shape variation, respectively. p_i is the eigenvector obtained from the covariance matrix, and s_i is the eigenvalue corresponding to p_i. m is the number of eigenvectors used in the model. The mean shape and four shapes with the four largest eigenvalues are shown in Fig. 5.

Fig. 5 The shape with the largest eigenvalues

One advantage of using ASM to extract face region is that it can remove insignificant region of facial expression recognition and decrease calculation amount.

3.2 FACS for FER

FACS[25] is the most widely used and validated method for measuring and describing facial behavior[26]. It is an efficient, objective and comprehensive approach to present facial expression without drawbacks[27], and it is accepted by many researchers in psychology and physics aiming at FER. Basically, the face can be divided into upper and lower parts of facial region according to FACS. The system is composed of 46 AUs and the expressions to represent human emotional states. They are produced by the movements of AUs or their combination based on these system proposed by two American psychologists Ekman and Friesen[25]. Different unifications of AUs movements generate different expressions by setting the corresponding muscles under its skin in motion. For example, AU 12 raises the cheeks.

On the basis of FACS, AUs combination (AUC) is extracted according to ASM. The expression regions are extracted by segmented face regions into two or three parts based on AUC using ASM, respectively. The two face parts are composed of upper face region including eyes and eyebrows, and down face part including nose and mouth. The three face parts include left eye, right eye and down face region. Therefore, the useless regions for recognition are removed. Some image samples and their face region extraction in different AUC extractions are shown in Fig. 6. Another advantage of ASM is its convenience of AUs extraction which is based on the landmarks representing facial feature in a definite sequence.

Fig. 6 Sample images after using ASM based on AUC

3.3 Facial feature representation

According to [13], the region image is divided into multiple blocks for extracting "uniform" LBP features, and then its LBP image is constructed by these sub-regions. After describing one image by LBP, the histogram of each block in the image $f_b(x, y)$ is defined as

$$H_i = \sum_{x,y} I\left(f_b\left(x, y\right) = i\right), \quad i = 0, \cdots, n-1 \tag{7}$$

where n is the number of the patterns produced by "uniform" LBP operator, and I is described as

$$I\left(A\right) = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false.} \end{cases} \tag{8}$$

Then, the whole histogram of one image is defined as

$$H_{i,j} = \sum_{x,y} I\left(f_b\left(x,y\right)=i\right), \quad I(f_b\left(x,y\right)\in R_j) \tag{9}$$

where j= 0, $\cdots$, k. Here, k is the number of the divided sub-regions. The feature vector is represented by LBP histogram with length of $k\times(P^{P-1}+3)$, which is composed of the "uniform" pattern library.

For (P, R) = (8, 1), the number of "uniform" patterns is 59, and the dimension of each feature vector is 34×59 with 34 blocks, which is much smaller than 34×2^8. The images represented by LBP with different sizes of neighborhood and radii are shown in Fig. 7. The image is divided into 64 blocks. For facial region extraction based on AUC, each feature region is divided into multiple blocks according to their size, and then the feature vector is formed.

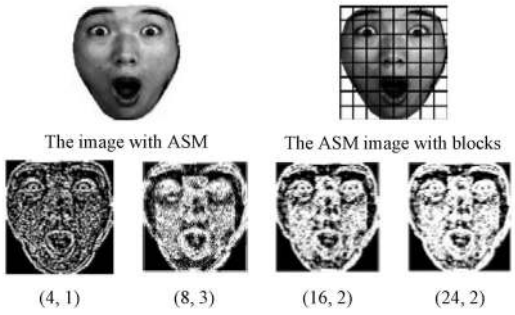

Fig. 7 LBP feature images with different radii and various sizes of local neighborhoods. Their values of (P, R) are described below each image after calculating LBP

3.4 Classification by using K-NN

There are many metrics for computing the dissimilarity of two histograms, such as log-likehood ratio, histogram intersection, and χ^2 statistic[28]. Here, χ^2 statistic is used to measure the dissimilarity between the test and the training images. It calculates the differences at each i-th bin, the histogram value of "uniform" LBP, as distinguished measurement. Chi square statistic metric is computed by

$$\chi^2\left(S,M\right)=\sum_i \frac{\left(S_i-M_i\right)^2}{S_i+M_i} \tag{10}$$

where S and M are two LBP histograms. K-NN classifier is used to recognize expression. It designates the type of expression of test sample images to the type of training images appearing among the K smallest χ^2 distance the most times. The aim of K-NN classifier is to find K nearest neighborhoods between one unknown sample and the known samples according to the distances between them[29]. In this paper, χ^2 metric is used to compute the distances. Given N known samples, there are N_1 samples from class w_1, N_2 samples from class w_2, $\cdots$, and N_c samples from class w_c. When the number of samples belonging to classes of w_1, w_2, $\cdots$, w_c are k_1, k_2, $\cdots$, k_c, the discrimination function is defined as

$$g_i(x)=k_i, \quad i=1,2,\cdots,c. \tag{11}$$

The fundamental determining rule is described as

$$g_j(x)=\max_i k_i.$$

Then, the class of x is considered as w_j.

According to [30], recognition vectors extracted by LBP are the histogram information, which are of lower dimensions and simple. Hence, K-NN classifier is employed for its convenience and easiness. In the work, K is chosen to be 1 for efficiency.

4 Experiments and analysis

In this paper, three databases are used to evaluate our method, including Japanese female facial expression (JAFFE), MMI and Cohn-Kanade, which are widespread and benchmark databases for FER testing[31−33]. The JAFFE database contains 213 images relative to ten female Japanese who devoted six predefined facial expressions respectively. The expressions are anger, disgust, fear, happiness, sadness and surprise. Each one posed 2 to 4 images with six expressions and neutral. The MMI database is created based on FACS from the man-machine interaction group of Imperial College London and has been developed as an easy research image database. It is composed of over 2 900 videos and high-resolution still images from 75 subjects of male and female, whose ages range from 19 to 62 years. It includes European, Asian, or South American from different ethnics. 602 images from 23 subjects are selected with 6 basic expressions and neutral in our test. The Cohn-Kanade database comprises 100 university students, whose ages are from 18 to 30. 1 752 images from 97 subjects are used in our experiment. There are also 6 expressions and neutral. Hence, 7 classes are classified in this section.

In this part, facial expression features are computed by LBP before accounting histogram to form feature vector for FER. LBP features of different radii and pixels in the neighborhood are extracted. There are 14 kinds of LBP features represented here classified by $K-$NN classifier, which are (P,R) = {(4, 1); (4, 2); (4, 3); (8, 1); (8, 2); (8, 3); (16, 1); (16, 2); (16, 3)}, called single LBP features and (P, R) = { (4, [1, 2, 3]); (8, [1, 2, 3]); (16, [1, 2, 3]; ([4, 3], [8, 2], [16, 1]); ([4, 1], [8, 2], [16, 3])), called multiple LBP features. To compare with other method for FER, 10-fold cross validation is performed. The database is divided into 10 groups averagely. One group is used to test while the rest groups are considered as the training samples. The whole face region is extracted by ASM labeled as $R_{\rm eg}$ = 1 for comparing with the method proposed in the paper. AUC with two face region parts is marked as $R_{\rm eg}$ = 2, and three parts as $R_{\rm eg}$ = 3.

4.1 Single LBP features for FFR

In this part, single LBP features are used as the differentiated features to classify expressions. The facial region is normalized into the same size. The images are preprocessed by Gaussian filter to reduce noise. The average recognition rates of the 10 groups from the 3 databases on AUC for $R_{\rm eg}$ = {2, 3} and $R_{\rm eg}$ = 1 are shown in Figs. 8–10. The horizontal ordinate describes the type of LBP features according

to $(P, R) = \{(4, 1); (4, 2); (4, 3); (8, 1); (8, 2); (8, 3); (16, 1); (16, 2); (16, 3)\}$ illustrated as in the above part and the vertical ordinate is the average recognition rate of 10-fold cross validation.

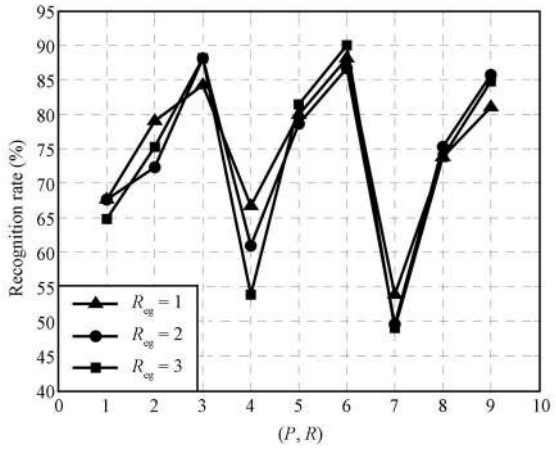

Fig. 8 Recognition rates with $R_{eg} = 1$ and two types of AUC using single LBP features on JAFFE data set

The recognition rates of the JAFFE database are shown in Fig. 8. It can be seen that the recognition performances are comparative for $R_{eg} = \{1, 2, 3\}$. The method of AUC with $R_{eg} = 3$ performs better than that with $R_{eg} = 2$. But the effects of our method on the MMI database shown in Fig. 9 are the best among the three databases. The combinations of AUs to present expressions proposed in this paper have higher recognition accuracy or comparable recognition effect in a certain extent. Furthermore, some redundant regions for recognizing expressions are removed in a convenient way. The representation capability with $R = 3$ is higher than those with $R = 1, 2$. Since the structures of nearer regions extracted by LBP are almost the same, the smaller neighborhoods contain less significant discriminant information for expression recognition.

4.2 Multiple features for FER

The experiments using multiple LBP features are performed in this section. The facial expression databases of the JAFFE, MMI and Cohn-Kanade are represented by J, M and C. The average rates with standard variances of the three data sets based on AUCs are exhibited in Table 1. The results are different from the three databases. The best recognition performance is indicated by the MMI data set, and the worst performance is indicated by the JAFFE data set. The complete correct results are obtained in many cases among the one group test in the 10-fold cross validation. The standard variances are different for different databases, from 9.8% on JAFFE data set to 1.2% on Cohn-Kanade data set. The most robust recognition performance appears on Cohn-Kanade database with the standard variance in the range 1.2%–1.8%, and the least stable case happens on the JAFFE database.

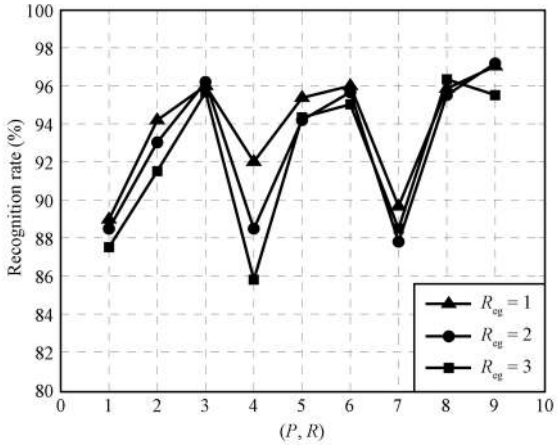

Fig. 9 Recognition rates with $R_{eg} = 1$ and two types of AUC using single LBP features on MMI data set

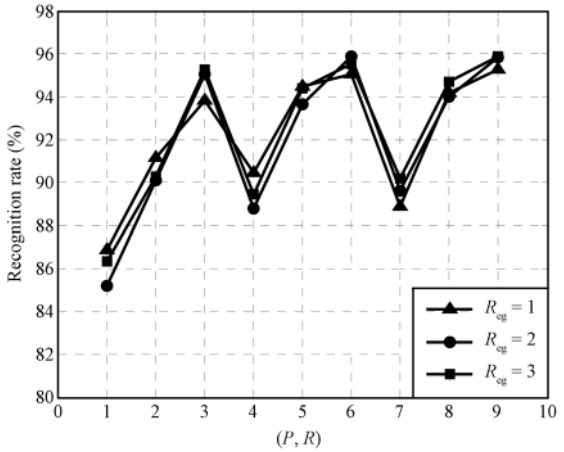

Fig. 10 Recognition rates with $R_{eg} = 1$ and two types of AUC using single LBP features on Cohn-Kanade data set

Table 1 The average recognition rates with standard variances for the proposed method on the three data sets with $R_{eg} = \{1, 2, 3\}$

R_{eg}	(P, R)	J	M	C
1	(4, [1,2, 3])	87.62±7.7	95.67±4.0	93.54±1.4
	(8, [1,2, 3])	87.14±9.8	95.83±5.1	95.26±1.3
	(16, [1,2, 3])	81.90±8.5	97.00±3.8	95.43±1.8
	([4, 1];[8, 2];[16, 3])	84.29±8.4	96.83±3.7	94.91±1.4
	([4, 3];[8, 2];[16, 1])	84.29±9.6	96.83±4.1	95.201.2
2	(4, [1,2, 3])	84.76±9.5	95.67±4.8	93.09±1.4
	(8, [1,2, 3])	86.67±6.3	95.67±5.1	95.77±1.3
	(16, [1,2, 3])	85.71±7.1	97.00±2.9	96.06±1.8
	([4, 1];[8, 2];[16, 3])	85.71±8.2	95.83±5.0	95.43±1.4
	([4, 3];[8, 2];[16, 1])	85.71±8.2	96.50±3.9	95.311.2
3	(4, [1,2, 3])	86.19±1.7	95.33±5.0	93.94±1.7
	(8, [1,2, 3])	89.05±6.0	95.33±4.9	95.77±1.3
	(16, [1,2, 3])	86.67±3.2	96.17±4.1	95.83±1.8
	([4, 1];[8, 2];[16, 3])	87.62±1.8	95.50±4.8	95.49±1.3
	([4, 3];[8, 2];[16, 1])	87.62±6.8	95.50±4.8	95.83±1.8

From the recognition results, it is notable that multiple LBP features own more robust representing capability than single LBP features. They represent facial expression features from different orientations and scales by using variables P and R. For example, when $P = 8$, the texture structures from angles $360^\circ \times \frac{p}{8}$ are extracted in 8-neighborhood, where $p = 1, 2, \cdots, P$. To obtain more detailed structure information, multiple local regions are used to describe LBP features. Although the differentiated effect is not improved evidently, it is comparable and comparative with $R_{\text{eg}} = 1$ by using less face regions. It indicates that our method has advantages of computational effectiveness and effective ability of expression recognition to a great extent.

Several other methods are shown in Table 2. It can be seen that our method performs better than the other feature representation methods for FER with higher and more robust recognition accuracy. This method reduces the useless face regions and improves recognition rate and speed by diminishing computation amount and memory. The rates combining LBP features are more robust than the other features, since they contain more significant and discriminative information, which is helpful or useful for expression recognition or analysis.

Table 2 The recognition rates of several other methods

Methods	JAFFE	MMI	Cohn-Kanade
LDN[34]	91.10±3.0	–	95.10±4.1
LDiP[35]	85.90±1.8	–	94.30±3.9
LBP[18]	81.00	86.70	85.00±4.5
LBP[36]	81.59±3.5	–	94.88±3.1
Gabor[37]	80.95	–	–

4.3 Result analysis

To analyze the performance of this method on each expression, the confusion matrices of different numbers of face regions are shown in Tables 3 to 8 with multiple LBP features from the three databases for $R_{\text{eg}} = \{2, 3\}$, respectively. They show the effect of AUC on expression recognition. The right parts of the tables represent the matrix of rates for $R_{\text{eg}} = 2$, while the left parts of the tables for $R_{\text{eg}} = 3$. Here, the confusion matrices are computed for all the images in the test of multiple LBP features for their high recognition rate and robustness. In these tables, the expressions are presented by the values of 1 to 7, which represent anger, disgust, fear, happiness, neutral, sadness and surprise, respectively. The second row of each table is the correct expressions, and the first column is the recognized expressions. The rates in the diagonal are higher than the other positions in each table, while the rates in the other positions are much smaller.

From Table 3, surprise is the worst expression for $R_{\text{eg}} = \{2, 3\}$ with only accuracy rates of 80.00% and 76.67%. Surprise is recognized as expressions of fear, happiness, and sadness when $R_{\text{eg}} = 2$. It is recognized as anger, disgust, fear, happiness, and sadness when $R_{\text{eg}} = 3$. Neutral can be completely recognized correctly for $R_{\text{eg}} = 3$, and is the most easily recognized expressions in the JAFFE database with all AUCs proposed in the paper. For the MMI database, the recognition rate of fear is the lowest for $R_{\text{eg}} = 2$ with 92.97% and 3.78% error rate for surprise. For $R_{\text{eg}} = 3$, disgust and fear are most easily confused with anger and surprise, respectively. The neutral is recognized completely right. Neutral is the most difficulty class to be discriminated for $R_{\text{eg}} = \{2, 3\}$ on the Cohn-Kanade database. It is recognized as fear or sadness in most cases. It is necessary that more work and efforts are to be made on improving discriminant ability by enhancing the describing function of LBP or combining multiple LBP features to strengthen some facial expression features.

Table 3 Confusion matrix of classification rates (%) on JAFFE data set for $R_{\text{eg}} = 2$

	1	2	3	4	5	6	7
1	85.33	2.76	0.00	0.00	0.00	2.00	0.00
2	3.33	89.66	7.74	1.94	0.00	0.00	0.00
3	0.67	6.90	85.81	0.65	0.00	4.00	3.33
4	0.00	0.00	0.00	81.94	0.00	1.33	6.67
5	4.00	0.00	0.65	10.32	96.55	11.33	4.00
6	6.67	0.69	4.52	4.52	3.45	81.33	6.00
7	0.00	0.00	1.29	0.65	0.00	0.00	80.00

Table 4 Confusion matrix of classification rates (%) on JAFFE data set for $R_{\text{eg}} = 3$

	1	2	3	4	5	6	7
1	90.67	3.45	1.29	0.65	0.00	1.33	0.67
2	2.67	91.03	6.45	0.00	0.00	0.00	2.67
3	3.33	2.76	82.58	0.65	0.00	3.33	2.67
4	0.00	0.00	0.00	86.45	0.00	1.33	5.33
5	1.33	0.00	0.00	9.03	99.31	8.00	6.67
6	2.00	2.76	8.39	2.58	0.69	86.00	5.33
7	0.00	0.00	1.29	0.65	0.00	0.00	76.67

Table 5 Confusion matrix of classification rates (%) on MMI data set for $R_{\text{eg}} = 2$

	1	2	3	4	5	6	7
1	97.62	1.07	0.00	0.00	0.00	0.00	0.19
2	0.95	96.00	1.89	0.00	0.00	0.24	0.00
3	0.00	1.07	92.97	0.37	0.00	0.00	1.92
4	0.00	1.07	0.27	98.52	0.00	0.71	0.00
5	1.43	0.00	0.81	0.37	100.00	0.95	0.19
6	0.00	0.80	0.27	0.74	0.00	98.10	0.00
7	0.00	0.00	3.78	0.00	0.00	0.00	97.69

From Figs. 11–16, the confused expressions from the three databases are shown respectively. The images in the left part of the first row are the testing samples and the recognized images with $R_{\text{eg}} = \{1, 2\}$ or $R_{\text{eg}} = \{1, 3\}$, and the expressions of the images are indicated in the following. The histograms are in the rest rows, in which the first row is for the testing samples for $R_{\text{eg}} = 1$ in the left part and the recognized samples in the right part, and the second rows are the testing samples for $R_{\text{eg}} = 2$ or $R_{\text{eg}} = 3$ in the

left part and the recognized samples in the right part. In the figures of histogram, the horizontal ordinate values are the values of "uniform" patterns, and the vertical ordinate values are the statistics of the patterns.

Table 6 Confusion matrix of classification rates (%) on MMI data set for $R_{\text{eg}} = 3$

	1	2	3	4	5	6	7
1	95.24	2.13	0.00	0.00	0.00	0.48	0.19
2	0.95	93.07	1.89	0.19	0.00	0.24	0.00
3	0.00	0.80	93.24	0.19	0.00	0.00	1.73
4	0.00	1.60	0.81	98.89	0.00	0.71	0.00
5	2.62	1.07	1.35	0.56	100.00	0.48	0.00
6	0.48	0.27	0.27	0.19	0.00	98.10	0.00
7	0.71	1.07	2.43	0.00	0.00	0.00	98.08

Table 7 Confusion matrix of classification rates (%) on Cohn-Kanade data set for $R_{\text{eg}} = 2$

	1	2	3	4	5	6	7
1	97.90	0.24	0.19	0.00	0.67	0.24	0.00
2	0.00	96.67	0.06	0.13	0.60	0.00	0.00
3	0.99	0.24	97.77	0.00	11.33	0.00	0.31
4	0.00	0.48	0.31	99.73	1.67	0.00	0.00
5	0.62	1.07	1.36	0.00	78.60	0.98	0.08
6	0.49	0.36	0.06	0.13	7.00	98.78	0.00
7	0.00	0.95	0.25	0.00	0.13	0.00	99.61

Table 8 Confusion matrix of classification rates (%) on Cohn-Kanade data set for $R_{\text{eg}} = 3$

	1	2	3	4	5	6	7
1	98.40	0.24	0.06	0.00	1.27	0.00	0.00
2	0.00	96.55	0.06	0.13	0.40	0.00	0.00
3	0.37	0.36	97.71	0.00	11.73	0.00	0.47
4	0.00	0.60	0.12	99.80	1.60	0.00	0.00
5	0.74	1.07	1.73	0.00	79.60	0.82	0.08
6	0.49	0.48	0.06	0.07	5.27	99.18	0.00
7	0.00	0.71	0.25	0.00	0.13	0.00	99.45

In Figs. 11 and 12, the confused expressions from the JAFFE database are shown for $R_{\text{eg}} = \{1, 2\}$ and $R_{\text{eg}} = \{1, 3\}$, respectively. There are subtle differences between the testing samples and the recognized images to discriminate expressions. For example, happiness with small movement amplitude of mouth corners is classified as neutral, as shown in Fig. 11. Fear is illustrated as mouth and eyes opening widely, eyebrows moving up, and nostrils flared, while anger is described as eyebrows downward, forehead wrinkled together, and tightening eyelids and lip. From Fig. 12, the expressions of fear and anger are very similar in the movements of mouths and eyes.

In Figs. 13 and 14, the confused expressions from the MMI database are shown. The confused expressions in the most cases are fear and surprise for $R_{\text{eg}} = \{1, 2, 3\}$. Surprise is represented as drooping lower jaw, relaxing lips and mouth, opening eyes wide, raising eyelids and eyebrows slightly up. The main differences between them are the movement of mouth, eyes and eyebrows. But sometimes, their movement combinations seem to be comparably similarly as shown in Figs. 13 and 14.

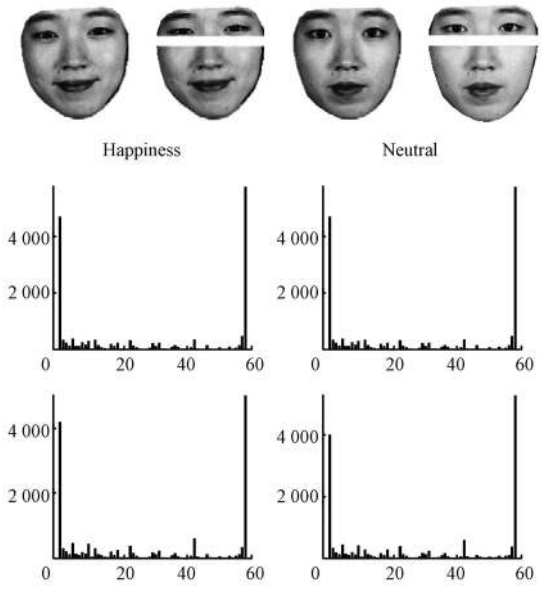

Fig. 11 The expression images of happiness and neutral easily confused from JAFFE data set for $R_{\text{eg}} = \{1, 2\}$

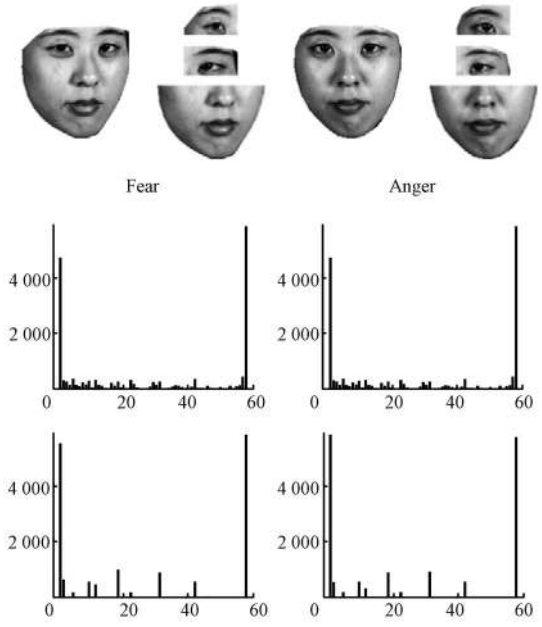

Fig. 12 The expression images of fear and anger easily confused from JAFFE data set for $R_{\text{eg}} = \{1, 3\}$

In Cohn-Kanade database, neutral is the most easily confused expression. The wrongly recognized expressions of neutral are fear and surprise, as shown in Figs. 15 and 16

for $R_{\text{eg}} = \{2, 3\}$. It can be seen that the action unit around mouth and eyes are similar compared among the three expressions. They are from the same subject with subtle movements to express fear and surprise, and are mix up from human and computer vision. More effective and finer feature representing method is crucial to discriminate these similar expressions.

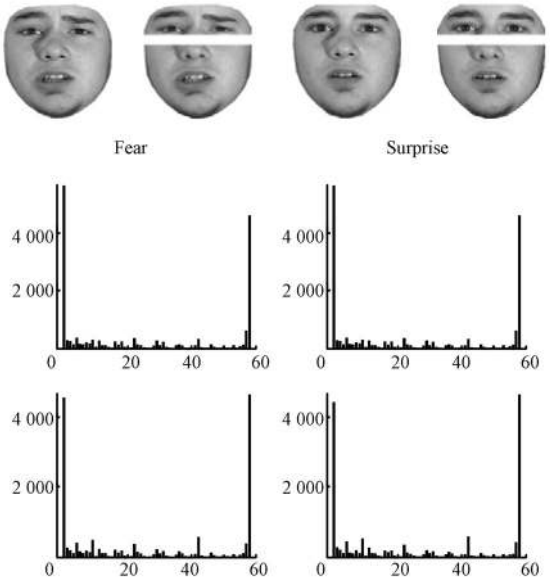

Fig. 13 The expression images of fear and surprise easily confused from MMI data set for $R_{\text{eg}} = \{1, 2\}$

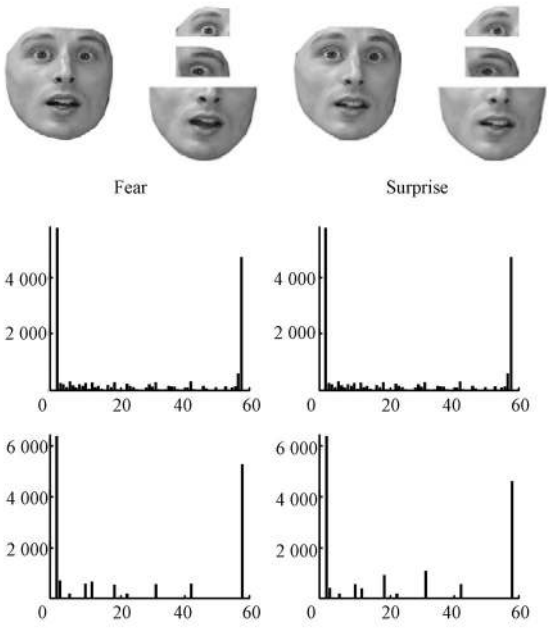

Fig. 14 The expression images of fear and surprise easily confused from MMI data set for $R_{\text{eg}} = \{1, 3\}$

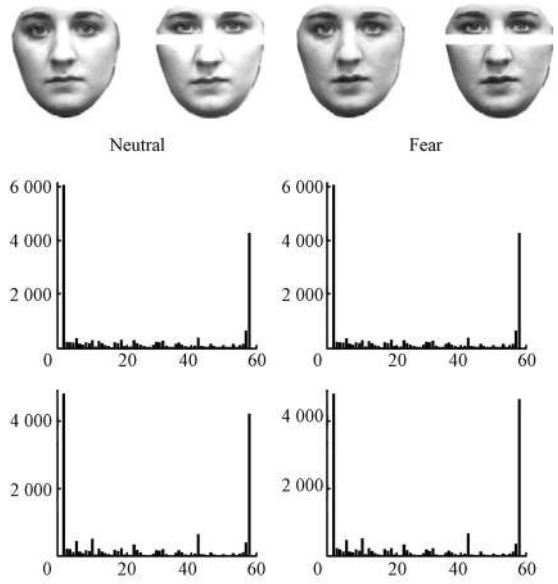

Fig. 15 The expression images easily confused from Cohn-Kanade data set for $R_{\text{eg}} = \{1, 2\}$

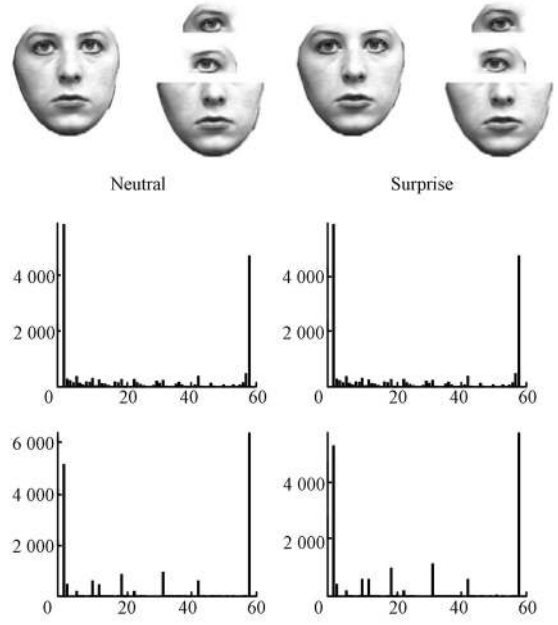

Fig. 16 The expression images easily confused from Cohn-Kanade data set for $R_{\text{eg}} = \{1, 3\}$

5 Conclusions

In this paper, a new feature representing method for FER combined LBP and FACS is proposed. The advantages of the proposed method include lower computational requirement and easier implementation. FER is implemented on facial regions to a maximum extent. The regions are segmented from the whole image using ASM. The different AUCs are acquired based on FACS. Highly-distinguished

facial expression vectors are obtained through "uniform" LBP for simplification. This makes it possible to employ a traditional and brute-force K-NN classifier for expression recognition.

To evaluate the performance of this approach, multiple tests have been executed. The recognition accuracy is acceptable and this method can be applied with the highest rate of 100% in some cases on the basis of AUC from the three data sets. By making comparisons between other recognition methods and our method of AUCs combing with LBP, 10-fold cross validation is carried out on the databases. The later method is better and more stable than the former for its simplification and effectiveness. From the results on the two types of AUC, it is notable that the highest rate can reach 100%. So, the method of critical facial feature extraction is capable of increasing recognition accuracy and decreasing computation memory.

In the future work, there are still some meaningful factors to improve the recognition performance. It is essential to extract facial expression information based on more meaningful LBP. Another way is to adopt an optimization method to select features and study more effective AUC.

References

[1] J. Li, N. M. Allinson. A comprehensive review of current local features for computer vision. *Neurocomputing*, vol. 71, no. 10–12, pp. 1771–1787, 2008.

[2] M. Pantic, J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Learning*, vol. 22, no. 12, pp. 1424–1445, 2000.

[3] R. Alemy, M. E Shiri, F. Didehvar, Z. Hajimohammadi. New facial feature localization algorithm using adaptive active shape model. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 26, no. 1, Article 1256003–1–1256003–18, 2012.

[4] F. Q. Tang, B. Z. Deng. Facial expression recognition using AAM and local facial features. In *Proceedings of the 3th International Conference on Natural Computation*, Haikou, China, pp. 632–635, 2007.

[5] C. Martin, U. Werne, H. M. Gross. A real-time facial expression recognition system based on active appearance models using gray images and edge images. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE, Amsterdam, Netherlands, pp. 837–842, 2008.

[6] C. F. Shan, S. G. Gong, P. W. McOwan. Appearance manifold of facial expression. *Computer Vision in Human-Computer Interaction*, Berlin, Heidelberg: Springer, pp. 221–230, 2005.

[7] W. Li, Q. Ruan, J. Wan. Fuzzy nearest feature line-based manifold embedding for facial expression recognition. *Journal of Information Science and Engineering*, vol. 29, no. 2, pp. 329–346, 2013.

[8] R. Xiao, Q. J. Zhao, D. Zhang, P. F. Shi. Facial expression recognition on multiple manifolds. *Pattern Recognition*, vol. 44, no. 1, pp. 107–116, 2011.

[9] M. J. Lyons, J. Budynek, S. Akamatsu. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

[10] L. L. Shen, L. Bai, Z. Ji. A SVM face recognition method based on optimized Gabor features. In *Proceedings of the 9th International Conference on Visual Information Systems*, *Lecture Notes in Computer Science* Springer, Shanghai, China, pp. 165–174, 2007.

[11] W. Jiang, J. Zhang, T. Z. Shen, X. H. Wang. A novel facial features extraction algorithm using Gabor wavelets. In *Proceedings of the 1st International Congress on Image and Signal Processing*, IEEE, Sanya, China, pp. 649–653, 2008.

[12] J. Satake, M. Chiba, J. Miura. Visual person identification using a distance-dependent appearance model for a person following robot. *International Journal of Automation and Computing*, vol. 10, no. 5, pp. 438–446, 2013.

[13] S. Y. Fu, X. Kuai, G. S. Yang. Facial expression recognition by independent log-Gabor component analysis. In *Proceedings of the 8th International Symposium on Neural Networks*, Guilin, China, pp. 305–312, 2011.

[14] T. Ahonen, A. Hadid, M. Pietikäinen. Face recognition with local binary patterns. In *Proceedings of the 8th European Conference on Computer Vision*, Prague, Czech Republic, pp. 469–481, 2004.

[15] T. Ahonen, A. Hadid, M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Learning*, vol. 28, no. 12, pp. 2037–2041, 2006.

[16] G. Zhao, M. Pietikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Learning*, vol. 29, no. 6, pp. 915–928, 2007.

[17] X. M. Zhao, S. Q. Zhang. Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding. *EURASIP Journal of Advances in Signal Processing*, pp. 1–9, 2012.

[18] C. F. Shan, S. G. Gong, P. W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[19] L. Liu, L. Zhao, Y. Long, G. Kuang, P. Fieguth. Extended local binary patterns for texture classification. *Image and Vision Computing*, vol. 30, no. 2, pp. 86–99, 2012.

[20] T. Ojala, M. Pietikäinen, D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[21] T. Ojala, V. Kimmo, E. Oia, M. Pietikäinen. Texture discrimination with multidimensional distributions of signed gray-level differences. *Pattern Recognition*, vol. 34, no. 3, pp. 727–739, 2001.

[22] T. Ojala, M. Pietikäinen, T. Maäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[23] T. F. Cootes, C. J. Taylor, D. H. Copper, J. Graham. Active shape models–their training and applications. *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[24] K. C. Huang, Y. H. Kuo, M. F. Horng. Emotion recognition by a novel triangular facial feature extraction method. *International Journal of Innovative Computing Information and Control*, vol. 8, no. 11, pp. 7729–7746, 2012.

[25] P. Ekman, W. Friensen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Palo Alto: Consulting Psychologists Press, 1978.

[26] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, T. J. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Learning*, vol. 21, no. 10, pp. 974–989, 1999.

[27] A. Ryan, J. F. Cohn, S. Lucey, J. Saragih. Automated facial expression recognition system. In *Proceedings of the 43rd Annual International Conference on Security Technology*, IEEE, Zurich, Germany, pp. 172–177, 2009.

[28] Z. H. Guo, L. Zhang, D. Zhang. Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recognition*, vol. 43, no. 3, pp. 706–719, 2010.

[29] Z. Bian, X. Zhang. *Pattern Recognition*, Beijing, China: Tsinghua University Press, pp. 136–142, 2000. (in Chinese)

[30] G. Y. Zhao, M. Pietikäinen. Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. *Pattern Recognition Letters*, vol. 30, no. 12, pp. 1117–1127, 2009.

[31] M. J. Lyons, M. Kamachi, J. Gyoba. Japanese Female Facial Expressions (JAFFE), Database of digital images, 1997.

[32] M. Pantic, M. F. Valstar, R. Rademaker, L. Maat. Web-based database for facial expression analysis. In *Proceedings of IEEE International Conference on Multimedia and Expo*, IEEE, Amsterdam, Netherlands, pp. 317–321, 2005.

[33] T. Kanade, J. F. Cohn, Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE, Grenoble, France, pp. 46–53, 2000.

[34] R. Rivera, J. R. Castillo, O. Chae. Local directional number pattern for face analysis: Face and expression recognition. *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1740–1752, 2013.

[35] T. Jabid, M. H. Kabir, O. Chae. Local directional pattern (LDP) for face recognition. In *Proceedings of International Conference on Consumer Electronics*, IEEE, Las Vegas, NV, USA, pp. 329–330, 2010.

[36] X. M. Zhao, S. Q. Zhang. Facial expression recognition based on local binary patterns and kernel discriminant isomap. *Sensors*, vol. 11, no. 12, pp. 9573–9588, 2011.

[37] J. Yu, B. Bhanu. Evolutionary feature synthesis for facial expression recognition. *Pattern Recognition Letters*, vol. 27, no. 11, pp. 1289–1298, 2006

Li Wang received the bachelor degree in mechanical engineering and automation from Hebei University of Engineering, China in 2005. She received the M. Sc. degree in mechanical design and theory from Shijiazhuang Tiedao University, China in 2009. Currently, he is a Ph. D. candidate in School of Mechatronics Engineering of Harbin Institute of Technology, China.

Her research interests include pattern recognition, image processing and computer vision.

E-mail: wangli-hb@163.com (Corresponding author)

Rui-Feng Li obtained the bachelor, M. Sc. and Ph. D. degrees in engineering from School of Mechatronics Engineering of Harbin Institute of Technology, China in 1988, 1991, 1997, respectively. He is currently a professor in the State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, China.

His research interests include intelligent service robot system, advanced industrial robot technology and humanoid robot mechanism.

E-mail: lrf100@hit.edu.cn

Ke Wang obtained the bachelor degree in automation and the Ph. D. degree in control theory and control engineering, from Dalian University of Technology, China in 2002 and 2008, respectively.

He is currently a lecturer in the State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, China.

His research interests include computer vision, image processing, human machine interaction, and estimation theory and their applications in robotics.

E-mail: wangke@hit.edu.cn

Jian Chen graduated from Shenyang University of Chemical Technology, China in 2006. He received the M. Sc. degree from Shenyang University of Chemical Technology, China in 2009. He is currently a Ph. D. candidate at School of Mechatronics Engineering of Harbin Institute of Technology, China.

His research interests include robust control of flexible joint robots and industrial robotic trajectory planning.

E-mail: phevanschen@gmail.com