



Fully automatic 3D facial expression recognition using polytypic multi-block local binary patterns

Xiaoli Li ^{a,b}, Qiuqi Ruan ^{a,b}, Yi Jin ^{a,b}, Gaoyun An ^{a,b}, Ruizhen Zhao ^{a,b}

^a Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

^b Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China



ARTICLE INFO

Article history:

Received 14 April 2014

Received in revised form

17 September 2014

Accepted 24 September 2014

Available online 6 October 2014

Keywords:

3D facial expression recognition

Automatic data normalization

P-MLBP

Feature-based irregular divisions

Feature fusion

ABSTRACT

3D facial expression recognition has been greatly promoted for overcoming the inherent drawbacks of 2D facial expression recognition and has achieved superior recognition accuracy to the 2D. In this paper, a novel holistic, full-automatic approach for 3D facial expression recognition is proposed. First, 3D face models are represented in 2D-image-like structure which makes it possible to take advantage of the wealth of 2D methods to analyze 3D models. Then an enhanced facial representation, namely polytypic multi-block local binary patterns (P-MLBP), is proposed. The P-MLBP involves both the feature-based irregular divisions to depict the facial expressions accurately and the fusion of depth and texture information of 3D models to enhance the facial feature. Based on the BU-3DFE database, three kinds of classifiers are employed to conduct 3D facial expression recognition for evaluation. Their experimental results outperform the state of the art and show the effectiveness of P-MLBP for 3D facial expression recognition. Therefore, the proposed strategy is validated for 3D facial expression recognition; and its simplicity opens a promising direction for fully automatic 3D facial expression recognition.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Facial expression recognition has attracted considerable attention in the past decades due to its potential applications in various fields, such as human-computer interaction, psychological studies, and facial animation. However, the performance of traditional algorithms for facial expression recognition degrades heavily with the illumination and head pose variations, as they are applied to 2D static images [1,2] or image sequences [3,4]. In order to address these limitations of 2D systems, many researchers explore this issue in 3D space which captures true facial surface

data and has better stability and robustness. Thus, how to represent 3D facial feature effectively remains a central problem in 3D facial expression recognition.

Up to date, a number of methods for 3D facial expression recognition have been proposed and have formed two main streams. One of the main streams is the feature-based approaches which require accurate localization of fiducial points. Wang et al. [5] marked 64 key points on face models to construct 7 expression local regions, where 12 primitive facial features were extracted for recognition. In [6,7], 6 distance vectors including eye opening, eyebrow height, mouth opening, mouth height, lip stretched and their normalization were generated as the facial features based on manual-labeled points. In [8], 24 normalized Euclidean distances were chosen as the best one from a pool of candidate features using the relative entropy. Tang et al. [9] took the properties of line segments into

E-mail addresses: 09112087@bjtu.edu.cn (X. Li), qruan@bjtu.edu.cn (Q. Ruan), yjin@bjtu.edu.cn (Y. Jin), gayan@bjtu.edu.cn (G. An), rzhzhao@bjtu.edu.cn (R. Zhao).

consideration to generate features. Similarly, based on the predefined key points in BU-3DFE database [10], six distances plus the angles and slopes that related to the shape of eyes and mouth were used for recognition in [11]; 46 distances and 27 angles were taken for classification in [12]; a complete set of distance vectors between all the landmarks were used in [13–16] and were additionally optimized based on NSGA II [13,17]. Besides those provided by BU-3DFE database, more points were manually marked [18,19] or sampled from the lines connecting with certain points [20]. Moreover, other efficient features were utilized for 3D facial expression recognition [20,21].

Another main stream of researches on 3D facial expression recognition is the model-based approaches. They always transformed a general neutral model to fit the facial expression model and subsequently took the fitting parameters to generate the feature vector. The statistical model was usually taken as the general model, and the morphable model which was raised in [22] was one of them. The Statistical Facial feAture Model (SFAM) was also employed as an alternative type of morphable model in [23]. The model was fitted by learning both global and local variations around landmarks and the fitting parameters were achieved to express facial features. Another alternative was the Basic Facial Shape Component (BFSC) model [24], which was able to model neutral faces, so the subtraction of BFSC from original mesh was used to form the feature vectors. Ref. [25] proposed the bilinear model to represent faces and facial expressions jointly. After establishing the correspondence among a set of faces using elastically deformation, they encoded faces and facial expressions at the same time based on both symmetric and asymmetric formulations.

Although the previous publications in both of the main streams have performed well in reducing the influence of illumination and pose variation and have achieved high recognition accuracies, we can still find several shortcomings addressed in both of them. On one hand, the feature-based approaches, which rely on the accurate localization of a great number of landmarks, have broken the automation of the recognition system. In addition, detection of landmarks still remains to be an open problem. Hence, automatic 3D facial expression recognition is one of the research directions [26]. Furthermore, the facial expressions are expressed in sparse pattern using the feature-based methods, which are possibly unable to express the facial expressions completely and precisely.

On the other hand, the statistic model (morphable model) employed in the model-based approaches is always constructed using a large set of training data, which needs the time-consuming alignment and fitting procedures to get point correspondence first, as well as the nonconvexity of the energy functions that trap them into a local optimum [27]. This procedure is complicated and is still an active field of research for 3D face data. Moreover, if the training data are not sufficient, the recognition performance will decrease [28]. So the size of the training set should be large to keep the performance, which makes it more complex to get the point correspondence. When the statistical model is constructed, it is involved as the intermediate to fit the new-coming scan, which is computationally expensive. As the amount of 3D face data increases, the problem will become even bigger. In addition, finding the optimal methods for model fitting is still an open area [29]. Accordingly, compared with the feature-based approaches, the model-based approaches are complex and computationally expensive.

In this paper, we take the overall information of 3D face models into consideration, so that the landmark detection is avoidable and the automation of the recognition system is well kept. Moreover, the complex procedure for model fitting is evitable. All these benefits are brought in by the 2D-image-like structure. This structure makes it possible to express both the depth values and the color information of 3D models in the structure similar to 2D images, so that the features of 3D face models can be extracted using the 2D methods.

Currently, lots of local descriptors are used to represent facial features. The local binary patterns (LBP) and its extension, local ternary patterns (LTP), extracted the small patterns of local gray-level differences to represent the higher-level image content [30]. Local quantized patterns (LQP) used a look-up table based vector quantization to code larger or deeper patterns [31]. Local phase quantization (LPQ) was based on quantizing the Fourier transform phase in local neighborhoods, which was often used for blurred face recognition [32]. The structured sparse representation based classification (SSRC) extended with the occlusion dictionary learning was specially proposed for face recognition with occlusion [33]. The common features discriminant analysis was employed to combat the modality gap, and it's successfully used for heterogeneous face images [34].

While LBP features have been successfully used in 2D facial expression recognition, we consider using LBP descriptors to

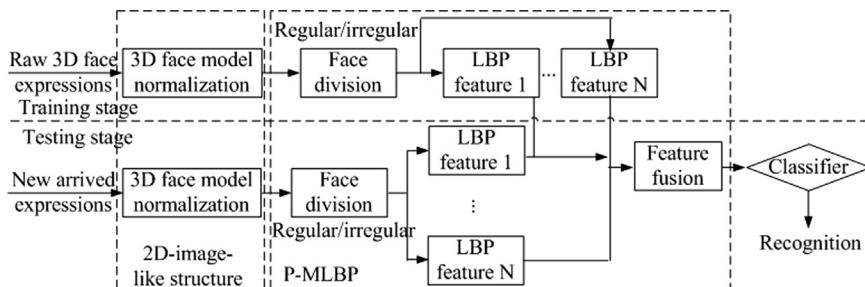


Fig. 1. The framework of the proposed algorithm.

express facial expressions. Motivated by this strategy, this paper proposes a novel facial feature, polytypic multi-block local binary patterns (P-MLBP), to depict facial expressions accurately. Based on the original LBP codes, this feature involves both the feature-based irregular divisions and the fusion of facial data to obtain a precise description of local faces. The framework of 3D facial expression recognition using the proposed strategy is illustrated in Fig. 1. If the classifier is a learning method, e.g. SVM, it can be divided into two stages: training and testing. In both of them, 3D face models are firstly normalized in 2D-image-like structure. Then the P-MLBP features are extracted. Finally, the classifier is applied to get the recognition results.

The rest of this paper is organized as follows. Section 2 presents the steps to normalize 3D face models. In Section 3, the polytypic multi-block LBP (P-MLBP) is introduced in details as the facial feature for 3D facial expression recognition. Section 4 gives the experimental results and discussions. And finally Section 5 concludes this paper.

2. 3D face model normalization

Feature extraction is critical to a recognition system. However, it's difficult to extract features directly from 3D face models as their depth change greatly with irregular triangulations. This paper proposes an automatic facial normalization method to represent 3D models in 2D-image-like structure, so that the facial features can be directly extracted from the 3D models. The 2D-image-like structure brings us three advantages:

1. It's easy-handled, not requiring model fitting;
2. The overall information of facial expressions can be completely considered to extract features, so that the landmark detection can be avoided and the automation of recognition system is well maintained;
3. It builds the bridge between 3D models and 2D methods, so that 2D methods can be employed to extract facial features from 3D face models.

The procedure to achieve normalized 3D models in 2D-image-like structure can be separated into two parts (Fig. 2): Face preprocessing and face normalization, both of them are detailed in the following sections.

2.1. Face preprocessing to present 3D models in 2D-image-like structure

Generally in the preprocessing stage, the raw face models are required to fill the holes using the cubic interpolation and smooth the face surface with a median filter. Then the smoothed models are aligned using iterative closest point (ICP) [35] algorithm in order to correct the minor pose variation. If the variation of poses is a major consideration during recognition, method in [36] can be employed to correct them firstly.

However, algebraic operations cannot be directly applied to the triangular patches in 3D models. To solve this problem, the patches are converted into 2D-image structure. The steps for this work are as follows: Given a 3D face model, locate its nose tip P_{nose} whose depth value is the highest in the $-1/2^{th}$ and $1/2^{th}$ mask of the central portion; shift all other nodes P_i to a new coordinate space P'_i by subtracting the nose tip; determine the resolution of the face model with a uniform space; using the Qhull algorithm [37], obtain the depth values cubically interpolated over the uninform nodes; each channel of the color information (R, G, B) can be respectively obtained following the same way as the depth value does. Thus, the nodes of the resulting 3D models are in 2D-image-like structure. And each node of them contains both depth value and its corresponding color information (R, G, B). Fig. 3 gives the comparison between original triangular patches and the transformed 2D-image-like structure, where the nose regions are magnified.

2.2. Face normalization using point-to-point mapping between depth and texture

To reduce the influence of different facial outline during facial expression recognition, we pay more attention on the face central regions (ROI, region-of-interest), such as the eyes, eyebrows, and mouth, so the face models are required to be cropped. However, the range of ROI is hard to determine as the landmarks on 3D models are difficult to locate only using the depth values. Referred to some previous work on 3D face and 3D facial expression analysis, the landmarks they employed were always manually defined [14,38]. Regardless of the fact that some of them could detect the landmarks automatically, the number of

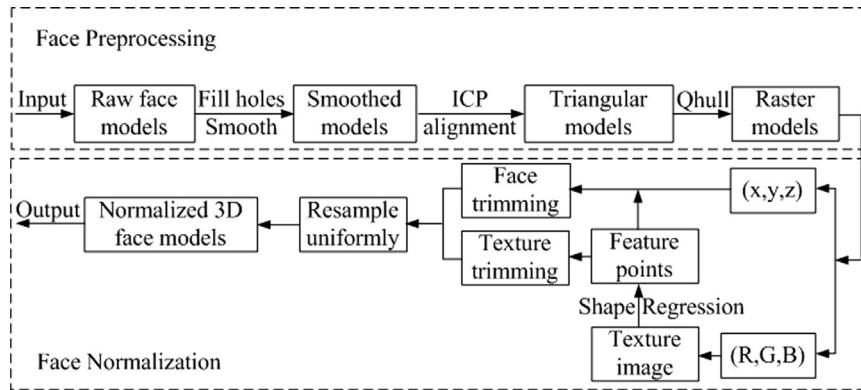


Fig. 2. The procedure to normalize 3D face models.

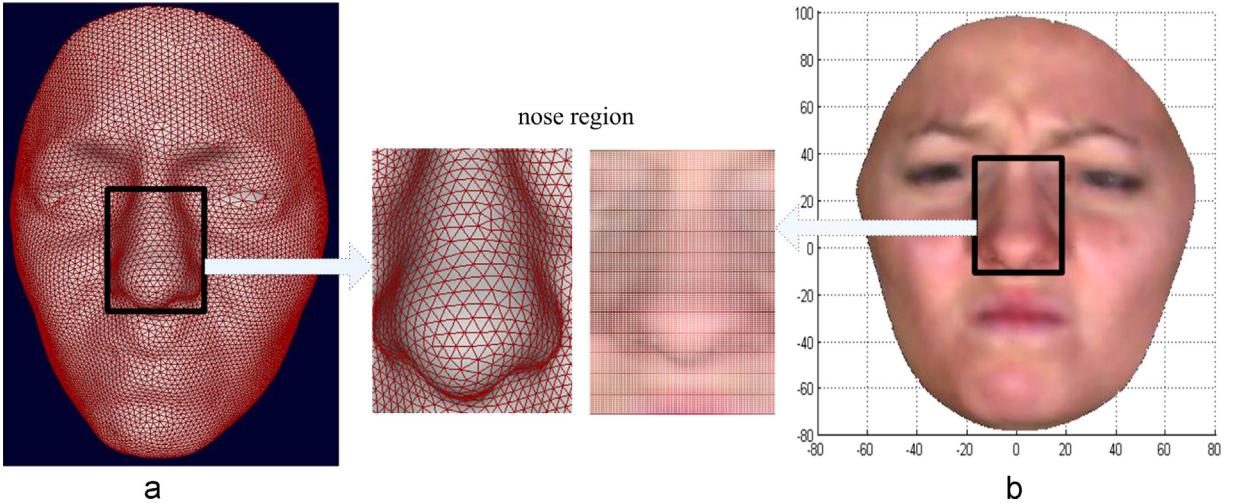


Fig. 3. The comparison of 3D face model in (a) triangular mesh and (b) 2D-image-like structure.

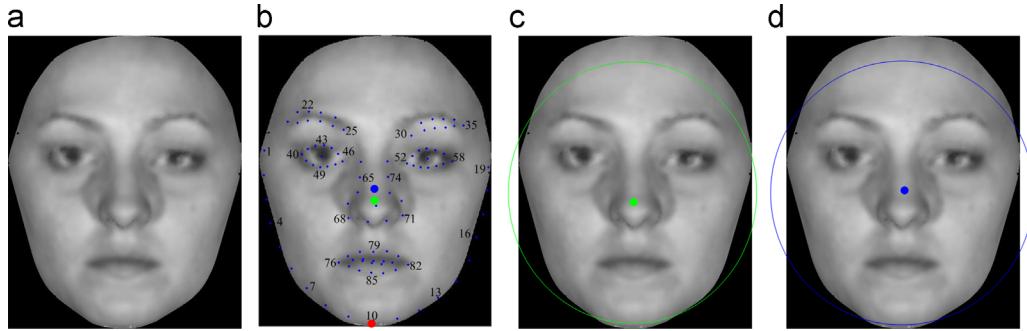


Fig. 4. The procedure to obtain the face central regions (ROI) (a) Generated texture image; (b) detected feature points; (c) the circle determined by the chin and nose tip; (d) the circle determined by the chin and nose bridge. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

detected landmarks was limited [39,40], which was insufficient for facial expression recognition. Moreover, lots of training data were needed in order to detect the landmarks, which were complex to implement [41]. So the methods for detecting landmarks are required to improve and remain to be an active research area [26]. Therefore, in this paper, we take the color information along with the depth values to detect the landmarks of 3D face models.

Actually, it's technically mature to detect landmarks automatically using color information; the point-to-point mapping between the color information and the depth values makes it possible to map the landmarks (obtained using color information) precisely onto the depth values. Therefore, we consider using the color information to locate these key points to determine the range of face ROI, and then depth values of face ROI can be cropped accordingly. To be simple and convenient, the 3-channel color information in face models is converted to YUV color space from the RGB space.

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

In addition, the component of luminance, Y value, can cover the complete graylevels of the image, so only the Y value is considered to generate the texture image.

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (2)$$

Therefore, the Y value for each pixel is obtained to generate the texture image. Then the texture image is used to detect these landmarks using the explicit shape regression [42]. The distribution of these landmarks is shown in Fig. 4(b) and numbered for the following description. Then these landmarks in the texture images are precisely mapped onto the depth values of face models using the point-to-point mapping.

In most of the existing methods for 3D face recognition, the face ROI is determined by a sphere whose center is the nose tip and radius is the distance between nose tip and eyebrow [43]. However, this strategy leads to the loss of useful expression information (see Fig. 8(c)). To solve this problem, we have designed two types of sphere for face ROI. The first one is to treat the nose tip as the center and the distance between nose tip and chin as the radius of the sphere. The circle is shown in Fig. 4(c) and the corresponding results are detailed in Fig. 8(d and e). The second one is

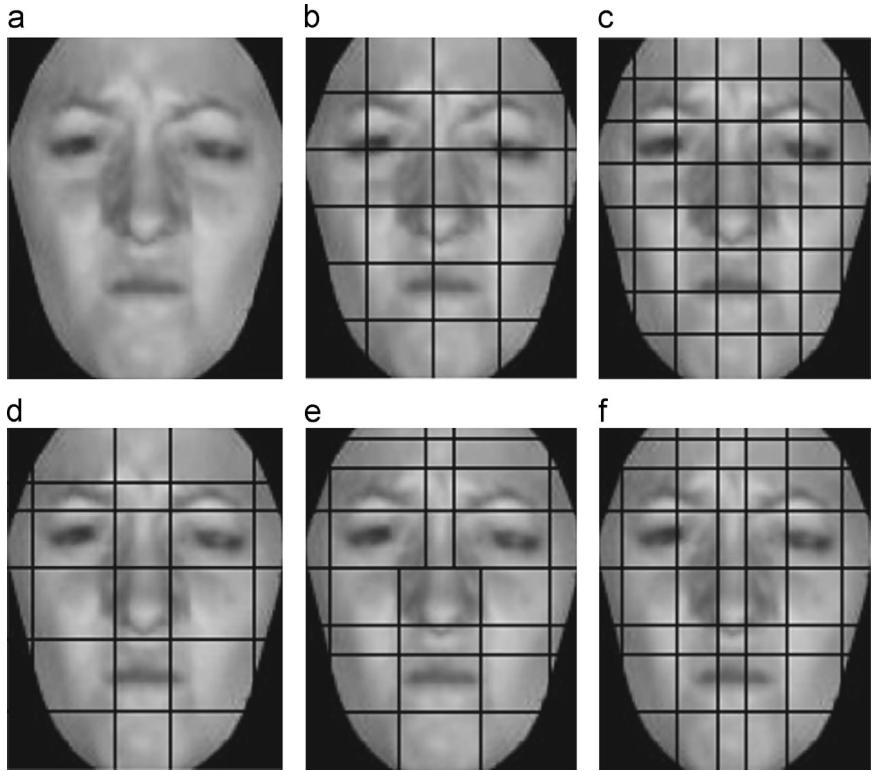


Fig. 5. Division schemes described based on a neutral texture image (a) Original texture image; (b) 6×5 regular division; (c) 8×8 regular division; (d) 6×5 irregular division; (e) 8×5 irregular division; (f) 8×7 irregular division.

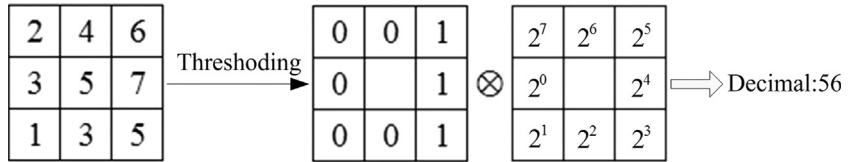


Fig. 6. The procedure of generating LBP code.

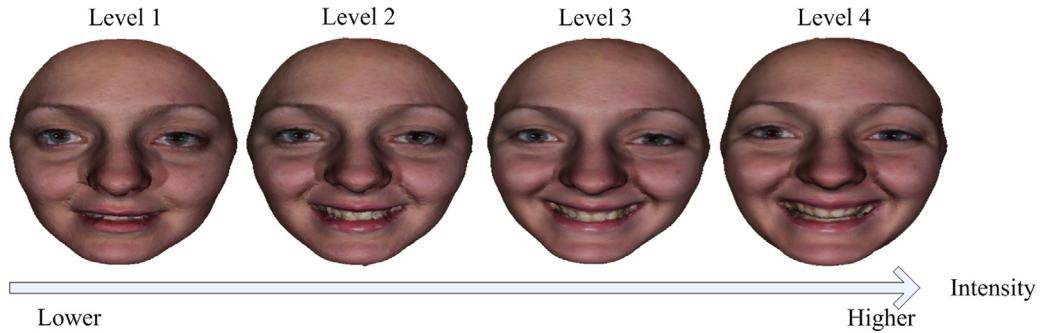


Fig. 7. Four different intensities of expression “happiness”.

to set the nose bridge as the center and the distance between the nose bridge and chin as the radius of the sphere. The point of the nose bridge is in the middle of 67th and 72nd points defined in Fig. 4(b) and it's also colored blue in Fig. 4(b). The circle is illustrated in Fig. 4(d) and the cropped face regions are detailed in Fig. 8(f and g).

Finally, texture images and the face models of the central region are unified to the same scale based on the cubic interpolation. Accordingly, all the faces used in the experiments are normalized using this method; not only the 3D face models but also their corresponding texture images are available to extract facial features.

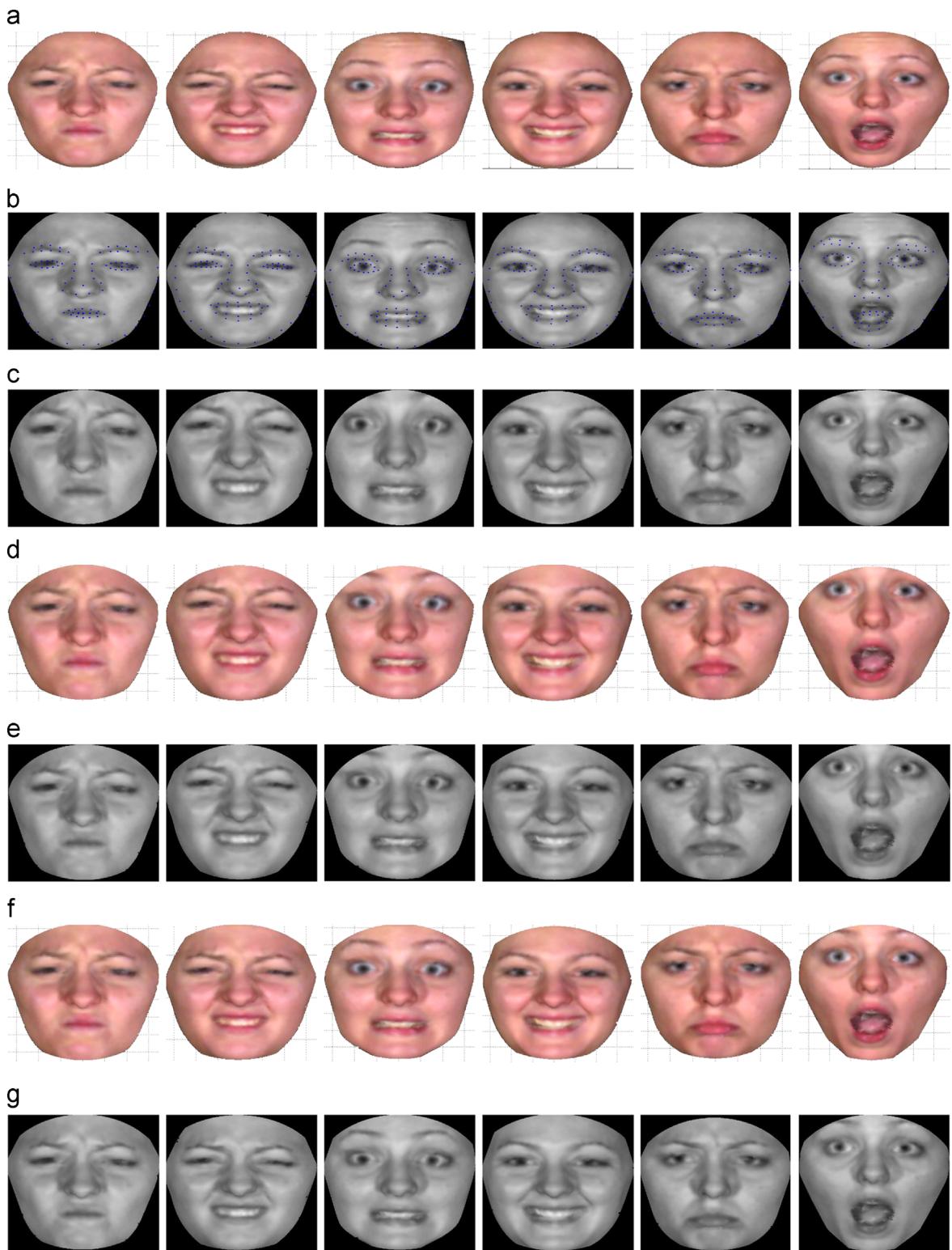


Fig. 8. Six universal expressions performed by the same subject in BU-3DFE database (a) The original models; (b) distribution of landmarks; (c) cropped faces referred to [43]; (d and e) cropped faces using the first type of sphere; (f and g) cropped faces using the second type of sphere.

3. The polytypic multi-block LBP (P-MLBP) to represent 3D facial expressions

As the 3D face models have been normalized in 2D-image-like structure, either the texture or the depth of the 3D models can be separately seen as 2D information, so we can take advantage of the wealth of 2D-image based features to represent the 3D models. For this work, we focus on the LBP (local binary patterns) descriptor which has aroused increasing interests both in computer vision and image processing [44]. LBP not only describes the relationship between adjacent pixels in image space, but also captures the responses of facial features in different scales and orientations. It has been successfully used in numerous applications. Therefore, this paper proposes a novel feature originated from the LBP descriptor to enhance its representation of 3D facial expressions. This brand new facial feature, which utilizes the feature-based irregular divisions and the fusion of depth value and texture information in 3D models, is named as polytypic multi-block LBP (P-MLBP) and it's detailed in the following sections.

3.1. The feature-based irregular divisions

The LBP descriptor computed over the whole face always neglects the local distortion caused by facial expressions. To get more details of facial expressions, the faces are divided into several blocks and then each of the blocks is encoded by LBP. However, the number of blocks is hard to determine as it should balance the face description against the memory-time consumption, so the best value of block size is obtained by practicing and it differs from one dataset to another, and thus we design different division schemes for comparison.

Generally, the face images are equally divided into blocks according to the scale of the image, which is named as regular divisions. According to some of them on face analysis using LBP descriptor [45–47], two regular divisions, 6×5 and 8×8 blocks (Fig. 5(b) and (c)), are employed in this paper. However, the regular divisions only consider the size of the face image, while neglecting the face structure. Usually, it may break the face feature distribution and thus results in inaccurate description of face images. This paper proposes the irregular divisions to keep the integrality of face local structure according to the feature points located on local organs (Fig. 4(b)). Referred to the regular division schemes, the 6×5 , 8×5 and 8×7 irregular division schemes are designed following the local organs and they are shown in Fig. 5(d), (e) and (f), respectively. Their details are described as follows (the numbers for the points refer to Fig. 4(b)).

The division lines in Fig. 5(d):

- (1) From top to down, the horizontal lines are determined by the 22nd, 25th, 65th, middle point between 69th and 79th, 85th point, respectively;
- (2) From left to right, the vertical lines are determined by the 20th, 68th, 71st, 35th point, respectively.

The division lines in Fig. 5 (e):

- (1) From top to down, the first horizontal line is 10 pixels away from the top margin of the image; the 2nd horizontal line is the middle between first line and 22nd point; the rest horizontal lines are determined by 25th, 65th, 69th, 79th, 85th point, respectively;
- (2) From left to right, the vertical lines are determined by 20th, 76th, 65th, 74th, 82nd, 35th point, respectively.

The division lines in Fig. 5 (f):

- (1) The design of horizontal lines is the same as that in Fig. 5 (e);
- (2) The design of vertical lines is similar to that in Fig. 5 (e), except the 2nd and 5th vertical lines. The 2nd vertical line is 10 pixels away from 76th point and the 5th vertical line is 10 pixels away from 82th point.

Three irregular divisions (Fig. 5(d)–(f)) are designed to divide the face images as well as to keep the integrity of the local facial structure. The design in Fig. 5(d) emphasizes to separate the nose region which is almost unaffected by the variation of facial expressions. The schemes in Fig. 5(e) and (f) divide the nose region following its shape. And they differ in whether the top half of the face is divided into small blocks. And the mouth region is expanded in Fig. 5(f) as the mouth is tremendously affected by facial expressions.

All these five division schemes shown in Fig. 5 can be adopted by both texture images (color information) and face models (depth values).

3.2. Polytypic Multi-block LBP (P-MLBP) for 3D face representation

Based on the division schemes proposed in the previous section, the face images are divided into multiple blocks. Considering the successful application of LBP representation for 2D facial expression recognition, we employ the LBP operator to calculate local detailed features for each divided block. Therefore, the Multi-block LBP is utilized to represent 3D facial expression for classification.

Fortunately, normalized 3D faces are already transformed into 2D-image-like structure, which makes it possible to apply LBP descriptor directly. Moreover, not only the depth values but also its corresponding texture images are feasible to extract LBP features.

Generally, the LBP descriptor is appropriate for the representation of 2D face images, as well as the texture images generated by this paper. The LBP operator labels the pixels of the texture image by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number. Given a pixel in the texture image, the LBP operator is defined as an ordered set of binary comparisons of pixel properties between the center pixel and its eight surrounding pixels. LBP code in decimal form is used to describe the property of the center pixel. The procedure of generating LBP code can be clearly seen in Fig. 6, so the property of the center

pixel is encoded in LBP as

$$LBP(p_c) = \sum_{i=0}^7 \text{sgn}(g(p_i) - g(p_c))2^i \quad (3)$$

$$\text{sgn}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (4)$$

where p_c is the center pixel, $g(p_i)$ is the gray value of the i th pixel in the 3×3 neighborhood, $g(p_c)$ is the gray value of the center pixel, $\text{sgn}(x)$ is the symbol function.

Likewise, the depth values of the 3D face models are also normalized in 2D-image-like structure, so they are supposed to be similar to the intensity values of the texture image. Therefore, the LBP feature of each vertex can be expressed as

$$LBP(v_c) = \sum_{i=0}^7 \text{sgn}(d(v_i) - d(v_c))2^i \quad (5)$$

where v_c is the center vertex, $d(v_i)$ is the depth value of the i th vertex in the 3×3 neighborhood, $d(v_c)$ is the depth value of the center vertex, $\text{sgn}(x)$ is still the symbol function defined in Eq. (4).

Therefore, LBP descriptors extracted from either the depth values or the texture images can be employed separately or jointly, to present 3D facial expressions. In our work, we employ the fusion of them for classification.

Accordingly, we simultaneously employ the irregular division schemes, and utilize the fusion of depth values and texture information to represent the LBP operator. These superimposed operations are intended to enhance the representation of LBP for 3D facial expression recognition. And this novel representation of LBP is named as polytypic Multi-block LBP (P-MLBP).

4. Experimental results and discussions

4.1. Experimental setup

Experiments in this paper are all carried out on the publicly available 3D facial expression database, BU-3DFE database [10], which contains 2500 3D face models from 100 subjects, 56 females and 44 males from different ethnic ancestries and ages. Each subject performed six universal expressions (shown in Fig. 8 (a)) namely angry, disgust, fear, happiness, sadness and surprise as defined by Ekman [48]. And each expression is of four different intensities which are shown in Fig. 7.

To date, most of the previous works only considered two highest levels of expression intensities, so we choose the same data too. Learned from their experimental setup, our experiments are constructed based on the following rules: 60 subjects are randomly chosen from each expression group without considering the gender; the selected subjects are randomly separated into two sets: one is the training set with 54 subjects, another is the testing set with the rest 6 subjects, which means any subject used for testing does not appear in the training set; each experiment is conducted for 100 times to obtain the average recognition results.

All the samples employed in the experiments are normalized following the steps introduced in Section 2. The results of one sample in BU-3DFE database are displayed in Fig. 8. Fig. 8(b) shows the distribution of detected landmarks. Fig. 8(c-g) are cropped results with different spheres introduced in Section 2. Their details can be found in Table 1.

4.2. Evaluation of the proposed strategy

In this paper, the nearest neighbor (NN), support vector machine (SVM) and the probabilistic neural network (PNN) are separately taken as the classifier to evaluate the performance of the proposed strategy. The facial features used for comparison are extracted from different facial data (including depth values, texture information and their fusion based on different divisions) based on different division schemes (including both regular divisions 6×5 , 8×8 and feature-based irregular divisions 6×5 , 8×5 and 8×7). Meanwhile, the global LBP descriptors extracted from the whole face without any divisions are also used as the baseline and the comparison can be found in Fig. 9. Moreover, performances of different face ROI with three kinds of classifiers are separately listed in Tables 2 and 3.

Facial expression recognition using LBP features extracted from (a) depth values; (b) texture images; (c) the fusion of both depth values and texture images; (d) average performances of (a)-(c), and 25 is chosen as the dimension of the principal components.

The NN classifier takes the Euclidean distance between different samples to measure their similarities. Moreover, PCA algorithm is additionally applied to the facial features to reduce the dimension of the similarity matrix. However, the dimension of principal components is difficult to decide, so the performance of features along with different dimensions is analyzed first. Take the performance of the global LBP features extracted from the first kind ROI as an example (Fig. 9). As the class of facial expressions is 6, the feature dimension begins from 5 (class-1) and increases with the uniform step (we set it as 5). Seen from Fig. 9, the performance of the global LBP features extracted from different facial data increases sharply at the beginning (lower than 25) and remains stable later (higher than 25). So we can get an inflection point and its x -coordinate is selected as the dimension of principal components to achieve the final recognition rates (The number, 25, is

Table 1
Details of Fig. 8(c-g).

| Fig. 9 | Data | Sphere | |
|--------|----------------|-------------|--------------------------------------|
| | | Center | Radius (distance between two points) |
| (c) | Texture images | Nose tip | Nose tip & eyebrow |
| (d) | 3D model | Nose tip | Nose tip & chin |
| (e) | Texture images | Nose tip | Nose tip & chin |
| (f) | 3D model | Nose bridge | Nose bridge & chin |
| (g) | Texture images | Nose bridge | Nose bridge & chin |

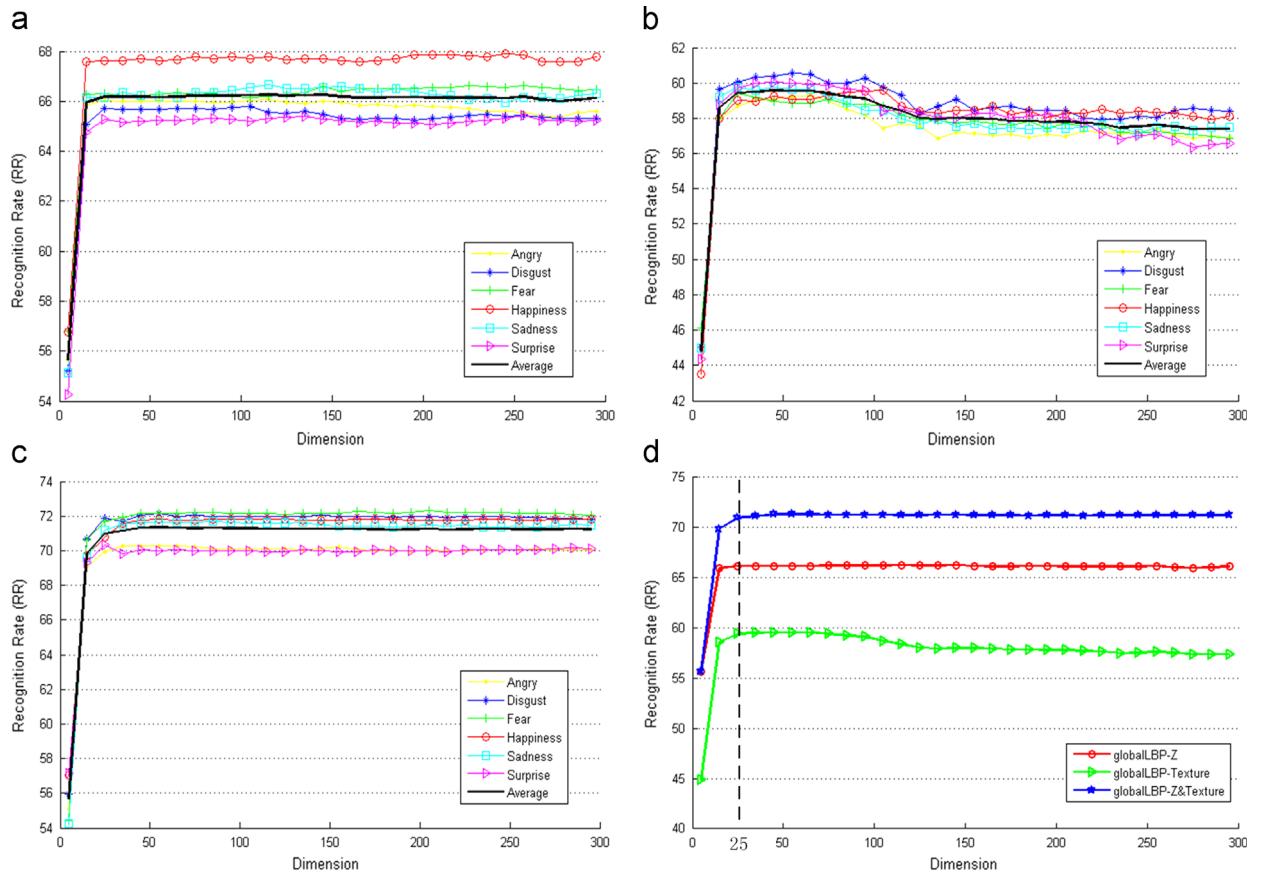


Fig. 9. Performance of the global LBP feature with different dimensions. Facial expression recognition using LBP features extracted from (a) depth values; (b) texture images; (c) the fusion of both depth values and texture images; (d) average performances of (a)–(c), and 25 is chosen as the dimension of the principal components.

selected as the dimension of the principal components in Fig. 9(d)). The dimension of principal components for other facial features is selected using the same criterion. Comparisons of their performances using NN classifier are listed in the 3rd column of Tables 2 and 3.

SVM is a successful supervised learning technique for classification and regression based on the structural risk minimization (SRM) [49]. We also take it as the classifier to evaluate the performance of P-MLBP for 3D facial expression recognition. Original SVM is suitable for binary classification, but it can be readily extended to a multi-class problem using one-against-one or one-against-all strategy [50]. As the one-against-one is time-consuming, we construct the multi-class SVM using the one-against-all strategy. The one-against-all SVM classifier schemes on a n -class classification; it considers n binary classifications, each of which labels one class as (+1) and all other $n - 1$ classes as (-1). In the experiments, the Gaussian function is taken as the kernel of the multi-class SVM. Recognition results using SVM is shown in the 4th column of Tables 2 and 3. To further verify the proposed strategy, they are also applied to the PNN architecture. Their detailed comparison results are shown in the 5th column of Tables 2 and 3.

Table 2
Recognition results using different classifiers based on the first kind ROI.

| Data origin | Division scheme | Recognition rate (%) | | |
|-------------|--------------------------|----------------------|--------------|------------------|
| | | NN classifier | Gaussian SVM | PNN architecture |
| Depth | Global | 65.7 | 85.7 | 85.0 |
| | 6 × 5 Regular division | 77.2 | 91.8 | 87.8 |
| | 8 × 8 Regular division | 79.3 | 93.0 | 88.5 |
| | 6 × 5 Irregular division | 78.9 | 92.0 | 88.6 |
| | 8 × 5 Irregular division | 80.0 | 92.6 | 88.8 |
| | 8 × 7 Irregular division | 81.5 | 92.6 | 89.2 |
| Texture | Global | 57.7 | 85.1 | 84.3 |
| | 6 × 5 Regular division | 77.7 | 91.4 | 88.0 |
| | 8 × 8 Regular division | 79.5 | 92.5 | 88.5 |
| | 6 × 5 Irregular division | 78.5 | 91.3 | 88.0 |
| | 8 × 5 Irregular division | 79.2 | 91.9 | 88.1 |
| | 8 × 7 Irregular division | 78.1 | 92.0 | 88.1 |
| Fusion | Global | 70.6 | 88.6 | 85.6 |
| | 6 × 5 Regular division | 78.7 | 93.2 | 89.7 |
| | 8 × 8 Regular division | 79.6 | 94.2 | 90.4 |
| | 6 × 5 Irregular division | 81.1 | 93.1 | 90.0 |
| | 8 × 5 Irregular division | 82.8 | 93.9 | 90.3 |
| | 8 × 7 Irregular division | 83.0 | 93.6 | 90.4 |

4.3. Results and discussion

Seen from the results in Tables 2 and 3, it can be concluded that:

1. Benefit from the 2D-image-like structure, it's possible to extract LBP features directly from 3D models. All the LBP features perform well, especially using SVM and PNN as the classifier. The 2D-image-like structure has built the bridge between 3D models and 2D methods, which makes it possible to take advantages of the wealth of 2D-image based features to represent the 3D models. Moreover, this structure makes it simple and automatic to implement 3D facial expression recognition, which opens a promising direction for automatic 3D facial expression recognition.
2. LBP features using local divisions including regular and irregular ones always perform better than the global one; the irregular divisions not only keep the integrity of local facial structure well but also preserve good performances or even better. Therefore, irregular divisions are effective to depict local facial features accurately.

Table 3

Recognition results using different classifiers based on the second kind ROI.

| Data origin | Division scheme | Recognition rate (%) | | |
|-------------|--------------------------|----------------------|--------------|------------------|
| | | NN classifier | Gaussian SVM | PNN architecture |
| Depth | Global | 64.1 | 86.1 | 84.7 |
| | 6 × 5 Regular division | 79.1 | 92.4 | 89 |
| | 8 × 8 Regular division | 82.6 | 93.1 | 89.1 |
| | 6 × 5 Irregular division | 78.6 | 91.8 | 88.5 |
| | 8 × 5 Irregular division | 81.1 | 92.7 | 89.1 |
| | 8 × 7 Irregular division | 81.4 | 92.8 | 89.5 |
| Texture | Global | 58.6 | 85.3 | 84.5 |
| | 6 × 5 Regular division | 77.0 | 92.2 | 87.5 |
| | 8 × 8 Regular division | 79.8 | 92.9 | 88.6 |
| | 6 × 5 Irregular division | 78.1 | 92.1 | 88.1 |
| | 8 × 5 Irregular division | 80.4 | 91.8 | 88.5 |
| | 8 × 7 Irregular division | 79.8 | 92.6 | 88.6 |
| Fusion | Global | 68.9 | 88.1 | 84.5 |
| | 6 × 5 Regular division | 81.0 | 93.6 | 90.2 |
| | 8 × 8 Regular division | 84.8 | 94.8 | 91.0 |
| | 6 × 5 Irregular division | 82.0 | 93.3 | 90.5 |
| | 8 × 5 Irregular division | 83.8 | 93.4 | 90.5 |
| | 8 × 7 Irregular division | 83.7 | 94.1 | 90.8 |

3. The features extracted from depth values always perform better than that extracted from texture images, and the performances of their fusion are the best. Therefore, the P-MLBP feature which involves the feature-based irregular division and the feature fusion strategy is effective to 3D facial expression recognition.
4. The results from both kinds of face ROI obtained using different classifiers are very close. Although their results are not completely equal to each other, the difference is very small. While the second kind ROI contains much more expression information and performs a bit better, the results of the first kind ROI are remarkable. Moreover, it is easier to detect the nose tip than the nose bridge. And it's tolerable to miss some information to reduce the influence of different facial outline, when a large expression is performed. So the first kind of ROI is competent to facial expression recognition.

Accordingly, based on the 2D-image-like structure, the automatic 3D facial expression recognition is easy-handled and P-MLBP features are verified to be remarkable for 3D facial expression recognition. Therefore, with the proposed strategy, the performance of 3D facial expression recognition system is enhanced, while being automatic and easy-handled.

4.4. Comparisons with state-of-the-art methods

In Table 4, we provide a comparison with several state-of-the-art algorithms. We can find that our methods achieve comparable results. Although the result of [25] is comparable, subdivision-mesh deformation used in [25] is partly based on predefined landmarks which destroy the automation of the recognition system. The proposed method in this paper which is mainly shown in Fig. 1 can be by passed altogether without any manual intervention. Therefore, it can be concluded that the proposed method outperforms previous work on 3D facial expression recognition, while being automatic and providing a possible way to address the problem of 3D facial expression recognition using 2D methods.

Table 4

Comparative results with previous reports on 3D facial expression recognition.

| Recent Work | Features + classifier | Automatic | Recognition rate (%) |
|--------------------|-------------------------------------|-----------|----------------------|
| Zeng [51] | CFI + MCI | Yes | 75.76 |
| Lemaire [52] | DMCM + SVM | Yes | 76.71 |
| Berretti [20] | SIFT Descriptors + SVM | No | 77.5 |
| Venkatesh [18] | Contour & Shape + Modified PCA | No | 81.7 |
| Wang [5] | Primitive Surface Label | No | 83.6 |
| Rabiu [12] | Distance & Slope Information | No | 87.1 |
| Mpiperis [25] | Bilinear model + Euclidean distance | No | 90.5 |
| Proposed framework | 6 × 5 P-MLBP + SVM | Yes | 93.1 |
| | 8 × 5 P-MLBP + SVM | Yes | 93.9 |
| | 8 × 7 P-MLBP + SVM | Yes | 93.6 |

5. Conclusion

In this paper, 3D face models are presented in 2D-image-like structure, which makes it possible to extract facial features using 2D methods. This operation utilizes the point-to-point mapping between the depth value and texture information in 3D models for facial data normalization. Then the polytypic multi-block local binary patterns (P-MLBP), which involves the feature-based irregular divisions and the fusion of facial data, is proposed to enhance the present of feature vectors. Finally, different classifiers are taken to conduct the 3D facial expression recognition for evaluation. The experimental results are remarkable, so they demonstrate the validity of the 3D models represented in 2D-image-like structure and the effectiveness of P-MLBP for 3D facial expression recognition. Therefore, the proposed method is effective to automatic 3D facial expression recognition and it builds a bridge between 3D face models and 2D analysis methods.

Acknowledgements

This work was supported partly by the National Natural Science Foundation of China (61172128), National Key Basic Research Program of China (2012CB316304), New Century Excellent Talents in University (NCET-12-0768), the fundamental research funds for the central universities (2013JBM020, 2013JBZ003), Program for Innovative Research Team in University of Ministry of Education of China (IRT201206), Beijing Higher Education Young Elite Teacher Project (YETP0544) and Research Fund for the Doctoral Program of Higher Education of China (20120009110008, 20120009120009).

References

- [1] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, *Image Vision Comput.* 27 (2009) 803–816.
- [2] T. Danisman, I.M. Bilasco, J. Martinet, C. Djeraba, Intelligent pixels of interest selection with application to facial expression recognition using multilayer perceptron, *Signal Process.* 93 (2013) 1547–1556.
- [3] Z. Zeng, Y. Fu, G.I. Roisman, Z. Wen, Y. Hu, Thomas S. Huang, Spontaneous emotional facial expression detection, *J. Multimedia* 10 (2006) 1–8.
- [4] S. Kumano, K. Otsuka, J. Yamato, E. Maeda, Y. Sato, Pose-invariant facial expression recognition using variable-intensity templates, in: 2007 Asian Conference on Computer Vision, 2007, 324–334.
- [5] J. Wang, L. Yin, X. Wei, Y. Sun, 3D facial expression recognition based on primitive surface feature distribution, in: 2006 IEEE Computer Society Conference on Computer Vision Pattern Recognition 2, 2006, 1399–1406.
- [6] H. Soyei, H. Demirel, 3D facial expression recognition with geometrically localized facial features, in: 23rd International Symposium on Computer and Information Science, 2008, 1–4.
- [7] H. Soyei, H. Demirel, Facial expression recognition using 3D facial feature distances, in: Fourth International Conference on Image Analysis and Recognition, 2007, 22–24.
- [8] H. Tang, T.S. Huang, 3D facial expression recognition based on automatically selected features, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, 1–8.
- [9] H. Tang, T.S. Huang, 3D facial expression recognition based on properties of line segments connecting facial feature points, in: IEEE International Conference on Automatic Face & Gesture Recognition, 2008, 1–6.
- [10] L. Yin, X. Wei, J. Wang, M.J. Rosato, A 3D facial expression database for facial behavior research, in: Seventh International Conference on Automatic Face and Gesture Recognition, 2006, 211–216.
- [11] X. Li, Q. Ruan, Y. Ming, 3D facial expression recognition based on basic geometric features, in: 2010 IEEE 10th International Conference on Signal Processing (ICSP), 2010, 1366–1369.
- [12] H. Rabiu, M.I. Saripan, S. Mashohor, M.H. Marhaban, 3D facial expression recognition using maximum relevance minimum redundancy geometrical features, *EURASIP J. Adv. Signal Process.* 213 (2012) (2012) 1–8.
- [13] U. Tekguc, H. Soyei, H. Demirel, Feature selection for person-independent 3D facial expression recognition using NSGA-II, in: 24th International Symposium on Computer and Information Sciences, 2009, 35–38.
- [14] T. Sha, M. Song, J. Bu, C. Chen, D. Tao, Feature level analysis for 3D facial expression recognition, *Neurocomputing* 74 (2011) 2135–2141.
- [15] H. Soyei, H. Demirel, Optimal feature selection for 3D facial expression recognition with geometrically localized facial features, in: Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, 2009, pp. 1–4.
- [16] H. Soyei, H. Demirel, Optimal feature selection for 3D facial expression recognition using coarse-to-fine classification, *Turk. J. Electr. Eng. Comput. Sci.* 18 (6) (2010) 1031–1040.
- [17] H. Soyei, U. Tekguc, H. Demirel, Application of NSGA-II to feature selection for facial expression recognition, *Comput. Electr. Eng.* 37 (2011) 1232–1240.
- [18] Y.V. Venkatesh, A.A. Kassim, O.V. Ramana Murthy, A novel approach to classification of facial expressions from 3D-mesh datasets using modified PCA, *Pattern Recogn. Lett.* 30 (2009) 1128–1137.
- [19] A. Maalej, B.B. Amor, M. Daoudi, A. Srivastava, S. Berretti, Shape analysis of local facial patches for 3D facial expression recognition, *Pattern Recogn.* 44 (2011) 1581–1589.
- [20] S. Berretti, A.D. Bimbo, P. Pala, A set of selected SIFT features for 3D facial expression recognition, in: 20th International Conference on Pattern Recognition, 2010, 4125–4128.
- [21] R. Srivastava, S. Roy, 3D facial expression recognition using residues, in: IEEE Tencon Conference for Region 10, 2009, 1–5.
- [22] S. Ramanathan, A. Kassim, Y.V. Venkatesh, W.S. Wah, Human facial expression recognition using a 3D morphable model, in: 2006 International Conference on Image Processing, 2006, 661–664.
- [23] X. Zhao, D. Huang, E. Dellandrea, L. Chen, Automatic 3D facial expression recognition based on a Bayesian Belief Net and a Statistical Facial Feature Model, in: 2010 International Conference on Pattern Recognition, 2010, 3724–3727.
- [24] B. Gong, Y. Wang, J. Liu, X. Tang, Automatic facial expression recognition on a single 3D face by exploring shape deformation, in: 17th ACM International Conference on Multimedia, 2009, 569–572.
- [25] I. Mpiperis, S. Malassiotis, M. Strintzis, Bilinear models for 3D face and facial expression recognition, *IEEE Trans. Inf. Forensics Secur.* 3 (3) (2008) 498–511.
- [26] T. Fang, X. Zhao, O. Ocegueda, S.K. Shah, 3D facial expression recognition: a perspective on promises and challenges, in: 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011, 603–610.
- [27] M. Song, D. Tao, X. Huang, C. Chen, J. Bu, Three-dimensional face reconstruction from a single image by a coupled RBF networks, *IEEE Trans. Image Process.* 21 (5) (2012) 2887–2897.
- [28] Dirk Smeets, Peter Claes, Jeroen Hermans, Dirk Vandermeulen, Paul Suetens, A comparative study of 3-D face recognition under expression variations, *IEEE Trans. Syst. Man Cybernetics, Part C: Appl. Rev.* 42 (5) (2012) 710–727.
- [29] Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, Lijun Yin, Static and dynamic 3D facial expression recognition: a comprehensive survey, *Image Vision Comput.* 30 (2012) 683–697.
- [30] S. Hussain, B. Triggs, Visual recognition using local quantized patterns, *Proceedings of the 10th European Conference on Computer Vision* (2012) 716–729.
- [31] S. Hussain, T. Napoleon, F. Jurie, Face recognition using Local Quantized Patterns, in: British Machine Vision Conference, 2012, 1–11.
- [32] T. Ahonen, E. Rahtu, V. Ojansivu, J. Heikkila, Recognition of blurred faces using Local Phase Quantization, in: 19th International Conference on Pattern Recognition, 2008, 1–4.
- [33] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, Z. Zhu, Robust face recognition via occlusion dictionary learning, *Pattern Recogn.* 47 (2014) 1559–1572.
- [34] Z. Li, D. Gong, Y. Qiao, D. Tao, Common feature discriminant analysis for matching infrared face images to optical face images, *IEEE Trans. Image Process.* 23 (6) (2014) 2436–2445.

- [35] T. Faltemier, K.W. Bowyer, P. Flynn, A region ensemble for 3D face recognition, *IEEE Trans. Inform. Forensics Secur.* 3 (1) (2008) 62–73.
- [36] G. Passalis, P. Perakis, T. Theoharis, I.A. Kakadiaris, Using facial symmetry to handle pose variations in real-world 3D face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (2011) 1938–1951.
- [37] Y.V. Venkatesh, A.K. Kassim, O.V.R. Murthy, Resampling approach to facial expression recognition using 3D meshes, in: 20th International Conference on Pattern Recognition (2010) 3772–3775.
- [38] Tianhong Fang, Xi Zhao, Omar Ocegued, Shishir K. Shah, Ioannis A. Kakadiaris, 3D/4D facial expression analysis: an advanced annotated face model approach, *Image Vision Comput.* 30 (2012) 738–749.
- [39] Stefano Berretti, Boulbaba Ben Amor, Mohamed Daoudi, Alberto del Bimbo, 3D facial expression recognition using SIFT descriptors of automatically detected keypoints, *Visual Comput.* 27 (11) (2011) 1021–1036.
- [40] Pierre Lemaire, Boulbaba Ben Amor, Mohsen Ardabilian, Liming Chen, Mohamed Daoudi, Fully automatic 3D facial expression recognition using a region-based approach, in: Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding, 2011, pp. 53–58.
- [41] Prathap Nair, Andrea Cavallaro, 3-D face detection, landmark localization, and registration using a point distribution model, *IEEE Trans. Multimedia* 11 (4) (2009) 611–623.
- [42] X. Cao, Y. Wei, F. Wen, J. Sun, Face alignment by explicit shape regression, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, 2887–2894.
- [43] H. Tang, B. Yin, Y. Sun, Y. Hu, 3D face recognition using local binary patterns, *Signal Process.* 93 (2013) 2190–2198.
- [44] D. Huang, C. Shan, M. Ardabilian, YPlease check the author names.. Wang, Member, Liming Chen, Local binary patterns and its application to facial image analysis: a survey, *IEEE Trans. Syst. Man Cybernetics-Part C: Appl. Rev.* 41 (2011) 765–781.
- [45] Timo Ahonen, Abdenour Hadid, Matti Pietikainen, Face description with local binary patterns: application to face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (12) (2006) 2037–2041.
- [46] Guoying Zhao, Matti Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (6) (2007) 915–928.
- [47] Baochang Zhang, Yongsheng Gao, Sanqiang Zhao, Jianzhuang Liu, Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor, *IEEE Trans. Image Process.* 19 (2) (2010) 533–544.
- [48] P. Ekman, W.V. Friesen, Constants across cultures in the face and emotion, *J. Personality Social Psychol.* 17 (2) (1971) 124–129.
- [49] A.S. Mian, M. Bennamoun, R. Owens, An efficient multi-modal 2D–3D hybrid approach to automatic face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (11) (2007) 1927–1943.
- [50] M.P. Segundo, C. Queirolo, O.R.P. Bellon, Automatic 3D facial segmentation and landmark detection, in: 14th International Conference on Image Analysis and Processing, 2007, 431–436.
- [51] Wei Zeng, Huibin Li, Liming Chen, J.-M. Morvan, X.D. Gu, An automatic 3D expression recognition framework based on sparse representation of conformal images, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, 1–8.
- [52] P. Lemaire, M. Ardabilian, Liming Chen, M. Daoudi, Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, 1–7.