The 3D Face Recognition Algorithm Fusing Multi-geometry Features

SUN Yan-Feng[1] TANG Heng-Liang[1] YIN Bao-Cai[1]

Abstract The 3D face recognition attracts more and more attention because of its insensitivity to the variance of illumination and pose. There are many crucial problems to be solved in this topic, such as 3D face representation and effective multi-feature fusion. In this paper, a novel 3D face recognition algorithm is proposed and its performance is demonstrated on BJUT-3D face database. This algorithm chooses face surface property and the principle component of relative relation matrix as the face representation features. The similarity metric measure for each feature is defined. A feature fusion strategy is proposed. It is a linear weighted strategy based on Fisher linear discriminant analysis. Finally, the presented algorithm is tested on the BJUT-3D face database. It is concluded that the performance of the algorithm and fusion strategy is satisfying.

Key words 3D face recognition, feature representation, feature fusion

Automatic face recognition (AFR) has been studied for over 40 years. It has become one of the most active research areas in pattern recognition and computer vision. Much progress has been made on its theories and algorithms in the past years. But most of them are based on two-dimensional face images. It has been demonstrated that the pose and illumination variances of face images influence seriously the recognition performance, and the images of the same individual with different poses or illuminations have less similarity than the images from different individuals with the same pose or illumination. Therefore, one of the greatest remaining research challenges in face recognition is to recognize faces across different poses and illuminations, and it has become a bottleneck for the progress of face recognition research.

With the development of 3D information acquisition technology, it is likely and convenient to acquire the 3D data of an object. Compared with the 2D face image, the 3D face data contains more spatial information, which is inherent property of the face and is robust to the uncontrollable environment where 2D appearance can be affected. In recent years, the 3D face recognition has attracted more and more attention, and has become one of the valuable research topics. Meanwhile, many 3D face recognition algorithms have been presented[1].

1 Related works

Like the 2D face recognition, the representation of the face features is obviously one of the crucial problems in 3D face recognition. The common methods of feature representation can be categorized as follows. The first category of methods are based on face surface analysis. Face is a complex surface. All face surfaces have nearly same structures but different details. The utilization of face surface curvature is a common method in 3D face recognition. Hallinan has recognized face by detecting face features and their relative locations[2]. The face surface curvatures were extracted and used for face recognition by Gordon[3]. But this method is sensitive to noises because of the relativity between the curvature and the second-order derivative. Samir has selected face surface contours as face feature and defined the contours distance as metric measure for face recognition[4]. Moreno has investigated the representation ability of face feature by experiments and concluded that face surface curvature can perform better than the surface area[5]. The second category of methods are based on the range image. The range image is usually formed from 3D coordinate data, and then, the subspace method is performed on range images. Hesher has explored principal component analysis (PCA) and independent component analysis (ICA) approaches with face range images[6]. The recognition results indicated that the performance from ICA is superior to that from PCA, but the method based on ICA needs much more training samples in gallery. Achermann has compared PCA and the hidden Markov model (HMM) approaches for 2D face recognition with range images, and demonstrated the PCA approach had better recognition performance than HMM[7]. The third category of methods are based on face profiles. Its key step is the extraction of face profile. Then, some reference points are defined to represent the face and to match with the gallery samples for face recognition[8−9]. The fourth category fuses multiple features for face representation. Every face representation feature has its limitation and advantage, so fusing multi-features or multi-approaches for face recognition is more reasonable and feasible. This representation can combine different statistic features, exert the advantages of every feature, and approach and result in more exact and robust recognition performance[10−12]. However, the fusion strategy is a crucial problem to be solved firstly.

These methods above can get good recognition results on a small scale face database. But they are restricted by the different face data formats, and can not be popularized easily. This means that these algorithms depend on the input 3D face data, and different features can be extracted from various input data. In addition, a face surface is not rigid and face recognition performance also depends largely on face expression and age variation. These factors must be considered when face feature is defined. Meanwhile, a 3D face image contains large amounts of data, and only the face features that lose little discriminating information but contain proper amounts of data can be used for face representation.

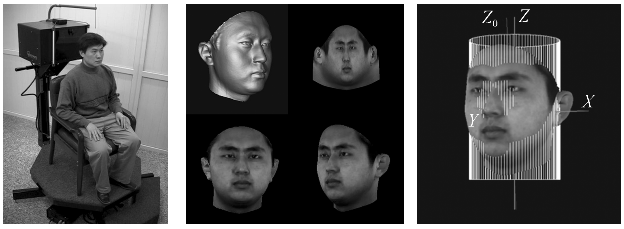
According to the idea above, this paper proposes a 3D face recognition algorithm whose 3D face data include vertices coordinates and their connected relation. This algorithm defines some face representation features and their similarity metric measures, and then, a novel linear weighted strategy based on Fisher linear discriminant analysis (FLDA) is proposed to fuse the different face features. Finally, the proposed algorithm is tested on the BJUT-3D face database[13]. The experimental results indicate that

Received November 12, 2007; in revised form May 6, 2008
Supported by National Natural Science Foundation of China (60533030) and Beijing Natural Science Foundation (4061001)
1. Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, P. R. China
DOI: 10.3724/SP.J.1004.2008.01483

the recognition performance of this method is satisfying, and the feature fusion strategy is effective.

2 3D face normalization

In this paper, the 3D face data are acquired by CyberWare 3030RGB/PS laser scanner (see Fig. 1 (a)). The data are so large that data of every 3D face are composed of more than 200 000 vertices and 400 000 triangular patches with a collection of connected edges and vertices.



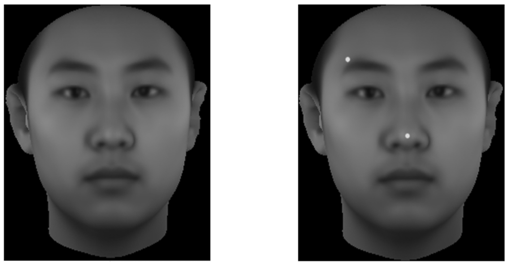
(a) 3D scanner (b) The cut 3D face (c) Face correction

Fig. 1 3D scanner and original face data process

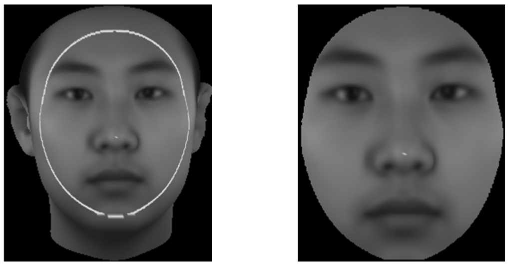
Because the acquired 3D face samples are noisy and spinose, a preprocessor is needed to remove these spike, to fill the holes that result from removing spikes by interpolation and to smooth the whole face surface. Then, the face is trimmed from the whole scanned data by cutting boundary and removing the 3D data lying on hair and shoulder (see Fig. 1 (b)). Lastly, all trimmed 3D face data are corrected to a uniform coordinate system. Here, the discrete 3D face vertices fit to a cylinder. The center axis of the cylinder is defined as Z-axis with the positive direction upward, and the direction through the nose tip forward and perpendicular to the Z-axis is the Y-axis, and the X-axis is obtained by cross product of the Z-axis and the Y-axis (see Fig. 1 (c)). As every 3D face sample does not have the same amounts of data (numbers of vertices and triangular patches), it is needed to align the 3D faces with pixel-wise correspondence with the human face properties by the mesh resampling method[14]. Moreover, the resulting 3D face samples keep alignment based on the face features, contain the same numbers of vertices and triangular patches. This means that the vertices and triangular patches of every sample can be recorded by the same rule, and the vertices with the same sequential number but of different samples are in the same location of face.

In face recognition, we pay attention to the face feature areas such as eye, brow, nose, mouth, and so on. In order to reduce unnecessary information, the aligned face samples are trimmed as follows. First, two key feature points (the nose tip and another outboard vertex on the left brow) are located (see Fig. 2 (b)). Because the Y-axis represents the face depth information, the nose tip vertex can be searched easily as the vertex with the maximum y-value from 3D face data. The outboard texture vertex on the left brow can be determined by face detection algorithm. Corresponding coordinate can be acquired by the mapping relationship between the texture and the shape information. It must be mentioned that all the 3D face samples used in this paper keep the alignment based on the face features. The process of searching the two key feature points need to be carried out only once, and the sequence numbers of the two key points on the face surface can be recorded. Given a new 3D face sample, the two key points can be located by their sequence numbers directly.

After locating the two key feature points, a sphere with radius r and center at the nose tip is then used to trim the 3D face geometry data and the corresponding registered 2D face texture data, where r is the Euclidean distance between the two key points. Then, the trimmed 3D face data are recorded and will be used in the experiment. All the 3D face samples are processed by this method, and the normalized 3D face can be obtained. Fig. 2 shows the aligned 3D, the searching result of the two key points, the trimmed boundary, and the final trimmed 3D face.



(a) Aligned 3D face (b) Key feature points



(c) The trimmed boundary (d) The final trimmed face

Fig. 2 The 3D face retrimming and final result

3 3D face feature extraction and fusion

The rimmed 3D face is a 3D surface. All face surfaces are different from each other. If a face is represented by its vertice coordinates, the dimension is so high that the computation and analysis of face recognition will be time-consuming. In this paper, we analyze the character of face surface, define some 3D face features, and extract these features for recognition.

3.1 3D face feature representation and extraction

In order to reflect the essential face attributes, we consider two kinds of face features for face representation. One kind is the face surface attribute features (e.g. face surface area, face surface volume, the normal vectors of face feature points, and so on). Another is the correlation of face key feature points. Because of the slight variations in face expression and corrected head pose, the Euclidean distances between the key feature points will be different for a same individual, but their relative location is invariable. So if we can construct relative invariable face features, then we can analyze their properties and fuse these features. It will be possible that the recognition performance might be improved while these invariable features are used for face representation.

The features describing face surface properties include the face surface area S, face surface volume V, normal vector N of key feature points, and relative distance matrix T between the key points on face surface. 59 key feature points distributing on brows (25), eyes (14), nose (8), and mouth (12) areas are marked. For more details of these feature points please refer to Fig. 3. Next, the surface feature extraction methods are introduced and their similarity measures are defined.

Fig. 3 Marked key feature points on face surface

All 3D face samples are arranged by discrete vertices and triangular patches, which provide us the coordinates, texture, and adjoining vertices information of every vertex. So the face surface area S can be calculated by the sum of all the triangular patch areas. It can be calculated with Helen formula as

$$S = \sum_{i=1}^{M} S_i \tag{1}$$

where

$$S_i = \sqrt{s\,(s-a)\,(s-b)\,(s-c)},\ s = \frac{a+b+c}{2}$$

$$a = \sqrt{\left(x_1^i - x_2^i\right)^2 + \left(y_1^i - y_2^i\right)^2 + \left(z_1^i - z_2^i\right)^2}$$

$$b = \sqrt{\left(x_1^i - x_3^i\right)^2 + \left(y_1^i - y_3^i\right)^2 + \left(z_1^i - z_3^i\right)^2}$$

$$c = \sqrt{\left(x_2^i - x_3^i\right)^2 + \left(y_2^i - y_3^i\right)^2 + \left(z_2^i - z_3^i\right)^2}$$

where M denotes the number of triangular patches for a given face sample, and $S_i (i = 1, \cdots, M)$ is the i-th triangular patch area. The spatial nodes (x_1^i, y_1^i, z_1^i), (x_2^i, y_2^i, z_2^i), and (x_3^i, y_3^i, z_3^i) are the three vertices of the i-th triangular patch. The area similarity measure D_S^{ij} of the two face samples i and j can be represented by the absolute value of difference of their surface areas.

Every triangular patch can be projected vertically to the reference plane α, which is spanned by X-axis and Z-axis. And the projection of every triangular patch can form a triangular column. Then, the volume feature V of face surface can easily be estimated as the sum of all the triangular columns volume approximately by

$$V = \sum_{i=1}^{M} V_i \tag{2}$$

where $V_i = S_i' \times H_i$, $S_i' = S_i \cos\theta_i$, $\cos\theta_i = \cos < \boldsymbol{n}_i, \boldsymbol{y} >= \dfrac{(\boldsymbol{n}_i \cdot \boldsymbol{y})}{\|\boldsymbol{n}_i\| \times \|\boldsymbol{y}\|}$, and $H_i = \dfrac{y_1^i + y_2^i + y_3^i}{3}$. V_i is the triangular column volume formed by the i-th triangular patch, S_i' denotes the projection areas of the i-th triangular patch on the plane α, H_i denotes the distance from the barycenter of the i-th triangular patch to the plane α, and θ_i stands for the angle between the i-th triangular patch plane and the reference plane α. The $\boldsymbol{n}_i$ denotes the normal vector of the i-th triangular patch, and $\boldsymbol{y}$ denotes the unit direction $(0, 1, 0)$. The volume similarity measure D_V^{ij} of the two face samples i and j can be represented by the absolute value of difference of their surface volumes.

The normal vector N at a feature point can be calculated by averaging the normal vectors of all the triangular patches adjoined with this feature point. For a given face sample, we mark the m feature points first. Then, the normal vectors of all the feature points forms a $3 \times m$ normal matrix, that is,

$$N^i = \left[n_1^i, n_2^i, \cdots, n_m^i\right] = \begin{pmatrix} n_{1x}^i & n_{2x}^i & \cdots & n_{mx}^i \\ n_{1y}^i & n_{2y}^i & \cdots & n_{my}^i \\ n_{1z}^i & n_{2z}^i & \cdots & n_{mz}^i \end{pmatrix} \tag{3}$$

where N^i denotes the normal matrix of feature points for the i-th sample, $\boldsymbol{n}_1^i$ denotes the normal vector of the first feature point, and n_{1x}^i, n_{1y}^i and n_{1z}^i are three components of vector $\boldsymbol{n}_1^i$. Then, every face sample can be represented directly as the matrix form above. The similarity measure of normal vector matrix of two face samples $D_{angle}(N^i, N^j)$ can be defined as the angle distance between normal vectors of all corresponding feature points, that is,

$$D_{angle}(N^i, N^j) = \sum_{k=1}^{m} \arccos \frac{(n_k^i \cdot n_k^j)}{\|n_k^i\| \times \|n_k^j\|} \tag{4}$$

In order to describe the correlation of all feature points on face surface, this paper uses the relative distance matrix (RDM) based on the face key feature points as one of the face representation features. In general, the Euclidean distance metric is used for the similarity measure of RDM. This measure will get just a scalar quantity metric and is impossible to take much more details into consideration. Much of the spatial information such as the relative location of the feature points will be neglected. Furthermore, all face samples have been aligned and corrected, but the Euclidean distance between key feature points can be affected frequently by the head pose and expression. This will result in much more errors to affect face recognition performance. In this paper, a novel similarity measure based on relative Euclidean distance matrix of the feature points is proposed. This measure takes the principal components of the relative Euclidean distance matrix as the eigenvector to represent the face sample. It not only takes the relative location of all the feature points into consideration, but also reduces the effect of head pose resulting from the head movement. The relative distance between two key feature points is defined as

$$d_{ij}^k = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{5}$$

$$(i, j = 1, \cdots, m)$$

(x_i, y_i, z_i) and (x_j, y_j, z_j) stand for the coordinate parameters of the i-th and the j-th feature points. The d_{ij}^k denotes the Euclidean distance between the i-th and the j-th feature points for the k-th face sample. Thus, the RDM T^k is defined as

$$T^k = \begin{bmatrix} d_{1\,1}^k & d_{1\,2}^k & \cdots & d_{1\,m}^k \\ d_{2\,1}^k & d_{2\,2}^k & \cdots & d_{2\,m}^k \\ \cdots & \cdots & \ddots & \cdots \\ d_{m\,1}^k & d_{m\,2}^k & \cdots & d_{m\,m}^k \end{bmatrix} \tag{6}$$

With the former definition, the k-th face sample can be represented by the matrix T^k with $m \times m$ dimensions. So the similarity measure of the RDM of two samples can be defined as

$$D^{ij} = \sum_{k=1}^{m}\sum_{l=1}^{m} |d_{kl}^{i} - d_{kl}^{j}| \tag{7}$$

If many key feature points are marked on a face surface, the RDM will be a large scale complex matrix. In this situation, the PCA will be used to reduce dimension of the RDM for further analysis. The dimension reduction not only reduces computation consumption but also denotes which vector (or key point) in T^k is of more classification information. By the PCA process, a series of optimal projective vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_p$ are obtained and the eigenvectors with the top d eigenvalues are denoted by $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\cdots$, $\boldsymbol{X}_d$. Thus, the RDM of a face sample can be projected in every direction X_k as

$$Y_k = TX_k \quad k = 1, 2, \cdots, d \tag{8}$$

Hence, the relative distance matrix T can be projected in the optimal directions $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_d$, that is,

$$Y = [Y_1 \quad Y_2 \quad \cdots \quad Y_d] = T\,[\boldsymbol{X}_1 \quad \boldsymbol{X}_2 \quad \cdots \quad \boldsymbol{X}_d] \tag{9}$$

It means that the RDM of every face sample can be represented by the matrix Y.

This 3D face representation model can be calculated easily and directly, and does not need the training set for learning. Compared with statistic learning methods, the recognition performance of this algorithm can not be decreased because of the data distribution difference between the training data set and the test data set. In addition, the feature of RDM makes the face recognition performance more robust to the facial tiny expression, time lapse, and pose correction errors.

3.2 The feature fusion strategy

A different face feature provides different discriminative information and also has different discriminative capability. It also plays a different role in face recognition. Single features can provide the limited discriminative information and local applicability. Face recognition based on a single face feature can not achieve satisfactory performance. So it is a reasonable approach to fuse more kinds of face features. The multi-feature fusion strategy can synthesize all the features information, reserve more original face data, and get satisfactory results. The popular fusion strategies are the feature series joint and the weighted joint and so on[15]. The former fusing strategy is doomed to deal with high dimensional feature vectors. If the dimension reduction is implemented, the real meanings of the new vector are ambiguous. The latter fusing strategy is the linear weighted strategy based on the recognition results of various face features. In this paper, the linear weighted fusion strategy is used. For the face recognition result $E(u_i)$ based on the i-th face feature, define

$$E(u) = \sum_{i=1}^{N} \omega_i E(u_i) \tag{10}$$

where the weighted coefficient ω_i denotes the contributive efficiency of the i-th face feature for the recognition. The final recognition results depend on the value $E(u)$.

Determining the weighted coefficient is the key step of this fusion strategy. Two methods are proposed in this section. One defines the contributive efficiency of every face feature as the weighted coefficient, such as the face recognition rate. Another is based on the FLDA. Its main idea is based on the following theory. The similarity between different samples from the same individual (or the same class samples) is superior to the similarity between different samples from different individuals (or different class samples). With the FLDA, the better discriminative feature should have better discriminative capability, and so, should have a large between-class similarity mean value, a little within-class similarity mean value, and a little within-class and between-class covariance mean values. So, we measure the discriminative capability of face features by the ratio of difference of within-class and between-class similarity mean values and the sum of within-class and between-class similarity covariances.

For a face feature, the similarity measure Ω_{ij} of two samples i and j is defined as

$$\Omega_{ij} = \bigcap(i, j) \quad (i \neq j) \tag{11}$$

where $\bigcap(i, j)$ denotes the difference between the two samples.

The face recognition is a multi-class classification problem. According to the FLDA, the within-class similarity measure is calculated on the different samples from the same individual and the between-class similarity measure is calculated on the different samples from different individuals for each feature defined in Section 3.1. The mean value and covariance of within-class and between-class similarity measures are formalized as

$$m_w^{\coprod} = \frac{1}{\mathrm{N}_w^{\coprod}} \sum \Omega_{ij} \quad \left(i, j \in \coprod\right) \tag{12}$$

and

$$S_w^{\coprod} = \frac{1}{\mathrm{N}_w^{\coprod}} \sum (\Omega_{ij} - m_w^{\coprod})^2 \quad \left(i, j \in \coprod\right) \tag{13}$$

where $m_w^{\coprod}$ and $S_w^{\coprod}$ are the within-class similarity measure mean value and covariance of the class $\coprod$, respectively. Samples i and j are all from class $\coprod$, and $\mathrm{N}_w^{\coprod}$ is the number of different samples from class $\coprod$.

The mean value and covariance of between-class similarity measure are defined, respectively, as

$$m_b^{\coprod} = \frac{1}{\mathrm{N}_b^{\coprod}} \sum \Omega_{ij} \quad \left(i \in \coprod, \quad j \notin \coprod\right) \tag{14}$$

and

$$S_b^{\coprod} = \frac{1}{\mathrm{N}_b^{\coprod}} \sum (\Omega_{ij} - m_b^{\coprod})^2 \quad \left(i \in \coprod, \quad j \notin \coprod\right) \tag{15}$$

where $m_b^{\coprod}$ and $S_b^{\coprod}$ are between-class similarity measure mean value and covariance between class $\coprod$ and any other class. Samples i and j are from different classes, but i is from class $\coprod$. $\mathrm{N}_b^{\coprod}$ is the number of different sample pairs from different classes, but one is from class $\coprod$.

Thus, the weighted coefficient ω can be defined as

$$\omega = \frac{(m_w - m_b)^2}{S_w + S_b} \tag{16}$$

This definition means that the larger is the difference of within-class and between-class similarity measure mean values, the larger is the ω and the better is the discriminative capability, but the smaller is the sum of within-class and between-class similarity measure covariances, the larger is the ω. So ω value reflects the discriminative capability of the face feature. Thus, the face feature with larger ω value can play a greater role in face recognition.

4 Experimental results and analysis

In order to test the performance of the feature representation and the fusing strategy developed in this paper, we have carried out the experiment on BJUT-3D face database. This 3D face database was built by our research group and released to the public during the International Conference on Computer Vision (ICCV) in 2005 with lots of efforts for many years. So far, this database has been used by more than 70 academic institutions from 14 countries. At the beginning, there were 500 face subjects with neutral expression and every subject had only one sample in the database. In order to be helpful for the evaluations of face recognition algorithms, we added another 50 subjects with neutral expression, and every subject had three samples (two for training and one for test). All of these 150 samples were normalized by the method mentioned in Section 2. Thus, they keep the alignment based on the face features and with the same number of vertices and triangle patches. All vertices of every sample were recorded in 3D coordinates, and structured by the triangle mesh for further use.

As the real 3D face samples are difficult to acquire, some virtual 3D faces based on the method in [16] were reconstructed to enlarge the training set and the testing set. We used some 2D face images from 2D face database CAS-PEAL released by the Institute of Computing Technology, Chinese Academy of Sciences to reconstruct the virtual 3D faces samples. These virtual samples were also recorded in 3D coordinates, and structured by the triangle mesh. They all are aligned based on face features. The following experiments were based on the real 3D face samples from the BJUT-3D face database and the virtual 3D face samples reconstructed from 2D face image from CAS-PEAL database.

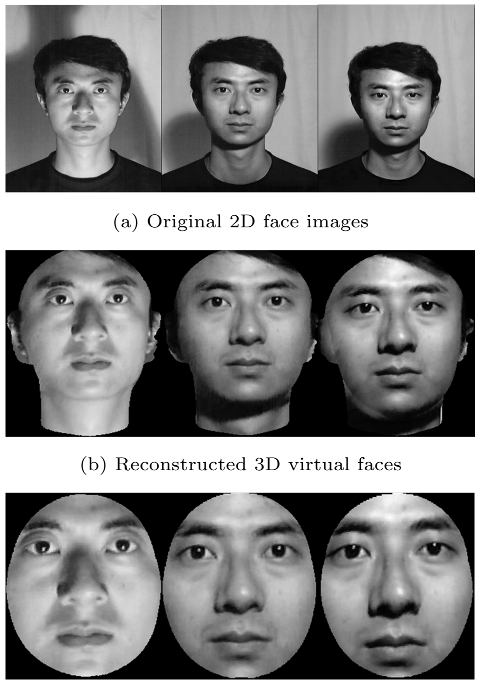
4.1 The recognition performance evaluation of 3D face representation feature

In order to test the representation efficiency of every face feature, the experiments on two data sets were carried out, respectively.

Set 1. All the testing samples are real 3D face data from BJUT-3D face database. There are 150 samples from 50 persons (37 males and 13 females), and each individual has 3 samples with neutral expression (two for training and one for testing).

Set 2. This set consists of two parts. One includes all the 150 real samples from Set 1, and the other is the virtual sample reconstructed from CAS-PEAL. There are 90 virtual samples of 30 persons and each individual has 3 virtual samples reconstructed from 3 different face images (two of three virtual samples for training and one for testing, see Fig. 4).

Four face geometry features are chosen for test. They are the face surface area S, the face surface volume V, the relative distance matrix of key feature point coordinates T, and the angle distance between the normal vectors of two key points NA. The selected key feature points are shown in Fig. 3. The recognition performances of different features tested on the two data sets are shown in Table 1.



(a) Original 2D face images

(b) Reconstructed 3D virtual faces

(c) Trimmed 3D virtual faces

Fig. 4 The virtual 3D face samples

Table 1 The recognition results of different features

Data set	S	V	T	NA
1	16.7 %	14.3 %	73.8 %	73.8 %
2	15.0 %	13.7 %	71.3 %	72.5 %

From the results shown in Table 1, we can conclude that recognition efficiencies of surface area S and volume V features are low, only about 15 %. This is because the surface area and volume are scalar, and they represent the holistic face features. Furthermore, the main face areas from different individuals are usually similar, so are the values of face surface area and volume. Hence, the recognition efficiencies of these features are very low. But the two features can distinguish samples with large difference in face shape, and they also have a certain discriminative capability. The face representation methods based on angle distance of the normal vector and the relative distance matrix show good recognition performance (about 73 %) and they are also positive face recognition features.

From these experimental results, we can see that the recognition performance of Set 1 is better than that of Set 2. Because some samples from Set 2 are virtual 3D data reconstructed from 2D images, the reconstructed error between the virtual sample and real sample is ineluctable. This will affect directly the veracity of virtual data and furthermore affect the face recognition performance. However, all samples from Set 1 are real 3D face data acquired by 3D scanner, and their precision and veracity are better. Therefore, the recognition performance tested on Set 1 is better than those on Set 2.

4.2 The performance evaluation of feature fusion strategy

In order to test the performance of the feature fusion strategy developed in this paper, we performed the feature fusion experiment on the two test sets mentioned in Section

Table 2 The recognition results fusing multi-geometry features

Data set	Weighted method	S	V	T	NA	Fusing results
1	RWM	0.090	0.080	0.415	0.415	84.4 %
1	TWM	0.109	0.179	0.382	0.330	86.7 %
2	RWM	0.067	0.058	0.442	0.433	76.3 %
2	TWM	0.161	0.221	0.430	0.188	81.3 %

4.1. The result weighted method (RWM) defines the weighted coefficient as the recognition efficiency of every single face feature. The training weighted method (TWM) defines the weighted coefficient as (16). The experiment was taken based on two kinds of the weighted methods. The four face geometry features (including surface area S, volume V, the relative distance matrix of key feature point coordinates T, and the angle distance between the normal vectors of two key points NA) were fused. The final recognition results are listed in Table 2.

Comparing the recognition results shown in Tables 1 and 2, we can conclude that the recognition results of fusing multiple face features are superior to that using a single face feature obviously. And for the two weighted coefficient methods, the fusing results of TWM proposed in this paper are a little better than those of RWM. This is because the RWM just fuses different features with the recognition results based on single feature, but neglects the relative relationship among samples. The TWM is a rational defining weighted coefficient method by analyzing the face data, especially the within-class and between-class relative relationship. The fusing experiment shows that the TWM fusing strategy is more scientific, and results in the better recognition performance.

Limited by the scale of 3D face database, the experiment was on a small scale data set, which is the reason why the advantage of TWM is not more obvious than the RWM. With the enlargement of 3D face database scale, the advantage of TWN will be more remarkable and the recognition performance will also be much better than the results based on RWM.

Xu has developed a method for 3D face recognition[17]. He has represented a face feature by the eigenvectors of face local regions and has used subspace method to analyze the face data. The reported experimental results were for the full 120 probed faces with the recognition rate 72 %. Godil has probed the face recognition method by fusing the 2D and 3D face recognition results[18]. He has tested for 200 face samples CAESAR database. The recognition rate is up to 82 %. Reference [19] has probed 3D face recognition performance using PCA feature on the FRGC2.0 data set. The verification rate at 0.001 FAR is 82 %. The recognition rate of the present method is approximately 85 %. This exciting performance is attributed to not only the effective description and representation of face features but also feature fusing strategy. As released, 3D face databases are few, the scale is small, and the most of them can not be obtained easily. The public data resource for probing 3D face recognition method does not exist. Therefore, it is impossible to compare with the performance of different algorithms. In fact, it also is insignificant to compare them in different database. Thus, most researchers just perform their algorithms and analyze the experimental results on a small scale 3D face data set.

5 Conclusions and future work

In this paper, multiple representation features of 3D face are defined for the 3D face recognition. An effective multi-feature fusion strategy is proposed with the contributive efficiency of every face feature. This is a linear weighted strategy. The similarity metric measures of face features are defined. The Fisher linear discriminant analysis is used to calculate the weighted coefficient of every face feature with the within-class and between-class similarity measures. The presented method was tested on BJUT-3D face database and the virtual 3D face samples reconstructed from CAS-PEAL 2D face images. The experimental results indicate that the method is satisfying. And the recognition performance based on the TWM fusion strategy is better than that based on the RWM.

There are limited public 3D face databases. Some available 3D face databases are small scale and the data formats are also different. These difficulties restrict research progress of 3D face recognition. Our research is also based on our own 3D face database, the BJUT-3D and 50 individuals, with multiple samples and were picked up for recognition performance test. In the future, we will keep on enlarging the experimental samples and perform experiments on the larger scale data set. Meanwhile, expanding the BJUT-3D face database and enriching the samples with illumination and expression variations are also the important work. It will be valuable for evaluating 3D face recognition method and providing public face data resource for researchers.

References

1 Gupta S, Markey M K, Bovik A C. *Advances and Challenges in 3D and 2D+3D Human Face Recognition*. New York: Nova Science Publishers, 2007

2 Hallinan P W, Gordon G G, Yuille A L, Giblin P, Mumford D. *Two- and Three-Dimensional Patterns of the Face*. Natick, USA: A K Peters, 1999. 262

3 Gordon G G. Face recognition based on depth and curvature features. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Champaign, USA: IEEE, 1992. 108−110

4 Samir C, Arivastava A, Daoudi M. Three-dimensional face recognition using shapes of facial curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(11): 1858−1863

5 Moreno A B, Sanchez A, Velez J F, Diaz F J. Face recognition using 3D surface-extracted descriptors. In: Proceedings of Irish Machine Vision and Image Processing Conference. Portrush, Antrim: Northern Ireland, 2003. 50–57

6 Hesher C, Srivasteva A, Erlebacher G. A novel technique for face recognition using range imaging. In: Proceedings of the 7th International Symposium on Signal Processing and Its Applications. Washington D. C., USA: IEEE, 2003. 201−204

7 Achermann B, Jiang X Y, Bunke H. Face recognition using range images. In: Proceedings of International Conference on Virtual Systems and Multimedia. Washington D. C., USA: IEEE, 1997. 129−136

8 Cartoux J Y, LaPreste J T, Richetin M. Face authentication or recognition by profile extraction from range images. In: Proceedings of the Workshop on Interpretation of 3D Scenes. Washington D. C., USA: IEEE, 1989. 194−199

9 Pan G, Wu Y J, Wu Z H, Liu W Y. 3D face recognition by profile and surface matching. In: Proceedings of the International Joint Conference on Neural Networks. Portland, USA: IEEE, 2003. 2169−2174

10 Tsutsumi S, Kikuchi S, Nakajima M. Face identification using a 3D gray-scale image — a method for lessening restrictions on facial direction. In: Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C., USA: IEEE, 1998. 306−311

11 Wang Y, Chua C S, Ho Y K. Facial feature detection and face recognition from 2D and 3D images. *Pattern Recognition Letters*, 2002, **23**(10): 1191−1202

12 Papatheodorou T, Rueckert D. Evaluation of automatic 4D face recognition using surface and texture registration. In: Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C., USA: IEEE, 2004. 321−326

13 BJUT-3D Face Database [Online], available: http://www.bjpu.edu.cn/sci/multimedia/mul-lab/3dface/face_database.htm, September 1, 2007

14 Gu C L, Yin B C, Hu Y L, Cheng S Q. Resampling based method for pixel-wise correspondence between 3D faces. In: Proceedings of International Conference on Information Technology: Coding and Computing. Washington D. C., USA: IEEE, 2004. 614

15 Jain A, Nandakumar K, Ross A. Score normalization in multimodal biometric systems. *Pattern Recognition*, 2005, **38**(12): 2270−2285

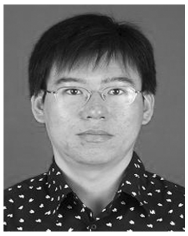
16 Wang Cheng-Zhang, Yin Bao-Cai, Sun Yan-Feng, Hu Yong-Li. An improved 3D face modeling method based on morphable model. *Acta Automatica Sinica*, 2007, **33**(3): 232−239 (in Chinese)

17 Xu C H, Wang Y H, Tan T N, Quan L. Automatic 3D face recognition combining global geometric features with local shape variation information. In: Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C., USA: IEEE, 2004. 308−313

18 Godil A, Ressler S, Grother P. Face recognition using 3D facial shape and color map information: comparison and combination. In: Proceedings of SPIE′s Biometric Technology for Human Identification. Orlando, USA: 2005. 351−361

19 Mian A S, Bennamoun M, Owens R A. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(11): 1927−1943

SUN Yan-Feng Professor at College of Computer Science and Technology, Beijing University of Technology. She received her Ph.D. degree from Dalian University of Technology in 1993. Her research interest covers pattern recognition and image process. Corresponding author of this paper.
E-mail: yfsun@bjut.edu.cn



TANG Heng-Liang Ph. D. candidate at College of Computer Science and Technology, Beijing University of Technology. He received his bachelor degree from Beijing University of Technology in 2005. His research interest covers face recognition and image process.
E-mail: tanghengliang@emails.bjut.edu.cn

YIN Bao-Cai Professor at College of Computer Science and Technology, Beijing University of Technology. He received his Ph. D. degree from Dalian University of Technology in 1993. His research interest covers digital multimedia, multi-functional perception, virtual reality, and computer graphics. E-mail: yinbc@bjut.edu.cn