Automatic 3D Face Recognition Using Discriminant Common Vectors

Cheng Zhong, Tieniu Tan, Chenghua Xu, and Jiangwei Li

National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, 100080, P.R. China
{czhong, tnt, chxu, jwli}@nlpr.ia.ac.cn

Abstract. In this paper we propose a fully automatic scheme for 3D face recognition. In our scheme, the original 3D data is automatically converted into the normalized 3D data, then the discriminant common vector (DCV) is introduced for 3D face recognition. We also compare DCV with two common methods, i.e., principal component analysis (PCA) and linear discriminant analysis (LDA). Our experiments are based on the CASIA 3D Face Database, a challenging database with complex variations. The experimental results show that DCV is superior to the other two methods.

1   Introduction

Automatic identification of human faces is a very challenging research topic, which has gained much attention during the last few years. Most of this work, however, is focused on intensity or color images of faces [1]. There is a commonly accepted claim that face recognition in 3D is superior to 2D because of the invariance of 3D sensors to illumination and pose variation. Recently with the development of 3D acquisition system, 3D face recognition has attracted more and more interest and a great deal of research effort has been devoted to this topic.

Many methods have been proposed for 3D face recognition over the last two decades. Some earlier research on curvatures analysis has been proposed for face recognition based on the high-quality 3D data, which can characterize delicate features [2] [3]; In [4], a 3D morphable model is described with a linear combination of the shape and texture of multiple exemplars. This model can be fitted to a single image to obtain the individual parameters, which are used to characterize the personal features; Chua et al. [5] treat face recognition as a 3D non-rigid surface matching problem and divided the human face into rigid and non-rigid regions. The rigid parts are represented by point signatures to identify the individual. Beumier et al. [6] develop a 3D acquisition prototype based on structure light and built a 3D face database. They also propose two methods of surface matching and central/lateral profiles to compare two instances. Chang et al. [7] use PCA on both 2D intensity images and 3D depth images, and fuse 2D and 3D results to obtain the final performance. Their results show that the combination

of 2D and 3D features is very effective for characterizing a person. However, it should be noted that the existing methods usually have a high computational cost [4] [6], involve small databases [3] [5] or depend on manual labeled points [7].

In this paper, we introduce a fully automatic 3D face recognition scheme. The flowchart is shown in Fig. 1. First, we preprocess the input 3D data. Second, we use DCV to project the normalized 3D data from the original high dimensional space to low dimensional subspace spanned by DCV. Third, we use nearest neighbor (NN) classifier to classify the 3D face images. We also make a detailed comparison between DCV, LDA and PCA to test their performance for 3D face recognition.

The main contributions of this paper are as follows: (1) We introduce the DCV method into 3D face recognition; (2) We make a detailed comparison between PCA, LDA and DCV.

The rest of this paper is organized as follows. In Section 2, we describe the 3D face data preprocessing. A detailed description on DCV is illustrated in Section 3. Section 4 shows the experimental results, and finally we conclude this paper in Section 5.

2   3D Face Data Preprocessing

Fig. 2 shows some examples from the CASIA 3D Face Database. The original images have many problems, such as different poses and much noise, so data preprocessing is necessary before recognition. Data preprocessing includes the following three steps:

The first step is nose location. In this step, we use local features to obtain some nose tip candidate points, and a trained SVM classifier is used to find the nose tip point [8].

The second step is the registration. In this step, we construct a mesh model corresponding to each 3D face and the ICP algorithm is applied to the mesh models to complete the registration [9].

The third step is data normalization. In this step, we follow the method as stated in [7], but here we use a double mask scheme. Because the margin region contains more noise than the region of interest, we first adopt a large mask. After we fill holes and smooth the data, we adopt a small mask to obtain the region of interest, which is the final output depth image. Fig. 3 shows some normalized 3D images after the data preprocessing.

3   3D Face Representation Using DCV

In this section, we mainly describe how to represent 3D face images using DCV. The main procedures can be summarized as follows: first we need to calculate common vector (CV) images from the given training set; second we calculate the DCV based on the obtained CV images; finally we represent the original 3D faces using DCV. Next we will describe these procedures in detail.
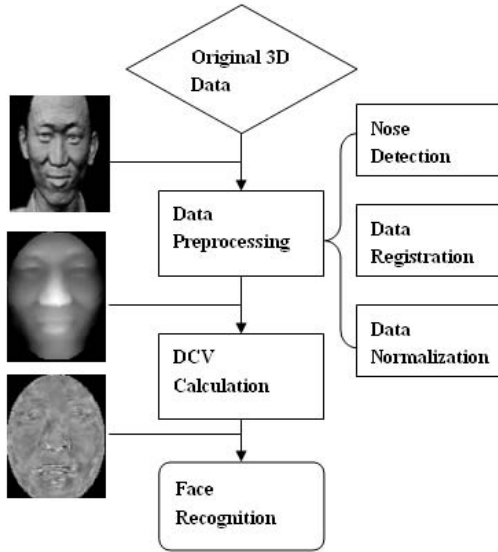
Fig. 1. The flowchart of our automatic 3D face recognition
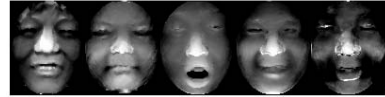

Fig. 2. Original 3D images

Fig. 3. Preprocessed 3D images

Fig. 5. Common vector images

Fig. 4. Comparison of different common vector images, the first one is common vector image of five images with neural expression, the second one is common vector image of five images with different expressions and the third one is common vector image of the above ten images
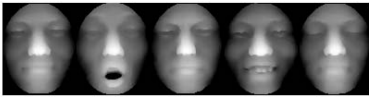
Fig. 6. Comparison of eigenfaces, fisherfaces and discriminant common vector images. The first row shows the eigenfaces, the second row shows the fisherfaces, and the third row shows discriminant common vector images

3.1   Common Vector Images

Supposed that in training set each person has m original images $\{a_1, a_2, \cdots, a_m\}$. We convert them into m original vectors , then we define the $(m-1)$ dimensional difference subspace B by taking the differences between the vectors, i.e.

$$b_i = a_{i+1} - a_1 (i = 1, 2, \cdots, m-1) \tag{1}$$

B is spanned by these difference vectors. Since $b_1, b_2, \cdots, b_{m-1}$ are not expected to be orthonormal, an orthonormal basis vector set can be obtained by using Gram-Schmit orthogonalization [10]. After that the basis vector set for B will be $\{z_1, z_2, \cdots, z_{m-1}\}$ in this case. If the common vector of one person is called as a_{com}, then each of the original vectors can be written as

$$a_i = a_{i,dif} + a_{com} \tag{2}$$

the difference vectors $a_{i,dif}$ are the projections of the original vectors onto the difference subspace B, that is

$$a_{i,dif} = < a_i, z_1 > z_1 + < a_i, z_2 > z_2 + \cdots + < a_i, z_{m-1} > z_{m-1} \tag{3}$$

We can obtain m difference vectors from m original vectors. The common vector a_{com} is chosen as

$$a_{com} = a_i - a_{i,dif} \qquad \forall i = 1, 2, \cdots, m \tag{4}$$

It can be seen as the projection of the original vectors onto the indifference subspace. As $a_{com} = a_1 - a_{1,dif} = a_2 - a_{2,dif} = \cdots = a_m - a_{m,dif}$, we can obtain only one common vector from the m original vectors of one person and more details on this may be found in [11]. Fig. 5 shows some common vector images.

3.2   Discriminant Common Vectors

After we obtain the common vector images of each person in the training set, we attempt to compute discriminant common vectors. DCV is the projection that maximizes the total scatter across the common vector images. We can use PCA to obtain the discriminant common vectors and more details on this may be found in [12].

After we obtain DCV, we can project the original high dimensional space into the low dimensional subspace spanned by DCV. Fig. 6 shows different eigenface, fisherface and the discriminant common vector images, respectively. From this figure, we can find that the discriminant common vector images contain more detailed information than eigenfaces or fisherfaces.

4   Experimental Results and Discussion

To make a comparison between PCA, LDA and DCV methods, we have done many experiments on CASIA 3D Face Database. There are 123 persons in the

database, and each person has 37 or 38 images. In our experiment we only use 5 images with neural expression and 5 images with different expressions (smile, laugh, anger, surprise, eye closed) for each person. First we construct a small 3D face database (DB1), which including 5 images with neural expression and 2 images with common expressions (smile and eye closed). Second we use the whole set of images to construct a larger 3D face database (DB2), which includes 5 images with neural expression and 5 images with different expressions. The comparisons of the three methods are all based on the same training sets and testing sets. In all experiments, we use the NN classifier with Mahalanobis cosine distance.

4.1   Experiments on DB1

We list the recognition rate in two cases. First, we use the first three images with neural expression as the training set (Experiment 1), and the remained images as the testing set. Second, we use one image with neural expression, one image with smile expression, one image with eye closed expression as the training set (Experiment 2), and the remaining images as the testing set. The results are shown in Table 1.

4.2   Experiments on DB2

We list the recognition rate in four cases. First, we use the first three images with neural expression as the training set (Experiment 3), and the remained images as the testing set. Second, we use five images with neural expression as the training set (Experiment 4), and the remained images as the testing set. Third, we use one image with neural expression, one image with laugh expression, one image with surprising expression as the training set (Experiment 5), and the remained images as the testing set. Fourth, we use the five images with different expressions as the training set (Experiment 6), and the remained images as the testing set. The results are shown in Table 2.

4.3   Experimental Results Analysis

From Table 1 and Table 2, we can make the following observations:

1) When the intra-class variation is large, we obtain a better performance;
2) In most cases, DCV obtains the best performance;
3) Although the size of training set 4 is larger than training set 3, its performance is worse.

Because the DCV performance mainly depends on the common vectors obtained, here we explain the reasons using common vectors. Fig. 4 shows the common vector images in different situations. We find when in training set one person contains much intra-class variation, the common vector image is almost the same as that of the whole set of images, which means the training set is a very

Table 1. Rank one recognition rates on DB1

Methods	Experiment1	Experiment2
DCV	99%	99.2%
LDA	97.4%	98.4%
PCA	92.9%	94.7%

Table 2. Rank one recognition rates on DB2

Methods	Experiment3	Experiment4	Experiment5	Experiment6
DCV	90.7%	84.6%	96.1%	98.5%
LDA	86.9%	87.4%	93.1%	97.4%
PCA	83.5%	80.8%	92.6%	94.0%

Table 3. Recognition rates on different size of training set

Size	2	3	4	5
Verification rate	90.2%	90.7%	87.4%	84.6%

good representation of the whole set of images, so all methods obtain better performance in this case. From the Section 3 we can find that if training set is a good representation of the whole set of images, DCV is a better choice. Not only it exploits the structures of the original high dimensional space, but also it is the best optimization of the Fisher linear discriminant rule. So in most cases, it performs better than the other two methods. But because DCV exploits more information than other methods from the training set, its recognition performance also depends more on training set. Table. 3 shows the recognition rates with different size of training sets only containing images with neural expression. Although the size of training set is increased, the recognition rate drops. We encounter the over fitting problem here. Because training set is not a representation of the whole set of images, the result we obtained is lack of the generalization ability. In this case, the projection cannot get a good performance in the testing set.

4.4   Discussion

As to computation cost, we only consider the eigen-analysis which is the most time-consuming procedure. Suppose we have N images in the training set, which can be divided into c classes ($N > c$). Then eigen-analysis is performed on one matrix in DCV ($c \times c$), one matrix in PCA ($N \times N$) and two matrices in LDA (one is ($N \times N$) , the other is (($N - c) \times (N - c)$)).The comparison illustrates DCV is the most efficient method of the three.

There are also some drawbacks of our experiments. Because of the limitation of the CASIA 3D Face Database, we only have the 3D face data in one session and we cannot test the influence on the DCV method due to session variations.There are also some other public 3D face databases, such as FRGC, but it is a manual labeled database and its 3D face data does not suit to our

preprocessing algorithm. Using the given points, our experimental results show that DCV also performs better than LDA and PCA on FRGC1.0.

5   Conclusions

In this paper, we have presented a fully automatic system integrating efficient DCV representation for 3D face recognition. We have also compared our proposed method with two other commonly used methods, i.e., PCA and LDA on a large 3D face database. All the experiments are performed in a fully automatic way. From the experimental results, we find that DCV obtains a better performance than LDA and PCA.

Acknowledgement

References

1. R. Chellapa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. In *Proceedings of the IEEE*, pages 705–740, May 1995.
2. G. Gordon. Face recognition based on depth and curvature features. In *Proc. CVPR*, pages 108–110, June 1992.
3. J. C. Lee and E. Milios. Matching range images of human faces. In *Proc. ICCV*, pages 722–726, 1990.
4. V. Blanz and T. Vetter. Face identification based on fitting a 3d morphable model. *IEEE Trans. PAMI*, (9):1063–1074, 2003.
5. C. S. Chua, F. Han, and Y. K. Ho. 3d human face recognition using point signature. In *Proc. FG*, pages 233–239, 2000.
6. C. Beumier and M. Acheroy. Automatic 3d face authentication. *Image and Vision Computing*, (4):315–321, 2000.
7. K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multi-model 2d+3d face biometrics. *IEEE Trans. PAMI*, (4):619–624, 2005.
8. Chenghua Xu, Yunhong Wang, Tieniu Tan, and Long Quan. Robust nose detection in 3d facial data using local characteristics. In *Proceedings of the IEEE, International Conference of Image Processing*, pages 1995–1998, 2004.
9. Chenghua Xu, Yunhong Wang, Tieniu Tan, and Long Quan. Automatic 3d face recognition combining global geometric features with local shape variation information. In *Proceedings of the IEEE, International Conference Automatic Face and Gesture Recognition*, pages 308–313, 2004.
10. M. Keskin. Orthogonalization process of vector space in work station and matlab medium. In *Elect. Electron. Eng. Dept., Osmangazi Univ., Eskisehir, Turkey*, July 1994.
11. M. B. Gulmezoglu, V. Dzhafarov, and A. Barkana. The common vector approach and its relation to principal component analysis. *IEEE Transactions on Speech and Audio Processing*, (6):655–662, 2001.
12. H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana. Discriminative common vectors for face recognition. *IEEE Trans. PAMI*, (1):4–13, 2005.