Front. Electr. Electron. Eng. China 2011, 6(2): 347–352
DOI 10.1007/s11460-011-0155-x

RESEARCH ARTICLE

Hongxia ZHANG, Yanning ZHANG, Zhe GUO, Zenggang LIN, Chao ZHANG

3D face recognition based on principal axes registration and fusing features

Abstract A 3D face recognition approach which uses principal axes registration (PAR) and three face representation features from the re-sampling depth image: Eigenfaces, Fisherfaces and Zernike moments is presented. The approach addresses the issue of 3D face registration instantly achieved by PAR. Because each facial feature has its own advantages, limitations and scope of use, different features will complement each other. Thus the fusing features can learn more expressive characterizations than a single feature. The support vector machine (SVM) is applied for classification. In this method, based on the complementarity between different features, weighted decision-level fusion makes the recognition system have certain fault tolerance. Experimental results show that the proposed approach achieves superior performance with the rank-1 recognition rate of 98.36% for GavabDB database.

Keywords 3D face recognition, principal axes registration (PAR), fusion feature, weighted voting

1 Introduction

3D face recognition is an emerging modality, trying to use 3D geometry [1] of the face for greater recognition accuracy than 2D face recognition. However, there are some challenges existing in current 3D face recognition system [2,3], which concern the field-deployable system. Two major challenges are: 1) efficiency; 2) accuracy.

One limitation of some existing approaches to 3D face recognition involves high time-cost, which makes it hard to incorporate into a practical recognition system. Iterative closest point (ICP) [4], which is the most popular alignment algorithm, has been widely used for 3D models registration. However, the iterative process makes ICP computationally expensive and the registration must be done for each model in the database. To deal with this problem, our registration method is based on face symmetry, which does not need iterative process, and also set each model under uniform correspondence.

Theoretically, combining multiple classifiers [5,6] will improve overall recognition rate. Inspired by selection in the real world, we choose the weighted voting principle [7]: different thresholds gained by training are employed for different classifiers. The effectiveness of the combination is proved by the weak learning theory. The experiments have been tested on the multi-pose subset and multi-expression subset of GavabDB.

The main contributions of our work are listed as follows: 1) A simple and fast 3D face registration method — principal axes registration (PAR) is used. The accuracy of the PAR registration works well on the result of the following comparison. 2) Based on the predictions of the classifier with their error, weighted voting for the fusion of multiple classifiers with statistic and geometry feature is presented.

2 PAR

We use the principal axes analysis algorithm as Zhang et al.'s in Ref. [8] to align each face in a standard coordinate system. A well-known standard algorithm to solve registration problem is the ICP, which in each step applies a motion in a least squares sense as close as possible to their closest points on the model shape. However, it is not suitable for real-time use and not available for registration with large gesture difference. To enable the registration, we follow the coarse-to-fine strategy using PAR.

Received March 23, 2011; accepted April 29, 2011

Hongxia ZHANG, Yanning ZHANG (✉), Zhe GUO,
Zenggang LIN, Chao ZHANG
Shaanxi Provincial Key Laboratory of Speech and Image Information Processing (SAIIP), School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
E-mail: ynzhang@nwpu.edu.cn

2.1 Principal axes coordinate systems

For any face model, the face principal axes coordinate systems are formed by three principal axes of the preprocessed face data. Figure 1 shows the principal axes coordinate, where Z-axis corresponds to the face orientation, Y-axis is perpendicular to vertical axis within the facial symmetry plane while X-axis is perpendicular to facial symmetry plane.

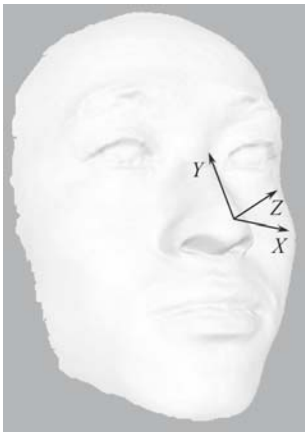


Fig. 1 Principal axes coordinate systems

2.2 Coarse registration

Since each intact face is frontal within its principal axes coordinate systems. We translate the model-reference to its principal axes. In the first step, we obtain the three principal axes (x_1, y_1, z_1) by applying the principal component analysis (PCA) on the face model. Subsequently, the P point set is translated to Q point set, which is in the principal axes (x_1, y_1, z_1) coordinate systems.

$$(x_1, y_1, z_1) = \begin{bmatrix} x_{11} & y_{11} & z_{11} \\ x_{12} & y_{12} & z_{12} \\ x_{13} & y_{13} & z_{13} \end{bmatrix}, \tag{1}$$

$$q_i = (x_1, y_1, z_1)^{\mathrm{T}} p_i. \tag{2}$$

2.3 Fine registration

Unfortunately, the actual 3D face models are not strictly symmetrical, owing to the collection of 3D face data. As a result, the Y-axis obtained by the PCA approach is not always the symmetry axis as we expected. Even though the data is not symmetric, the region $R(r_1, r_2, \ldots, r_t)$ covered by a certain radius centered in the nose tip remains symmetric. So the region of interest $T(t_1, t_2, \ldots, t_H)$ is opted for areas where Z value is larger than a threshold. We choose 50 as the threshold and the region T mainly includes the nose.

Second, the PCA operation upon T is applied on X and Y coordinates. The final face principal axis is obtained by Eq. (5). The position of each point in the principal axes coordinate systems is defined as

$$(x_2, y_2, z_2) = \begin{bmatrix} x_{21} & y_{21} & z_{21} \\ x_{22} & y_{22} & z_{22} \\ x_{23} & y_{23} & z_{23} \end{bmatrix}, \tag{3}$$

$$(x, y, z) = \begin{bmatrix} x_{11} & y_{11} & z_{11} \\ x_{12} & y_{12} & z_{12} \\ x_{13} & y_{13} & z_{13} \end{bmatrix} \begin{bmatrix} x_{21} & y_{21} & z_{21} \\ x_{22} & y_{22} & z_{22} \\ x_{23} & y_{23} & z_{23} \end{bmatrix}, \tag{4}$$

$$f = (x, y, z)^{\mathrm{T}} p. \tag{5}$$

Finally, the nose of each face is obtained by the geometric properties. All points of the current face are translated to the new position with the nose tip as the origin. The nose tip is finally located in the zero point of common consistent coordinate system.

Table 1 shows the average time cost between different facial models of one individual using PCA and PAR. In the experiment, the face models contain 16000 points, 32000 triangular patches on average. Beside the reduction of the registration computational time, another attractive attribute of this approach is that in the matching paradigm when searching for an input face in a gallery database, the input face does not have to be aligned with each face in the gallery for matching to occur, enabling fast database search. As a result, PAR is a nice solution to registration of 3D facial data.

Table 1 Average time cost of different registration methods in Matlab circumstance

registration	average time/s
ICP	36.81
PAR	3.98

3 Modified voting principle

In pattern recognition, the multiple classifier fusion is one of the most promising directions. The ensemble classifier can not only reduce variance but also reduce bias. The results are less dependent on peculiarities of a single training set. And the combination of multiple classifiers may learn a more expressive characterization than a single classifier. There are many methods of combining strategies: product, average, max, min, majority voting, rank-based voting, weighted voting, and weighted probability. In our system, weighted voting, the most frequently used combing rules, is adopted to merge three support vector machine (SVM) classifiers with statistic and geometry features.

Since all surfaces are registered (pose normalized), the facial region is extracted from each 3D face by ellipse cropping, which discards at the same time the irrelevant points (e.g., hair, ears, etc.). We resample each face model $(M_1, M_2, \ldots, M_k)$ using a rectangular grid $(H \times W)$ upon the xoy plane. The nose tip N is located in the center of the rectangular grid. The selected reference size of xoy plane is chosen as x: $[-80, 80]$, y: $[-90, 120]$ considering the statistic value of the face frame size of 48 facial surfaces. The resample value is determined by the mean depth value of all the vertices within the mesh. After resampling, all the facial models are in frontal pose with the same size.

In order to reduce the computational time, each facial surface is stored as orientation normalized 50×40 depth map, represented as vectors of length 2000. By now the facial regions are not only cropped into region of interested but also simplified. Next, image processing methods can be performed on the cropped 3D faces, we extract statistic features: Eigenfaces, Fisherfaces (as with Belhumier et al.'s Eigenfaces and Fisherfaces approach) and the geometry feature: Zernike moments.

3.1 Feature extraction

3.1.1 Eigenfaces

Formally, given training set of depth image $X = \{x_1, x_2, \ldots, x_m\}$, each of which is a vector of the same size. An initial set of training vectors X is used to construct the face space. The optional basis W can be determined by computing the eigenvectors of the covariance matrix C, obtained using the singular value decomposition of A.

$$\bar{u} = \frac{1}{m}\sum_{i=1}^{m} x_i, \tag{6}$$

$$u_i = x_i - \bar{u}, \tag{7}$$

$$A = \{u_1, u_2, \ldots, u_m\}, \tag{8}$$

$$C = AA^{\mathrm{T}}. \tag{9}$$

The new feature vector Y is defined by the following linear transformation:

$$y_k = W^{\mathrm{T}} x_k, \quad k = 1, 2, \ldots, M. \tag{10}$$

Orthogonal basis W can be chosen differently depending on different consideration. These eigenvectors span a linear space. Many eigenvectors are devoted to individual differences in face structure, while noise is represented orthogonal to these eigenvectors. So we can reconstruct the face if proper eigenvectors are selected. In our system, the first 55 eigenvectors are chosen to build the orthogonal space.

3.1.2 Fisherfaces

Although the previous PCA approach, calculating projection directions that maximize the total scatter across all classes' projection, is available for reconstruction from a low-dimensional basis, it may not be optimal for the issue of classification. The reason is that the scatter is maximized due to not only the intra-class scatter that is useful for classification, but also the inner-class scatter that, for classification purposes, is redundant information. Rather the previous PCA, Fisher's linear discriminate (FLD) tries to "shape" the scatter in order to make it more reliable for classification. This method selects W in such a way that the ratio of the inner-class scatter matrix S_B and the intra-class scatter matrix S_W is maximized. S_B and S_W are defined as

$$S_B = \sum_{i=1}^{c} M_i \left(u_i - \bar{u}\right)\left(u_i - \bar{u}\right)^{\mathrm{T}}, \tag{11}$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} \left(x_k - u_i\right)\left(x_k - u_i\right)^{\mathrm{T}}, \tag{12}$$

where u_i is the center of class X_i, M_i is the number of samples in class X_i, $\{w_i | i = 1, 2, \ldots, m\}$ is the general feature vectors corresponding to the m maximal general feature values $\{\lambda_i | i = 1, 2, \ldots, m\}$ of S_B and S_W, i.e.,

$$S_B w_i = \lambda_i S_W w_i, \quad i = 1, 2, \ldots, m. \tag{13}$$

Note that, there exist $c - 1$ nonzero general feature values, as a result, the upper limit of m is $c - 1$, and c is the number of classes.

It has been proved that if w satisfies the following Eq. (15), the optimum projection will make criterion function gain the maximum value.

$$W_{\mathrm{opt}} = \arg\max_{W} \frac{\left|W^{\mathrm{T}} S_B W\right|}{\left|W^{\mathrm{T}} S_W W\right|} = \left[w_1, w_2, \ldots, w_m\right], \tag{14}$$

$$S_W^{-1} S_B w = \lambda w. \tag{15}$$

3.1.3 Zernike

To calculate the Zernike moments [9] of the depth map $f(x,y)$, the image is first mapped to the unit disk using polar coordinates, where the center of the image is the origin of the unit disk. Those depth pixels falling outside the unit disk are not used in the calculation. The coordinates are then described by ρ which is the length of the vector from the origin to the coordinate point and θ which is the angle from the x-axis to the vector ρ, by convention measured from the positive x-axis in a counter clockwise direction. The Zernike moment descriptor has many desirable properties: rotation invariance, robustness to noise, expression efficiency, fast computation and multi-level representation for describing the various shapes of pattern.

3.2 Weighted voting

Known in literature conclusions, the majority vote [10], which is by far the simplest combination methods, does not assume prior knowledge of the behavior of the individual classifiers, and it does not require training on large quantities of representative recognition results from the classifiers. On the other hand, the weighted vote is taken into consideration [11]. The weighted voting (WV), called linear combination of classifiers, is one of the most common combination methods in machine learning. It is based on the idea that not all voters are equal. Instead, it can be desirable to recognize differences by giving voters different amount of say (weights) concerning the outcome of vote.

In our recognition system, SVM classifiers [12] are deployed, and that for each input different kinds of facial feature, each classifier produces a unique decision regarding the identity of the facial feature. The optimal weight value is inversely proportional to errors made by three SVM component classifiers. Voting each ensemble member votes for one of the classes, WV ultimately make a weighted sum of the votes of the ensemble members and predict the class with the highest number of vote.

4 Experiments

To test our system we employ GavabDB dataset [13], which contains 549 3D scans of facial surfaces corresponding to 61 different individuals (45 male and 16 female). There are seven different scans for each person, two neutral frontal, two neutral with pose (looking down and up), and three frontal scans in which the subject presents different and accentuated facial expressions. We divide these 3D faces into disjoint training and testing sets. Each training and test set consists of different images of 61 people respectively.

It should be noted that combing multi-classifiers is not exactly a substitute for designing better classifiers. But it is a truism that combinations of better algorithms tend to produce better results. This is graphically depicted in Figs. 2 and 3, in which the performance of ensemble

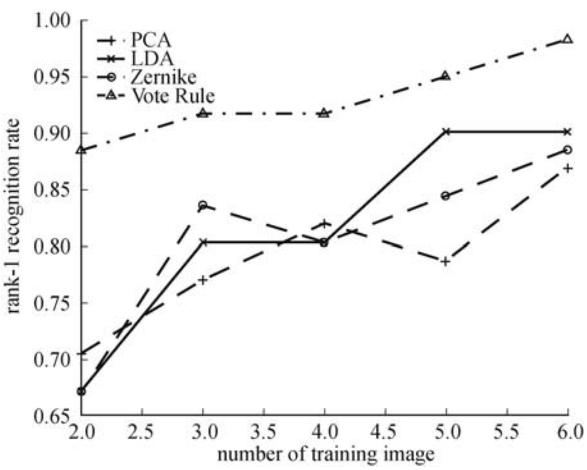


Fig. 2 Rank-1 of neutral frontal (NF)

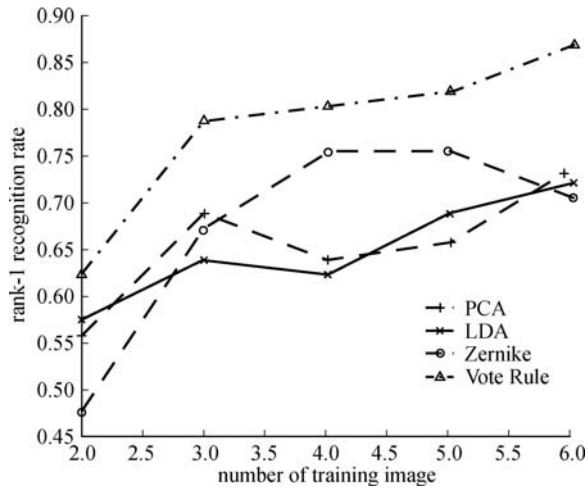


Fig. 3 Rank-1 of multi-pose (MP)

classifier and three individual classifiers are shown.

In the experiment, one model is randomly selected for test while other six models for training. At the same time, we gradually reduce the training model number by one. Figures 2 and 3 show the comparison of rank-1 recognition rates with different facial features. The x-axis is the training facial number, and the y-axis represents the rank-1 recognition rate. The identification rate becomes smaller while the training set becomes smaller, that is to say, enough training set is a premise to the superior performance of the SVM classifiers.

Each feature has its own advantage although they are in different spaces, the subspace features: Eigenfaces and Fisherfaces, produce a low-dimensional subspace projection matrix with less calculation. Especially owing to the orthogonal transformation the Eigenface is the furthest explanation feature while the Fisherface is the most determinant one for the suitable projection direction. And the Zernike moments feature has a rotation invariant property which can offset the registration error to some extent. From Figs. 2 and 3, it can be seen that different features have complementary, faces which cannot be recognize in one space may easily be known in other space. The recognition system achieves higher recognition rate than single feature: 98.4% and 86.9% recognition rate on NF and MP set of 3D GavabDB database, which shows certain fault tolerance through weighted decision-level fusion

5 Conclusion

In this paper, a new scheme for automatic 3D face recognition was proposed by using PAR and multi-classifier fusion. In this scheme, the 3D facial surface is first iteratively registered by principal component analysis. This new approach largely reduces the registration time. In addition, it works well on the following recognition stage.

Table 2 Identification rates of different methods on GavabDB/%

rank-1 rate	Guan's [14]	ter Haar's [15]	Mahoor's [16]	ICP	our method
NF	93	91.1	93.5	83.7	98.4
MP	83	NA	NA	72.6	86.9

Furthermore, the resemble SVM classifier together with the WV largely improve the recognition rate. As the experimental results shown in Table 2, our method achieves the rank-1 recognition rate 86.9% on multi-pose dataset, which has the obvious advantages and outperforms the state-of-the-art [14–16] methods.

Acknowledgements The authors would like to acknowledge the use of the GavabDB face database in this paper due to Moreno and Sanchez. This work was supported in part by the National Natural Science Foundation of China (Grant No. 60872145), the National High Technology Research and Development Program of China (No. 2009AA01Z315), and the Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (No. 708085).

References

1. Llonch R S, Kokiopoulou E, Tosic I, Frossard P. 3D face recognition with sparse spherical representations. Pattern Recognition, 2010, 43(3): 824–834
2. Phillips P J, Flynn P J, Scruggs T, Bowyer K W, Chang J, Hoffman K, Marques J, Min J, Worek W. Overview of the face recognition grand challenge. In: Proceedings of IEEE Computer Society Conference on Computer Vision. 2005, 1: 947–954
3. Bowyer K W, Chang K, Flynn P. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. Computer Vision and Image Understanding, 2006, 101(1): 1–15
4. Yan P, Bowyer K W. A fast algorithm for ICP-based 3D shape biometrics. In: Proceedings of the fourth IEEE Workshop on Automatic Identification Advanced Technologies. 2005, 213–218
5. Zhang P, Bui T D, Suen C Y. A novel cascade ensemble classifier system with a high recognition performance on handwritten digits. Pattern Recognition, 2007, 40(12): 3415–3429
6. Hullermeier E, Vanderlooy S. Combining predictions in pair wise classification: an optimal adaptive voting strategy and its relation to weighted voting. Pattern Recognition, 2010, 43(1): 128–142
7. Fu Z L. Effective property and best combination of classifier linear combination. Journal of Computer Research and Development, 2009, 46(7): 1206–1216 (in Chinese)
8. Zhang G P, Zhang Y N, Guo Z. 3D face registration based on accurate principal axes analysis and ICP. Journal of Computer Engineering and Applications, 2006, 42(29): 62–64 (in Chinese)
9. Khotanzad A, Hong Y H. Invariant image recognition by Zernike moments. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(5): 489–497
10. Lam L, Suen C Y. Application of majority voting to pattern recognition: an analysis of its behavior and performance. Systems and Humans, 1997, 27(5): 553–568
11. Kuncheva L I. Combining pattern classifiers: methods and algorithms. IEEE Transactions on Neural Networks, 2007, 18(3): 964–964
12. Chang C C, Lin C J. Training v-support vector classifiers: theory and algorithms. Neural Computation, 2001, 13(9): 2119–2147
13. Moreno A B, Sanchez A. GavabDB: a 3D face database. In: Proceedings of the 2nd COST Workshop on Biometrics on the Internet. 2004, 75–80
14. Guan P, Zhang L M. 3D face recognition based on facial structural angle and local region map. In: Proceedings of IEEE International Conference on Multimedia and Expo. 2008, 41–44
15. ter Haar F B, Veltkamp R C. 3D face recognition using facial contour curves. In: Proceedings of IEEE International Conference on Shape Modeling and Applications. 2008, 259–260
16. Mahoor M H, Abdel-Mottaleb M. 3D face recognition based on 3D ridge lines in range data. In: Proceedings of IEE International Conference on Image Processing. 2007, 1: 137–140

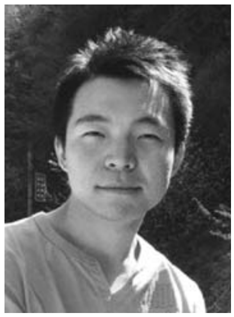
Hongxia ZHANG was awarded the M.S. degree in computer science from Northwestern Polytechnical University in 2011. Her research interests include computer vision, bioinformatics, pattern recognition and 3D mesh pre-processing.

Yanning ZHANG received her B.S. degree from Dalian University of Science and Engineering in 1988, and her M.S. and Ph.D degrees from Northwestern Polytechnical University in 1993 and 1996, respectively. She is currently the assistant dean and professor of School of Computer Science, Northwestern Polytechnical University. In 2009, she served as the organizing chair of the Ninth Asian Conference on Computer Vision (ACCV2009). She has won two first class awards from Shaanxi province and several science and technology innovation awards from the Commission of Science, Technology and Industry for National Defense.

Her research interests focus on signal and image processing, computer vision, and pattern recognition. She has published more than 200 papers in international journals, conference proceedings, and Chinese key journals.

Zhe GUO was awarded the B.S. and M.S. degrees in computer science from Northwestern Polytechnical University in 2005 and 2008, and studied for her Ph.D degree in computer science (3D face feature extraction) at Northwestern Polytechnical University. Her research interests include computer vision, bioinformatics, pattern recognition and 3D mesh pre-processing. Her current research focuses on 3D face recognition and 3D face feature extraction.

Zenggang LIN was awarded the B.S. degree in computer science from Nanjing University of Aeronautics and Astronautics in 1999, and M.S. and Ph.D degrees in computer science from Northwest University in 2003 and 2011, and now he works in computer science at Northwestern Polytechnical University. His research interests include computer vision, bioinformatics, pattern recognition and 3D mesh pre-processing. His current research focuses on 3D+2D face recognition.



Chao ZHANG was awarded the B.S. degree in computer science from Northwestern Polytechnical University in 2008, and studied for his M.S. degree in computer science at Northwestern Polytechnical University. His research interests include computer vision, bioinformatics, pattern recognition and 3D mesh pre-processing.