Automatic 3D Face Recognition Combining Global Geometric Features with Local Shape Variation Information

Chenghua Xu[1], Yunhong Wang[1], Tieniu Tan[1], Long Quan[2]
[1]*Center for Biometric Authentication and Testing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, P. R. China, 100080*
E-mails: {chxu, wangyh, tnt }@nlpr.ia.ac.cn
[2] *Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*
E-mails: quan@cs.ust.hk

Abstract

Face recognition is a focused issue in pattern recognition over the past decades. In this paper, we have proposed a new scheme for face recognition using 3D information. In this scheme, the scattered 3D point cloud is first represented with a regular mesh using hierarchical mesh fitting. Then the local shape variation information is extracted to characterize the individual together with the global geometric features. Experimental results on 3D_RMA, a likely largest 3D face database available currently, demonstrate that the local shape variation information is very important to improve the recognition accuracy and that the proposed algorithm has promising performance with a low computational cost.

Keywords: *3D face recognition, shape variation, mesh model, Gaussian-Hermite moments*

1. Introduction

Nowadays biometric identification has obtained much attention due to the urgent need for more reliable personal identification. Of all the biometrics features, face is among the most common and most reachable so that face recognition remains one of the most active research issues in pattern recognition. In the past decades, most work focuses on the source of 2D intensity or color images. Since the accuracy of 2D face recognition is influenced by variations of poses, expressions, illumination and subordinates, it is still difficult to develop a robust automatic 2D face recognition system.

The 3D facial data can provide a promising way to understand the feature of the human face in 3D space and has potential possibility to improve the performance of the system. There are some distinct advantages in using 3D information: sufficient geometrical information, invariance of measured features relative to transformation and capture process by laser scanners being immune to illumination variation.

With the development of 3D acquisition system, 3D capture is becoming faster and cheaper, and face recognition based on 3D information is attracting more and more attention. Some earlier researches on curvature analysis [1,2,3] were proposed for face recognition based on the high-quality range data acquired from 3D laser scanners. Recently, Blanz *et al.* [4,5] constructed a 3D morphable model with a linear combination of the shape and texture of multiple exemplars. That model could be fitted to a single image to obtain the individual parameters, which were used to characterize the personal features. Their results seemed very promising except that the modeling process incurred a high computational cost. Chen *et al.* [6] treated 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. Bronstein *et al.* [7] represented facial surface based on geometric invariants to isometric deformations. They realized multi-model recognition by integrating flattened textures and canonical images. Their algorithm was robust to some expression variations. Beumier *et al.* [8,9] developed a 3D acquisition prototype based on structured light and built a 3D face database. They also proposed two methods of surface matching and central/lateral profiles to compare two instances. Both of them constructed some central and lateral profiles to represent the individual, and obtained the matching value by minimizing the distance of the profiles. It should be noted that there are two main difficulties facing 3D face recognition: high computational and spatial cost and inconvenient 3D capture. The existing methods usually have a high computational cost [4,5,8,9] or are tested on a small database [1,2,3,6].

In our previous work [10], we used the global geometric feature to realize the face recognition. Further, we observed that the shape variation of the local areas (e.g. mouth, nose, etc.) was also crucial for characterizing the individual. In this paper, we develop an automatic face recognition method combining the global geometric features with local shape variation information.

The main contributions of this paper are as follows: 1) A robust method is developed to build the regular mesh model based on the scattered point cloud. 2) The local shape variation information is extracted to represent the face feature together with the global geometric features. Here we first define a metric to quantify the local shape and then Gaussian-Hermite moments [11,12] are applied to describe the shape variation.

The remainder of this paper is organized as follows. In Section 2, we introduce how to obtain the regular mesh model from the 3D point cloud. The process of feature extraction is described in Section 3. Section 4 illustrates the classifiers for face recognition. Section 5 reports the experimental results and gives some comparisons with existing methods. Finally, Section 6 summarizes this paper and future work.

2. Face modeling

In this work, we use the face database 3D_RMA [8], in which each sample is represented with one 3D scattered point cloud. We intend to build a regular mesh with a fixed number of nodes and facets to represent the shape of one human face. Moreover, the different meshes have the corresponding nodes and the same pose to the average model.

Our modeling process includes three steps: pre-modeling, pose acquisition and re-modeling as outlined in Fig.1. This process is described in [10], and here we only describe it concisely.

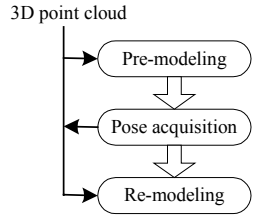


Figure 1. Modeling process

2.1. Pre-modeling

Beginning with a simple basic mesh (see Fig.2a), a regular and dense mesh model is generated to fit the 3D scattered point cloud. We develop a universal fitting algorithm for regulating the hierarchical mesh to be conformed to the 3D points. This process includes two steps: initialization of the basic mesh and fitting of the hierarchy meshes.

Due to the limited quality, the nose seems to be the only facial feature providing robust geometrical features for preliminary effort. We localize the prominent nose in the point cloud and utilize it to initialize the basic mesh.

After initialization, the basic mesh is aligned with the point cloud. Nevertheless, the basic mesh is so coarse that the basic contour of the human face cannot be described. The non-linear subdivision scheme [13] is utilized to refine the basic mesh, and at the same time the refined mesh is regulated according to the data at each level. With the proceeding of refinement and regulation, the mesh can represent the individual well level by level.

Fig.2 shows the mesh after regulation in different refining levels. The coarse mesh does not describe the human face well though it attempts to approach the point cloud. The mesh of level four is dense enough to represent the face surface. Of course, the denser the mesh is, the better the face is represented. Obviously, the denser mesh costs more time and space. In this paper, we use the mesh refined four times.

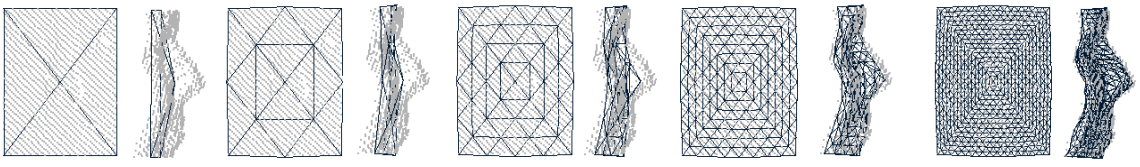
2.2. Pose acquisition by regulating mesh models

The different point clouds have different position and rotation relative to the 3D equipment. This difference is usually called as pose variation. The mesh model obtained from the previous step has the same pose to its corresponding point cloud. Thus we can get the pose parameters from the mesh models rather than from the point clouds directly, which will save much time.

First, an average mesh model is obtained by averaging the mesh models from pre-modeling process. This average model is considered as the ground model and all the models are rotated and translated to align with it. Finally, we obtain the result that the models have the best overlap with the average model, as well as the values of pose.

2.3. Re-modeling

We transform the original point clouds using the obtained values from previous section so that they have the same pose with the average mesh model. The transformed point cloud is modeled again in the same way as the pre-modeling stage. Thus after this stage, a



(a) (b) (c) (d) (e)

Figure 2. The regulated mesh models in different levels. (a) Basic mesh. (b) Level one. (c) Level two. (d) Level three. (e) Level four. Each mesh is showed in front and profile views.

regular mesh model for each point cloud is built. Moreover, all these mesh models have the same pose and represent the facial geometric shape realistically. Next we will use this kind of model to extract the individual features.

3. Feature extraction

Our feature vector to characterize the individual includes two parts: the global geometric features and local shape variation information. In the following, we will discuss them respectively.

3.1 Global geometric features

From the analysis of the above modeling process, each point cloud is described with a regular mesh. All these mesh models have the same pose and the corresponding nodes, which have the same position in X-Y plane and different values along Z-axis. Thus we can build a feature vector, which describes the global geometric features as follows

$$V_{geometric} = \{Z(v_1), Z(v_2), \cdots, Z(v_n)\} \quad (1)$$

where $Z(v_i)$ is the Z-coordinate of node v_i of the mesh model. We used this normalized vector to characterize the individual previously [10] and it has limited ability to improve the recognition accuracy.

3.2 Local shape variation information

With our observation, we find that shape variation, especially near the areas such as mouth, nose and eyes, is the important information to characterize the individual. In signal processing, Gaussian-Hermite (G-H) moments provide an effective way to quantify the signal variation and have wide applications in signal and image processing [11,12]. Here we first define a metric to describe the shape of the principle areas with a 1-D vector and then use the G-H moments to analyze the shape variation.

To reduce redundance, we only consider the areas with larger shape variation. We estimate the position of the four areas (mouth, nose, left eye, right eye) in the average mesh model and mark the same areas in the individual mesh model at the same position as shown in the left image of Fig.3. Although the marked areas can only label the similar corresponding areas in the individual model, it is enough for the following process.

To transfer the 3D shape into 1D vector, we first define a metric to describe the shape of one vertex. To each vertex p_e in the marked area, its neighboring vertices $\{p_{e1}, p_{e2}, \cdots, p_{en}\}$ can be obtained easily as shown in the right image of Fig.3. In our regular mesh model, the number of neighboring vertices of the common vertex (not the edge vertex) is always six. The shape metric of this vertex can be described with a vector whose component is the distance from p_e to its neighboring vertices counterclockwise from the top left vertex, i.e.

$$s_e = \{d_{e1}, d_{e2}, \cdots, d_{e6}\} \quad (2)$$

where d_{ei} is the distance from p_e to p_{ei}. This vector describes the shape near this vertex. According to this metric, we can describe the shape of the four marked areas with the following vectors respectively

$$\begin{aligned} S_{mouth} &= \{s_{m1}, s_{m2}, \cdots, s_{mn}\} \\ S_{nose} &= \{s_{n1}, s_{n2}, \cdots, s_{nn}\} \\ S_{leye} &= \{s_{le1}, s_{le2}, \cdots, s_{len}\} \\ S_{reye} &= \{s_{re1}, s_{re2}, \cdots, s_{ren}\} \end{aligned} \quad (3)$$

where s_i is the shape vector of one vertex in its corresponding marked areas.

The *n*th order 1-D G-H moment $M_n(x,S(x))$ of a signal $S(x)$ is defined as [12]:

$$M_n(x) = \int_{-\infty}^{\infty} B_n(t) S(x+t)\, dt \qquad n = 0,1,2,..... \quad (4)$$

with

$$\begin{aligned} B_n(t) &= g(t,\sigma) H_n(t/\sigma) \\ H_n(t) &= (-1)^n \exp(t^2) \frac{d^n \exp(-t^2)}{dt^n} \\ g(t,\sigma) &= (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2) \end{aligned} \quad (5)$$

where $g(t, \sigma)$ is a Gaussian function and $H_n(t)$ is a scaled Hermite polynomial function of order n. G-H moments have many excellent performances, especially insensitive to noise generated during differential operations. The parameter σ and the order of G-H moments need to be determined by experiments. Here we use *1*st and *2*nd order G-H moments to analyze the shape variation when $\sigma = 2.0$.

To each shape vector in Eq.3, we calculate its *1*st and *2*nd order G-H moments, thus obtaining eight 1-D vectors, i.e. $M_{m,1}, M_{m,2}, M_{n,1}, M_{n,2}, M_{le,1}, M_{le,2}, M_{re,1}, M_{re,2}$. Each vector describes the shape variation of one marked area. After being normalized, these vectors are connected together to form one 1-D feature vector.

$$M = \{M_{m,1}, M_{m,2}, \cdots, M_{re,1}, M_{re,2}\} \quad (6)$$

We use this feature vector to describe the shape variation information of the marked areas.

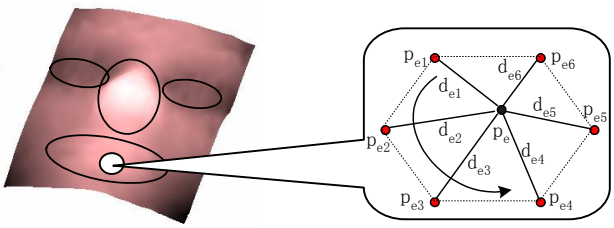


Figure 3. The marked areas and shape representation of one vertex

3.3 Feature vector

We connect the geometric vector $V_{geometric}$ and shape variation vector M together to form the feature vector

$$F = \{V_{geometric}, M\} \quad (7)$$

It not only describes the global geometric feature, but also contains the local shape variation information. In our case that the mesh model has 545 nodes, the geometric vector $V_{geometric}$ contains 545 components as well. The shape variation vector M contains 156*6*2 components since the four marked areas have 156 vertices. Thus the total number of components in the feature vector F is 2417.

4. Matching

The point cloud is represented with a regular mesh model in Section 2 and then the mesh model is characterized by a 1-D feature vector F as described in Section 3. To reduce the computational cost and improve the recognition performance, we use the principal component analysis (PCA) to obtain a lower-dimension vector and then nearest neighbor classifier (NN) is used for classification.

There exist two popular methods for linear dimensionality reduction, i.e. PCA [14] and Fisher linear discriminant (FLD) [15]. As discussed in [15], the FLD method usually needs more training samples to obtain the better result. In our case, we have limited samples so that we adopt the PCA to transform the higher-dimension vector F into the lower-dimension vector G.

We use the nearest neighbor classifier (NN) to solve the classification problem in the lower-dimension space. The similarity between two feature vectors is measured with Euclidean distance $d_i = (G - G_i)(G - G_i)^T$. Here our focus is to validate the separability of the proposed features and only use the simple classifiers. More sophisticated classifiers can be used to improve the recognition accuracy.

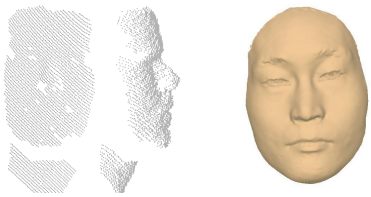
5. Experiments

To demonstrate the performance of our proposed method, we implement it on the 3D database 3D_RMA [8,9]. All these tests are finished under the hardware environment of PⅣ 1.3G CPU and 128M DRAM. The modeling experiments are done with C++ and OpenGL and the others are finished with Matlab (6.1 platform).

5.1. 3D face databases

Our proposed method is tested on the 3D face database 3D_RMA [8,9], where each face is described with a scattered 3D point cloud obtained by structured light. Compared with the data obtained from the laser scanner (one example showed in Fig.4b), these point clouds are of limited quality (Fig.4a).

The database includes 120 persons and two sessions: Nov. 97(session1) and Jan. 98 (session2). In each session, each person is sampled three shots, corresponding to central, limited left/right and up/down poses. People sometimes wear their spectacles, and beards and moustaches are also represented. Some people smile in some shots. From these sessions, two databases are built: Automatic Database (ADB, 120 persons) and Manual Database (MDB, 30 persons). The data in MDB has better quality than that in ADB. In this paper, we test our proposed method on the data set of session1, session2 and session1-2 (blending two sessions) in ADB and MDB.



(a) (b)

Figure 4. 3D data. (a) From 3D_RMA; (b) From laser scanners.

5.2. Experimental results

During our processing, the dimensionality of the original feature vector F is reduced using PCA. The dimensionality of the reduced feature vector affects the recognition rate strongly. Fig.5 describes variations of the recognition rate with the increasing dimensionality of the reduced feature vector based on the session2 set of MDB. From it, we can find that with the increase of dimensionality of the reduced feature vector, the recognition rate also rapidly increases. But when the dimensionality is up to 50 or much higher, the recognition rate nearly stabilizes at a very certain level (about 94.4%). Thus, we use only 50 features in the following experiments.

Identification accuracy is evaluated with the different sets in 3D_RMA. Considering the limited quantity of the samples, we use the strategy of Leave-one-out Cross Validation. Each time we leave one sample out as a test sample and train on the remainder. After computing the similarity differences between the test sample and the training data using Euclidean distance, the nearest neighbor (NN) is then applied to classification. To validate the effectiveness of the shape variation information, we estimate the recognition rate using the feature vector $V_{geometric}$ only including global geometric features (GGF) and the feature vector F containing GGF and the shape variation information (GGF+SVI). Table 1

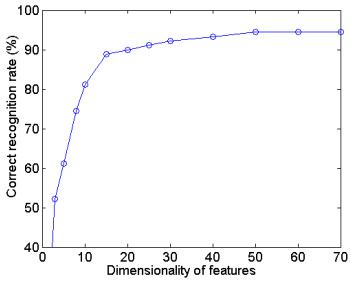


Figure 5. Recognition performance under different dimensionality of features

Table 1. CCR in 3D_RMA (%)

Database	GGF	GGF+SVI
Manual DB, session1 (30 persons, 3 instances for each)	92.2	95.6
Manual DB, session2 (30 persons, 3 instances for each)	84.4	94.4
Manual DB, session1-2 (30 persons, 6 instances for each)	93.9	96.1
Automatic DB, session1 (120 persons,3 instances for each)	59.2	66.9
Automatic DB, session2 (120 persons,3 instances for each)	59.2	66.7
Automatic DB, session1-2 (120 persons,6 instances for each)	69.4	72.4

summarizes the Correct Classification Rate (CCR) with these two different manners.

In addition, we use one more familiar method, Cumulative Match Score (CMS) [16], to evaluate the identification performance of GGF+SVI. Fig.6 shows the CMS curves using the NN classifier on three data sets of MDB. In fact, the CCR is equal to the case that Rank=1.

We test the verification performance using leave-one-out scheme as well. On each test, one is the probe sample and the remaining samples are trained. The probe sample is classified with the training set. In each iteration, there is only one true test since we know the classification of the probe sample. Fig. 7 shows the ROC curves for different data sets in MDB.

From an overall view of Table 1, Fig.6-7, we can draw the following conclusions:

a) The highest recognition is up to 96.1% (30 persons) and 72.4% (120 persons). Although the testing database is not big enough, this result is obtained in the fully automatic way, which is fairly encouraging.

b) Shape variation is the important information to characterize the individual. The feature vector containing the shape variation information improves the CCR distinctly (see Table 1).

c) The increase of the training samples can improve the verification and identification performance (see Table 1 and Fig.6-7). The sets blending two sessions always have better performance.

d) Noise and volume of the tested database affect the CCR strongly. In Table 1, the CCR in ADB is lower than that in MDB.

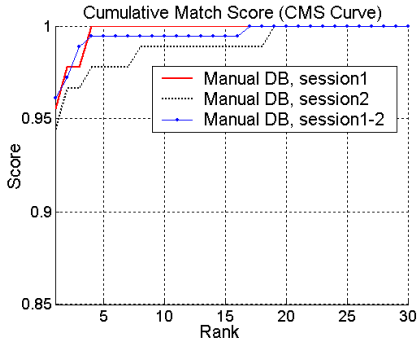


Figure 6. CMS curves for identification performance

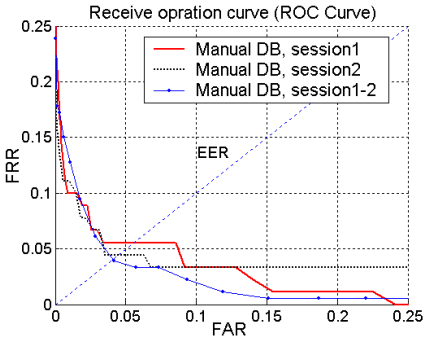


Figure 7. ROC curves for verification performance.

5.3. Comparisons and discussions

We make detailed comparisons with some existing methods for 3D face recognition to show the performance of the proposed algorithm.

(1) Our method has a lower computational cost. Our modeling process costs more time (about 2s for each) and the feature extraction costs about 0.2s. However, the matching process costs little time due to only calculating the Euclidean distance between two points in a lower-dimensional space. Beumier *et al.* [8] built the database 3D_RMA and developed surface matching (SURF) and central/lateral profiles (CLP) to realize face authentication. Their reported verification performances of automatic matching algorithms were close to ours as shown in Table 2. However, their matching process was an optimization process, which incurred a high computational cost (at least 0.5s for each matching). The less matching cost means that it takes less time to search for the corresponding object in a big database.

Blanz *et al.* [4,5] developed an excellent method to fit a 3D deformable model to the image and used the obtained shape and texture coefficients for face recognition. Due to the largely different background, it is no sense to compare our recognition performances directly. But it should be noted that their fitting procedure is slow and requires some manual interaction.

(2) Our algorithm is tested on a bigger and more

complex database. Gordon [2] obtained the higher recognition rate (100%) using depth and curvature features since they adopted a small database (only 8 persons) with high-quality range data (similar to Fig.4b) and without eyeglasses, beards or pose variations.

Table 2. EER of our algorithm and automatic matching algorithms in Beumier *et al.* [8] on manual DB

Algorithms	Session1	Session2	Session1-2
Beumier [8], SURF	8.0%	7.0%	13.0%
Beumier [8], CLP	4.75%	6.75%	7.25%
Ours	5.5%	4.5%	4.0%

Chua *et al.* [6] used the rigid region to characterize the individual in order to conquer the influence of the expressions. They tested their algorithm with only six objects (four expressions for each, without pose variations) and obtained promising results. Our algorithm is performed on 3D_RMA, which contains 120 persons with different quality and limited pose and expression variations.

(3) From the analysis in Section 2, our modeling scheme can overcome the pose variation effectively. However, the feature extraction is strongly influenced by expressions. Chua *et al.* [6] had done some work to deal with the expression variation on a small database. This is also our focused direction to further improve the recognition accuracy in the future work. In addition, some mesh models cannot describe the correct shape of the real person due to noise, thus resulting in the false feature vector. So increasing the precision of mesh model is another avenue to improve the recognition performance.

In Eq.7, we connect the geometric vector $V_{geometric}$ and shape variation vector M directly. Intuitively, it is not the best way to fuse them since they have different properties. In the future, we can carry out a deep research on this topic and obtain a better fusing algorithm.

6. Conclusions

In this paper, we have proposed a new scheme for 3D face recognition. In this scheme, the 3D point cloud is first represented with a regular mesh using hierarchical mesh fitting. Based on the observation that local shape variation is the important information to characterize the individual, our feature vector is constructed by combining the global geometric features and the local shape variation information together. We test the proposed algorithm on 3D_RMA and the encouraging results show the importance of the shape variation to characterize the individual. Compared with previous works, our algorithm demonstrates its outstanding performance. In the future, we will focus on searching for some invariant features to expressions and build larger 3D database for estimating the performance of the 3D recognition algorithm.

Acknowledgements

This work is supported by research funds from the Natural Science Foundation of China (Grant No. 60121302 and 60332010) and the Outstanding Overseas Chinese Scholars Fund of CAS (No.2001-2-8). We would like to thank Dr. C. Beumier for the database 3D_RAM. Also thanks to our colleagues for constructive suggestions.

References

[1] J.C. Lee, and E. Milios, "Matching Range Images of Human Faces", *Proc. ICCV'90*, pp.722-726, 1990.

[2] G.G. Gordon, "Face Recognition Based on Depth and Curvature Features", *Proc. CVPR'92*, pp.108-110, 1992.

[3] Y. Yacoob and L.S. Davis, "Labeling of Human Face Components from Range Data", *CVGIP: Image Understanding*, 60(2):168-178, 1994.

[4] V. Blanz and T. Vetter, "Face Identification Based on Fitting a 3D Morphable Model", *IEEE Trans. on PAMI,* Vol.25, No.9, pp.1063-1074, 2003.

[5] V. Blanz, S. Romdhani and T. Vetter, "Face Identification across Different Poses and Illumination with a 3D Morphable Model", *Proc. FG'02*, pp.202-207, 2002.

[6] C.S. Chua, F. Han, and Y.K. Ho, "3D Human Face Recognition Using Point Signiture", *Proc. FG'00*, pp.233-239, 2000.

[7] A.M. Bronstein, M.M. Bronstein, and R. Kimmel, "Expression-Invariant 3D Face Recognition", Auto- and Video- Based Person Authentication (AVBPA 2003), LCNS 2688, pp.62-70, 2003.

[8] C. Beumier and M. Acheroy, "Automatic 3D Face Authentication", *Image and Vision Computing*, 18(4):315-321, 2000.

[9] C. Beumier and M. Acheroy, "Automatic Face Authentication from 3D Surface", *Proc. BMVC*, pp.449-458, 1998.

[10] C. Xu, Y. Wang, T. Tan, and L. Quan, "A New Attempt to Face Recognition Using 3D Eigenfaces", *ACCV'04*, pp.884-889, 2004.

[11] S. Liao, M. Pawlak, "On Image Analysis by Moments", *IEEE Trans. on PAMI*, Vol.18, No.3, pp.254-266, 1996.

[12] J. Shen, W. Shen and D. Shen, "On Geometric and Orthogonal Moments", *Inter. Journal of Pattern Recognition and Artificial Intelligence*, Vol.14, No.7, pp.875-894, 2000.

[13] C. Xu, L. Quan, Y. Wang, T. Tan, M. Lhuillier, "Adaptive Multi-resolution Fitting and its Application to Realistic Head Modeling", *Geometric Modeling and Processing 2004*, to be appeared.

[14] M. Turk, and A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, 3(1): 71-86, 1991.

[15] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Trans. on PAMI*, Vol.19, No.7, pp.711-720, 1997.

[16] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The Feret Evaluation Methodology for Face-Recognition Algorithm", *IEEE Trans. on PAMI*, Vol.22, No.10, pp.1090-1104, 2000.