Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Sample based 3D face reconstruction from a single frontal image by adaptive locally linear embedding*

ZHANG Jian, ZHUANG Yue-ting‡

(*School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*)

E-mail: zhangsdust@yahoo.com.cn; yzhuang@cs.zju.edu.cn

Received Feb. 10, 2006; revision accepted Nov. 28, 2006

Abstract: In this paper, we propose a highly automatic approach for 3D photorealistic face reconstruction from a single frontal image. The key point of our work is the implementation of adaptive manifold learning approach. Beforehand, an active appearance model (AAM) is trained for automatic feature extraction and adaptive locally linear embedding (ALLE) algorithm is utilized to reduce the dimensionality of the 3D database. Then, given an input frontal face image, the corresponding weights between 3D samples and the image are synthesized adaptively according to the AAM selected facial features. Finally, geometry reconstruction is achieved by linear weighted combination of adaptively selected samples. Radial basis function (RBF) is adopted to map facial texture from the frontal image to the reconstructed face geometry. The texture of invisible regions between the face and the ears is interpolated by sampling from the frontal image. This approach has several advantages: (1) Only a single frontal face image is needed for highly automatic face reconstruction; (2) Compared with former works, our reconstruction approach provides higher accuracy; (3) Constraint based RBF texture mapping provides natural appearance for reconstructed face.

Key words: Face reconstruction, Manifold learning, RBF interpolation, Reconstruction error rate
doi:10.1631/jzus.2007.A0550 **Document code:** A **CLC number:** TP391

INTRODUCTION

Realistic 3D face synthesis technique has improved tremendously in the last three decades and plays an important role in the fields of virtual conference, film-making and game-entertainment. Prevalent photorealistic 3D face synthesis techniques are based on images or videos, with the differences between these techniques resting on whether 3D reconstruction work is done with the support of a 3D face sample base or not.

The most impressive work was proposed by Blanz and Vetter (1999). Utilizing a 3D face sample base, they derived a morphable face model by transforming the shape and texture of the examples into vector space representation. New faces and expressions can be well modelled by forming linear combinations of the examples. However, the computation of correspondences and face model parameters is burdensome work. According to (Blanz and Vetter, 1999), the whole procedure was performed on SGI R10000 processor, with computation time being 50 min. Zhang *et al*.(2003) developed a geometry-driven facial expression synthesis system which can generate 3D photorealistic facial ex pressions through blending sub-region texture images according to the facial feature positions, although this work may need many interactive steps during blending. Stereo vision reconstruction method is also used when two or multiviews of a human subject are available. Pighin *et al*.(1998) presented a framework to create photo-realistic textured 3D facial models from several uncalibrated views of a human subject. A few feature points are chosen manually from these images to

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 60533090, 60525108), the National Basic Research Program (973) of China (No. 2002CB312101), the Science and Technology Project of Zhejiang Province, China (Nos. 2005C13032, 2005C11001-05), and the China-US Million Book Digital Library Project

recover the camera poses as well as the 3D coordinates of the feature points on the subject's face. Then, scattered data interpolation is adopted to deform a generic face mesh to obtain the personalized geometry of the subject's face. Similar technique was utilized by Park *et al*.(2004), whose reconstruction work is based on only frontal and profile images. Multiview reconstruction needs two or more informative views which are sometimes not available. Shape from shading (Zhang *et al*., 1999) has been explored by many researchers. Since face image is subject to illumination, camera pose and face geometry, it is very hard to recover a human face's 3D shape without prior knowledge. Besides image based techniques, recovering 3D face from video sequences is another focus in computer vision domain. It is closely related to one sort of problem, i.e. "Shape from Motion" or "Structure from Motion" (Tomasi and Kanade, 1992; Sclaroff and Alon, 1999). In (Torresani and Hertzmann, 2003; Dimitrijevic *et al*., 2004; Xiao and Kanade, 2004), reconstruction of non-rigid shape (including human face) is achieved by use of shape from motion (SFM) technique from uncalibrated video sequences. With the help of a 3D face database, Dimitrijevic *et al*.(2004) obtained good result from video sequence, although the correspondence between a pair of images should be established in advance. Zhao *et al*.(2006) proposed using the 2D morphable shape model to estimate the 3D face model and extract texture from multiple images. Yoshiki *et al*.(2006) reconstructed 3D face model from a single frontal shading face image by using a 3D face database based on the optimization of the error value. Reiter *et al*.(2006) adopted Canonical Correlation Analysis to build 3D face depth maps and predicted near-infrared face texture according to the input RGB face image. These works indicate that 3D face modelling is still an open research area in computer vision and graphics.

In this paper, we present a highly automatic photorealistic face reconstruction approach based on 3D face samples from one frontal image. The novelty lies in the implementation of adaptive manifold learning technique during the analysis and synthesis procedure. The approach needs little human interaction. Beforehand, the proposed adaptive locally linear embedding (ALLE) algorithm is adopted to reduce the dimensionality of 3D face samples, with this process being done offline. Given an input image, first, a number of facial feature points are automatically aligned by a well trained active appearance model (AAM); then, the low-dimensional coordinate of the input face is synthesized by that of the 3D samples, which are adaptively selected according to the aligned facial features; after that, the neighbors as well as the reconstruction weight of input face image in the low-dimensional sub-space are adaptively chosen, then personalized 3D face is reconstructed according to the weight and corresponding neighbors in the original sample space; at last, radial basis function (RBF) technique is adopted to do constraint based texture mapping and obtain fine realistic 3D face from a single frontal image. Experiment showed that manifold learning technique is superior to principal component analysis (PCA) method (Turk and Pentland, 1991) adopted by Hu *et al*.(2004) in face reconstruction.

The following contents are organized as follows. In Section 2, several related key techniques used in the reconstruction are briefly introduced. In Section 3, we discuss in detail the approach to reconstruct 3D personalized face geometry from image. In Section 4, a constraint based RBF texture mapping method is given. Section 5 provides face reconstruction experimental results as well as comparison with the former methods. Section 6 concludes this paper.

RELATED TECHNIQUES

In this section, we will briefly introduce several key techniques used to achieve the face reconstruction.

Adaptive manifold learning

Manifold learning techniques can be regarded as a class of nonlinear dimensionality reduction methods which are superior to traditional linear subspace methods such as PCA and ICA (independent component analysis). Manifold learning uncovers the most intrinsic characteristics of high-dimensional observations distributing in a non-linear subspace. Generally used manifold learning methods consist of ISOMAP (Tenenbaum *et al*., 2000), LLE (Roweis and Saul, 2000), LTSA, etc. An important extension of manifold learning algorithms was proposed in

(Wang *et al*., 2005), which is to adaptively select the neighborhood size in the *k*-nearest neighbor computation to construct the local connectivity. Adaptive neighborhood selection is initially utilized in LTSA, and as it is independent of specific learning algorithm, it can be combined with other manifold learning algorithms.

Locally linear embedding (LLE)

LLE is an unsupervised manifold learning algorithm that computes low-dimensional, neighborhood preserving embeddings of high-dimensional input (Roweis and Saul, 2000). According to LLE, each high-dimensional datapoint can be reconstructed by linear weighted combination of its neighbors. Essentially, the reconstruction weight reflects intrinsic geometric properties of the data that are invariant when high-dimensional datapoints are transformed to low-dimensional coordinates. Since the neighbor datapoints cannot be adaptively selected, traditional LLE is not suitable for non-rigid face reconstruction because each datapoint is dynamically influenced by different numbers of neighbor points. In this paper, traditional LLE method is combined with the adaptive neighborhood selection method to build a novel "Adaptive LLE (ALLE)" algorithm which contributes much to face reconstruction.

Radial basis function (RBF) network

RBF network is a type of artificial neural network for application to problems of supervised learning (e.g., regression, classification and time series prediction). RBF network is well known for its global and optimal approximation performance. Owing to its good performance, RBF network has been widely used in the fields of computer graphics and computer vision (Arad *et al*., 1994; Turk and O'Brien, 1999; Huang *et al*., 2002). In this paper, we propose a constraint based RBF texture mapping method after the reconstruction of face geometry. Single frontal image can be smoothly mapped to the face geometry by use of RBF network.

3D RECONSTRUCTION OF FACE GEOMETRY

In this section, we present in detail the sample learning based 3D face geometry reconstruction approach. The approach consists of two main parts: (1) dimensionality reduction of 3D face samples by adaptive manifold learning approach; (2) automatic reconstruction of 3D personalized face geometry initialized by a few automatically selected feature points from the given face image.

Dimensionality reduction by ALLE

To support the reconstruction work, we use 3D head samples generated by 3D sculpture software FaceGen Modeller. The database includes 55 males and 45 females, ranging in age from 25 to 50, 80 are eastern Asian and 20 are Caucasian. After preprocessing for good performance, each head model has 6174 vertices which constitute 6054 quadrangles. What was more, the topology relationships of these vertices as well as the point by point correspondences between different head samples are already known during the preprocessing.

According to our work, personalized 3D face reconstruction needs not only the features selected from 2D image but also the statistical information from the 3D face samples. Each 3D face sample could be seen as a high-dimensional datapoint in nonlinear embedded subspace. Manifold learning technique contributes to learning corresponding low-dimensional coordinate in a manifold space. These low-dimensional data reflect the most intrinsic properties of the original samples. LLE algorithm indicates that high-dimensional data and low-dimensional coordinates can be reconstructed by their neighbors with the same weight (Roweis and Saul, 2000). So, once the low-dimensional coordinate of the image subject can be synthesized by that of the samples, corresponding high-dimensional data can be reconstructed by linear combination of these samples. According to (Wang *et al*., 2005), the effectiveness of manifold learning depends on the manner in which the nearby neighborhoods overlap with each other. However, because each face is a datapoint lying in a nonlinear embedded subspace, selecting its neighborhood size appropriately by hand is very difficult. So we use ALLE algorithm to perform an offline dimensionality reduction process on the 3D face samples. ALLE is used to adaptively select the neighborhood size of each face sample. Essentially, the adaptive neighborhood selection algorithm is an iterative process of optimization. Two parts are included in the algorithm, that is, neighborhood contraction and neighborhood expansion (Wang *et al*., 2005). The detailed implementation is described below.

(1) Neighborhood contraction

Step 1: Empirically choose the initial neighborhood size K and K-NN neighbor $\boldsymbol{X}_i^k=[\boldsymbol{x}_{i1},\dots,\boldsymbol{x}_{ik}]$ for each data $\boldsymbol{x}_i$, ordered in non-decreasing distances to $\boldsymbol{x}_i$. Assign K to k.

Step 2: Subtract the mean value: $\boldsymbol{X}'=\boldsymbol{X}_i^k-\bar{\boldsymbol{X}}_i^k$; compute the d largest singular vectors $\boldsymbol{Q}_i^k$ of $\boldsymbol{X}'$ corresponding to d non-zero singular values and project $\boldsymbol{X}'$ to $\boldsymbol{Q}_i^k$: $\boldsymbol{E}_i^k=(\boldsymbol{Q}_i^k)^{\mathrm{T}}\boldsymbol{X}'$.

Step 3: η is constant;

If $\|\boldsymbol{X}'-\boldsymbol{Q}_i^k\boldsymbol{E}_i^k\|_{\mathrm{F}}<\eta\|\boldsymbol{E}_i^k\|_{\mathrm{F}}$ then

$\boldsymbol{X}_i=\boldsymbol{X}_i^k$; $\boldsymbol{E}_i=\boldsymbol{E}_i^k$; end.

Step 4: Assume $0<k_0<1$.

If $k>d+k_0$ then

delete the last column of $\boldsymbol{X}_i^k$ to get $\boldsymbol{X}_i^{k-1}$;

set $k=k-1$ and return to Step 2;

else go to Step 5.

Step 5: Let

$k=\arg\min_{d+k_0\le j\le k}(\|\boldsymbol{X}_i^j-\bar{\boldsymbol{X}}_i^j-\boldsymbol{Q}_i^j\boldsymbol{E}_i^j\|_{\mathrm{F}}/\|\boldsymbol{E}_i^j\|_{\mathrm{F}})$ and set $\boldsymbol{X}_i=\boldsymbol{X}_i^k$, $\boldsymbol{E}_i=\boldsymbol{E}_i^k$;

Then $\boldsymbol{X}_i$ represents the contracted neighbors of $\boldsymbol{x}_i$.

(2) Neighborhood expansion

Set k_i to be the contracted neighborhood size, $\boldsymbol{x}_{i_j}$ is the jth neighbor of $\boldsymbol{x}_i$, $\bar{\boldsymbol{x}}_i$ is the column mean of $\boldsymbol{X}_i$, $\boldsymbol{Q}_i$ represents the singular vectors.

for $j=k_i+1, \dots, K$

compute $\boldsymbol{e}_j^i=\boldsymbol{Q}_i^{\mathrm{T}}(\boldsymbol{x}_{i_j}-\bar{\boldsymbol{x}}_i)$;

if $\|(\boldsymbol{I}-\boldsymbol{Q}_i\boldsymbol{Q}_i^{\mathrm{T}})(\boldsymbol{x}_{i_j}-\bar{\boldsymbol{x}}_i)\|_2\le\|\boldsymbol{e}_j^i\|_2$, then

expand $\boldsymbol{X}_i$ by adding $\boldsymbol{x}_{i_j}$;

end; end;

In brief, the above adaptive neighborhood selection algorithm is validated by minimizing the ratio as objective function:

$$\|\boldsymbol{X}_i-\bar{\boldsymbol{X}}_i-\boldsymbol{Q}_i\boldsymbol{E}_i\|_{\mathrm{F}}/\|\boldsymbol{E}_i\|_{\mathrm{F}}.$$

Given 3D face geometry, it can be described as a shape vector: $\boldsymbol{S}=[X_1,Y_1,Z_1,X_2,Y_2,Z_2,\dots,X_n,Y_n,X_n]^{\mathrm{T}}\in\mathbb{R}^{3n}$. X_i, Y_i, Z_i ($i=1, \dots, n$) represent the X, Y, Z coordinates of the face vertices respectively. Thus, the N 3D face samples can be represented as:

$$\boldsymbol{Sp}=\begin{bmatrix} X_{1,1} & Y_{1,1} & Z_{1,1} & \dots & X_{1,n} & Y_{1,n} & Z_{1,n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_{N,1} & Y_{N,1} & Z_{N,1} & \dots & X_{N,n} & Y_{N,n} & Z_{N,n} \end{bmatrix}^{\mathrm{T}}.$$

Through ALLE, we get low-dimensional coordinate of each face sample:

$$\boldsymbol{Y}_{\mathrm{origin}}=ALLE(\boldsymbol{Sp},numk,d),$$

where $\boldsymbol{Y}_{\mathrm{origin}}\in\mathbb{R}^{d\times N}$ will be used to synthesize the low-dimensional coordinate of the new face subject, $\boldsymbol{Sp}\in\mathbb{R}^{3n\times N}$ is the matrix form of 3D face samples, $numk\in\mathbb{R}^N$ means adaptively selected neighborhood size of all samples, d is the size of the low-dimensional coordinate. This process is done offline, then we can perform the reconstruction process efficiently based on the dimensionality reduced samples.

Automatic face reconstruction

After dimensionality reduction of 3D samples, given an input image, the feature points to be selected from the 2D image can be expressed as a feature vector:

$$\boldsymbol{Fp}=[X_1,Y_1,X_2,Y_2,\dots,X_m,Y_m]^{\mathrm{T}}\in\mathbb{R}^{2m},$$

where m means the number of feature points. Then the reconstruction process can be performed in three steps.

Step 1: Automatically align facial features $\boldsymbol{Fp}$ from the input image by AAM, the facial features are shown in Fig.1b.

(a) (b)

Fig.1 (a) Input frontal face image of one subject; (b) Automatically aligned face image for reconstruction

After translation, rotation and scaling, $\boldsymbol{Fp}$ is transformed into sample space. Transformed image feature points are denoted as $\boldsymbol{Fp}'$. From $\boldsymbol{Fp}'$, we extract X, Y coordinates of m feature vertices from all face samples. This is done by measuring the Euclidean distances between $\boldsymbol{Fp}'$ and vertices of each face sample. The extracted features are represented as subspace of the original face samples:

$$\boldsymbol{S}_f = \begin{bmatrix} X_{1,1} & \dots & X_{N,1} \\ Y_{1,1} & \dots & Y_{N,1} \\ \dots & \dots & \dots \\ X_{1,m} & \dots & X_{N,m} \\ Y_{1,m} & \dots & Y_{N,m} \end{bmatrix} \in \mathbb{R}^{2m\times N}.$$

Then, we use adaptive neighborhood selection algorithm to select k_i samples from $\boldsymbol{S}_f$ which best reconstruct $\boldsymbol{Fp}'$. These samples are denoted as $\boldsymbol{S}_i \in \mathbb{R}^{2m\times k_i}$.

$$[\boldsymbol{S}_i, K_i] = AdaptivelySelectNeighbor([\boldsymbol{Fp}', \boldsymbol{S}_f]).$$

Step 2: Using LLE algorithm to compute weight $\boldsymbol{W}_{\text{extract}}$ for reconstruction of $\boldsymbol{Fp}'$ by $\boldsymbol{S}_i$. That is, $\boldsymbol{W}_{\text{extract}}$ satisfies $\boldsymbol{Fp}' = \boldsymbol{S}_i \times \boldsymbol{W}_{\text{extract}}$.

The computation of weight can be described as:

$$[\boldsymbol{W}_{\text{extract}}, \boldsymbol{Y}_{\text{extract}}] = LLE([\boldsymbol{Fp}', \boldsymbol{S}_i], k_i, d),$$

where k_i is the neighborhood size computed in Step 1, d is the size of low-dimensional coordinate, $\boldsymbol{W}_{\text{extract}} \in \mathbb{R}^{k_i}$ denotes reconstruction weight.

Then, given $\boldsymbol{S}_i$, we find the corresponding low-dimensional coordinates $\boldsymbol{Y}'_{\text{origin}} \in \mathbb{R}^{d\times k_i}$ from $\boldsymbol{Y}_{\text{origin}}$ through one by one correspondence. Thus, the low-dimensional coordinate of the input image subject can be synthesized by linear weighted combination of $\boldsymbol{Y}'_{\text{origin}}$:

$$\boldsymbol{Y}_{\text{recon}} = \boldsymbol{Y}'_{\text{origin}} \times \boldsymbol{W}_{\text{extract}}.$$

In Step 3, $\boldsymbol{Y}_{\text{recon}}$ will be used to reconstruct personalized 3D face geometry from image.

Step 3: Adaptively select K_r coordinates from $\boldsymbol{Y}_{\text{origin}}$ which best reconstruct $\boldsymbol{Y}_{\text{recon}}$:

$$[\boldsymbol{Y}_r, K_r] = AdaptivelySelectNeighbor([\boldsymbol{Y}_{\text{recon}}, \boldsymbol{Y}_{\text{origin}}]).$$

These coordinates are represented as $\boldsymbol{Y}_r \in \mathbb{R}^{d\times K_i}$. Similarly, given $\boldsymbol{Y}_r$, we can find the corresponding original face samples $\boldsymbol{S}_r \in \mathbb{R}^{3n\times K_r}$ because of the one by one correspondence between high- and low-dimensional data.

Now, we use LLE algorithm with neighborhood size K_r to get the reconstruction weight $\boldsymbol{W}_{\text{recon}} \in \mathbb{R}^{K_r}$ which satisfies $\boldsymbol{Y}_{\text{recon}} = \boldsymbol{Y}_r \times \boldsymbol{W}_{\text{recon}}$.

The computation of weight is described as:

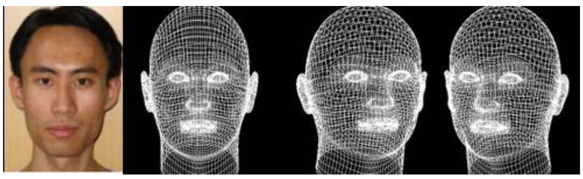
$$[\boldsymbol{W}_{\text{recon}}, \boldsymbol{Y}'] = LLE([\boldsymbol{Y}_{\text{recon}}, \boldsymbol{Y}_{\text{origin}}], K_r, d).$$

According to (Roweis and Saul, 2000), $\boldsymbol{W}_{\text{recon}}$ is also suitable for reconstructing high-dimensional datapoints in nonlinear space. Thus, the personalized 3D face geometry $\boldsymbol{S}_n \in \mathbb{R}^{3n\times 1}$ is reconstructed from a single image.

$$\boldsymbol{S}_n = \boldsymbol{S}_r \times \boldsymbol{W}_{\text{recon}}.$$

The key point of our work is to automatically select the most appropriate samples $\boldsymbol{S}_r$ from the database and compute the most appropriate weight $\boldsymbol{W}_{\text{recon}}$ to reconstruct personalized 3D face geometry $\boldsymbol{S}_n$.

The reconstructed 3D face geometry is shown in Fig.2.



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

Fig.2 Reconstructed face geometry from a single frontal image. (a) Input image; (b), (c), (d) are reconstructed face geometries from three different views

Though the above adaptive manifold subspace learning reconstruction can be performed by manually setting the reconstruction neighborhood size alternatively, the adaptive method performs better than the latter. A diagram showed in Fig.3 could be helpful in understanding this section.

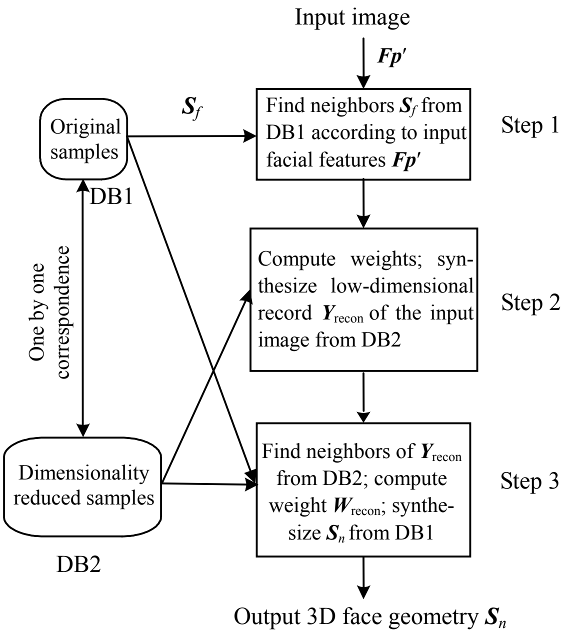


Fig.3 Diagram for automatic face reconstruction

CONSTRAINT BASED RBF TEXTURE MAPPING

By nature, texture mapping is a parameterized process of curved surface, namely, finding the corresponding 2D texture coordinates of 3D vertices:

$$S(x_i, y_i, z_i) = (u_i(x_i, y_i, z_i), v_i(x_i, y_i, z_i)), \quad i=1, \ldots, n,$$

where S is the mapping function, (x_i, y_i, z_i) is the 3D coordinate of a vertex, (u_i, v_i) is texture coordinate of a vertex, n is the number of vertices.

In order to build the photorealistic human face, we first manually assign texture coordinates to some feature vertices of newly reconstructed face geometry. The correspondence between feature coordinates and feature vertices is defined offline as constraints. The texture coordinates can be obtained by our previously built AAM and are described as $\boldsymbol{T}_i\{u_i, v_i\}_{i=1}^{n} \subset \mathbb{R}^2$, the corresponding feature vertices on the reconstructed face geometry are depicted as $\boldsymbol{P}_i\{x_i, y_i, z_i\}_{i=1}^{n} \subset \mathbb{R}^3$. To obtain a mapping function from vertices to image coordinates, we use $\boldsymbol{T}$ and $\boldsymbol{P}$ to train RBF network:

$$S(\boldsymbol{P}) = \sum_{i=1}^{n} \alpha_i \phi(\boldsymbol{P} - \boldsymbol{P}_i) = \boldsymbol{T}.$$

Once α is known, we obtain the texture coordinates of other vertices by RBF interpolation:

$$\boldsymbol{T}_{\text{new}} = \sum_{i=1}^{n} \alpha_i \phi(\boldsymbol{S}_{\text{new}} - \boldsymbol{S}_{\text{new}\,i}),$$

where $\boldsymbol{S}_{\text{new}} = [X_1, Y_1, Z_1, X_2, Y_2, Z_2, \ldots, X_n, Y_n, Z_n]^{\text{T}} \in \mathbb{R}^{3n}$ is the reconstructed 3D face geometry using the method discussed in Section 3.

Texture information between face and ears is unavailable in the frontal image. We solve this problem by sampling from face margin region on the frontal image for RBF interpolation. The texture mapped face as shown in Fig.4 looks fine and natural.

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

Fig.4 Texture mapped 3D face from different views

EXPERIMENTS AND DISCUSSIONS

We implement our reconstruction algorithm on a Pentium IV 2.4 GHz PC with 512 M RAM. To reconstruct the face geometry, 52 feature points on the input image are aligned by a well trained AAM and the system adaptively selects the optimized neighborhood size. In terms of texture mapping, the feature points also act as constraints. We choose multi-quadric function as basis function and implement the reconstruction process on the arbitrary images. The reconstruction results using adaptive algorithm are depicted in Fig.5.

To verify the superiority of ALLE algorithm, we implement the "cross validation" process (randomly pick one out of the training data) 10 times and for each time we perform two sets of experiments. In the first 5 experiments, the reconstruction neighborhood size is manually assigned as 4, 5, 10, 15 and 20 respectively. In the second experiment, the reconstruction neighborhood size is adaptively selected by algorithm. Reconstruction error is defined as the average deviation of reconstructed vertices from the ground truth of the test sample. Thus, reconstruction error rate can be computed as the ratio of recon-

Fig.5 Reconstructed multiview 3D faces with adaptive manifold learning

struction error and the ground truth. The average reconstruction error rate of "cross validation" tests are shown in Fig.6.

We learn from Fig.6 that the reconstruction error rate amounts to 3.3888% when 4 neighbors are used for reconstruction, but it drops dramatically when 5 neighbors take part in the reconstruction; when the number of neighbors surpasses 10, the error rate remains stable at about 2.43%. However, when the number of neighbors is automatically selected by the adaptive algorithm, the error rate remains at a lower value of 1.756%. In our experiment, the neighborhood size for each of the 10 cross-validation tests is adaptively selected as 8, 23, 25, 29, 30, 15, 22, 28, 32, 10 respectively, and the corresponding reconstruction error rates are 2.486%, 2.029%, 1.458%, 1.252%, 1.027%, 2.43%, 2.041%, 1.364%, 1.043%, 2.43% with the average value of 1.756%. This means that to our 3D sample set, the optimized neighborhood size surpasses 22 in most cases, and manual selection is inadequate to determine the exact neighborhood size. Note that the adaptive neighborhood size selection algorithm does not necessarily work better than the manual method in every experiment, the superiority lies in that it offers an elegant way to take the place of manual work and ensures better performance in most cases (if one happens to select the right neighborhood size each time, manual work is as effective as the adaptive algorithm).

The most relevant work is proposed by Hu *et al.*(2004), where PCA method was utilized to obtain the shape eigenvectors, and then 3D reconstruction was well done by computing the coefficients of eigenvectors. The superiority of our algorithm lies in the manifold learning technique. We also use PCA to perform 10 "cross validation" reconstructions based on our samples. The average reconstruction error rate is shown in Fig.7 with different numbers of eigenvectors.

Fig.7 indicates that the reconstruction error rate

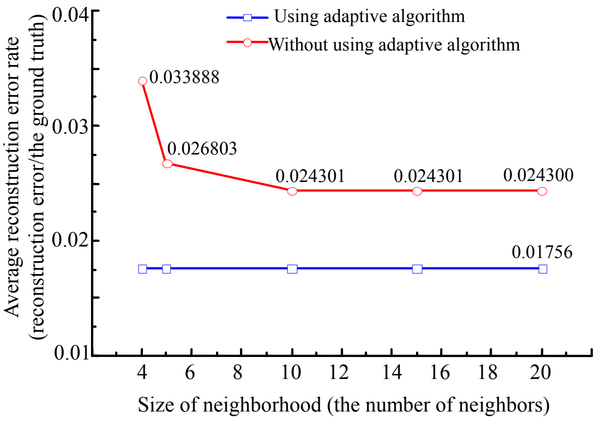


Fig.6 Comparison of average reconstruction error rates by using adaptive algorithm or not

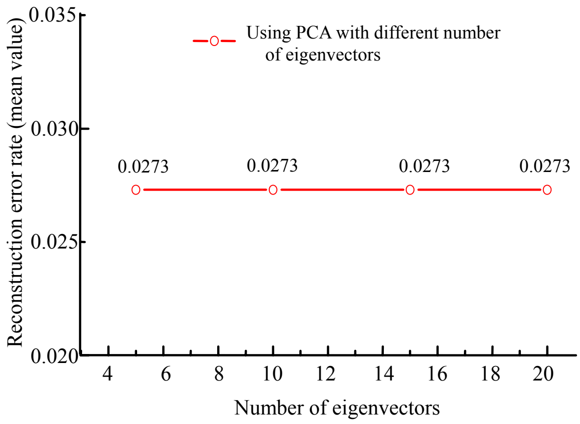


Fig.7 Average reconstruction error rate using PCA. The number of the eigenvectors is empirically selected

under PCA is not sensitive to the number of eigenvectors. Though we adjust the number of eigenvectors from 5 to 20, the error rate remains at about 2.73%. The reason mainly lies in two points: on one hand, PCA is a kind of linear approach designed to extract the principal features hidden in the sample set and eliminate the inter-correlations between sample data; on the other hand, our sample set contains only 3D face meshes with point-to-point correspondence, and similar structure reduces the diversity of sample data distribution. In our sample set, 5 eigenvectors are enough to capture the most principal features, so the reconstruction error rates in Fig.7 saturate as we adjust the number of eigenvectors from 5 to 20. Comparing Fig.7 with Fig.6, we find that with manifold learning, only when reconstruction neighbors are less than 5, the accuracy of our method is inferior to PCA. In any other case, our reconstruction method performs better than PCA based method in terms of reconstruction accuracy under the same experimental condition.

Owing to the computational convenience, PCA reconstruction with 4 eigenvectors (including texture mapping and rendering) lasts about 10 s. The proposed reconstruction work without adaptive algorithm is done in 20 s. When adaptive algorithm is adopted, the time needed is no more than 1 min.

CONCLUSION

In this paper, we propose a highly automatic 3D face reconstruction approach. The highlight of our work is the adoption of adaptive manifold learning algorithm which provides higher accuracy compared with other related works. In brief, the approach has the following advantages: (1) Only a single frontal face image is required for face reconstruction and this is easily satisfied; (2) The system is highly automatic; (3) The accuracy of reconstruction is higher than that of other 3D reconstruction approachs; (4) Constraint based RBF texture mapping provides natural appearance for reconstructed face. Future work will focus on two points: (1) Enlarging the 3D database in terms of facial expression. With more prior information, we could synthesize parameterized 3D face. Then, different potential facial expressions of a subject could be synthesized by adjusting the parameters. (2) Expression reconstruction. Given multiple images or a video sequence with expression changes, with the help of enlarged 3D database, we intend to recover personalized 3D face with realistic facial expressions.

References

Arad, N., Dyn, N., Reisfeld, D., Yeshurun, Y., 1994. Image warping by radial basis functions: application to facial expressions. *CVGIP: Graphical Models and Image Processing*, **56**(2):161-172. [doi:10.1006/cgip.1994.1015]

Blanz, V., Vetter, T., 1999. A Morphable Model for the Synthesis of 3D Faces. Proc. SIGGRAPH'99. Los Angeles, p.187-194.

Dimitrijevic, M., Ilic, S., Fua, P., 2004. Accurate Face Models from Uncalibrated and ILL-Lit Video Sequences. Proc. 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **2**:1034-1041. [doi:10.1109/CVPR.2004.1315278]

Hu, Y., Jiang, D., Yan, S., Zhang, L., Zhang, H., 2004. Automatic 3D Reconstruction for Face Recognition. Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition, p.843.

Huang, L., Shimizu, A., Kobatake, H., 2002. Face Detection Using a Modified Radial Basis Function Neural Network. Proc. IEEE International Conference on Pattern Recognition, **2**:342-345.

Park, K., Zhang, H., Vezhnevets, V., Choh, H.K., 2004. Image-based Photorealistic 3D Face Modeling. Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition, p.49-54. [doi:10.1109/AFGR.2004.1301508]

Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., Salesin, D., 1998. Synthesizing Realistic Facial Expressions from Photographs. Proc. SIGGRAPH'98. Orlando, p.75-84.

Reiter, D., Donner, R., Langs, G., Bischof, H., 2006. 3D and Infrared Face Reconstruction from RGB Data Using Canonical Correlation Analysis. Proc. 18th International Conference on Pattern Recognition, p.425-428.

Roweis, S., Saul, L., 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**(5500):2323-2326. [doi:10.1126/science.290.5500.2323]

Sclaroff, S., Alon, J., 1999. Non-rigid Shape from Image Streams. Boston University, Computer Science, Technical Report, No. 99-006.

Tenenbaum, J., Silva, V., Langford, J., 2000. A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**(5500):2319-2323. [doi:10.1126/science.290.5500.2319]

Tomasi, C., Kanade, T., 1992. Shape and Motion from Image Streams: A Factorization Method Full Report on the Orthographic Case. Cornell TR 92-1270 and Carnegie Mellon CMU-CS-92-104.

Torresani, L., Hertzmann, A., 2003. Learning Non-rigid 3D Shape from 2D Motion. Proc. Neural Information Processing Systems. Vancouver, BC, p.1555-1562.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, **3**(1):71-86.

Turk, G., O'Brien, J., 1999. Shape Transformation Using Variational Implicit Functions. Proc. SIGGRAPH'99, p.335-342.

Wang, J., Zhang, Z., Zha, H., 2005. Adaptive Manifold Learning. Advances in Neural Information Processing Systems 17. MIT Press, Cambridge, MA, p.1473-1480.

Xiao, J., Kanade, T., 2004. Non-rigid Shape and Motion Recovery: Degenerate Deformations. Proc. 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **1**:668-675. [doi:10.1109/CVPR.2004.1315096]

Yoshiki, K., Saito, H., Mochimaru, M., 2006. Reconstruction of 3D Face Model from Single Shading Image Based on Anatomical Database. Proc. 18th International Conference on Pattern Recognition, p.350-353.

Zhang, R., Tsai, P., Cryer, J., Shah, M., 1999. Shape from shading: a survey. *IEEE Tran. Pattern Anal. Machine Intell.*, **21**(8):690-706. [doi:10.1109/34.784284]

Zhang, Q., Liu, Z., Guo, B., Shum, H., 2003. Geometry-Driven Photorealistic Facial Expression Synthesis. Eurographics/SIGGRAPH Symposium on Computer Animation. San Diego, CA, p.177-186.

Zhao, M., Chua, T., Sim, T., 2006. Morphable Face Reconstruction with Multiple Images. Proc. 7th International Conference on Automatic Face and Gesture Recognition, p.597-602. [doi:10.1109/FGR.2006.79]

APPENDIX A: PROCESS OF LLE/ALLE ALGORITHM

Step 1: The K closest neighbors are selected for each point using a distance measure such as the Euclidean distance (The neighborhood can be adaptively selected by our proposed method to build ALLE algorithm).

Step 2: Solve for the manifold reconstruction weights. The reconstruction errors are measured by the cost function:

$$\varepsilon(w) = \sum_{i=1}^{n} \left\| \boldsymbol{X}_i - \sum_{j=1}^{n} W_{ij} X_{ij} \right\|^2,$$

where $\boldsymbol{X}_i$ is a datapoint and $\varepsilon(w)$ is the sum of the squared distances between all datapoints and their reconstruction neighbors. The weight W_{ij} represents the contribution of the jth data onto the ith reconstruction.

Two constraints should be obeyed:

(1) $\sum W_{ij}=1$;
(2) $W_{ij}=0$ if x_j is not a neighbor of x_i.

The weights are then determined by a least squares minimization of the reconstruction errors.

Step 3: Map each high-dimensional data $\boldsymbol{X}_i$ to a low-dimensional coordinate $\boldsymbol{Y}_i$. This is done by minimizing the cost function representing locally linear reconstruction errors:

$$\Phi(\boldsymbol{Y}) = \sum_{i=1}^{n} \left\| \boldsymbol{Y}_i - \sum_{j=1}^{n} W_{ij} Y_{i_j} \right\|^2.$$

The cost function defines a quadratic form:

$$\Phi(\boldsymbol{Y}) = \sum_{i=1}^{N} \sum_{j=1}^{N} M_{ij} \boldsymbol{y}_i^{\mathrm{T}} \boldsymbol{y}_j.$$