

A NOVEL FACIAL FEATURE POINT LOCALIZATION METHOD ON 3D FACES

Peng Guan, Yaoliang Yu, Liming Zhang

Department of Electronic Engineering, Fudan University, Shanghai 200433, China

pengguan1983@gmail.com, yaoliangyu1983@gmail.com, lmzhang@fudan.edu.cn

ABSTRACT

Although 2D-based face recognition methods have made great progress in the past decades, there are also some unsolved problems such as PIE. Recently, more and more researchers have focused on 3D-based face recognition approaches. Among these techniques, facial feature point localization plays an important role in representing and matching 3D faces. In this paper, we present a novel feature point localization method on 3D faces combining global shape model and local surface model. Bezier Surface is introduced to represent local structure of different feature points and global shape model is utilized to constrain the local search result. Experimental results based on comparison of our method and curvature analysis show the feasibility and efficiency of the new idea.

Index Terms— facial feature point localization, global shape model, local surface model

1. INTRODUCTION

With the development of 3D acquisition system, the capture of 3D faces is becoming faster and cheaper, and face recognition based on 3D information is attracting more attention. One reason is that 3D face recognition technologies are believed to have greater potential for higher recognition rate than 2D [1]. Generally, a 3D face includes 10^5 - 10^6 vertices given by their coordinates (x, y, z) and each vertex has its texture information (R, G, B) . Since it is hard to handle the entire set of data, we always extract some key points (far less than the number of vertices) called feature points, and use them to construct a visual model to describe and represent the expected structure of the object. Over the past decades, model-based approaches have been applied successfully in image understanding and computer vision. The famous examples are probably Active Shape Model (ASM) and Active Appearance Model (AAM). More information about them can be found in [8][9].

Due to the reason above, most 3D-based recognition approaches are concentrated on feature point extraction and using the relationship between feature points in face matching process [1][2][6]. Curved Surface Analysis [7] attempts to represent a local shape on the surface of 3D object using curvature analysis including Gaussian, minimum and maximum curvature etc. The main idea of the curvature analysis is to calculate the first-order and second-order derivatives of a certain feature point on object surface, and then using a function of these derivatives to describe the local structure. J.T. Lapreste and coworkers [3] presented a system that identifies facial feature points, such as center of eyes,

tip of nose, lips, and chin. This method is based on the local curvature which is determined by means of the range data in a neighborhood. Gordon [4] gave a detailed study of range images in the human face recognition based on depth and face surface curvature features. The curvatures on a face surface are computed to find prominent features such as nose ridge and eye corners. Moreno et al. [5] proposed a 3D face recognition method by firstly performing a segmentation based on Gaussian curvature and then creating a feature vector based on the segmented regions. Lee and Shim [6] considered an approach by using “depth-weighted Hausdorff distance” and surface curvature information for 3D face recognition. However, when several feature points have similar local structure (curvature feature) but different positions e.g feature points on cheek, how to localize them is still a problem.

In this paper, we propose a novel feature point localization method integrating both global shape model (GSM) and local surface model (LSM). The idea of combining global and local information is inspired by ASM[8][9]. However our method is quite different from ASM. Our contribution is: 1) propose a totally 3D method to localize feature points on 3D object using the positions of vertices. 2) propose a new feature pattern—Bezier Surface Descriptor to describe and represent the geometric information on 3D object surface.

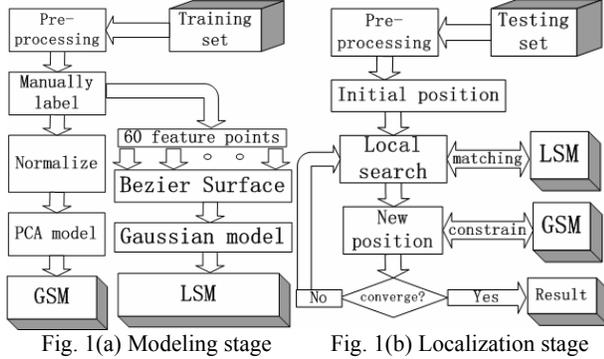
2. OVERVIEW

The proposed method is briefly overviewed shown as Fig1 (a), (b), which has two stages: modeling and localization. All of the training and testing 3D human faces are from “BJUT-3D Large-Scale Chinese Face Database” [11]. For all the 3D faces, pre-processing is required which includes smoothing, moving burrs, filling missing data, evenly re-sampling, and rectifying coordinates.

There is no particular restriction of the number of feature points. However they should be sufficient to represent a 3D face. In this paper, we define 60 feature points, e.g eye corners, nose tips, nose ridges, lips and so on. The distribution of these feature points is shown in Fig.6. In the modeling stage, 60 feature points, which are semantically well-defined, are manually labeled. Then we normalize the global face shapes given by the feature points and then apply PCA analysis to get the GSM. The establishment of LSM for each feature point is exactly the same. We collect $m \times m$ vertices on the 3D face around a certain feature point for all the training samples and regard them as control points of Bezier surface. The descriptive points on the smooth curved surface are obtained by Bezier functions. Finally, “local structure feature” which is derived by these descriptive points is used to set up LSM.

Localization is an iterative procedure in which the first step is to roughly estimate the initial positions of feature points. Then we search the neighboring vertices around each feature point on 3D

face within a small range. The best-matching point is given by LSM and the feature point positions are updated. However, the new positions should be constrained by GSM so that the geometrical relationship of these points is reasonable [8]. This procedure is repeated until convergence.



3. GLOBAL SHAPE MODEL (GSM)

Since we have labeled the feature points for each image in the training set, the shape of 3D human face can be described as a vector $\mathbf{X} = (x_1, \dots, x_{60}, y_1, \dots, y_{60}, z_1, \dots, z_{60})^T$, where (x_j, y_j, z_j) is the coordinates of the j th feature point. Procrustes analysis [8][10] is used to eliminate the differences among face shapes caused by translation, rotation, and scaling parameters through minimizing (1).

$$\sum_{i=1}^N \|T(\mathbf{X}_i) - \mathbf{Y}\|^2 \longrightarrow \min \quad (1)$$

In (1), N is the number of training samples, \mathbf{X}_i is the shape of the i th sample, \mathbf{Y} is a standard shape, and $T(\bullet)$ represents similarity transformation. ‘‘Normalized Shapes’’ are obtained by Procrustes analysis. Then, PCA analysis is introduced to setup GSM. Thus, we can approximate any shape of the training set using $\mathbf{X}_i = \bar{\mathbf{X}} + \boldsymbol{\psi}\mathbf{b}$. Here $\bar{\mathbf{X}}$ is the mean shape of 3D face shown in Fig.2. The matrix $\boldsymbol{\psi}$ consists of eigenvectors for covariance matrix of \mathbf{X}_i , and vector \mathbf{b} defines a set of parameters of deformable model. The variance of the i th parameter, b_i , over the training set is given by σ_i . By applying limits of $\pm 3\sqrt{\sigma_i}$ to the parameter b_i , we can keep that the generated shape is similar to those in the training set.

4. LOCAL SURFACE MODEL (LSM)

4.1 Bezier Surface Descriptor

The neighboring points in a small range of each feature point form a discrete surface and these points are called control points. The discrete surface is a descriptor of the local structure of different feature point. Fig3 (a) and (b) show the situations of cheek and nose tip respectively, where 11×11 control points on the face surface are selected and the corresponding feature point is located at the center of these points. Since different feature points have different local structure, we can use this information to distinguish them.

In order to describe local surface as accurate as possible, we introduce the notion of Bezier Surface. Given $n_0 \times n_1$ discrete control points, a smooth curved surface called Bezier Surface is structured by (2), (3) where P_{i_0, i_1} ($0 \leq i_0 \leq n_0, 0 \leq i_1 \leq n_1$) is the distance from a control point (i_0, i_1) to a user-defined base plane W and $X(s, t)$ is the distance from a descriptive point (s, t) on Bezier Surface to the base plane. Here, s and t are normalized to $0 \sim 1$.

$$X(s, t) = \sum_{i_0=0}^{n_0} \sum_{i_1=0}^{n_1} B_{n_0, i_0}(s) B_{n_1, i_1}(t) P_{i_0, i_1}, \quad (s, t) \in [0, 1]^2 \quad (2)$$

$$B_{n, i}(z) = \binom{n}{i} z^i (1-z)^{n-i} = \frac{n!}{i!(n-i)!} z^i (1-z)^{n-i} \quad z \in \{s, t\} \quad (3)$$

In figure 4, $n_0 = n_1 = 3$ and the base plane is X-Y. Initially, there are 16 control points shown as larger dots forming a transparent grid. Using (2), (3) we can obtain 49 descriptive points shown as smaller dots forming a gray grid. For example, $P_{1,0}$ denotes the distance from control point (1,0) to the base plane. $X(0, 0.5)$ represents the distance from descriptive point (0, 0.5) to the base plane, where 0 and 0.5 are normalized coordinates. Theoretically, there is no limit of the number of descriptive points because s, t have continuous values. This is the reason why Bezier surface can approximate a set of discrete points accurately. However, the more the descriptive points the heavier the computational cost. In the next section we will focus on how to define a uniform base plane.

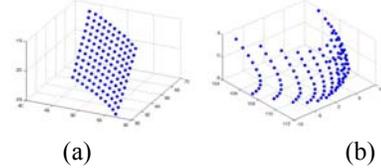


Fig.3 The local geometrical feature of (a) cheek, (b) nose tip

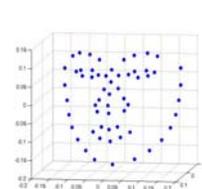


Fig.2 mean shape

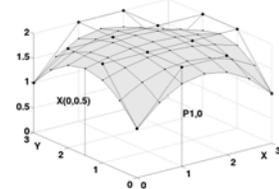


Fig.4 Bezier surface

4.2 Base Plane

Let $n_0 = n_1 = m$ and m is an odd number. The corresponding feature point is one member of control points and located at the center of the control points. In order to decide a unique base plane, another two conditions are satisfied: 1) The base plane pass through the feature point, 2) $E = \sum_{i=0}^{m-1} \|P_i\|^2$ is minimized, where P_i is the distance from the i th control point to base plane W . Define the base plane W as $x \cdot n_x + y \cdot n_y + z \cdot n_z - D = 0$, where $\bar{n}_0 = (n_x, n_y, n_z)^T$ is the normal direction of W and $(x, y, z)^T$ is the coordinates of points on W . \bar{n}_0 and D are parameters to be

decided. Given the feature point $\bar{Q}_0 = (x_0, y_0, z_0)^T$ and the i th control point $\bar{Q}_i = (x_i, y_i, z_i)^T$, the distance from \bar{Q}_i to W is calculated by (4) according to solid geometry.

$$P_i = \left\| x_i \cdot n_x + y_i \cdot n_y + z_i \cdot n_z - D \right\| = \left\| \bar{Q}_i^T \cdot \bar{n}_0 - D \right\| \quad (4)$$

From the first condition—the feature point lies on the base plane W , we have $d_0 = \left\| \bar{Q}_0^T \cdot \bar{n}_0 - D \right\| = 0$, so $D = \bar{Q}_0^T \cdot \bar{n}_0$. The second condition can be converted to the following problem:

$$E = \sum_{i=0}^{m^2-1} \left\| \bar{Q}_i^T \cdot \bar{n}_0 - \bar{Q}_0^T \cdot \bar{n}_0 \right\|^2 \text{ is minimized s.t. } \left\| \bar{n}_0 \right\| = 1 \quad (5)$$

Lemma: When (5) is satisfied, the corresponding \bar{n}_0 is the eigenvector associated with the smallest eigen-value of $C_Q = \sum_{i=0}^{m^2-1} (\bar{Q}_i - \bar{Q}_0)(\bar{Q}_i - \bar{Q}_0)^T$.

Proof:

$$\begin{aligned} E &= \sum_{i=0}^{m^2-1} \left\| \bar{Q}_i^T \cdot \bar{n}_0 - \bar{Q}_0^T \cdot \bar{n}_0 \right\|^2 = \sum_{i=0}^{m^2-1} \left\| (\bar{Q}_i - \bar{Q}_0)^T \cdot \bar{n}_0 \right\|^2 \\ &= \sum_{i=0}^{m^2-1} ((\bar{Q}_i - \bar{Q}_0)^T \cdot \bar{n}_0)^T ((\bar{Q}_i - \bar{Q}_0)^T \cdot \bar{n}_0) \\ &= \bar{n}_0^T \sum_{i=0}^{m^2-1} (\bar{Q}_i - \bar{Q}_0)(\bar{Q}_i - \bar{Q}_0)^T \bar{n}_0 = \bar{n}_0^T \cdot C_Q \cdot \bar{n}_0 \end{aligned}$$

When \bar{n}_0 is available by the lemma, the base plane is obtained.

4.3 Gaussian Model

We build a statistical model of the local structure around each feature point and during search period simply find the point which best matches the model. Here we use “nose tip” as an example to explain the modeling procedure. Suppose there are N 3D faces in the training set, we pick up $m \times m$ points on the 3D face surface around “nose tip” (e.g Fig3) and regard them as “control points” of Bezier Surface. In this paper, $m=11$. Then base plane W is obtained by section 4.2. In order to avoid heavy computational cost, we calculate $(2m-1)^2$ descriptive points by section 4.1 (add one more descriptive point between two control points). Each of these descriptive points is described by $X(s, t)$. These descriptive points are then reshaped to form a vector $Z_i^j \in R^{(2m-1)^2}$, where $i \in [1, N]$ $j \in [1, 60]$, which means the local feature of the j th point in the i th face (in this example, j associate with “nose tip”). For each j , PCA analysis is applied across training faces to get a low dimensional representation of Z_i^j where 95% energy of the covariance matrix is retained. The dimensions of the representation vary across different feature points according to the complexity of different local surface. If Z_i^j is projected to a k dimensional space, a feature set $\mathbf{Z}^j = \{Z_1^j, Z_2^j, \dots, Z_N^j\}$, where $Z_i^j \in R^k$, is obtained for a given model point e.g “nose tip”. We assume that Z_i^j is distributed as a multivariate Gaussian as (6), and estimate their mean \bar{Z}^j and covariance Σ^j :

$$p(\mathbf{Z}^j) = \frac{1}{(2\pi)^{k/2} |\Sigma^j|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{Z}^j - \bar{\mathbf{Z}}^j)^T \Sigma^{-1} (\mathbf{Z}^j - \bar{\mathbf{Z}}^j) \right\} \quad (6)$$

The matching criterion of a new feature pattern, Z_s^j , to the j th model is given by (7)

$$f(Z_s^j) = (Z_s^j - \bar{\mathbf{Z}}^j)^T \Sigma^{-1} (Z_s^j - \bar{\mathbf{Z}}^j) \quad (7)$$

This is the Mahalanobis distance and minimizing (7) is equivalent to maximizing the probability that Z_s^j comes from the distribution.

5. LOCALIZATION

There are 3 steps in the localization stage: 1) Initial positions estimation. 2) Local search 3) Global shape constrain.

1) The initial positions of feature points are roughly estimated shown as Fig. 5(a) with the help of the position of nose tip which is detected by curvature feature. 2) Local search is performed to find the “best-matching” point to every feature point. In this process, Z_s^j is computed for all candidate points around the j th feature point on 3D human face and the “best-matching” point will minimize (7). 3) The positions of “best-matching” points should be constrained by GSM, in which the shape parameter b_i is limited in the scope of $[-3\sqrt{\sigma_i}, 3\sqrt{\sigma_i}]$. The whole process mentioned above will be repeated until convergence. Typically, the algorithm will converge after 10-15 iterations. The localization process is illustrated in Fig5.

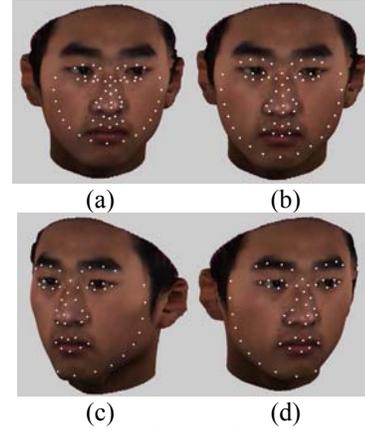


Fig.5 Localization process from different viewpoint. (a) initial position (b) after 3 iterations (c) after 6 iterations (d) after 10 iterations (converged)

6. EXPERIMENTAL RESULTS

We select 160 3D face images from “The BJUT-3D Large-Scale Chinese Face Database” and all of them are manually labeled using a consistent standard. Then 60 faces as training set are used to setup the model and the rest 100 images are used to test the accuracy of our method.

In order to explain the accuracy of our method, we define “Relative Error”

$$\varepsilon = \left\| C - C_{search} \right\|_2 / Dis_{eye} \times 100\% \quad (8)$$

In (8), $C = (x, y, z)^T$ is the position of a certain feature point which is manually labeled while C_{search} is the position of the same feature point automatically localized by our method. Dis_{eye} is the distance

between two outer eye corners of this person. The distribution of feature points is shown in a 2D image (Fig.6) with serial numbers labeled aside. Comparisons of our method, Gaussian Curvature, Minimum Curvature, and Maximum Curvature are given in Fig.7, where X axis stands for serial number of feature points and Y axis for the average relative error among 100 testing images. It should be emphasized that, for the sake of equity, we introduce GSM into these curvature methods. That means, the new positions gained by local search using curvature features are constrained by GSM. Actually, curvature methods alone can not localize the feature points on the cheek, because the curvature analysis of these feature points will yield extremely similar results. So they can not distinguish with each other.

Feature points 1-8 are on the eyebrow. The local surfaces of these points have no definite structure and they vary drastically for different people. Accordingly, the accuracy of points 1-8 is the worst compared to other feature points. Feature points 9-25 are on the cheek. The curvature descriptors of these points are prone to be similar. Nevertheless, LSM is better at distinguishing the small variations of local patterns. The local structures of 26-60 are much easier to identify. These points include eye corners, nose ridge, nose-wing, nose-tip, and mouth tips. It can be shown from Fig.7 that our method outperforms other methods.

7. CONCLUSIONS

In this paper, a novel feature point extraction method on 3D face image is proposed. In the modeling stage, GSM and LSM are setup. GSM is used to represent the global shape of 3D face. This information helps us to discard those non-face shapes easily, and thus constrains the new positions given by local search result. LSM is a descriptor of how the local structure looks like. With the help of LSM, small variations are detected and different kinds of local surfaces are distinguished. Mahalanobis distance is adapted to measure the distance from a new pattern to a local surface model. Experimental results show that our method is better than the conventional curvature-based methods. In the LSM, users should define 2 parameters: 1) how many control points? 2) how many descriptive points? In this paper, 11*11 control points are selected and approximately 4 times amount of descriptive points are calculated. There is no strict criterion for the selection of the two parameters, however control points should be enough to describe the local structure and too many descriptive points should be avoided regarding the computational complexity. Further, in order to achieve illumination invariance, we do not use textual features. But from the experiment above, structure information is sufficient to localize the feature points correctly.

Acknowledgement

This work is supported by a grant from the National Science Foundation (NSF 60571052), China.

REFERENCE

[1] K.W. Bowyer, K. Chang, P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding* 101 (1) (2006) 1-15.
 [2] Berk Gokberk, M.Okan Irfanoglu, Lale Akarun, "3D shape-based face representation and feature extraction for face recognition," *Image and Vision Computing* 24 (2006) 857-869

[3] J.T.Lapreste, J.Y. Cartoux, and M. Richetin, "Face Recognition from Range Image Data by Structural Analysis," *Syntactic and Structural Pattern Recognition, NATO ASI Series*, pp, 303-314, 1988
 [4] G.G.Gordon, "Face Recognition based on depth and curvature feature," *Proceeding of IEEE Computer Society Conference on CVPR*, pp. 808-810, 1992
 [5] A.B. Moreno, A. Sanchez, J.F. Velez, F.J. Diaz, "Face recognition using 3D surface-extracted descriptors," *Proceedings of the Irish Machine Vision and Image Processing Conference*, 2003.
 [6] Y. Lee, J. Shim, "Curvature-based human face recognition using depth-weighted Hausdorff distance," *International Conference on Image Processing (ICIP)*, 2004, pp. 1429 - 1432.
 [7] F. G. Peel and T. S. Sahota, "Surface Curvature as a Measure of Image Texture," *IEEE PAMI*, Vol. PAMI-7, No.6, pp. 734-738, November 1985.
 [8] T.F. Cootes and C.J. Taylor, "Statistical models of appearance for medical image analysis and computer vision", *Proc. SPIE Medical Imaging* 2001
 [9] R.H.Davies, C.J.Twining, P.D.Allen, T.F.Cootes and C.J.Taylor, "Building optimal 2D Statistical Shape Models," *Image and Vision Computing*, Vol.21, pp.117-82, 2003
 [10] Shinji Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. PAMI*, vol 13, No 4, April 1991.
 [11] Multimedia and Intelligent Software Technology Beijing Municipal Key Laboratory Beijing University of Technology, "The BJUT-3D Large-Scale Chinese Face Database.," 2005.8

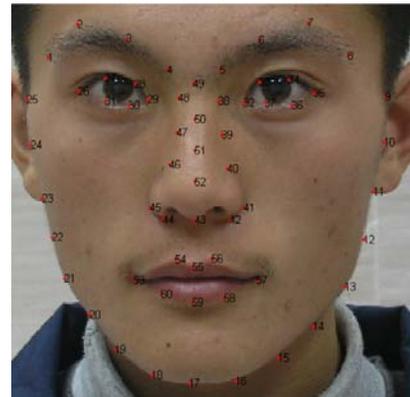


Fig.6 Distribution of feature points

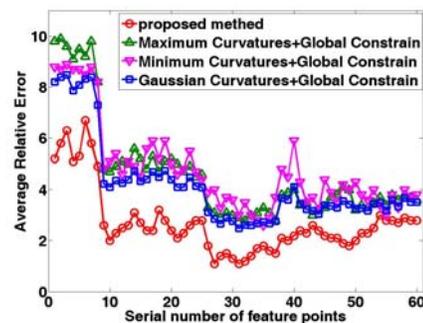


Fig.7 Accuracy of localization