A 3D FACIAL FEATURE POINT LOCALIZATION METHOD BASED ON STATISTICAL SHAPE MODEL

Guangpeng Zhang, Yunhong Wang

School of Computer Science and Engineering
Beihang University, Beijing, 100083, P. R. China
gpzhang@cse.buaa.edu.cn, yhwang@buaa.edu.cn

ABSTRACT

Registration is a necessary step for automatic 3D face recognition systems, and feature point localization is usually used to find the correspondence in registration. Traditional localization methods are sensitive to pose changes, and can only deal with frontal or limited pose variations. In this paper we propose a new 3D facial feature point localization method that is insensitive to pose variation. Feature regions are firstly segmented out based on Shape Index features, and then selected by a statistical shape model. Point nearest to the region center is chosen as a feature point. Experimental results show that the localization accuracy is comparable to manually labeled feature points.

***Index Terms*—** Feature point localization, Shape index, Statistical Shape Model

1. INTRODUCTION

Registration is necessary for automatic 3D face recognition systems, which transforms different face models into the same orientation and position. Feature point localization is usually used in registration to find the correspondence between two models, and an efficient localization method can help to achieve fast and accurate registration. Facial feature point localization also plays an important role in accurate facial feature extraction, 3D face animation [1] and so on.

While 2D facial feature point localization methods were well studied, few people have conducted research on 3D facial feature point localization. In 3D cases, pose variation is a problem to be solved, and different from 2D images, 3D range scans are often stored in unequally sampled mesh format, which causes some traditional methods such as template matching hard to use. Gordon [2] segmented facial surface based on curvature features, and then used constraints to search for the facial features. The constraints she used somewhat limited pose changes. For example, the face symmetric plane was assumed to be roughly vertical. The method was just tested on a small database of 24 range scans. Wang et al. [3] extended jet bunch [4] to 3D facial feature point localization. In their method, they first manually labeled a training set and compute point signature [5] at feature points to construct jet bunch, then for a new face model, based on the average layout and position, point which corresponds to the minimum distance with jet bunch was chosen as the final feature point. This method allows limited pose variations. Recently Lu proposed a method [6] in which yaw angle range was quantized and a feature extractor was proposed to estimate the nose tip location. A nose profile model represented by subspaces was used to select the best candidates for the nose tip. A multimodal scheme was presented to extract eye and mouth corners. His method can be extended to arbitrary pose changes, but the pose quantization will cause time cost increases exponentially. Dirk Colbry et al. [7] firstly detected a set of candidate points which satisfy a series of criteria, then eliminated impossible labeling by a series of constraints, and finally ICP algorithm was run to determine the fitness of the points. They achieved 99% success in frontal images and 82% success with large variations in pose and expression, however the arbitrary pose anchor point detector takes around 15 seconds to complete, which is time consuming. Template matching was used for inner eye point detection, and it couldn't be applied directly in the case of unequally sampled mesh. As we can see from the above, most of the former methods assumed pose to be frontal or evaluate the pose in some way so that the average layout of the feature points could be fixed. However, when the pose isn't known or cost much time to evaluate, those methods are not suitable.

In order to restrain the influence of pose variance and achieve robust and effective feature point localization, no assumption about pose should be made. Also no assumption about the mesh format should be made so that the method can be used for the unequally sampled mesh. A method based on shape index [8] and statistical shape model is proposed for the purpose of developing a feature point localization algorithm robust to the variation of pose and format.

In section 2 the Shape Index feature is briefly introduced, which is a pose invariant representation of the surface. In section 3 we show how candidate feature regions are segmented out based on Shape Index. In section 4, candidate regions are further selected by a trained statistical shape model, point

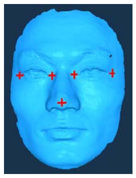

Fig. 1. Feature points' location.

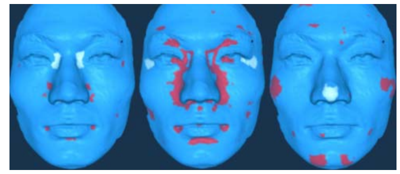

Fig. 2. Candidate regions segmented out. The white regions are the exact feature regions.

nearest to the region center is chosen as a feature point. In section 5, the efficiency of this method is verified through experiments. As feature points are used for coarse alignment, they must be salient and robust to various pose and expression changes. So we choose to locate five different feature points, including left and right inner eye corners, left and right outer eye corners, and nose tip, which are shown in Figure 1.

2. FEATURE POINTS AND SHAPE INDEX

To robustly locate feature points on a 3D face model, the feature used should be pose-invariant. Curvatures are independent of coordinate systems, and do not change with rigid transformation, so they are usually used as 3D object features. Dorai et al. [8] proposed Shape Index feature to represent surface concave and convex attributes. Shape Index has been used in [6, 7] to locate feature points. The Shape Index at point p is calculated using maximum (κ_1) and minimum (κ_2) local curvature (Equation 1).

$$S(p) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \tag{1}$$

From Equation (1), we can see that the range of Shape Index is $[0, 1]$. Local shape at point p is a spherical cup when $S(p) = 0$, and a spherical cap when $S(p) = 1$. When Shape Index changes from 0 to 1, local shape changes from spherical cup to spherical cap.

3. CANDIDATE REGION SEGMENTATION

Shape Index can be used to effectively separate out feature regions which have prominent concave and convex characteristics. In these regions (inner eyes, outer eyes, nose tip) Shape Index clusters together and can easily be separated from surrounding regions. For example, the inner eye regions have very low Shape Index, and nose tip region has very high Shape Index, while their surrounding regions have different Shape Index. This property can help segment out those feature regions using simple threshold method.

Shape Index images are firstly smoothed by an average filter to restrain the noise and emphasize Shape Index clustering. In order to get thresholds to segment out feature regions, we manually label a training set of frontal scans, and get the Shape Index values at feature points. The means $\mu_i, i = 1, 2, 3$ and variances $\sigma_i, i = 1, 2, 3$ of Shape Index at three kinds of feature points (inner eyes, outer eyes, and nose tip) are calculated, and the corresponding intervals are set to be $[\mu_i - 2 * \sigma_i, \mu_i + 2 * \sigma_i]$ to segment out feature regions.

Using these intervals to segment Shape Index images, we can get candidate feature regions. Figure 2 shows segmented candidate regions. As can be seen from the resulting images, there are many false candidate regions colored red except the exact feature regions colored white, which must be eliminated.

4. REGION SELECTION

As Figure 2 shows, we must select the exact feature regions out of candidate regions and eliminate disturbing regions. Because relative distribution of facial feature regions are stable and obey certain geometric constraint, while disturbing regions distribution are noisy and unstable, feature regions can be selected using relative distribution constraint. Some of the methods proposed earlier also use geometric constraints to select feature points, but their constraints were pose related. Different from traditional methods, we just use scalar measures which are invariant with pose variation. Our method contains two steps, in the training step, a statistical shape model is trained and constructed, in the test step, for a segmented Shape Index image, feature regions are selected as those maximize the fitness with the shape model.

4.1. Statistical Shape Model Construction

Although facial feature region distribution are basically similar, various differences still exist, such as differences between male and female, younger and adult, Asian and European, etc., the distribution model must be able to cover these differences. A set of typical frontal face scans are selected as training samples to construct a statistical shape model. The model is represented in the form of a graph whose vertexes are region centers and edges connect different regions, as is shown in Figure 3.

Geometric measures are calculated after manually labeling feature regions of training scans. To ensure pose invariant, measures used must be robust to rigid transformation. Different from traditional methods which mainly use distances as constraints, we use angles together with distances, as we

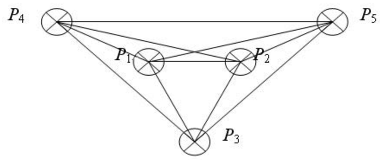

Fig. 3. Feature regions distribution model.P_1, P_2 are inner eyes, P_4, P_5 are outer eyes, and P_3 is nose tip.

find that distance constraints are not sufficient, and there are points which satisfy distance constraint but not the exact feature points. Specifically, we calculate the following 21 measures:

1. Length of edges P_1P_2, P_1P_3, P_2P_3, P_4P_1, P_4P_3, P_5P_2, P_5P_3, P_4P_2, P_5P_1, P_4P_5;

2. Angles $\angle P_2P_1P_3$, $\angle P_1P_2P_3$, $\angle P_1P_3P_2$, $\angle P_3P_1P_4$, $\angle P_3P_2P_5$, $\angle P_4P_1P_3$, $\angle P_3P_5P_2$, $\angle P_3P_1P_5$, $\angle P_3P_2P_4$, $\angle P_3P_4P_5$, $\angle P_3P_5P_4$.

Assume we have three points $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$, and $P_3(x_3, y_3, z_3)$, the length of edge P_1P_2 and angles $\angle P_2P_1P_3$ can be defined as

$$\begin{aligned} \overrightarrow{v_1} = P_2 - P_1, \overrightarrow{v_2} &= p_3 - P_1 \\ L(P_1P_2) &= \|\overrightarrow{v_1}\| \\ \angle P_2P_1P_3 &= \arccos(\frac{\overrightarrow{v_1} \bullet \overrightarrow{v_2}}{\|\overrightarrow{v_1}\|\|\overrightarrow{v_2}\|}) \end{aligned} \quad (2)$$

After getting these measures $x_{ij}, i = 1, \ldots, m, j = 1, \ldots, 21$, ($m$ is the size of the training set), we can calculate their mean values as the model parameters.

$$\overline{x_j} = \frac{1}{m} \sum_{i=1}^{m} x_{ij} \quad (3)$$

For a new set of candidate feature points, once n measures $x_{j'}^s$ are computed, their fitness f with the distribution model is defined as

$$f = -\frac{1}{n} \sum_{j'} \frac{(x_{j'}^s - \overline{x_{j'}})^2}{\overline{x_{j'}}^2} \quad (4)$$

where $x_{j'}^s$ and $\overline{x_{j'}}$ are the corresponding measures.

4.2. Feature Region Selection and Feature Points Localization

For a new face scan, we firstly calculate its Shape Index and segment out candidate regions, and then select regions that best fit the distribution model as the feature regions. Although there are not so many candidate regions, an exhaustive search of all candidate regions is still time consuming and ineffective. Here we select feature regions sequentially in four steps:

1. Through analysis of candidate inner eyes, the real inner eye regions are found to be almost always the largest two among candidate regions, so candidate inner eyes' area are first compared, and the largest two regions are treated as the real inner eyes regions;

2. Search in all the candidate regions of nose tip, and for each one calculate model measures related with P_3, P_1 and P_2, and get the fitness with the distribution model by Equation 4. The one with the highest fitness is treated as the real nose tip region;

3. Select outer eyes regions in the same way as in step 2, while the measures to be computed in this case are all the 21 measures;

4. For the above selected regions, if the corresponding fitness is less than a threshold , then the region is regarded as a wrong region and discarded.

By sequentially selecting feature regions, the computation complexity is largely reduced.

After the real feature regions are selected, their central points are selected as feature points. The reason for locating regions firstly is that the noise influence can be reduced in this way. The central point of a region is defined as the point in the region that is nearest to the region center.

5. EXPERIMENTS AND RESULTS

In order to verify our method's validity, experiments were carried out on a 3D face database scanned by Minolta Vivid 910. The database is divided into a training set and a testing set. The training set contains 51 frontal scans and is manually labeled to get the segmenting thresholds and train the statistical shape model. The testing set contains 362 scans of 52 objects under 7 different poses of frontal, looking up, looking down, in plane left rotating, in plane right rotating, out-of-plane left rotating and out-of-plane right rotating. In order to evaluate the quality of the feature points, we manually labeled feature points in the testing set as the ground truth, and each located feature point is compared to the manually labeled feature point. The locating error represents the distance from the final located point to the manually labeled point.

The histograms of the error for all five feature points are shown in Figure 4. As was shown in [7], the ICP algorithm can tolerate up to 20 mm locating errors for registration, so the located points with errors less than 20 mm are considered as correctly located points. As can be seen from Figure 4, approximately 90% of the scans are below 20 mm error. We can also see that localization of inner eyes and nose tip is more accurate than that of outer eyes. This is because Shape Index at inner eyes and nose tip is more stable than outer eyes, and occlusion and distortion affect outer eyes more seriously.

To quantify the success rate of our method, for each test scan, if more than 3 feature points are correctly located, then

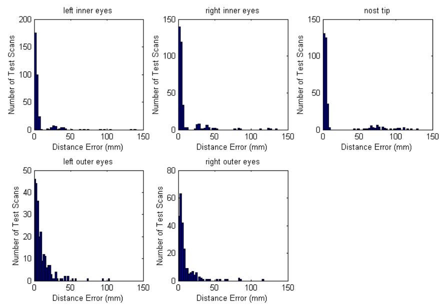

Fig. 4. Error histograms for each of the five feature points.

Pose	Success Rate	Pose	Success Rate
FT	98.1%	LD	86.3%
OL	55.8%	IL	84.6%
OR	59.6%	IR	80.4%
LU	90.4%	Total	84.2%

Table 1. Feature point localization success rate under arbitrary pose. FT, OL, OR, LU, LD, IL, IR respectively stand for Frontal, Out-of-plane Left rotating, Out-of-plane Right rotating, Looking Up, Looking Down, In plane Left rotating, In plane Right rotating.

the localization is labeled as success. The success rate under each pose is shown in Table 1.

As is shown in Table 1, feature points localization under frontal pose has the highest accuracy, near 100%. Localization under poses of rotating in plane and looking up and down has a lower accuracy. Localization when rotating out-of-plane has the lowest accuracy. This is mainly because rotating out-of-plane causes serious occlusion and distortion, and many scans have only one inner eye corner, so the condition of our method is not satisfied. For scans that have out-of-plane rotating angles less than 30 degrees the localization is still very accurate.

The database used here contains more pose variations than those formerly used [2, 3], which contained only some of the pose variations. The method has no assumption about pose and mesh format, and achieves competitive and even higher localization accuracy. This definitely proves the advantage of the proposed method.

6. CONCLUSION

A method for localizing 3D facial feature points under various poses is proposed in this paper. The method is based on Shape Index and a statistical shape model. Experimental results show that the proposed method can achieve relatively high localization accuracy under arbitrary poses. The localization accuracy under poses of rotating out-of-plane is not as high as that of other poses. The future work will focus on employing a view based shape model to enhance the accuracy.

7. ACKNOWLEDGEMENTS

This work was supported by Program of NCET, National Natural Science Foundation of China (No. 60575003, 60332010), and Hi-Tech Research and Development Program of China (2006AA01Z133). The authors acknowledge Professor Yanning Zhang [9] and Zenggang Lin from ShaanXi Key Laboratory of Speech and Image Information Processing, Northwestern Polytechnical University for supplying their 3D Chinese Face Database. We also acknowledge Kui Jia and Caifeng Shan from Queen Mary, London University for their kind suggestion.

8. REFERENCES

[1] T. Fujiwara, T. Nishihara, M. Tominaga, K. Kato, K. Murakami, and H. Koshirnizu, "On the detection of feature points of 3D facial image and its application to 3D facial caricature," in *3DIM99*, 1999, pp. 490–496.

[2] G. G. Gordon, *Face Recognition from Depth and Curvature*, Ph.D. thesis, Harvard University, Division of Applied Sciences, 1991.

[3] Y. Wang, C. S. Chua, and Y.-K Ho, "Facial feature detection and face recognition from 2D and 3D images," *Pattern Recognition Letters*, vol. 23, pp. 1191–1202, 2002.

[4] L. Wiskott, J. M. Fellous, and C. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on PAMI*, vol. 19, no. 7, pp. 775–779, 1996.

[5] C. S. Chua and R. Jarvis, "Point signature: A new representation for 3D object recognition," *International Journal on Computer Vision*, vol. 25, no. 1, pp. 63–85, 1997.

[6] X. Lu and A. K. Jain, "Automatic feature extraction for multiview 3D face recognition," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 585–590.

[7] D. Colbry, G. Stockman, and A. K. Jain, "Detection of anchor points for 3D face verification," in *Proc. IEEE Workshop on Advanced 3D Imaging for Safety and Security*, 2005.

[8] C. Dorai and A. K. Jain, "Cosmos - a representation scheme for 3D free-form objects," *IEEE Transactions on PAMI*, vol. 19, no. 10, pp. 1115–1130, 1997.

[9] J. Pan, Y. Zhang, H. Zhou, Y. Hou, L. Long, and L. Yao, "The research of the match of corresponding points in multi-view and the realization by evolutionary programming," in *Proc. International Conference on Signal Processing*, 2004, pp. 2202–2205.