Normalization for Unconstrained Pose-Invariant 3D Face Recognition

Xun Gong[1,2], Jun Luo[3], and Zehua Fu[1]

[1] School of Information Science and Technology,
South West Jiaotong University,
Chengdu 600031, P.R. China
[2] ChongqingKey Laboratory of Computational Intelligence,
Chongqing 400065, P.R. China
xgong@home.swjtu.edu.cn
[3] Sichuan Academy of Medical Sciences &Sichuan Provincial People's Hospital,
Chengdu 600072, P.R. China

Abstract. This paper presents a framework for 3D face representation, including pose and depth image normalization. Different than a 2D image, a 3D face itself contains sufficient discriminant information. We propose to map the original 3D coordinates to a depth image using a specific resolution, hence, we can remain the original information in 3D space. 1) Posture correction, we propose 2 simple but effective methods to standardize a face model that is appropriate to handle in following steps; 2) create depth image which remain original measurement information. Tests on a large 3D face dataset containing 2700 3D faces from 450 subjects show that, the proposed normalization provides higher recognition accuracies over other representations.

Keywords: 3D face recognition, depth image, pose correction, normalization.

1 Introduction

Boston Marathon bombing events, in 2013, has drawn a lot of public attention to automatic face recognition problem again. A.K. Jain et al. [1] have conducted a case study on automated face recognition under unconstrained condition using the two Boston Marathon bombing suspects. Results indicate that there is still a room for processing images under unconstrained scenes.

With the rapid development and dropping cost of 3D data acquisition devices, 3D face data, which represents faces as 3D point sets or range data, can be captured more quickly and accurately. The use of 3D information in face recognition has attracted great attention and various techniques have been presented in recent years [2-4]. Since 3D face data contain explicit 3D geometry, more clues can be used to handle the variations of face pose and expression.

Even though 3D data potentially benefit to face recognition (FR), many 3D face recognition algorithms in the literature still suffer from the intrinsic complexity in representing and processing 3D facial data. 3D data bring challenges for data

Z. Sun et al. (Eds.): CCBR 2013, LNCS 8232, pp. 1–8, 2013.

preprocessing, like spike removing, hole filling, pose correction and feature representation. The motivation of this paper is to propose a practical 3D face pose normalization method. And what's more, we argue that 3D face itself contains abundant discriminant information. The question is that how to find a good 3D data representation manner. The main contributions are: (1) we propose 2 effective pose correction methods for an arbitrary 3D face obtained by a 3D data acquisition device; (2) create a depth image which can keep original measurement information.

2 3D Face Posture Normalization

For any practical 3D face recognition systems, posture normalization is always an indispensable step, which can also be called as pose correction. In this paper, we propose 2 pose normalization methods, self-dependent pose correction, shorted as SD, and generalized iterative closest point pose correction (GICP).

2.1 Self-dependent Pose Correction

The key step of the proposed correction method SD is carried out by finding a synthesized plane that is parallel to the face plane (as shown in Fig. 1(a)). Then, rectify the plane to make it perpendicular to the direction of Z-axis. In that way, the angle of yaw (around Y-axis) and pitch (around X-axis) can be correctly compensated. At last the roll angle (around Z-axis) can be easily corrected by the connecting line between two pupils or the center line of the mouth, see Fig. 1 for more details.

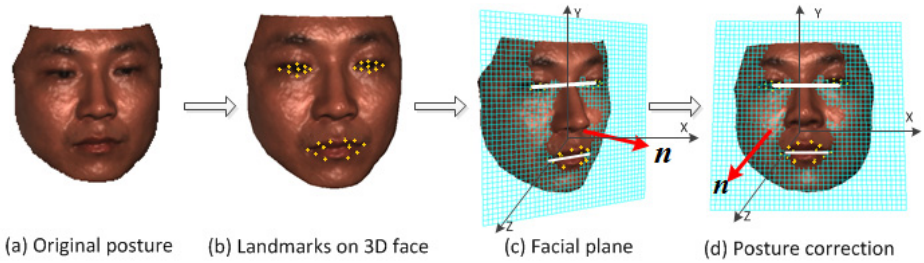

Fig. 1. Main idea of SD pose correction. Cyan grid in (c) and (d) illustrates the virtual face plane, normal direction of which is denoted by the arrow $\boldsymbol{n}$. And the connecting lines between two eyes and mouth denote the auxiliary line used for roll angle correction.

Current commercial 3D scanners can generate a 3D point cloud and a registered 2D texture image simultaneously, just as published by some 3D databases like Texas 3DFR[5] and BU_3DFE[6]. Since the texture channel and the 3D points correspond well, 2D information can assist to find the face regions and key features. Chang et al. [7] applied a skin detection method on the texture channel to help 3D facial region extraction. Wang et al. [8] preform 3D facial region cropping with the help of the texture channel. For feature extraction, we also apply the method ASM work on the texture image, as shown in Fig. 1(b). The steps of SD are described as follows:

1) 3D Facial landmarks detection. At first, the face and its characteristic points ("landmarks") on 2D texture image are located through the approach presented in [9], namely, the extended Active Shape Model (STASM) algorithm which is widely used in academic area [10, 11]. The algorithm locates 76 interest points. The precision of the location procedure depends on the amount of face distortion. However, as aforementioned, those key feature points are only used for fitting the face plane, so our system has tolerance of inaccuracy in extend. Once obtain the 2D landmarks on the texture, 3D landmarks V_{76} on 3D face can be obtained according to their corresponding relationship.
2) Face plane Σ_f fitting. Points on eyes and mouth can be generally seen as placed on the same virtual plane. So we choose the landmarks on two eyes and mouth, 32 points in total (see Fig. 1(b)), to synthesize the facial plane. Those 32 points denote as V_f. Excluding landmarks around the facial contour because those points are always more inaccurate than V_f. A plane can be defined as

$$ax + by + cz + d = 0 \tag{1}$$

With V_f, parameters $[a,b,c,d]$ in (1) can be approximated by using least square method, thus Σ_f is determined.
3) Pose correction. With Σ_f and its normal $\boldsymbol{n}$, the pitch and yaw angles can be easily obtained by compute the angle between $\boldsymbol{n}$ and Y-axis and Z-axis, respectively. By calculating the angel α_1 between X-axis and connecting line of two pupils, the angel α_2 between X-axis and connecting line of two mouth corners, then, the roll angle is straightforward by average α_1 and α_2. Correction example is shown in Fig. 1(d).

2.2 Generalized Iterative Closest Point Pose Rectification

Iterative Closest Point (ICP) [12] is an effective tool for 3D model registration. But its defects are also well known, like time consuming. For ICP, how to find the matching point pairs is always the key problem, which influent the final result significantly. Another common technique, named as Generalized Procrustes Analysis (GPA) [13], is frequently used for aligning a group of 2D shapes. However, in the problem of 3D model registration, scale in GPA is not needed.

Inspired by GPA, we propose a novel generalized ICP, denoted as GICP, for pose correction problem, which is summarized in Algorithms 1.

Algorithm 1**.** Generalized ICP
Input: A set of *n* 3D faces $F = \{f_1,...,f_n\}$

Procedure:
1) Similar to the SD, 76 landmarks are extracted by STASM, and then the 3D landmarks are obtained for each 3D face in F. All of the 3D landmarks are concatenated as a vector $\boldsymbol{s}_i$, which represents the shape of each face f_i:

$$\boldsymbol{s}_i = \left(x_1, y_1, z_1, ...x_j, y_j, z_j, ...x_{76}, y_{76}, z_{76}\right),\ 1 \le j \le 76\ , \tag{2}$$

where, $\boldsymbol{v}_j = \left(x_j, y_j, z_j\right)$ is the *j*-th landmark.

2) Removing the translational component for each shape by subtracting the mean of all landmarks (i.e., $\boldsymbol{v}_j \leftarrow \boldsymbol{v}_j - (1/76)\sum_{j=1}^{76} \boldsymbol{v}_j$).

3) Choose the 1st shape $\boldsymbol{s}_1$ as the reference, i.e., $\boldsymbol{s}_r \leftarrow s_1$.

4) As relationship between $\boldsymbol{s}_r$ and $\boldsymbol{s}_i$ $(1 \le i \le n)$ is known, compute translational and rotational matrix $\boldsymbol{R}_i, \boldsymbol{T}_i$ of $\boldsymbol{s}_i$ by ICP. Update each $\boldsymbol{s}_i = \boldsymbol{R}_i \cdot \boldsymbol{s}_i + \boldsymbol{T}_i$.

5) Update the reference shape by $\boldsymbol{s}_r = (1/n)\sum_{i=1}^{n} \boldsymbol{s}_i$.

6) Compute the difference e between previous reference shape $\boldsymbol{s}_r$ and updated $\boldsymbol{s}_r^{'}$, if $e = \left|\boldsymbol{s}_r - \boldsymbol{s}_r\right|_2$ larger than a given threshold ε then repeat step 4 and 5, or goto step 7.

7) Compute the final average shape $\bar{\boldsymbol{s}} = (1/n)\sum_{i=1}^{n} \boldsymbol{s}_i$.

8) Using ICP once again to compute the final translational matrix $\boldsymbol{R}_i$ and rotational matrix $\boldsymbol{T}_i$ for each shape $\boldsymbol{s}_i$, which are used to adjust each face f_i .

Output: The corrected 3D faces $F' = \left\{f_1^{'}, ..., f_n^{'}\right\}$ and average shape $\bar{\boldsymbol{s}}$.

It's worth to mention that, with the average shape $\bar{\boldsymbol{s}}$, a new input 3D face can be quickly corrected by the simple ICP with one step after landmarks annotated. For a novel input, GICP can be run as quickly as SD method dose without any iteration.

3 Depth Image Normalization

For a 3D face, a fast and effective way to use 3D information is to create a depth image [2]. And in our experiments, depth image is effective enough for FR if created correctly. If not, the depth image will introduce errors in matching. From examples shown in the second column of Fig. 2(b), we can see that the top face is apparently smaller than the bottom one, which will cause miss matching, due to normalization. The main concern of this section is to find a right way to align depth images.

A common approach alignment in 2D uses the centers of the two eyes. The face is geometrically normalized using the eye locations to (i) scale the face so the inter-pupillary distance (IPD) between eyes is *l* pixels, and (ii) crop the face to *m***n* pixels.

We argue that IPD based alignment method is ill-suited for depth image. There are two apparent defects to normalize the depth image by IPD:

(1) Different persons have different IPDs, normalize the depth image by IPD will inevitably cause loss of the original metrics contained in 3D data.
(2) In terms of current technology, the same to 2D face recognition, pupil detection is sensitive to illumination or other factors.

We propose to map the original coordinates (X, Y directions) to a 2D space with fixed resolutions, e.g., 0.5mm per pixel. Since the original measurement data is

remained, we denote this image as measurement depth image (MDI). As shown in Fig. 2, IPD based depth image (shorted as IDI) and MDI have different appearances. As for IDI, the results are sensitive to the accuracy of pupils' location, which is always effected by illumination, see Fig. 2(b), e.g., reflection of glass may cause pupil detection fail. For MDI, however, we just put the nose tip to the center of the image, and the final image need not to scale to a specified resolution. In this way, MDI remain the real dimension of one subject. Then we crop it to a predetermined size.

In general, the position of nose tip is easy to obtain by finding the vertex with the largest Z value. Note that this method sometimes may fail to find the actual nose tip due to the burrs, which can be easily wiped off by a filter. So it is still the most significant geometrical feature in 3D space that is widely used in 3D FR area. Even we can develop more sophisticated method to find nose tip, but this is not the main concern of this paper. One can refer more detail from the references [4, 8].

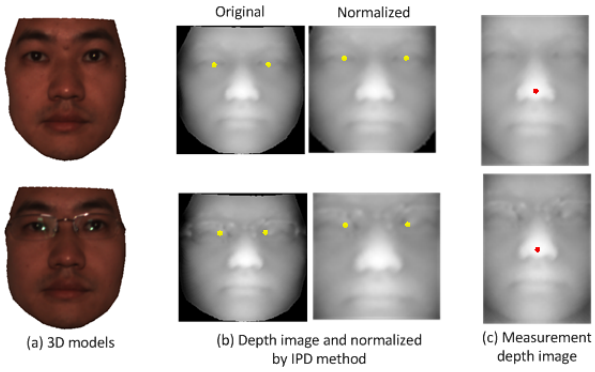

Fig. 2. Depth image normalized by two different manners. The yellow points in (b) denotes the pupils detected automatically. The red points in (c) are the nose tips.

4 Experiments and Discussion

4.1 3D Face Database

This section carries out experiments to validate our normalization method. At first, we create a 3D face database that consists of 450 persons with 6 models per subject.

Fig. 3. Capture system and one captured 3D face model with different views

The capture system setup is shown in Fig. 3 (a) and, the 3D face example obtained is shown in Fig. 3 (b). Our database consists of 450 persons due that some faces are not correctly segmented in the post-processing. Each person in the database has 6 models, including 5 different poses (frontal, up, down, left and right) and one model with random expression & lighting condition.

4.2 Evaluation of Posture Correction

This paper has proposed 2 kinds of posture correction methods in section 2, i.e., SD and GICP. As we known, without a special system to measure the actual pose of the face obtained, it's hard to assess the accuracy of correction results. We propose here to consider the pose evaluation as a texture matching problem. Local correlation matching (LCM) [10] is used as the similarity measurement. Matching function value closer to 1 once the two images are similar enough. After pose correction, 3D face model is mapped to a 64*64 depth image at first, and then we use LCM to measure the similarity between each pair of depth images of one person. Both SD and GICP methods are tested for all persons in the database, the result is illustrated in Fig. 4, where (a) is the average similarity of every person measured by LCM and (b) is its variance. It's clear that these two approaches are performing basically very similar. What's more important, the average similarity, for both methods, is larger than 0.98 and variance is less than 10^{-3}, which is summarized in Table 1. On conclusion can be drawn that both SD and GICP are appropriate for posture normalization.

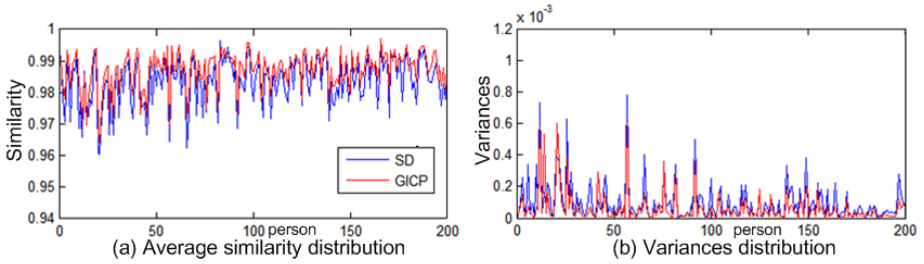

Fig. 4. Comparison the similarity and variance of depth image by 3D models after corrected by SD and GICP, respectively. Only 300 persons are shown in order to see the details more clearly.

Table 1. Average similarity and variance of data in Fig. 4

Methods	Average similarity	Average variance
SD	0.9835	**1.5704e-004**
GICP	**0.9863**	2.4624e-004

4.3 Evaluation of Two Depth Image in FR

In this part, we will illustrate the advantage of measurement depth image (MDI) over IPD based depth image (IDI). In our tests, Local Binary Patterns (LBP), Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM) and their combinations are used: LBP+LDA(LL), LBP+SVM(LS), and LBP+LDA+SVM(LLS).

At first, we compare 3 different methods to normalize the depth image: (1) IPD based depth image (IDI) with automatic pupil detection [14], denoted as IDI-A; (2) IPD based depth image (IDI) with manually selected pupil positions, denoted as IDI-M; (3) measurement depth image (MDI). Since the 2nd type of depth image need tedious manually annotation works, we just choose 100 persons in the test. Even in this case, we have to annotate 600 (100*6) images manually in total. Rank-1 recognition accuracy by LS and LLS are compared in Table 2 and Table 3. As we can see that MDI performs significantly better than the other two depth image.

Table 2. Recognition results of 3 different normalized depth images. FR method: LS

Depth image	Number of images per subject in training set				
	1	2	3	4	5
IDI-A	73.20%	90.50%	93.33%	91.00%	84.00%
IDI-M	72.80%	93.00%	97.67%	98.00%	97.00%
MDI	**81.60%**	**96.75%**	**99.00%**	**99.50%**	**99.00%**

Table 3. Recognition results of 3 different normalized depth images. FR method: LLS

Depth image	Number of images per subject in training set				
	1	2	3	4	5
IDI-A	93.00%	92.50%	94.67%	91.50%	84.00%
IDI-M	95.40%	96.75%	97.67%	98.00%	95.00%
MDI	**98.80%**	**99.25%**	**99.00%**	**98.50%**	**99.00%**

Performance of different FR methods is compared in Fig. 5, as we can see that LLS outperforms the other two methods. Bu LS achieves nearly the same accuracy when more than 1 images for training. Based on SVM theory, it can be easily understood that SVM could not get a valid hyper-plane for classification with a single training sample for each class. As the training number grows, LL performs as well as LLS.

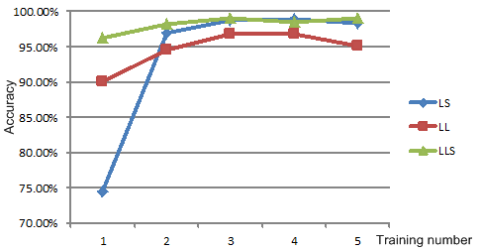

Fig. 5. Comparison of 3 FR method using MDI

5 Conclusions

This work is motivated by fact that 3D data capture system is capable to capture actual space coordinate. A 3D face itself contains sufficient discriminant information.

The main objective of this work is to demonstrate the potential of take full advantage of 3D data in face recognition. Without any feature alignment, FR accuracies using our depth images really exceed the accuracy using other normalized depth image.

In conclusion, this paper proves that even depth image could represent the 3D face information well if it is created in a right way.

Acknowledgments. The authors wish to thank Jiawei Sun for his constructive comments. This work is partially supported by the National Natural Science Foundation of China(61202191), the Fundamental Research Funds for the Central Universities(SWJTU12CX095), and Chongqing Key Laboratory of Computational Intelligence(CQ-LCI-2013-06).

References

1. Klontz, J.C., Jain, A.K.: A Case Study on Unconstrained Facial Recognition Using the Boston Marathon Bombings Suspects 2013, pp. 1–8 (2013)
2. Wang, Y., Liu, J., Tang, X.: Robust 3D Face Recognition by Local Shape Difference Boosting. IEEE T Pattern Anal, 32, 1858-1870 (2010)
3. Guo, Z., Zhang, Y., Xia, Y., Lin, Z., Fan, Y., Feng, D.D.: Multi-pose 3D face recognition based on 2D sparse representation. J. Vis. Commun. Image R 24, 117–126 (2012)
4. Mohammadzade, H., Hatzinakos, D.: Iterative Closest Normal Point for 3D Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(2), 381–397 (2013)
5. Gupta, S., Castleman, K.R., Markey, M.K., Bovik, A.C.: Texas 3D Face Recognition Database. In: Proc. 2010 IEEE Southwest Symposium on Image Analysis Interpretation (SSIAI), TX 2010, Austin, pp. 97–100 (2010)
6. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3D Facial Expression Database For Facial Behavior Research, pp. 211–216 (2006)
7. Chang, K., Bowyer, K.W., Flynn, P.: Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression. IEEE Trans. Pattern Analysis and Machine Intelligence 28(10), 1695–1700 (2006)
8. Wang, Y., Liu, J., Tang, X.: Robust 3D Face Recognition by Local Shape Difference Boosting. IEEE T Pattern Anal. 32, 1858–1870 (2010)
9. Milborrow, S., Nicolls, F.: Locating Facial Features with an Extended Active Shape Model. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 504–513. Springer, Heidelberg (2008)
10. De Marsico, M., Nappi, M., Riccio, D., Wechsler, H.: Robust Face Recognition for Uncontrolled Pose and Illumination Changes. IEEE Transactions on Systems, Man, and Cybernetics: Systems 43(1), 149–163 (2013)
11. Bonnen, K., Klare, B.F., Jain, A.K.: Component-Based Representation in Automated Face Recognition. IEEE Transactions on Information Forensics and Security 8(1), 239–253 (2013)
12. Besl, P.J., McKay, N.D.: A Method for Registration of 3-D Shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14, 239–256 (1992)
13. Gower, J.C.: Generalized procrustes analysis. Psychometrika 40(1), 33–51 (1975)
14. Valenti, R., Gevers, T.: Accurate Eye Center Location through Invariant Isocentric Patterns. IEEE T Pattern Anal. 34, 1785–1798 (2012)