Building Large Scale 3D Face Database for Face Analysis*

Yuxiao Hu, Zhenqiu Zhang, Xun Xu, Yun Fu, and Thomas S. Huang

Beckman Institute, University of Illinois at Urbana-Champaign
405 North Mathews Avenue, Urbana, IL 61801 USA
{hu3, zzhang6, xunxu, yunfu2, huang}@ifp.uiuc.edu

Abstract. We propose to build a large scale 3D face database with dense correspondence for variant face analysis research purposes. Large scale means that the number of subjects in the database is more than 400, which is, to our best knowledge, the biggest[1] one at this time. 3D face means that we provide both the texture and shape of human faces, which is also balanced in gender and race. Dense correspondence means that the key facials points with semantic meanings are carefully labeled and aligned among different faces, which can be used for a broad range of face analysis tasks. We provide the data description, data collection schema and the post-processing methods to help the usage of the data and future extension. More and more data is still being collected and processed to enlarge the extensive 3D face database. The proposed face database provides solid ground truth for human face related tasks such as alignment, tracking, recognition and animation, etc.

1 Introduction

Human faces are the most popular and important objects in the real world. Computer graphics researchers are interested in face modeling, synthesis, and animation, while computer vision researchers are investigating the methods for face analysis, including face detection, alignment, tracking and recognition. Though research on frontal detection and recognition has achieved a lot, non-frontal face analysis, especially face alignment and recognition still remains open problems. In order to address these issues, more and more attentions are attracted to 3D face analysis. As the foundation to all these tasks, building a comprehensive 3D face databases is necessary.

There are many existing databases containing 3D face information. According to their construction methods, these databases can be categorized to three types: built from multi-view geometry, built from structure lights and captured by 3D scanners. The CMU Face In Action (FIA) Database [1] consists of 20-second videos of face data from 180 participants mimicking a passport checking scenario. The data is captured by six synchronized cameras from three different angles. Since it only includes

* This work is supported by NSF Grant CCF 04-26627.

[1] Since the data format is also compatible to some other database of the same type. The total number of the available subject is almost 700 now.

N. Sebe, Y. Liu, and Y. Zhuang (Eds.): MCAM 2007, LNCS 4577, pp. 343–350, 2007.

unmarked face video sequences, sophisticated face tracking algorithms need to be applied to recover the head poses and face shapes. Some other 3D face databases are collected by structure lights. For example, in 3D_RMA [2], it contains 120 individuals. The 3D face shape was constructed using a camera and a projector. The 3D coordinates of the surface points were represented with high precision. The limitation of this method is that glasses and dark parts of faces could not be captured, resulting the shapes of the face are not complete. 3D scanners are widely used recently, which are able to provide precise shape and texture information of an object. In GavabDB [3], 61 individuals are scanned from different views with variant expressions. Since the Minolta VI-700 digitizer in [3] only captured one view of the face, there are some invisible parts in a single view due to self occlusion. Hence extra post-processing must be carried out to concatenate the shape and texture map from different views to get a complete face surface. Compared with the aforementioned scanners, a Cyberware Scanner [4] **Error! Reference source not found.**is able to scan the complete view of human heads in one scan. It is convenient both for the data collection and post processing. Vetters et al. provided a 3D morphable face database based on USF 3D face database, which included 200 subjects [5][6]. However, comparing to the high dimension of the 3D shape and texture data, the number of the samples is still far from enough for future face analysis tasks.

In this paper, we propose to build a larger 3D face database with Cyberware 3D scanner which includes more than 400 individuals. Different to the registration method used by Vetters, the raw 3D face shapes are aligned manually instead of by optical flow, which provide more accurate results. Currently, there are already 475 subjects in our 3D face database with both neutral and smile expressions available for most of them. The ages of the subjects range from 19 to 55 and balanced in ethnic group. The database can be used for a broad range of face analysis tasks including alignment, tracking, recognition and animation, etc. The data description, data collection schema and the post-processing methods are also provided to help the usage and future extension of the database.

2 The Data Collection Schema

In this section, we first brief our data collection schema. About 500 subjects were invited for the 3D face scanning during February to May in 2006, most of them are students. The subjects were carefully selected to balance gender. About half of the subjects were western people including Caucasia and African American ethnic groups and the other half was Asian. The ages of the subjects were also recorded for future reference, which ranged from 19 to 55. The subjects were asked to show neutral expression during the scanning and for most of them, another scan for smile expression was collected for possible future recognition purposes.

A 3D human face was first scanned by Cyberware scanner, which output the raw data including the head shape and corresponded texture map. Then 65 pre-defined key facial points, such as eye/mouth corners and nose tip, were labeled on the texture map manually. According to these key points and the reference mean face model, the face region was cropped out and the dense triangular mesh shape model was created by interpolation so that the vertices in different face model corresponded with semantic meaning. Finally, the poses of the 3D faces were normalized so that their 3D coordinates were in the same reference frame.

The steps of the 3D face capture and post-processing are detailed in section 3.

3 Three Dimension Face Capture

3.1 Cyberware Scanner

To capture the 3D human face data, we used the Cyberware Head & Face Color Scanner bundle, including a Cyberware 3030 scanhead mounted on a Cyberware PS cylindrical motion platform. The 3030 shines an off-axis low-intensity laser illuminating a profile captured from two video sensors to triangulate its spatial location. Thousands of these profiles are combined to construct a single range scan dataset. The scanner additionally captures a 24-bit RGB texture map using a cylindrical surface parameterization.

3.2 Data Capture Environment

The CyberWare3030 scanner provides an ambient illumination system. We shielded all the other lights from the window and in the room to get consistent and uniform ambient rays so that the illumination condition is reasonable and consistent at each time we scan a new subject.

3.3 Accessories

The result of the scanner is not good when it scans complex structure objects such as hair and black objects. Since only face region is concerned, every subject was asked to wear a wig cap and take off glasses. The cap is white, blue or black which is different from the skin color of the subject. The cap shields the hair above the forehead and behind the ears, and keeps the face region at the most.

Some examples about the raw 3D data are shown in Figure 1.

4 Data Labeling and Post-Processing

The raw 3D data from the Cyberware scanner is in resolution of 734*457 for both the shape and texture map. The raw shape contains 734*457 vertices and each vertices is represented by x, y and z coordinates. The raw texture is represented with 734*457 .tiff image. Because the absolute position and orientation of the head are different for each subject, the raw data need to be labeled and post-processed to extract the useful human face information. In this section, we detailed the steps of labeling the key points, cropping the face region, building sparse/dense correspondence and aligning the face model.

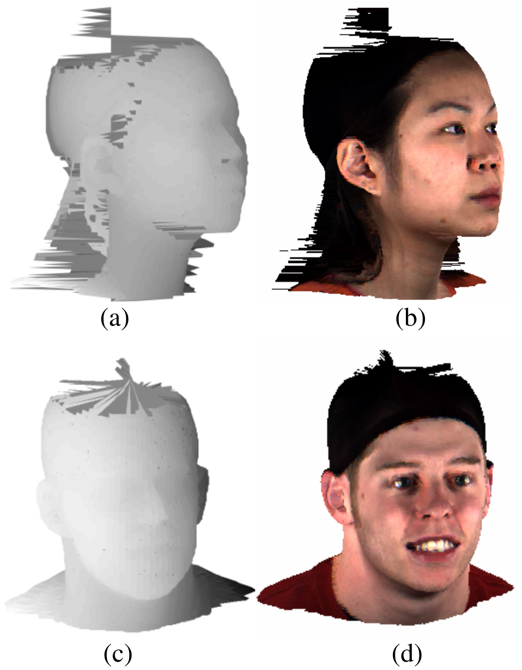

Fig. 1. Samples of the raw 3D data from Cyberware scanner. (a)(b) Shape and Texture of Sample#304, Female, in neutral expression (c)(d) Shape and Texture of Sample#310, Male, in smile expression.

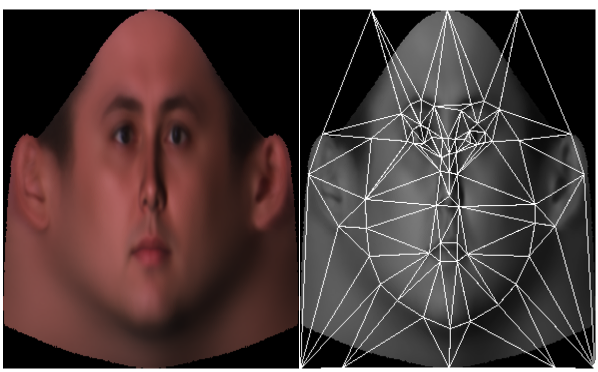

Fig. 2. The Reference Face and the Key Points

4.1 Data Labeling and Cropping

Typically the face region occupies an area about 500*400 within the whole scanning cylinder. In order to specify the face regions and build the correspondence between

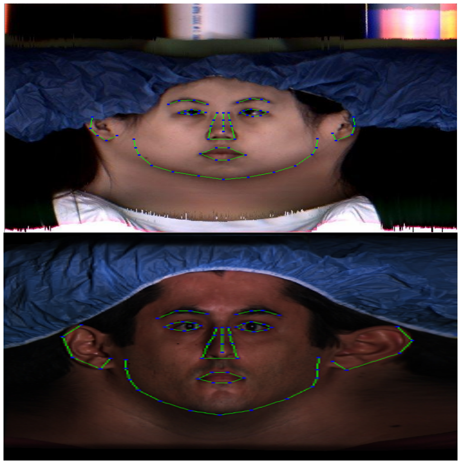

Fig. 3. The texture map with the key points labeled with blue dots. The key points are connected with green lines in different components.

different faces, a reference 3D face model is used. 75 key facial points on the reference face model is defined as anchor points, including eye corners, mouth corners, nose tip and face contour, etc. The sample data are also labeled on these anchor points manually. The reference face and the pre-defined key points are shown in Figure2. The triangular configuration is also defined for future warping as stated in Section 4.2. Again, since only face regions are concerned, the vertices and texture corresponding to the region outside the reference face are cropped. Some labeled face texture map are shown in Figure 3. We notice that some of the key points are not labeled compared to the reference face. These points will be calculated in the post-processing steps according to other labeled points and will be used for face warping in the next step.

4.2 Building the Correspondence

Because different subjects' heads are in different position and orientation when they are scanned, in the raw 3D data, the vertices and texture information of different subjects are unrelated. For example, the *i-th* vertex in one 3D face is on the nose tip, but the *i-th* vertex in another 3D face is on the cheek, i.e., these two vertices have different semantic meanings. Since vertices correspondence are preferred in typical face research tasks, especially for facial animation and subspace analysis, re-sampling the vertices and texture map is required to make the vertices "*corresponded*", i.e., the *i-th* vertex in different 3D faces are all nose tips and the *j-th* vertex in different 3D faces are all left corner of the mouth.

Since the 75 key facial points with semantic meanings have been located manually in the texture map, a sparse correspondence is already established. Other points within the face region can be interpolated and warped to the reference face. The points which can be warped to the same points on the reference face are corresponded. Because

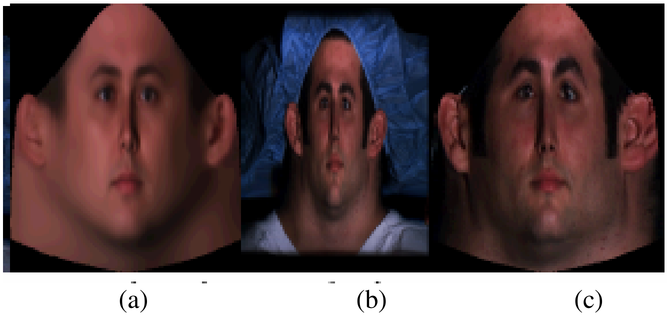

Fig. 4. Verification the result of texture correspondence: (a) ref. face (b) input sample face (c) warped sample face

Fig. 5. Verification of result of shape correspondence: The first row is the original raw face shape, the second row is the reconstructed face shape with dense correspondence.

each point in the texture map corresponds to a vertex in the shape model, the correspondence between two vertices from different 3D faces are also established under the help of reference face.

There are two aspects to be verified for the result of our data labeling and the dense correspondence. On texture maps, we warp a sample face to the reference face according to the established correspondence. If the facial points are aligned perfect according to their semantic meanings, the result warped face should looked in the same shape to the reference face, but the texture is the same to the original sample face. On the contrary, if the facial points are not aligned well, there will be some distortions. A sample of the face warping is shown in Figure 4. For shape vertex, the vertices actually are re-arranged to represent the 3D face shape. If the shape correspondence is perfect, the 3D shape constructed with the re-arranged 3D vertices should be the same shape to the original raw 3D face shape. On the contrary, the reconstructed 3D shape will be different to the original data. As shown in Figure 5, the 3D face shape remains the same before and after the processes.

Fig. 6. Key facial points plots before/after the Procrustes Analysis. Left: before, Right: after.

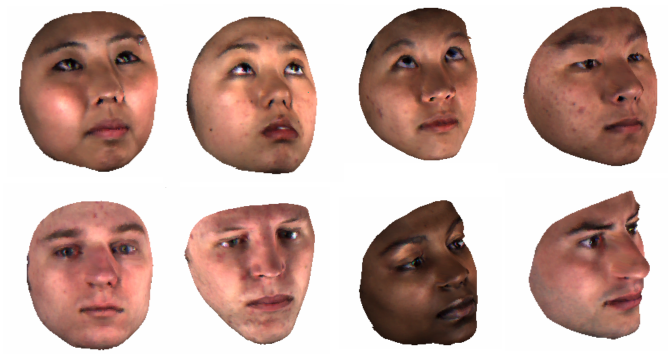

Fig. 7. Samples of the 3D face after post-processing

From above verification, we may claim that building the dense correspondence did not affect the original 3D face shape and texture.

4.3 Vertex Coordination Normalization

Although the vertices are already in the same semantic meaning for different 3D faces after the dense correspondence has been established, their coordinate systems are still not consistent, i.e., the faces are not in the same poses if we put them in a 3D space. In order to normalize the 3D pose (location, rotation and scale) of the collected shape data so that they achieve the best alignment with each other, 3D Procrustes analysis [7] is applied. After the 3D Procrustes analysis, all the 3D face shapes vertices are transformed to the same coordinate system, in which all the corresponded face vertices are well aligned. The 3D face shape vertices before and after the 3D Procrustes analysis are compared in Figure 6. In the left plot, we plot the key facial points of several faces together, the 3D points scattered around because are not well alignment. The same set of points is plotted again in the right plot after Procrustes analysis, which is more consistent. (Because of the individual variance, they will not completely overlap.)

Samples of the 3D face data after post-processing are shown in Figure 7.

5 Data Organization

Both the original raw data and the normalized 3D face data with dense correspondence are available with the information of the subjects, including age, gender and ethic group. The shape and texture map of the raw data are stored separately, which occupied about 3GB disk space. Each file pair is named by the subject number with suffix "-N" or "-S" for different expression. After post-processing, there are 33420 vertices/points for each face, which is about 400MB in file size.

6 Conclusion

A large scale 3D face database with dense correspondence was described in this paper. The database contains more than 400 subject which balances in gender and race. Dense correspondence and alignment among different faces are also established, which makes the proposed database is useful for a broad range of face analysis tasks. The data description, collection schema and the post-processing methods are detailed to help its usage and for future extension. Since the data format is compatible to some existing databases [8], more subjects can be added and processed to further enlarge it. In Conclusion, the proposed face database provides solid ground truth for human face related tasks such as alignment, tracking, recognition and animation, etc.

References

1. Goh, R., Liu, L., Liu, X., Chen, T.: The CMU Face. In Action (FIA) Database, In: the Proceedings of the 2005 IEEE International Workshop on Analysis and Modeling of Face and Gestures, in conjunction with the 2005 International Conference of Computer Vision (ICCV '05), Beijing, China, October (2005)
2. http://www.sic.rma.ac.be/~beumier/DB/3d_rma.html#1
3. Moreno, A.B., Sanchez, A.: GavabDB: A 3D Face Database. In: Proc. 2nd COST275 Workshop on Biometrics on the Internet, Vigo (Spain) (2004)
4. http://www.cyberware.com/products/scanners/ps.html
5. Blanz, V., Vetter, T.: A Morphable Model for the Synthesis of 3D Faces, In: SIGGRAPH'99 Conference Proceedings, pp. 187-194 (1999)
6. USF DARPA HumanID 3D Face Database, Courtesy of Prof. Sudeep Sarkar, University of South Florida
7. http://www.dcs.gla.ac.uk/ mc/1stYearReport/2.5_GPA.htm
8. The BJUT-3D Large-Scale Chinese Face Database, Technical Report of The Multimedia and Intelligent Software Technology Beijing Municipal Key Laboratory in Beijing University of Technology (2005)