Face Shape Reconstruction from Image Sequence Taken with Monocular Camera using Shape Database

Hideo Saito, Yosuke Ito
Dept. Information and Computer Science, Keio University
Yokohama 223-8522, Japan
saito@ozawa.ics.keio.ac.jp

Masaaki Mochimaru
Digital Human Research Center, AIST
Tokyo 135-0064, Japan
m-mochimaru@aist.go.jp

Abstract

We propose a method for reconstructing 3D face shape from a camera, which captures the object face from various viewing angles. In this method, we do not directly reconstruct the shape, but estimate a small number of parameters which represent the face shape. The parameter space is constructed with Principal Component Analysis of database of a large number of face shapes collected for different people. By the PCA, the parameter space can represent the shape difference for the faces of various persons. From the input image sequence that is captured by the moving camera, the parameters of the object face can be estimated based on optimization framework. The experiments based on the proposed method demonstrate that the proposed method can reconstruct the facial shape with the accuracy of 2.5mm averaged error.

1. Introduction

In this paper, we propose a method for 3D shape reconstruction of face using a handy camera that captures the face from multiple viewpoints. A supposed application of the reconstructed face shapes is order-made manufacturing of products worn on the face, such as glasses, masks, etc. Conventional way to obtain 3D shapes for such applications is based on special 3D shape scanning devices. However, scanning the face shape via such 3D scanning devices is not easy for all customers. Especially, we need to avoid using the device with the laser radiated onto the face surface of normal customers, because of safety reasons. If 3D shape can be easily acquired via general purpose cameras such as web-cameras, mobile-phone-cameras, the order-made manufacturing products can be easily available for normal customers.

In this paper, we propose a method for 3D shape reconstruction of face that can be applied to such purpose. In this method, a handy camera is used for capturing multiple-view images of the face, which is used as input information to reconstruct 3D facial shape. The shape reconstruction is performed based on database that includes 3D shapes of a number of real faces that are previously collected. The database contributes to reduce the degree of freedoms of the facial shape, so that accuracy of 3D shape reconstruction can be sufficient to be used for the supposed applications.

Since the proposed method can reconstruct 3D shape with only a handy camera, it is very easy for all customers to acquire their own 3D facial shapes. The personal 3D facial shape acquired by the proposed method can also be used to even on-line shopping of face-worn products.

2. Related Works

In most of the conventional method for facial shape acquisition, special devices, such as multiple cameras [18, 12, 15, 6], laser range scanners [3], and cameras with projectors [13, 18], are normally used as input devices. However, such special data input devices are not easy to be prepared just for fitting of face-worn products such as glasses and masks. Our proposed method aims to acquire 3D shape of faces with just a general purpose camera, but we do not want to compromise in terms of accuracy. To perform accurate 3D shape acquisition with a handy camera input, we employ the database of 3D facial shapes collected for a number of persons. The database enables to accurate shape acquisi-

14th International Conference on Image Analysis and Processing (ICIAP 2007)
0-7695-2877-5/07 $25.00

tion even with a handy camera, because PCA analysis of the database reduces the degree of freedoms of the reconstructed facial shape [8, 7, 10].

Since the database of our method is the dataset of anatomical 3D shape of the face, our method is especially suitable to the application to fitting face-worn products to the individual face. If we collect the database of facial 3D shape by simply scanning faces, we do not have any information of relation among the scanned data. Our database already has such relation among the data as the anatomical correspondence, which also represents important individual facial shape features in terms of osteology. Therefore, the data is suitable to fitting face worn products. The same database is also used in [16], but the only one input image is used for face shape reconstruction, so it is difficult to obtain accurate facial shape reconstruction with easy operation.

The face shape measurement with camera input is a popular topic in the field of image analysis. Blanz et al. proposed a method of face shape recovery from monocular camera for face recognition[1]. They applied PCA to the database of a number of 3D facial shape with 2D texture for defining the morphorable model, and then the parameters of the morphorable model is estimated from the input image for recovery of facial shape. Blanz et al. also proposed a method for face recognition [2] using the face shape recovery. Jiang et al. proposed a similar method with the Blanz method, in which they recover the facial shape from about 30 feature points extracted from an input 2D image[8]. Even though facial shape recovery method based on PCA has already been proposed in those papers, they do not give any accuracy evaluation, but just demonstrate qualitative performance.

Moghaddam et al. proposed a method for facial shape recovery based on 2D silhouettes of face from image sequence captured with a camera[11, 10]. This method tracks the face position and the pose for every input image based on feature point tracking, which might not be stable and accurate. Chowdhury et al. [4] proposed shape recovery method based on a sort of structure from motion, which might not easy to achieve sufficient accuracy and robustness.

3. Proposed Method

3.1. Capturing Environment

In this method, we assume that a marker is attached onto the top of the object face area, which is captured with a handy camera as shown in Fig. 1. It is not easy to robustly estimate the camera motion from the image sequence taken with a handy camera without any information of geometrical structure of the object scene. Therefore we compromise to use the marker attached on the head, but it is not so difficult procedure for the user to simply put the maker onto his/her head. The marker used in this paper is the marker of the ARToolkit[9], which helps to estimate camera motion relative to the marker at each frame in the input image sequence. Figure 8 shows some examples of the captured image in this method.

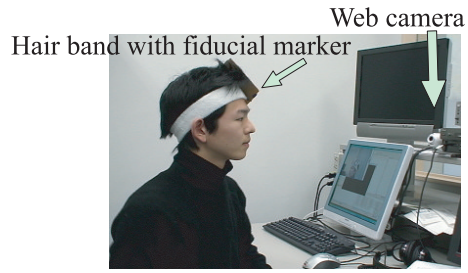


Figure 1. Experimental environment.

We also need to estimate the geometrical relationship between the marker and the object face coordinate, because it is difficult to guarantee that the marker is always attached at the same position and the pose. For estimating this relationship, the user should manually detect four feature points in two key-frames. The feature points are also used for recovery of the face shape.

3.2. 3D Active Shape Model

In this method, we use a database of a number of the actually measured facial shapes as a dataset for determining the shape parameter space via PCA of the database. In the parameter space, the face shape can be represented by 3D Active Shape Model [5, 14] with small dimension parameter space.

The 3D Active Shape Model (ASM) represents the object shape as the following equation.

$$v = \bar{v} + PB, \tag{1}$$

where

$$v = (x_1, y_1, z_1 \cdots x_s, y_s, z_s)\,,$$

which is the object shape represented by the 3D positions of the vertices of the shape. $\bar{v}$ indicates the averaged shape of in the database. $P = (p_1, p_2, \cdots p_q)$ and $B = (b_1, b_2, \cdots b_q)$ are the principal component shapes and the coefficient up to qth dimension, respectively. According to ASM, we can represent the facial shape using just q parameters.

The database in this paper consists of 52 shapes of faces for 20's Japanese males that are represented with 430 vertices and 840 meshes. Each vertex is corresponded for all the shape data. 100 vertices from 430 are determined by touching diagnosis of experts of anatomy.

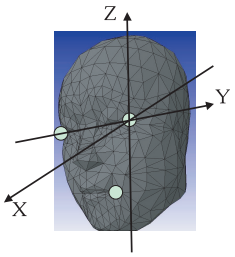


Figure 2. Definition of database coordinate system.

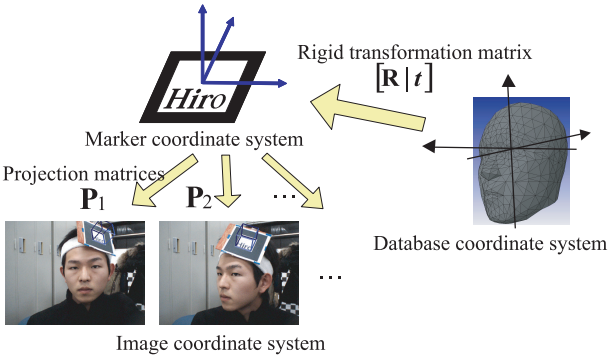
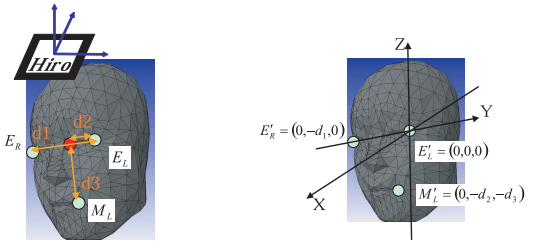


Figure 3. Two steps for projecting reconstructed model onto input image plane.



(a) Marker coord. system (b) Database coord. system

Figure 4. Three feature points used to estimate rigid transformation matrix.

For the database, we determine q as 20, with which the total contribution rate is over 90%. This means that we represent the facial shape with 20 parameters, which are determined via optimization procedure in this method. In this optimization process, the candidate parameter vector is evaluated by comparing the silhouettes of the input image sequence with the projected shape represented by the parameter vector.

All the data in the database are aligned to the database coordinate, which is defined based on the end points of the both eyes, and left end point of the mouth. The database coordinate is defined as shown in figure 2.

3.3. Registration of coordinates

In the optimization procedure, the shape model represented by the parameter vector should be projected onto the image so that we can compare the silhouette. For this purpose, we need to know the rigid transformation from the database coordinate to the marker coordinate, and then we can project the shape onto the image coordinate according to the projection matrix from the marker coordinate to the image coordinate. Figure 3 shows the relationship among the three coordinates.

The projection matrix from the marker coordinate to the image coordinate is represented by the extrinsic parameter matrix $[R|t]$ and the intrinsic parameter matrix A. We employ Zhang's method [17] to estimate the intrinsic parameter matrix A. Since we can assume that the intrinsic parameters do not change as long as the same camera is used, the estimation of A is only required once as a preprocessing. On the other hand, the extrinsic parameters should be estimated frame-by-frame as an on-line process, because they dynamically change according to the camera motion. Therefore, we compute the extrinsic parameter matrix $[R|t]$ by using the function implemented in the software of ARToolKit[9].

The rigid transformation between the database coordinate and the marker coordinate is estimated based on three feature points that are corresponded by manually. The selection of features is performed on two key frames in the input image sequence. The three feature points are indicated in fugure 4.

3.4. Optimization of shape parameter

The purpose of this method is to estimate optimum parameters that represent the shape of the object face according to the 3D Active Shape Model. As described before, we use the top 20th principal component shapes. This means that 20 dimension parameters should be estimated with the optimization framework. We consider that the face shape represented by the optimum parameters provides the fittest silhouette with the input images.

The optimization starts with the initial estimated parameters. The face shape represented with the estimated parameters B is projected onto the input image plane based on the camera pose and positions computed with the AR Toolkit markers as described in the previous section. By comparing the shape of silhouette projected from the represented shape model and the silhouette captured in every input images, the

0-7695-2877-5/07 $25.00

Figure 5. No truncation of silhouettes

Figure 6. Truncation of Silhouettes

parameters B is evaluated. By repeating the evaluation of various B, we can estimate the optimum parameters.

3.4.1 Evaluation of estimated parameters

We use the combination of two different evaluations. The first evaluation is the fitness of the silhouette shape projected from the represented shape with the silhouette in the input image. The second evaluation is the fitness of the position of the feature points between the projected image from the shape and the input image. The total evaluation E_{total} can be indicated by the following equation.

$$E_{total}(B) = w_1 E_{silhouette}(B) + w_2 E_{feature}(B) \quad (2)$$

where, $E_{silhouette}$ and $E_{feature}$ represent the first evaluation based on the fitness of the silhouette shape, and the second evaluation based on the fitness of the position of the feature points, respectively. w_1 and w_2 indicate the weight coefficients that balance both evaluations. $B = (b_1, b_2, \cdots b_{20})$ shows the estimated shape parameters, which is the vector of the top 20th coefficient for the principal component shapes.

The first evaluation based on the fitness of the silhouette, $E_{silhouette}$, is computed by the Boundary-weighted XOR-based cost function (B-XOR) proposed by Baback et al. The area of the XOR can properly evaluate the similarity between two silhouette shapes if there is no truncation in both silhouettes as shown in figure 5. However, if there are some truncated areas in the two silhouettes, simple XOR cannot properly represent the similarity of the silhouette shapes, because the un-overlapped area will give un-expected penalty to the XOR as shown in figure 6. To reduce such penalty of XOR for non-overlapped area, B-XOR changes the XOR value according to the distance from the boundary.

The second evaluation is based on the fitness of the positions of the four feature points, which are manually specified by the user in two key frames as a pre-process of the proposed method. For the evaluation, the 3D positions for

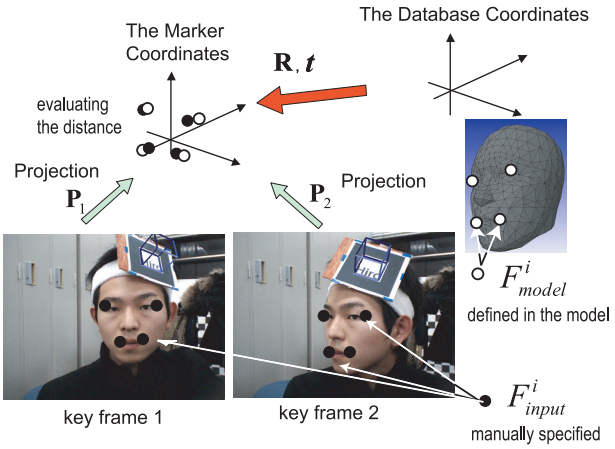


Figure 7. Computing the 3D positions of four feature points in the marker coordinates.

the four feature points $F^1_{input}, \cdots, F^4_{input}$ in the marker coordinate are computed, which are indicated as the black points in figure 7. On the other hand, the feature points on the shape represented by the evaluated parameters are represented $F^1_{model}, \cdots, F^4_{model}$, which are indicated as the white points in figure 7. The average of the Euclid distances of the four feature points between $F^1_{input}, \cdots, F^4_{input}$ and $F^1_{model}, \cdots, F^4_{model}$ is finally computed as the evaluation $E_{feature}$.

3.4.2 Optimization algorithm

For optimization of the evaluation function, we employ the downhill simplex method, which does not need the gradient of the evaluation function that cannot easily be computed for our evaluation function. The downhill simplex method needs to keep $N + 1$ parameter sets for blanketing the estimated parameters with N dimension in optimization. In this method, we use 20 dimensions, so we need to keep 21 parameter sets from the initialization of the optimization. The initial parameter sets are given by randomly selected 21 face shapes in the database consists of 52 face shapes.

4. Experiments

In order to demonstrate effectiveness of the proposed method, we performed experiments to reconstruct the real face shape using handy camera.

The shape database in this experiment includes 52 face shapes of 20's Japanese males. The object face shape for the shape reconstruction is not included in the database, but 20's Japanese males same as the database.

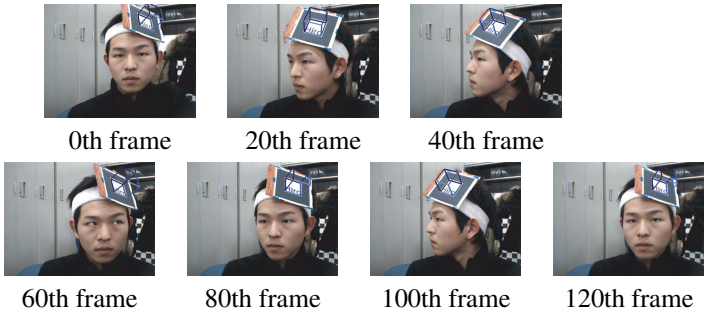


Figure 8. Selected seven images for estimating the parameters.

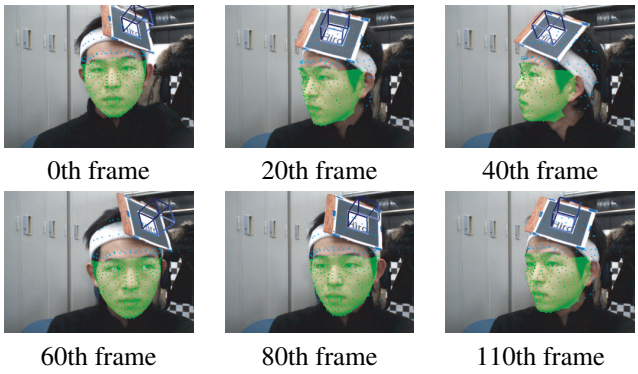


Figure 9. Projection of the shape represented by the estimated parameters.

A marker is attached onto the head area as shown in figure 8, which is a video sequence captured with handy video camera. The number of frames of this experiment was 130. From the input images, seven frames including the two key frames shown in figure 8 are selected for estimating the parameters representing the object shape based on the optimization. The shape represented by the estimated parameters are projected onto the input images, which are shown in figure 9.

Figure 10 indicates the transition of the evaluation value during the iteration of computing evaluation for optimizing the parameters. As shown in this figure, the evaluation value converges until 500 iterations.

Next, we evaluate the accuracy of the shape represented by the estimated parameters. As a reference shape, we measure the shape of the object face by using a 3D shape scanner, Cartesia FACE SYSTEM manufactured by Spacevision Inc [19].

Table 1 shows the shape error of the shape represented by the estimated parameters compared with the scanned shape.

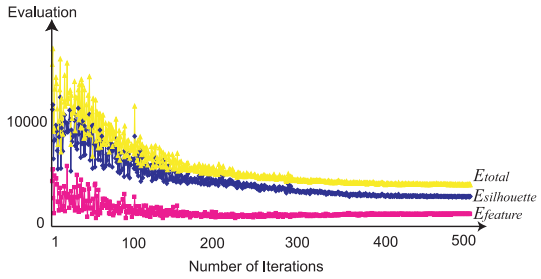


Figure 10. Transition of the evaluation value.

Table 1. Comparison of reconstruction errors before and after optimization of shape parameters.

	Average absolute (mm)	Variance (mm^2)
initial	4.015	3.014
estimated	2.518	2.475

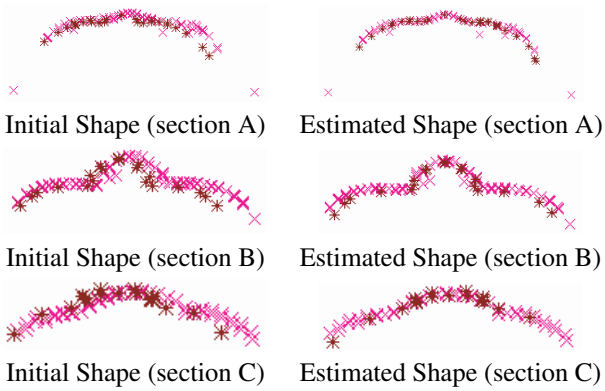
The error is much smaller than for the initial face shape in optimization, which means that the proposed method provides accurate shape of the object face.

The error with the scanned 3D shape is also represented by the cross-section plots as shown in figure 11. We can also observe that the estimated shape is much closer to the scanned shape than the initial shape.

Figure 12 indicates the distribution of the error with the scanned shape for the estimated shape. The plots are colored by the absolute errors for the plots. The error around the face area of the face is especially small (0-3mm), while the error around the only boundary area is larger. This error is sufficiently small to use the proposed method for fitting glasses, or other face-worn products.

5. Conclusion

In this paper, we propose a novel method for reconstructing face shape from multiple images taken with a handy camera. This method estimates the parameters represent the object face shape in the parameter space spanned by the principle component shapes computed with PCA from the database including a number of 3D facial shape data. We demonstrate that the proposed method can reconstruct object face shape within the average error of 2.5mm by just using a camera and a marker for camera pose and position estimation. The accuracy achieved by the proposed method is sufficient for the practical use as the measurement of the individual face shape for order-making facial worn products



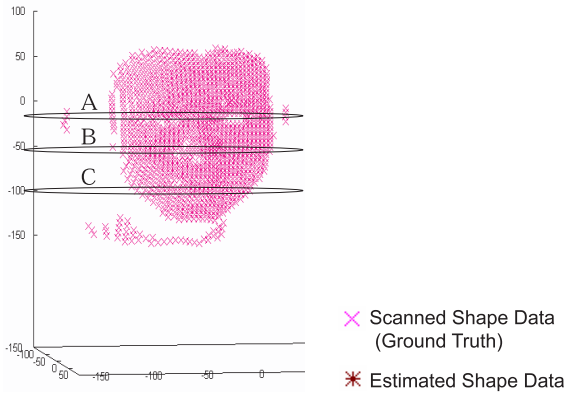


Figure 11. Comparison with the scanned shape data for the cross sections.

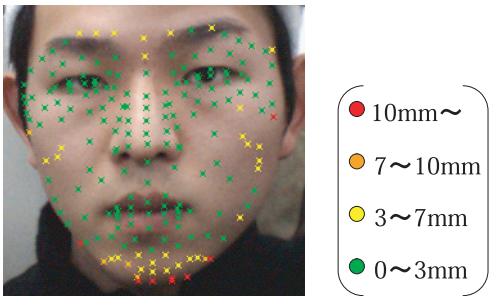


Figure 12. Error distribution of estimated shape.

such as glasses.

References

[1] Volker Blanz and Thomas Vetter: "A Morphable Model for the Synthesis of 3D Faces". Proceedings of the SIGGRAPH'99, Los Angeles, USA, pp.187-194, August 1999

[2] Volker Blanz and Thomas Vetter: " Face Recognition Based on Fitting a 3D Morphable Model". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, pp. 1063-1074, September 2003.

[3] M. A. Brunsman, H. Daanen, K. M. Robinette: "Optimal postures and positioning for human body scanning". Proceedings of 3DIM 1997, pp.266-273, 1997

[4] Amit K. Roy Chowdhury and Rama Chellappa: "Face reconstruction from monocular video using uncertainty analysis and a generic model" Computer Vision and Image Understanding 91, pp.188-213, 2003

[5] T. F. Cootes, C. J. Taylor, D. H. Cooper: " Active shape models their training and application". Computer Vision and Image Understanding, 61(1), pp.38-59, 1995

[6] H. Ip , L. Yin: "Constructing 3D Individualized Head Model From Two Orthogonal Views". International Journal in Computer Graphics, 12(5), pp.254-266, 1996

[7] P. A. Griffin, N. Redlich: "Statistical Approach to Shape from Shading : Reconstruction of 3D Face Surfaces from Single 2D Images". Neural Computation, 8(6), pp.1321-1340, 1996

[8] D. Jiang, Y. Hu, S. Yan, L.Zhang, H. Zhang, W.Gao: "Efficient 3D reconstruction for face recognition" Journal of Pattern Recognition 38, pp.787-798, 2005

[9] H. Kato, M. Billinghurst: "Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System". Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99), 1999.

[10] Jinho Lee, Baback Moghaddam, Hanspeter Pfister, Raghu Machiraju: "Finding Optimal Views for 3D Face Shape Model". Proc. FGR2004, Seoul, Korea, pp.31-36, 2004.

[11] Baback Moghaddam, Jinho Lee, Hanspeter Pfister, Raghu Machiraju: "Model-Based 3D Face Capture with Shape-from-Silhouettes". Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG), pp. 20-27, 2003

[12] Davide Onofrio, Stefano Tubaro, Antonio Rama, Francesc Tarres: "3D Face Reconstruction with a Four Camera Acquisition System". Proceedings of International Workshop on Very Low Bitrate Video Coding (VLBV05), 2005

[13] J. Siebert, S. Marshall: "Human body 3D imaging by speckle texture projection photogrammetry". Sensor Review, 20(3), pp.218-226, 2000

[14] Jiahui Wang, Hideo Saito, Makoto Kimura, Masaaki Mochimaru, Takeo Kanade: " Shape Reconstruction of Human Foot from Multi-Camera Images Based on PCA of Human Shape Database". Proc. 3DIM2005, pp.424-431, 2005.

[15] S. Weik: "A Passive Full Body Scanner Using Shape from Silhouettes". Proc. ICPR2000, pp.1750-1753, 2000

[16] K.Yoshiki, H.Saito, M. Mochimaru: "Reconstruction of 3D Face Model from Single Shading Image Based on Anatomical Database". Proc. ICPR2006, Vol. 4, pp.350 - 353, 2006.

[17] Z. Zhang: "A flexible new technique for camera calibration". IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11): pp.1330-1334, 2000

[18] L. Zhang, N. Snavely, B. Curless, S. Seitz: "Spacetime faces: high resolution capture for modeling and animation". ACM Transaction on Graphcs 23(3): pp.548-558, 2004

[19] Cartesia FACE SYSTEM, http://www.space-vision.jp