3D Reconstruction from a Single Image for a Chinese Talking Face

Ning Liu[1], Ning Fang[2]
School of Electronic, Information and Electric Engineering[1]
School of International Education [2]
Shanghai Jiao Tong University
Shanghai, China
{lnaji2001[1], nfang[2]}@sjtu.edu.cn

Seiichiro Kamata
Graduate School of Information, Productions and System
Waseda University
Kitakyushu, Japan
kam@waseda.jp

***Abstract*-This paper proposes an automatic 3D reconstruction approach for a Chinese talking face by a generic model and a single image. Firstly, an improved color-based ASM method is used to detect the face area and get the 2D face feature points automatically from the given image, which is not restricted to full frontal one. Then, color information is used to correct the location of face feature points. Finally, after text mapping, a particular and realistic 3D face model is deformed from a generic model. Using ASM face feature points extraction and correction based on skin color model, the problem of side face information missing is successfully resolved. Depending on using only one image and one generic model, the computing cost of memory and time is largely reduced. The 3D face reconstructed can be easily deformed to form different expressions and mouth shapes. Experiments show that this approach is fast and efficient and has an output of a lifelike Chinese talking face.**

***Keywords*- Talking Face; 3D Reconstruction; Skin Model; Feature Points Extraction; Texture Mapping**

I. INTRODUCTION

Recently, "talking face" becomes a very active research area in HCI (Human-Computer-Interface) field. A "talking face" gives people a lifelike face when listen to the computer, which can make HCI more friendly. Research shows that add visual information into speech can increase the rate of understanding for people with hearing problem.

Though many researchers have made deeply work on 3D face reconstruction, it is still a difficult and interesting field because of the complication of a face. So far, there are mainly two ways to reconstruct a 3D face: one is based on a 3D model while another is based on a face image database. Using the latter method you have to create a human face database which includes parts of a face and put together the needed parts to compose a face. This method has a lifelike output face but it needs great computation and the output face can not be changed freely. Using the former method you should create a generic 3D face model first. Then, the model can be deformed into a particular 3D face according to the input face image. The 3D face can be changed easily and can be synthesized at real time. But the output face is not lifelike [1, 2].

Blanz and Vetter [3] demonstrate a morphable model method based on a 3D face database: firstly, they create a morphable model, which includes a shape model and a texture model, from a set of 3D faces in the database; then new faces can be modeled by forming linear combinations of the prototypes. The output face is lifelike but the computation based on the 3D database is too complicated and takes too much time. Jiang [4] presents an improved morphable model method: face feature points instead of dense face model are used. It decreases computation greatly. Chen [5] and Tu [10] improved Jiang's method by using sparse morphable model to get depth information. It can get depth information from a single image by prior statistic knowledge instead of complicated dense calculate, so that the process was simplified. Though Jiang's method [4] and Chen's method [5] both need only one image, the input face is restricted to full frontal ones. Otherwise, the side face information will be lost.

In this paper, we propose a model based 3D face reconstruction approach from a single image which allows a rotation angle ranging from -15 degrees to 15 degrees. Firstly, an improved color-based ASM method is used to detect the face area and get the 2D face feature points automatically from the given image, which is not restricted to a full frontal one. Then, color information is used to correct the location of face feature points. Finally, after text mapping, a particular and realistic 3D face model is deformed from a generic model. In section II, we introduce the improved method for 3D face reconstruction. In section III, we describe the synthesis of a talking face.

II. 3D FACE RECONSTRUCTION

In this section, we present an automatic approach for 3D face reconstruction. The framework consists of three parts: a) face feature points extraction; b) feature points correction; c) generic model deforming and texture mapping.

A. Face Feature Points Extraction

Due to a great number of experiments on ASM method [6], it is shown that the location of the initial model plays a key role in the speed and accuracy of the feature points' extraction process. So we improved Du's ASM method [7]: firstly, we located skin area based on skin model, which can effectively reduce the area that ASM need to search; then, we precisely located pupils of the eyes, which can remarkable increase the accuracy of the location of the initial model; finally, face feature points are got by searching the whole skin area using ASM method. Moreover, to make mouth shapes more nature we add 9 points in the mouth area; so that there are totally 69 points in the whole face area.

1) Skin area detection:

We detect skin area by using color information based on YUV and YIQ space and adding Gamma Correction [8] to decrease illuminant affection.

Firstly, we get R_{gamma}, G_{gamma} and B_{gamma} after Gamma Correction. Then, U, V can be obtained as follows:

$$U= -0.147*R_{gamma} - 0.289*G_{gamma} +0.436*B_{gamma} \tag{1}$$

$$V=0.615*R_{gamma} - 0.515*G_{gamma} - 0.100*B_{gamma} \tag{2}$$

In YUV space, each color corresponds to a chrominance, which is the sum of U and V. Chrominance is described by saturation (Ch) and hue (θ), where θ is calculated by:

$$\theta = \tan^{-1}(|U|/|V|) \tag{3}$$

As shown in Fig.1, if the hue of the pixel meets the condition: $\theta \in [100, 155]$, then it is considered as a skin point.

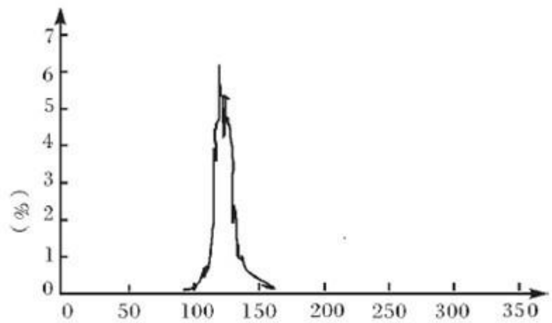

Figure 1. Distribution of θ in YUV space

I, given by (4), represents for color between yellow and blue-green. Fig.2 shows that skin points have a value of I ranging from 15 to 85 approximately.

$$I=0.596*R_{gamma} - 0.274*G_{gamma} - 0.322*B_{gamma} \tag{4}$$

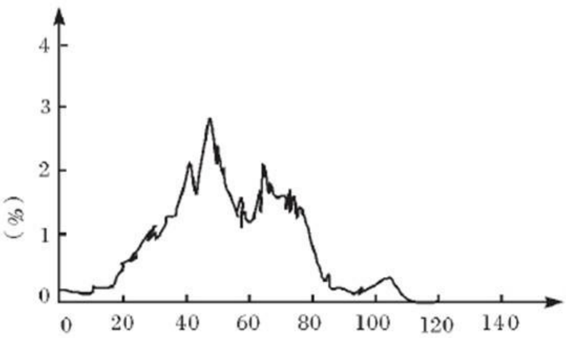

Figure 2. Distribution of I in YIQ space

Then, according to θ and I got from (3) and (4), we can judge whether the point is a skin point by (5):

$$\begin{cases} 100 \le \theta \le 155 \\ 15 \le I \le 85 \end{cases} \tag{5}$$

2) Locating pupils of the eyes

There are many ways to locate pupils. Here, we use Harris corners detection method [9], which can fast and precisely locate pupils.

3) Feature points extraction results:

With the skin area detected and the pupils located, ASM searching is used to extract feature points in that area. Testing more than 100 face images shows that it takes 0.7 seconds in average to extract all feature points. The time is less than directly using Du's ASM method which takes 1.4 seconds in average. Fig.3 shows some extraction results.

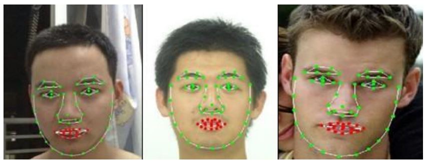

Figure 3. Face feature points extraction

B. *Face Feature Points Correction*

However, since ASM method using a symmetric template, some side face feature points will be out of face area when the input face is not a full frontal one. In this case, some side texture information will be lost. To solve this problem we judge side feature points whether they are skin points or not first. If they are not skin points then put them towards face area by parallel translation until all side feature points are in the skin area.

C. *Generic Model Deforming and Texture Mapping*

Firstly, 200 of 3D face data, 100 male and 100 female respectively, are selected to form an average generic 3D face model by (6), in which S_i represents for each 3D data and a_i, which ranges from 0 to 1, satisfies the condition (7). Then, we use feature points before correction to deform generic model into particular model. Finally, we use feature points after correction in texture mapping. Here, we simply describe our approach based on Chen's method [5] and Tu's method [10].

$$S_{avg} = \sum_{i=1}^{n} a_i S_i \tag{6}$$

$$\sum_{i=1}^{n} a_i = 1 \tag{7}$$

Since a particular shape model is given, we may acquire a particular face by texture mapping. Correspondent texture was extracted from the input image by affine transformation. Affine transformation is calculated as below, where c represents for scale and o for offset.

$$I' = c * I + o \quad (8)$$

D. *Reconstruction Results*

Using this method, we have tested a set of 100 human face images which are 640 × 480 big and have a rotation angle between -15 degrees and 15 degrees. Fig.4 shows some reconstruction results.

Figure 4. 3D face reconstruction results

Comparing to Jiang's method [4] and Chen's method [5], as shown in Table I, our approach effectively resolve the problem of side information missing. And using our approach, it will take about 3.1 seconds in average for each image to reconstruct, which is less than Chen's [5] and Jiang's [4] and less than Blanz and Vetter's [3], as shown in Table II.

TABLE I. RECONSTRUCTION COMPARISON I

	Jiang [4]	Chen [5]	Our method
Left side information missing	21	20	4
Right side information missing	10	8	2
Successfully reconstruction	69	72	94
Side information missing	31	28	6
Rate of successfully reconstruction	69%	72%	96%

TABLE II. RECONSTRUCTION COMPARISON II

	Blanz and Vetter [3]	Jiang [4]	Chen[5]	our method
CPU	P4-2GHz	P4-2GHz	P4-2GHz	P4-2GHz
Computation time	4.5 minutes	4.6 seconds	3.8 seconds	3.1 seconds

III. SYNTHESIS OF A CHINESE TALKING FACE

After reconstruction of a lifelike 3D face, we have to synthesize mouth shapes according to different phonemes to get a Chinese talking face. To make the Chinese talking face more nature when it "talks" we also synthesized some expressions.

A. *Synthesis of Different Mouth Shapes*

We can get different mouth shapes by moving mouth feature points (19 in total) from their originally location to another by TPS (Thin Plate Splines) [11]. Here, we use an improved TPS algorithm. The new coordinates can be got by (9) shown below.

$$f(x,y,z) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} w_i U(P_i - (x,y,z)) \quad (9)$$

And we divide Chinese phonemes into 16 categories as shown in Table III.

TABLE III. CHINESE VISUAL PHONEMES

Number	Phonemes	Number	Phonemes
1	a	9	o
2	b, p, m	10	r
3	d, t, l	11	u, v
4	e	12	z, c, s
5	f	13	zh, ch, sh
6	g, k, h	14	n
7	i	15	ng
8	j, q, x	16	silent

Then, according to the 16 categories we synthesized different mouth shapes for the talking face. Fig.5 shows some mouth shape images we have got.

(a)"a" (b)"o" (c) "e"

Figure 5. Mouth shapes

After got these mouth shapes, we have to insert some transitional shapes between two mouth shapes to make the animation more nature and continuous. Here, we use linear inserting function (10), in which α values between 0 and 1 and V_f represents for the first visual phoneme while V_s for the second, to synthesis these transitional shapes.

$$V_i = (1-\alpha)V_f + \alpha V_s \tag{10}$$

B. *Expressions*

According to FACS (Facial Action Coding System), we can get different expressions by combining different AU (Action Unit). So we use TPS algorithm [11] to transform 3D feature points to get different expressions, such as smile, anger, sadness, surprise, fear and so on. Fig.6 shows some expressions we have got:

(a)normal (b)smile (c)sadness (d)wink

Figure 6. Expressions

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we propose an automatic 3D reconstruction approach for a Chinese talking face. Our approach only needs a single input image and the image is not restricted to a full frontal one. Experiments show that this approach can create a lifelike Chinese talking face from only one normal face image in a really short time. And the 3D face reconstructed can be changed easily to form different mouth shapes and expressions for synthesizing the Chinese talking.

We are working on synchronizing speaking (both Chinese and English) with face animation (mainly mouth shapes) using the 3D face model reconstructed. We are also planning to add different expressions and emotions into the 3D face when it "speaks" in different environments.

ACKNOWLEDGMENT

In this paper, we choose 100 male and 100 female 3D faces data from "BJUT-3D Face Database" to form a generic face model.

REFERENCES

[1] F.I.Parke, "A Parametric Model of Human Faces," PhD thesis, University of Utah, Salt Lake City, 1974.

[2] S.Platt, N.Badler, "Animating Facial Expression," Computer Graphics, 1981, pp.245–252.

[3] Volker Blanz, Thomas Vetter, "A Morphable Model for the Synthesis of 3D Faces," SIGGRAPH, 1999, pp. 187–194

[4] Dalong Jiang, Yuxiao Hu, Shuicheng Yan, Lei Zhang, Hongjiang Zhang and Wen Gao, "Efficient 3D Reconstruction for Face Recognition," IEEE Pattern Recognition, 2005, pp. 787–798

[5] Lu Chen and Jie Yang, "Automatic 3D Face Model Reconstruction Using One Image," IWICPAS 2006, Lecture Notes in Computer Science(LNCS), ISSN 0302-9743, Springer, Vol. 4153/2006, pp.235-243.

[6] T.F.Cootes, C.J.Taylor, D.H.Cooper and J.Graham, "Active Shape Models-Their Training and Application," Computer Vision and Image Understanding, Jan 1995, Vol.61(1), pp.38-59.

[7] Chunhua Du, Jie Yang, Qiang Wu, Tianhao Zhang, Huahua Wang, Lu Chen and Zheng Wu, "Extended Fitting Methods of Active Shape Model for the Location of Facial Feature Points," Lecture Notes in Computer Science, ICVGIP 2006, pp. 610–618.

[8] Huaixun Zhao, Min Zheng and Zhiqiang Li, "A Skin-color Model Adapting Light Actively," Microelectronics and Computer, 2007.

[9] Wang li, Tang jin and Xu haizhu, "Fast Pupils Localization Method Based on Harris Corner Detection," Chinese Journal of Liquid Crystals and Displays, Vol.23, No.1, Feb 2002.

[10] Tu yi, Zhao qunfei and Lu fangfang, "3D Face Reconstruction Based on generic model and single image," Application Research of Computers, Vol.26, No.10, Oct 2002.

[11] Fred L.Bookstein, "Principle Warps: Thin-Plate Splines and the Decomposition of Edformation," IEEE Transactions on PAMI, June 1989, pp. 567–585.