Fast Texture Mapping of Photographs on a 3D Facial Model

Yuya Iwakiri Keisuke Yorioka Toyohisa Kaneko
Department of Information and Computer Sciences
Toyohashi University of Technology, Toyohashi, Aichi, Japan
kaneko@ics.tut.ac.jp

Abstract

This paper presents a new fast method for adding texture on a 3D geometric facial model derived from CT for the application of surgery simulation. Registration of three photographs(front, left-, and right-side), and the 3D geometric model is carried out automatically using edge patterns of landmarks. For reliable registration, those edges (e.g. eyebrow, beard) on the photographs not presented on the CT image are removed in advance. A graphics board with a geometry engine is used for a faster calculation of projections from 3D to 2D images. The computation speed is approximately two minutes per person for registration and texture mapping.
Keywords: 3D CG, 3D facial model, edge extraction, texture mapping

1 Introduction

This paper presents a new fast method for adding texture on a 3D geometric facial model derived from CT for surgery simulation applications. The face is the most concerned part of the body for a patient. Therefore, facial surgery simulation, for instance maxillo-facial surgery[1, 2], has attracted much attention.

CT (Computing Tomography) is becoming readily available in many hospitals. A geometrical model derived from CT data that lacks in color is not suited for presenting simulated results to a patient. This paper presents a practical method to add texture on the 3D geometric model by mapping texture acquired with a digital camera. Based upon estimated camera parameters (i.e. camera position and direction), a corresponding inverse projection is carried out for mapping.

1.1 Related Work

Most past works [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] addressing the problem of facial texturing utilize manual selection of feature points for registration.

Ali and Kaneko[14] proposed a method that employs 3D edge features from eyes, nose, ears, and facial profiles extracted with illuminating the 3D geometry model and that maps extracted features on the surface of the 3D model. The position and direction of a human head with respect to those of a digital camera are adjusted iteratively so that two 2D edge images are as close as possible. This method is, however, unpractical since it required computation time of more than six hours. The method presented here is similar to this method but makes substantial improvement in terms of computation time and registration accuracy.

Although in this paper we are not concerned with hair, there are works by Ali and Kaneko[15, 16] on a 3D hair model to be constructed based upon several photographs.

This paper describes a fast method for automating the registration for a 3D facial model derived from CT and three photographs, where edges are employed but not point features.

The main reason for automating registration is ease of use for users who are medical doctors or dentists. With our approach, users are not required to select feature points but obliged to obey specific requirements such as inclusion of the entire ears on the right-side and left-side photograph, the distance between the camera and patient, diffused light(e.g. fluorescence light), etc. as typically shown in the examples to follow.

2 Method

2.1 Principle of Texture Mapping

The principle of mapping photographs on the 3D geometric model is explained using Figure 1. Consider two optical systems: a real system and a virtual system. The real system consists of a real object (a human head in this case) and a real digital camera. The virtual system is a virtual version realized in a computer. The 3D geometric head model is obtained from CT data. If these two systems are identically set, then the texture on the head model of the virtual system can be set to be identical to the real object by inverse projection of photographs.

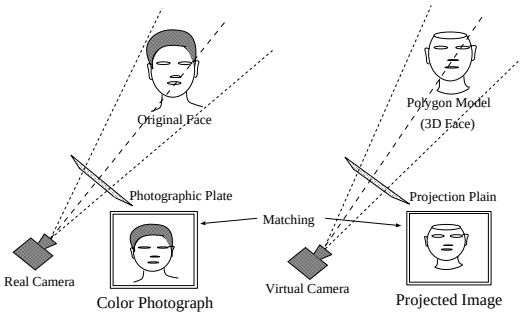
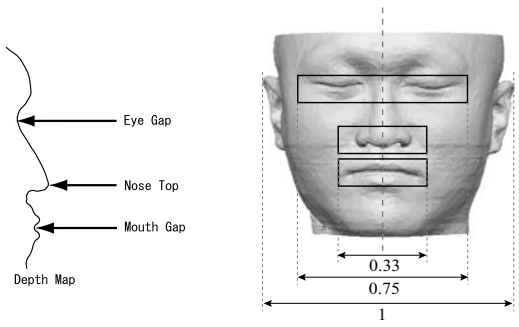


Figure 1: Real and Virtual Camera System



(a) (b)

Figure 2: Edge detection scheme from 3D geometric model

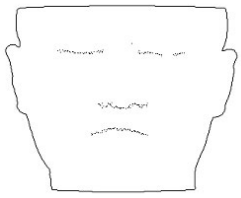


Figure 3: Edges extracted from frontal depth buffer

Figure 4: Edges extracted from side view depth buffer

For aligning two optical systems, edges from landmarks on the face such as eyes, nose, mouth etc. are used. Special attention needs to be paid here. Firstly, timing of CT scanning is different from that of photographing. Also patient's posture is different (i.e. a lying position for CT versus a standing position for photographing). The difference causes slight dissimilarity in geometry between the real and virtual object. Secondly, such landmarks as hair, eyebrows, eyelashes, moustache (if any), spots etc. are not present in CT image and therefore not on the 3D geometry model, although they appear on the photograph. For better registration, the edges from photographs have to be strictly restricted only to those present on the 3D geometric model.

Three photographs are employed: front, right-, and left-side views. From a frontal view photograph, we extract edges from eyes, nose, mouth, ears, and facial outline. From right and left-side view photographs, the profile outline and edges from ears are extracted.

2.2 Edge Extraction

Our strategy is to first extract edges present on the 3D geometric model of the head. Edges from a photograph are then restricted to the counterpart of these edges.

2.2.1 Edge Extraction from 3D Geometric Model

The 3D geometric facial model extracted from CT can provide the following stable edges:

1) The eye gap just below the eyelid under a closed eye situation, 2) the nose profile resulting from the abrupt transition from the nose head to the flat part above the upper lip, 3) the gap under the upper lip and above the lower lip, 4) the external perimeter of ears, and 5) the facial outline.

Based upon the extracted 3D geometric model, we compute three depth buffer maps of 480×480 corresponding to three camera positions.

On the frontal depth buffer map (see Figures 2(a) and (b)), we (1) identify the nose top point which is usually the highest, (2) then identify the eye point and the lip point as shown in Figure 2 (a), (3) set up three diagonal regions as shown in Figure 2(b), where widths are normalized using the distance between two ears, (4) then detect local minima as edges of the closed eyes and the mouth, and (5) for the nose, edges are set as the points of sudden sharp difference in depth which occurs between the nose and the flat region above the upper lip. The resulting edge image is shown in Figure 3.

On the left or right-side depth buffer map, we (1) identify the side facial profile as the edge line of the background which has a large value, and (2) detect the ear drum outline since there is sudden difference in the depth buffer around the ear hole. The resulting edge image for the right-side view is shown in Figure 4.

The important point is that the edges of four landmarks (eyes, nose, lip, and ear drums) are projected back to the 3D geometric model and fixed, while the facial profiles are not.

For the estimation of the camera parameters, we must vary the position of the 3D geometric model with respect to the camera position which is fixed as the coordinate origin. The facial profiles can be derived very rapidly from the depth buffer map using a graphics board. However, the detection of edges from the landmarks on the depth buffer map requires considerable processing time. Additionally these edges extracted using geometric features are irrelevant to the position of the 3D geometric model. This method can effectively shorten the computation time.

2.2.2 Edge Extraction from Photographs

Pictures taken (against a blue sheet) are 640×480 from which a central part of 480×480 is used for processing here. The distance from the camera to the head is set to be about 150[cm], which is one of user requirements.

(1) Frontal Photograph

Firstly the skin region is extracted based upon the average color at the center region of a given photograph. On the extracted skin region, a Kirsh edge operator[17, 18] is applied to detect edges and a profile based upon the horizontal sum of Kirsh edge values (see the left figure of Figure 5) is derived. Usually two peaks that correspond to low (black) values are easy to identify: the closed eyes which are along the straight line connecting the left and right ear and the mouth which is the firt peak from the bottom. Based upon the two peak positions, three regions for detecting eyes, nose, and lips are set up as shown on the right figure of Figure 5. The Sobel operator on these regions will provide necessary edges.

Figure 7 shows the resulting edges after applying a Percentile method. This figure contains double edges and unwanted edges that do not correspond to the edges (see Figure 3) extracted from the CT data. So we carry the following operations. In the eye and nose regions, a vertical downward (i.e. top to bottom) scanning is performed. In the eye region, the firstly encountered edges along this scanning direction that correspond to the upper eyelid or the upper part of the eyelash are kept. The next or the following edges that correspond to the lower part of the eyelash are removed since the eyelash is not present on the CT data. In the nose region, the first edges that correspond to the nose tip profile are kept but those edges that appear as the shadow of the nose are removed. In the mouth region, an upward scanning(i.e. bottom to top) is done to obtain the first edges that appear between the upper and lower lips. But unwanted edges sometimes detectable at the top part of the upper lip are not taken. Finally a thinning operation on edges is carried out to obtain their essential content (see Figure 8). The edge image gathers from these three regions are matched against that shown in Figure 3.

(2) Edge extraction from side view photographs

Here we extract two items: ear edges and the facial profile. For ear edges, we set up a circular region of radius r (=100 pixels) at the center. Then a Sobel edge operator is applied in the region. Ear edges are obtained based upon a P-tile method with a threshold of 1%. The side view profile is defined from the point just below the eyebrow to the sharply curved point at the chin. The eyebrow should not be included since it is not present in the 3D geometric model.

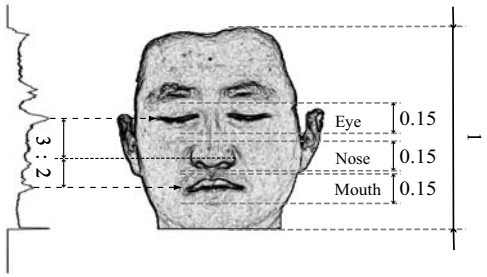


Figure 5: Horizontally projected sum

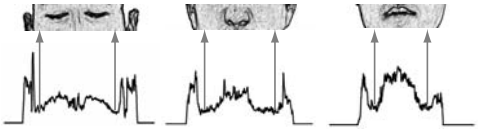


Figure 6: Vertically projected sum

Figure 7: Extracted edges from a frontal photograph

Figure 8: Final edges from the frontal photograph

Figure 9: Extracted edges from a left-side photograph

2.3 Matching

Matching is carried out for identifying the camera parameters. Iteratively changing the location and direction of the 3D head model with respect to the camera position and direction, the sum of edge to edge distances based upon the ICP principle[19] is minimized. Here the summation is done for each region separately and the final total is their weighted sum. The weight allocation is largely dependent upon the reliability of edges. On the frontal photograph we allocate heavier weights to the eyes and nose region while lighter weight to the mouth region.

The matching is carried out in the following steps.

(1) View angle

The distance from the left to the right ear is computed on the 3D model and on the front photograph. Then the view angle (which corresponds to the focal depth) is adjusted to make the two lengths identical.

(2) Alignment of gravity centers

The center points of landmark edges (right and left eye edges, nose edges, and mouth) are computed from the 3D geometric model and the frontal photograph. Then the initial values of two camera parameters T_x and T_y are set accordingly to equate the two gravity centers. (Note here that the translation along the z-axis T_z is not included since the distance to the camera is kept constant (150[cm] in our case)).

(3) Coarse matching

Three parameters (translations T_x, T_y and rotation R_x) are found to have more effects in matching than the other two parameters R_y and R_z. Note here that R_x determines the vertical rotation (i.e. up and down) of the head.

(4) Fine Matching

The final fine matching is performed by varying all the six parameters. This is done in two steps to increase computational speed. The first step is done for the view angle and two translation parameters T_x and T_y. The second step involves three rotation parameters R_x, R_y, and R_z. This separate computation increased the speed about four times.

2.4 Texture Mapping

The final process is texture mapping. As Figure 1 illustrates, the real and virtual system are aligned based upon the estimated camera position. The real color photograph in the real system is aligned on the virtual colorless photograph in the virtual system as exactly as possible based upon edge features as described in the above. After alignment, each polygon on the former is copied to the corresponding polygon on the latter as long as it is visible from the camera position. When the orientation of a surface polygon with respect to the camera's direction exceeds a threshold (75 degree in our case), the polygon is not used for texture mapping. Most likely there is another photograph. For example, some polygons on the cheek in a slanted angle on the front photograph may be seen in a less slanted angle on the right-side or left-side photograph.

It should be noted that the front photograph overlaps with the left- or right-side photograph somewhere near at -45 degree and 45 degree angle where the frontal straight direction is set to be at 0 degree angle. The blending function is mutually linear as shown in Figure 10. The starting and ending angle θ_1 and θ_2 should be chosen so that the blending is carried out solely on the cheek.

3 Experiments

All the required computer programs are written in C and executed on a PC with CPU Pentium 2.8GHz

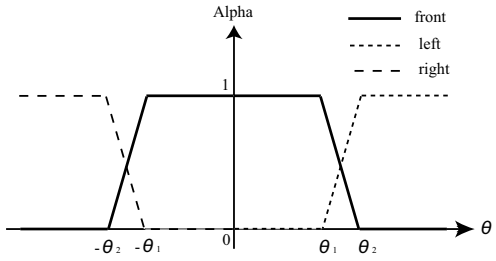


Figure 10: Blending function

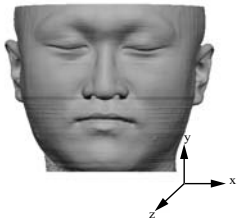


Figure 11: 3D geometric model of human head (Person #1)


Figure 12: 3D head after polygon reduction (Person #1)

Table 1: Computation time for edge extraction [sec]

	Front	Right	Left
3D model edges	0.80	0.39	0.38
Photo edges	4.3	0.27	0.27

and a graphics board with NVIDIA Quadro2 Pro. (A faster graphic board (e.g.Quadro4) will accelerate the computational speed.) The digital camera used is a SONY Cybershot DSC-F505 with a resolution of 640×480[pixel]. The center part of 480×480[pixel] was used as mentioned already.

We employed here CT data of a person. This is $512 \times 512 \times 168$ with a resolution of $0.468 \times 0.468 \times 1.0$[mm]. The CT data were converted into surface polygon data with the marching cubes algorithm. The resulting texture-less geometric models are shown in Figure 11. We then applied a polygon reduction algorithm in the VTK library[20], obtaining $32,861$ polygons. Figures 12 show the polygon reduced versions which show no visible difference.

Three pictures (front, left-side, and right-side) were taken for two persons. Edges were extracted from photographs and from the 3D geometric models based upon the algorithms described in Section 2.2. Computation time is tabulated on Table 1 which shows time of .3 to 4 seconds. The threshold level for a P-tile method was set to be 3% for detecting edges of eyes, nose, and mouth from a frontal photograph and 1% for those of ears from a side view photograph. Then matching was carried out. The weights on the frontal photographs were set as .25,.25,.25,.05, .1, and .1 for the left eye, right eye, nose, mouth, left ear, and right ear, respectively. On the left- and right-side view

Table 2: Computation time for matching [sec]

	(1)View angle	(2)Gravity Ctr.	(3)Coarse Matching	(4)Fine Matching
front	1.5	0.08	13.3	26.2
right	1.6	0.093	14.1	27.5
left	1.6	0.093	14.3	27.7

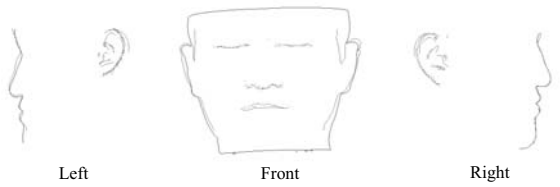


Figure 13: Edge matching result (Person #1)

Figure 14: Registration error at the eye and lip

photographs, .3 and .7 were chosen as the weights for the facial profile and the ear, respectively. As expected, the most time-consuming part was at the matching process, which is tabulated in Table 2. It took about 40 seconds for each matching, providing total time of two minutes.

The time required for texture mapping is about a second, which is considerably shorter than that of the matching process. The blending of the left or right image with the frontal image was done from 45 to 55 degree (i.e. θ_1=45 and θ_2 =55 degree).

Based upon three photographs shown in Figure 15(a)(b)(c), the textured result is shown in Figure15(e)(f)(g). In order to judge the result, we show Figure 15(h) which is a 45 degree angle view. Figure 16, and 17 show the result for another person. Generally the results are judged to be excellent.

In order to evaluate the registration accuracy, we show an image (Figure 14) which was made by superimposing the edges (indicated by red color) extracted from the 3D geometric model (see Figure 8 for person #1) on the textured model. It appears that registration is quite good. The edge matching results are shown in Figures 13. The accuracy around the ears is not as good as that around the eyes and the nose where the accuracy is estimated to be one to three pixels.

4 Conclusion

We proposed here a fast practical automatic method for acquiring textures on a 3D geometric model derived from CT data. Three photographic images (front, left- and right-side) are used as the basis of textures. The proposed system easily implemented on a PC will be particularly useful in a hospital where a CT scanner is readily available.

There are two main technical contributions. The first is to use edges common to a 3D model and photographs. Edges from other landmarks not to be seen on the 3D model are to be removed. This enables faster and more accurate registration or matching. The second is a fast matching process that exploits an affordable graphic board with a geometry engine. Consequently the total computation time was reduced to about two minutes.

Future work includes 3D hair modeling to have a more natural model.

References

[1] E. Keeve, S. Girod, and B. Girod. Craniofaceil surgery simulation. In *Proc. of 3D Image Analysis and Synthesis 96*, pages 219–224, 1996.

[2] E. Keeve, S. Girod, P. Pfeifle, and B. Girod. Anatomy-based facial tissue modeling using the finite element method. In *Proc. of IEEE Visualization 96*, pages 21–28, 1996.

[3] T. Akimoto and Y. Suenaga. Automatic creation of 3d facial model. *IEEE Computer Graphics and Applications*, 13(4):16–22, September 1993.

[4] J.F.S. Yau and N.D. Duffy. A texture mapping approach to 3-d facial image synthesis. *Computer Graphics Forum*, 7:129–134, 1988.

[5] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. In *Proc. of ICCV '95*, pages 16–23, June 1995.

[6] R.M. Koch, M.H. Gross, F.R. Carls D.F. von Büren, G. Fankhauser, and Y.I.H. Parish. Simulating facial surgery using finite element models. In *Proc. of SIGGRAPH '96*, pages 421–428, 1996.

[7] F. Betting and J. Feldmar. 3d-2d projective registration of anatomical surfaces with their projections. In *Information Proceeding in Medical Imaging*, pages 275–286. Kluwer Academic Publishers, June 1995.

[8] J. Feldmar, N. Ayache, and F. Betting. 3D-2D projective registration of free-form curves and surfaces. *Computer Vision and Image Understanding*, 65(3):403–424, March 1997.

[9] W. Lee and N.Magnenat-Thalmann. Fast head modeling for animation. *J. Image and Vision Computing*, 18(4):355–364, 1996.

(a)Left (b)Front (c)Right (d)Rotated by 45 deg.

Photographs

(e)Left (f)Front (g)Right (h)Rotated by 45 deg.

Textured Model

Figure 15: The original photographs and resulting textured model (Person #1)

(a)Photograph

(b)Textured Model

Figure 16: The original photographs and resulting textured model (Person #2)

(a)Photograph

(b)Textured Model

Figure 17: Resulting textured model rotated by 45 degree (Person #2)

[10] H.H.S. Ip and L. Yin. Constructing a 3d individualized head model from two orthogonal views. *The Visual Computer*, 12(5):254–266, 1996.

[11] S. Marschner, B. Guenter, and S. Raghupathy. Modeling and rendering for realistic facial animation. In *Proc. of the 11th Eurographics Rendering Workshop*, pages 231–242, 2000.

[12] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D.H.Salesin. Synthesizing realistic facial expressions from photographs. In *Proc. of SIGGRAPH '98*, pages 75–84, 1998.

[13] M. Tarinin, H. Yamauchi, J. Haber, and H-P. Seidel. Texturing faces. In *Proc. of Graphics Interface 2002*, pages 89–98, 2002.

[14] Ali Md. Haider and Toyohisa Kaneko. Automatic reconstruction of 3d human face from ct and color photographs. *IEICE trans. on Information and Systems*, E82–D(9):1287–1293, September 1999.

[15] Ali Md. Haider and Toyohisa Kaneko. Hair shape modeling from video captured images and ct data. In *The 10th International Conference on Artifical Reality and Teleristence*, pages 52–57, August 2000.

[16] Ali Md. Haider and Toyohisa Kaneko. Realistic 3d head modeling from video captured images and ct data. In *Third IEEE EMBS International Conference on Information Technology Applications in Biomedicine*, pages 238–243, November 2000.

[17] J. C. Russ. *The Image Processing Handbook, 2nd Edition*. IEEE Press, 1994.

[18] K.R.Castleman. *Digital Image Processing*. Prentice Hall, 1996.

[19] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Tr. on PAMI*, PAMI-14(2):239–257, 1992.

[20] Visualization Toolkit. http://www.vtk.org.