3D Face Modeling from Perspective-Views and Contour-Based Generic-Model

Kam-sum LEE, Kin-hong WONG, Siu-hang OR, Yiu-fai FUNG
Department of Computer Science and Engineering
The Chinese University of Hong Kong
email: {kslee, khwong, shor, yffung}@cse.cuhk.edu.hk

Abstract

Three-dimensional human head modeling is useful in video-conferencing or other virtual reality applications. However, manual construction of 3D models using CAD tools is often expensive and time-consuming. Here we present a robust and efficient method for the construction of a 3D human head model from perspective images viewing from different angles. In our system, A generic head model is first used, then three images of the head are required to adjust the deformable contours on the generic model to make it more closer to the target head. Our contributions are as follows. Our system uses perspective images that are more realistic than orthographic projection approximation used in earlier works. Also for shaping and positioning face organs, we present a method for estimating the camera focal length and the 3D coordinates of facial landmarks when the camera transformation is known. We also provide an alternative for the 3D coordinates estimation using epipolar geometry when the extrinsic parameters are absent. Our experiments demonstrated that our approach produces good and realistic results.

Keywords: facial images, face model, model reconstruction, facial landmarks, perspective projection, generic model, epipolar geometry, essential matrix

1 Introduction

Realistic modeling of 3D objects is essential to many multimedia applications. For example in very-low-bit-rate video compression, model-based image coding is one of the most effective techniques[1]. In this method, the user's face is first modeled geometrically. Then the 3D model is transmitted to the receiver at the beginning of a communication session. During the communication, only the movement of the user is transmitted, such as rotation and translation of the head, the movement of eye balls, opening or closing of the mouth, etc. The images of the user at the receiving side are reconstructed by using the 3D model, the movement information and texture-mapping. In this coding method, one of the important problems is the construction of the user's 3D head model, not only the outward shape, but also the structural information of the face organs for animation. The 3D model of an existing object can be constructed manually by using CAD tools, however, it is expensive and time consuming. The manual production of large amount of 3D models, i.e., a large population database with a 3D face model of each record, becomes very expensive. Thus there is a need for highly automatic method of human head model construction. Our objective is to build a robust and cost-effective prototype system to re-construct a 3D human head model of a user from perspective camera images without tremendous amount of image sequences and computation.

Automatic 3D object model construction involves the acquisition of object views at different angles. "Shape from silhouettes" or volume modeling[2][3] is a widely used approach in the reconstruction of general 3D objects from multiple 2D camera images. But most of them depend on an accurate and long image sequence up to 30-40 frames to yield satisfactory result[3]. They also require a turntable or other special machinery to produce an accurate image sequence. The acquiring of the image sequence is time-consuming and the object cannot have any movement during the process. In general, they are more suitable for static objects rather than human heads. Another problem is that the concavity of the object is difficult to be recovered. These problems limit the system to convex objects only[4]. To overcome these problems, some approaches using stereo-based techniques[4]. However, they require dense features extraction and correspondences that are difficult and expensive to obtain.

In this paper, we narrow down the problem to the reconstruction of a specific object-human head. An efficient method to create a complete model of a human head is developed. It employs a generic head model, and adjusts the model to fit the frontal view and the side view(s) of the target head. Our approach makes use of the prior knowledge of general human head shape and face organs, with 3D coordinates estimation techniques to reconstruct 3D human head models. In our generic model, the face organs and the head shape are handled separately. The prior information, i.e., a generic prototype of human head, reduces the concavity modeling error. It also reduces the computation load in features selection and correspondences recovery. The minimum number of images required is reduced to three, or two while assuming symmetry of the head shape. To make the 3D model obtained realistic enough for applications, it will be texture-mapped by using the images used in the model reconstruction. Satisfactory experimental result on real targets demonstrated the feasibility of our system.

This paper is organized as follow. We will first present the previous approaches in Section 2. Then, we will discuss the acquisition of camera images from different viewpoints in Section 3. In Section 4, the techniques and mathematical issues in the estimation of the head shape and the deformation of the generic model will be described. The shaping and positioning of face organs, as well as the integration of the head shape and face organs will be also presented in this section. In Section 5, we will discuss about the texture-mapping process. The result of our implementation and discussion will be presented in Section 6. Finally, Section 7 consists of the conclusion and the future directions.

2 Previous Work

Many researchers have been working on the model reconstruction via multiple views. Busch[2] proposed a method to construct the 3D model of an existing objects from multiple camera images. In similar works, "Shape from silhouettes" is usually used in the first step to construct a volume model by applying the method of occluding contours[5][6]. Then a triangular mesh is used to approximate the surface of the volume model. Finally the images of the real object are used to texture-map the surface model.

When the object to be modeled has been already known as a human head, some prior information can be used, such as the use of generic face models. For example, there are several 2D models based on spline or deformable templates [7][8] which are obtained by tracking the contours of a face in an image sequence. Essa and Pentland[9] used a 3D mesh with higher degrees of freedom. DeCarlo and Metaxas[10] proposed an method which integrates optic flow and deformable models with applications to human face shape and motion estimation. Most of the models used belong to the deformable ones, including the one used in our design. The generic 3D model of human head used

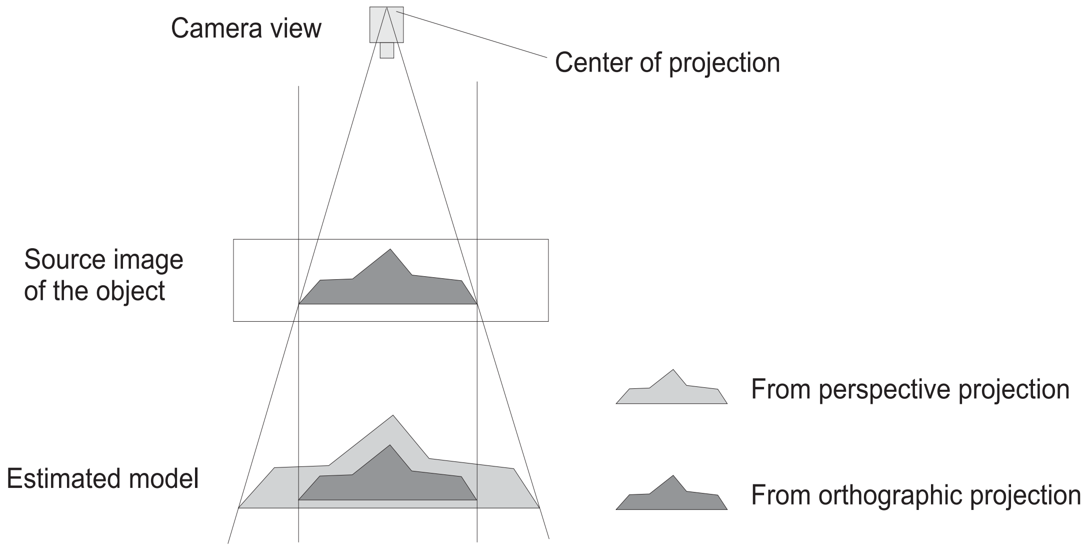

Figure 1: Orthographic and perspective projection

in our design is based on the studies in anthropometry studies[11]. The face organs and the head shape are handled separately and the deformation of the face organ relies on those anthropometry measurements.

There already existed some developments involving the use of generic model in the reconstruction of 3D head models from a couple of camera images. G. Xu et al.[12] proposed their work as an integral component of a virtual space teleconferencing system, which generate 3D facial models from facial images and to synthesize images of the model virtually viewed from different angles. T. Akimoto and Y. Suenaga[13] presented a method of automatic creation of 3D facial models which are needed for facial image generation. L. Tang and T. S. Huang[14] proposed a similar approach, and a template matching based algorithm is developed to automatically extract all necessary facial features from the front and side profile face images. C. J. Kuo, R. S. Huang and T. G. Lin[15] proposed a method using only the frontal image to synthesize a face model. The depth information is estimated statistically from a facial parameter database, according to the anthropometric definition on face and head. L. Yin and A. Basu[16] proposed an approach for modeling a person's face for model-based coding (MPEG4).

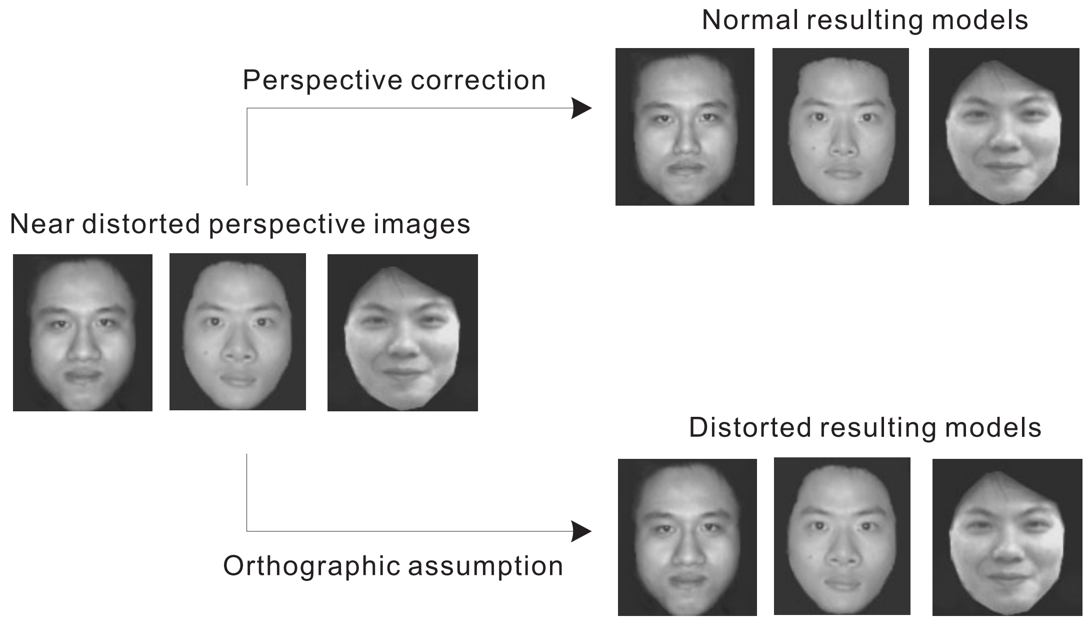

Figure 2: Distortion from orthographic assumption and perspective projection

For all the work mentioned above, long focal length is assumed to approximate orthographic projection. However, this assumption is not valid for most practical CCD cameras, and the short distance between the face and the camera makes the problem worse. In fact, the scaling effect from perspective projection cannot be ignored. Figure 1 shows the difference between orthogonal projection and perspective projection. As shown in Figure 2, the images are distorted in short distance due to perspective projection. If we made the assumption of orthographic projection, then the resulting models will be distorted, too. Therefore we have to take account of the perspective projection information, i.e., the distance and the focal length, into our calculation to correct the distortion. If the camera transformation is known, the focal length of the camera can be found within the framework of our system, without pre-calibration before the process. We also use a contour-based generic model rather than a triangular mesh model to reduce the complexity in the calculation of new positions of vertices.

3 Image Capturing

In our approach, three camera images of the target head should be captured. The three viewpoints are the front-view and the two side-views of the target head. The front-view is assumed to be in roughly perpendicular viewing direction with respect

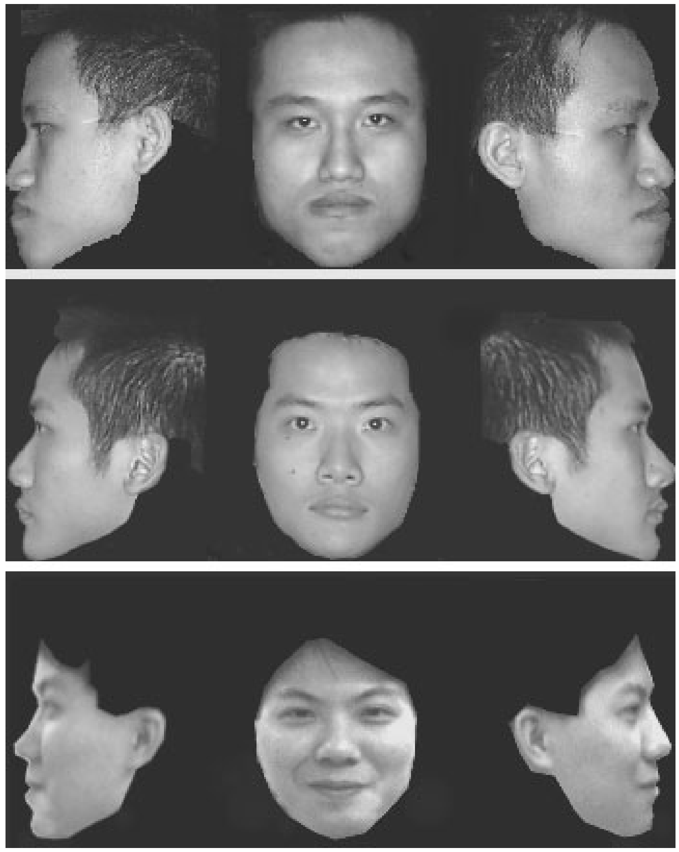

Figure 3: Images captured

to those of the side-views, so as to obtain better images for the use of texturing of the resulting model. Another advantage of selecting three orthogonal views is that the computational efforts in the determination of landmark coordinates in 3D space can be reduced. In case the views are not orthogonal to each other, we can use epipolar geometry [17] to recover the camera transformation. However, each facial feature must be visible in two adjacent views no matter the views are orthogonal or not. We must ensure that the entire head shape falls into all of the three views. During the capturing process, other objects should not occlude the facial features. To extract the head shape from the image, we will need a unique background color and the segmentation of the head shape can be done simply by the color keying method. Figure 3 shows the images captured in our experiment.

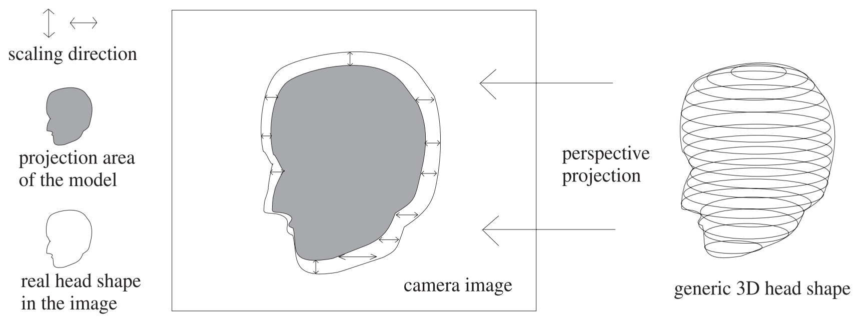

Figure 4: Contours and scaling of model

4 Shape Estimation & Model Deformation

4.1 Head shape estimation & model deformation

The shape of the head can be estimated by the deformation of a generic human model. In our design, the head shape is considered as a volume model, which is bounded by a number of horizontal contours. In the implementation, a number of vertices are connected to form each contour. Each contour can be stretched in certain directions at several control points to deform the model. The generic model is deformed until its projection areas are fully coinciding with the corresponding real head shape areas in the source images. Using this contour-based approach, we only need to calculate the deformation parameters for each contour, and do not need to perform the calculation of the deformation of all vertices in the model. In the stretching process, the face organs with complex and convex surfaces such as nose and ears are neglected. They will be handled in later processes. Figure 4 shows the construction of the generic model and the scaling/stretching method. The arrows show the directions of scaling. The shaded area is the projection area of the generic model before the scaling and the outline is the outline of real head shape in the image. The projection areas of face organs in both model and real image are neglected in the calculation. The image from the front viewpoint will be used to calculate the scaling in the x-axis direction, that is to adjust the "width" of the contours in the generic model. Similarly, to adjust the "thickness" or "depth" of the contours in the generic model, the images from side viewpoints will be used to calculate the scaling in the z-axis direction. Any one of the three images can be used to calculate the scaling in the y-axis direction, i.e. to adjust the "height" of the entire generic model.

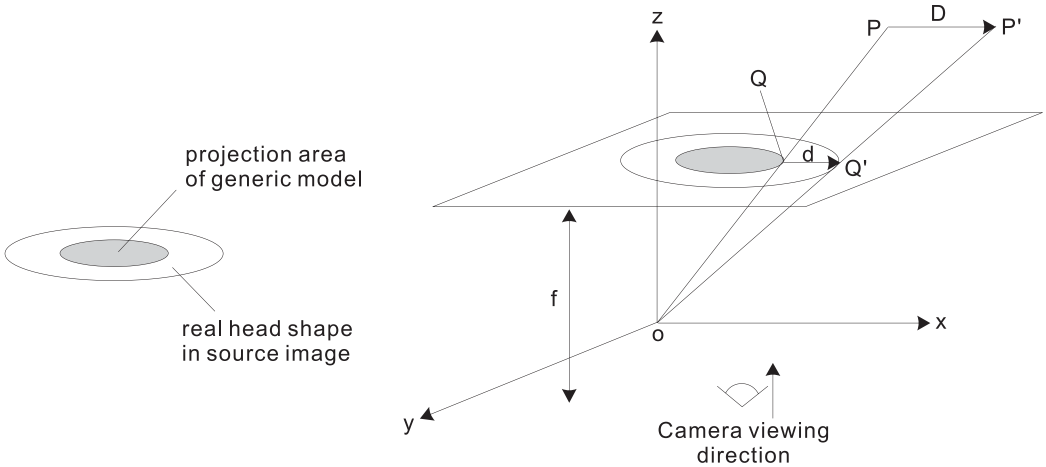

Figure 5: Calculation of model scaling

Figure 5 shows the calculation of the deformation parameters. Similar to the previous figure, the arrows show the directions of scaling. The shaded area is the projection area of the generic model before the scaling and the outline is the outline of real head shape in the image. Here we show only the calculation of scaling in the x-axis direction and the calculation in other directions can be done similarly by the same method.

The point $P(x_P, y_P, z_P)$ in Figure 5 is the 3D position of a scaling point of the generic model. The $P'(x_{P'}, y_{P'}, z_{P'})$ is the new position of point P after scaling. We denote the distance between P and P' by D. The point $Q(x_Q, y_Q, f)$ is the 2D projection of scaling point P and $Q'(x_{Q'}, y_{Q'}, f)$ is the new position of the projection of P' after scaling. We denote the distance between Q and Q' by d. The focal length of the camera is known as f, and we will discuss the method for finding f in the next section. By using similar triangles relation, we obtain the value of D:

$$D = \frac{d \cdot z_P}{f} = \frac{(x_{Q'} - x_Q) \cdot z_P}{f} \tag{1}$$

The sign of D represents the direction of scaling. Then the scaling factor, denoted by s, can be calculated by the following equation:

$$s = \frac{x_{P'}}{x_P} = \frac{x_P + D}{x_P} = 1 + \frac{(x_{Q'} - x_Q) \cdot z_P}{x_P \cdot f} \tag{2}$$

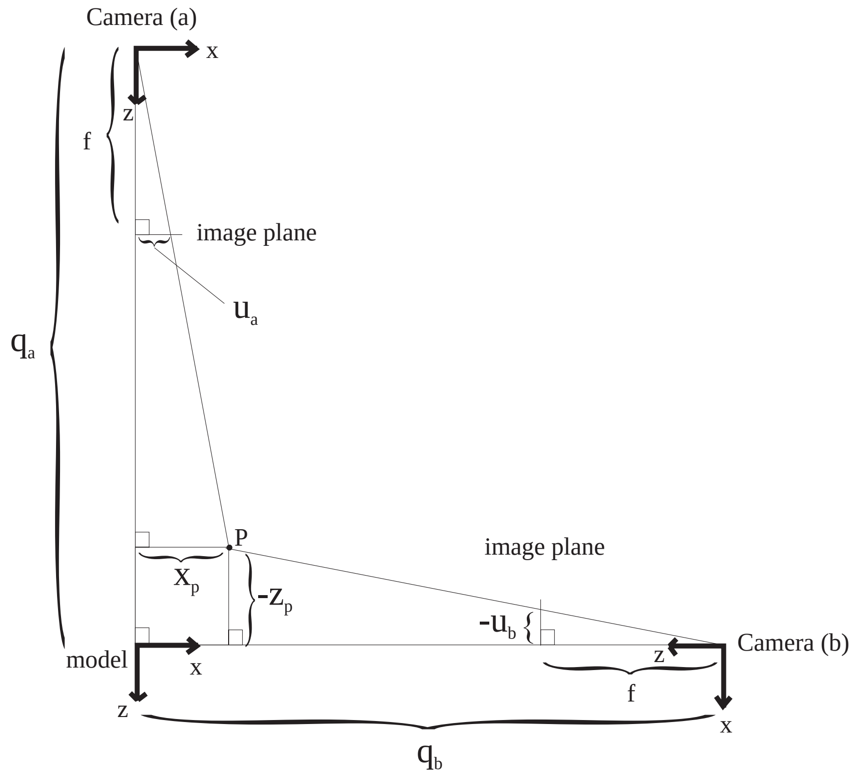

Figure 6: Determination of 3D coordinates

4.2 Face organs shaping and positioning

The face organs such as ears, mouth and nose are handled separately from the head shape. These parts are complicated in shape, and thus make direct modeling more difficult. We simplify the process by using a number of simplified generic organs and manual selection of some control points on the images. The control points chosen will determine the scale, dimension and location of the generic models. The calculation of scaling is the same as the previous section. We also have to determine the 3D positions of the face organs and integrate the face organ models into the coordinate system of the head model. Thus the camera parameters and the 3D coordinates of the landmarks in the coordinate system of head shape model must be found.

4.2.1 Reconstruction with both intrinsic and extrinsic parameters

First we start with the case that both the intrinsic parameter (focal length), and the extrinsic parameters (transformation between the two views) are required. The camera transformation is assumed to be given and the intrinsic parameter, the focal length, can be found during our calculation. The reconstruction of the 3D coordinates is simply done by triangulation. Figure 6 shows our method in the determination of the 3D coordinates of a landmark with orthogonal camera positions.

For a landmark point $P(x_p, y_p, z_p)$, let the projection of P on the image plane of camera (a) be (u_a, v_a), and the projection of P on the image plane of camera (b) be (u_b, v_b). In our implementation, both cameras will have the same focal length f, which will be found in later calculations. Since the camera positions are known, then q_a and q_b are assumed to be given, which are the distances between the model center and the optical centers of camera (a) and camera (b) respectively. As the principle axes of the cameras are on the same plane with the center of the model, i.e., $y = 0$, so that we have the following relations:

$$v_a = v_b = 0 \Leftrightarrow y_p = 0 \tag{3}$$

$$u_a = 0 \Leftrightarrow x_p = 0 \tag{4}$$

$$u_b = 0 \Leftrightarrow z_p = 0 \tag{5}$$

The symbol $\Leftrightarrow$ represents the if and only if relation. From the above relations, we can concentrate on those cases that the unknowns are non-zero and the calculation of f. Now there are four unknowns, x_p, y_p, z_p and f, in our problem. From the side ratio of similar triangles, we can obtain four equations as shown here relating the unknowns:

$$\frac{u_a}{x_p} = \frac{f}{q_a + z_p} \tag{6}$$

$$\frac{u_b}{z_p} = \frac{f}{q_b - x_p} \tag{7}$$

$$\frac{v_a}{y_p} = \frac{f}{q_a + z_p} \tag{8}$$

$$\frac{v_b}{y_p} = \frac{f}{q_b - x_p} \tag{9}$$

With four constraints and four unknowns, we can solve the problem as follow. Rearranging the above equations, we can get:

$$x_p = \frac{u_a}{v_a} \cdot y_p \tag{10}$$

$$z_p = \frac{u_b}{v_b} \cdot y_p \tag{11}$$

With the above equations together with equations (6) and (7), we can get two possible solutions of y_p:

$$y_p = 0 \tag{12}$$

or

$$y_p = \frac{v_b^2 \cdot q_b \cdot v_a - v_a^2 \cdot q_a \cdot v_b}{v_a^2 \cdot u_b + v_b^2 \cdot u_a} \tag{13}$$

For the case of non-zero value of y_p, we can easily find f. From equation (10) and equation (11), we can find x_p and z_p respectively. However, to find f for the case that $y_p = 0$. Moreover, from equation (13), the following condition must be satisfied to solve the problem:

$$v_a^2 \cdot u_b + v_b^2 \cdot u_a \neq 0 \tag{14}$$

In our implementation, we will first choose one of the facial landmarks which can satisfy the above condition to solve for the non-zero value of y_p and the focal length f. Once we can get the value of y_p, we can easily solve x_p and z_p from equations (10) and (11). In fact, if it is the case that $y_p \neq 0$, we even do not need to solve the focal length f for finding x_p and z_p. For the remaining points which do not satisfy condition (14), since we have already obtained the value of f, the calculation is trivial by solving equations (6), (7), (10) and (11) to find x_p, y_p and z_p.

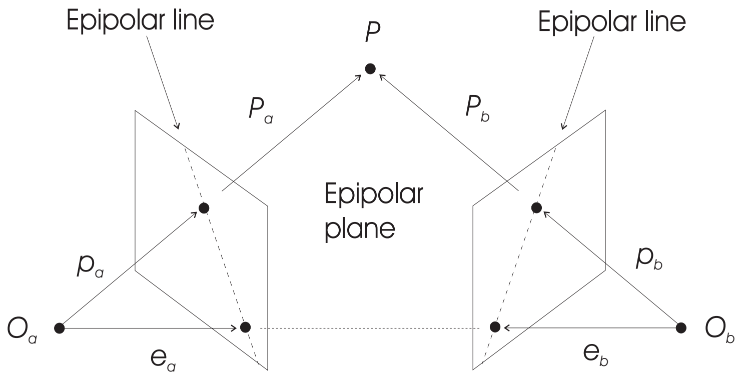

Figure 7: Epipolar geometry

4.2.2 Reconstruction with only intrinsic parameter

The reconstruction by triangulation method mentioned above relies on the prior knowledge of extrinsic parameters and the recovery of intrinsic parameter. In a real situation, we may not be able to tell the exact camera positions. However, the case in which only the intrinsic parameters are known, we can still solve the problem up to a scale factor by using epipolar geometry [17]. We now first give a brief introduction of epipolar geometry and explain how it works to solve the 3D coordinates of the facial landmarks.

Figure 7 shows the two cameras and their epipolar geometry. The projection centers of the camera (a) and camera (b) are O_a and O_b respectively. The vector $P_a = (X_a, Y_a, Z_a)$ and $P_b = (X_b, Y_b, Z_b)$ are the extended projection vectors $p_a = (x_a, y_a, z_a)$ and $p_b = (x_b, y_b, z_b)$ of the point P on the image planes of the cameras, which are expressed in the corresponding reference frame. For the image points, which are lying on the image planes, we have $z_a = f_a$ and $z_b = f_b$. If the two cameras have the same focal length, we may simply have $f_a = f_b = f$. The relation between the 3D points and their projections can be given as:

$$p_a = \frac{f_a}{Z_a} P_a \tag{15}$$

$$p_b = \frac{f_b}{Z_b} P_b \tag{16}$$

The two reference frames of the cameras are related by the extrinsic parameters, which are the rotation matrix R and translation vector $T = (T_x, T_y, T_z)$ in 3D space. Then we can have the following relation between the vectors P_a and P_b:

$$P_b = R(P_a - T) \tag{17}$$

4.2.3 Recovery of 3D coordinates from epipolar geometry

The 3D coordinates can be obtained from the essential matrix E in epipolar geometry. The essential matrix E can be estimated by the linear least square method or the eight-points algorithm by Longuet-Higgins[18]. After the estimation of the essential matrix which is up to an arbitrary scale factor, the calculation of the 3D coordinates of the landmark points can be done by the algorithm proposed by Longuet-Higgins[18]. This algorithm first decomposes the essential matrix and estimates the extrinsic parameters R and T. With R and T known, P can be found easily. The recovered position is unique only up to a scaling factor. This is not a problem in our application since we do not need the true scale of the target, i.e. we can arbitrarily scale up and down the size of the reconstructed model without affecting its appearance. Moreover, the scaling factor can be easily determined if we know the distance between two landmarks.

As we have mentioned before, if only extrinsic parameters (camera positions) are known, we can use triangulation method to recover the intrinsic parameter (focal length). On the other hand, if only the focal length is known, we can estimate the camera transformation by epipolar geometry. In case both intrinsic and extrinsic parameters are known, we will use both triangulation method and epipolar geometry to generate two sets of results, which back up each other when one of them fails.

Figure 8 shows the selected landmarks on the human head. Each face organ has 3 to 4 landmarks selected on it, from each captured image. Since the 3D coordinates of the landmarks are now known, the deformation of the generic model is simplified by just scaling of the model by a ratio of the original model coordinates to the landmark coordinates.

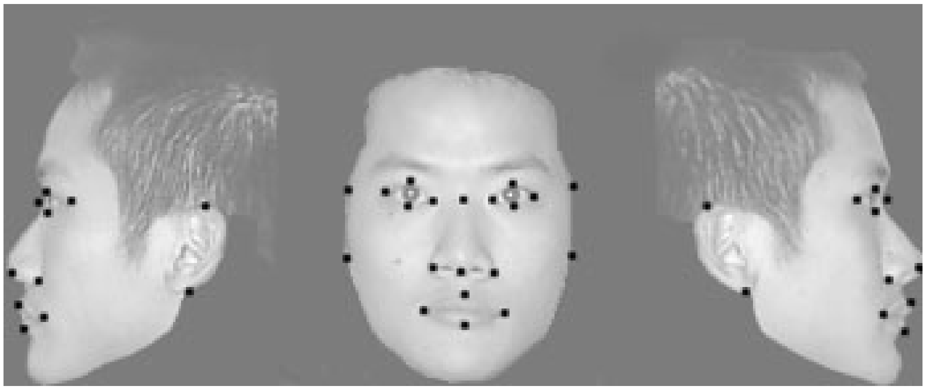

Figure 8: Selected facial points shown in black dots

4.3 Integration of head shape and face organs

After shaping the head and the face organs, we can integrate them together according to the relative locations of the face organs. These locations are obtained in the previous part. Since the 3D coordinates obtained are already in the same coordinate system of the head shape model, the merging process is simply placing them together. After the integration, we now have a complete wire-frame model of the target head without texture-mapping.

5 Texture-Mapping

To improve the realism of the result, the 3D model obtained in the previous section will be texture-mapped. Before the texture-mapping, a polygon mesh is formed on the model obtained. The texture-mapping is making use of the three original images. For the 3D model obtained, each composing polygon face will be projected onto the image plane with one of the original images. The overlapping 2D polygon on the image will be clipped out. The clipped part is wrapped and stretched to fit the polygon shape, and then served as the texture map of the polygon face. The blending of the images is determined by the angle between the normal vector of the polygon face and the viewing directions of the images. As shown in Figure 9, for a point p on a polygon face with normal vector n, the projection of point p has intensity I_a and I_b on images (a) and image (b) respectively. The d_a and d_b are the viewing directions of the image (a) and image (b) respectively. We define the angle between two vectors x and y as $\phi_{(x,y)}$. Then the resulting intensity of point p is:

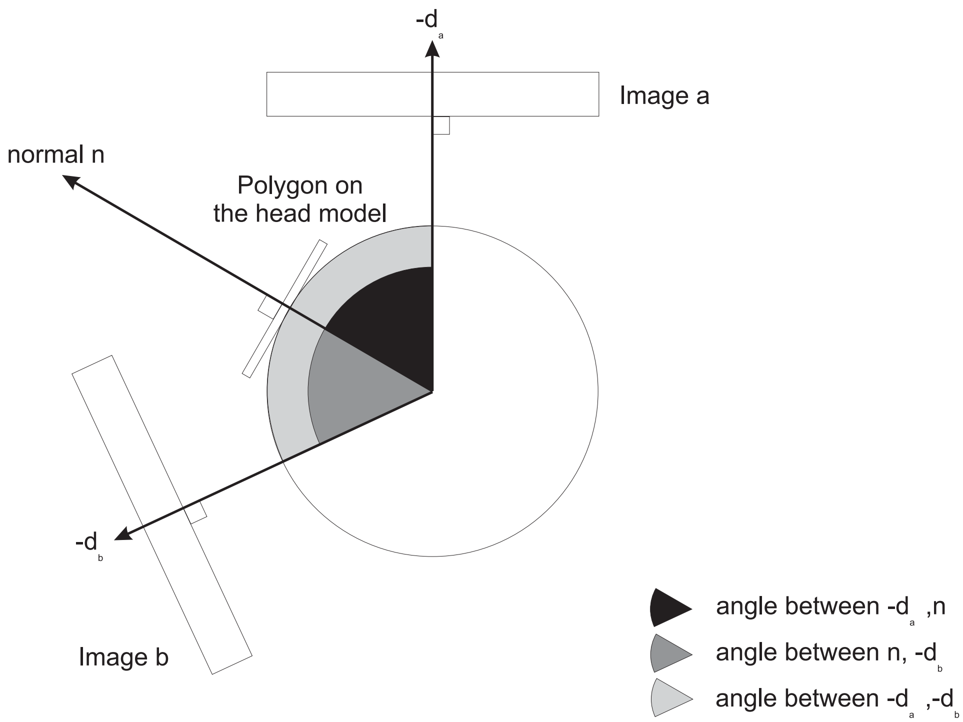

Figure 9: Blending of texture images

$$I_p = \frac{1}{\phi_{(-d_a,-d_b)}}[I_a(\phi_{(-d_a,-d_b)} - \phi_{(-d_a,n)}) + I_b(\phi_{(-d_a,-d_b)} - \phi_{(n,-d_b)})] \qquad (18)$$

However, for some of the face organs, the same image will be used for the texture-mapping of the entire organ rather than independently selecting image for each polygon face to avoid distortion.

6 Experiment Result & Discussion

We have tested our prototype system on real images using a CCD digital camera. Figure 10 shows some different views of the generated human head model, named as example 1, 2 and 3 from top down. The number of images used in our design is three. This is a consideration of both quality of result and the cost of the implementation. Three images of roughly orthogonal views provide enough information for the modeling and also good quality of textures for the texture-mapping of the wire-frame model. In fact, more images can improve the quality of textures and provide more information for modeling, with higher cost, of course. The extraction of landmarks from images is

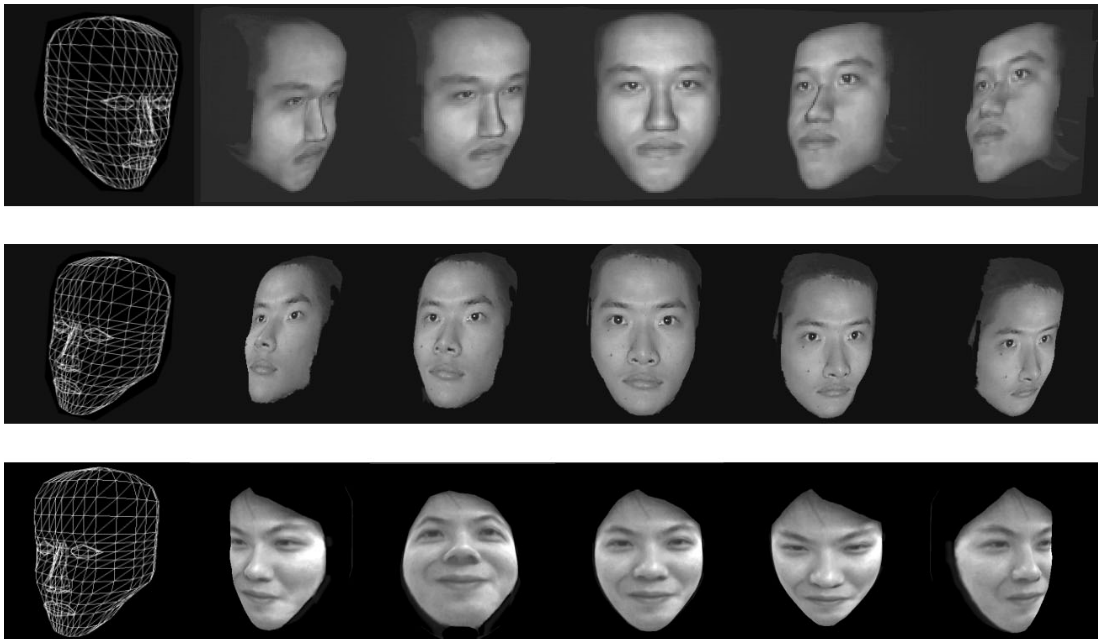

Figure 10: Multiple views of the generated models
(The VRML version of these face models can be downloaded from http://www.cse.cuhk.edu.hk/~khwong/demo/face/face.html)

done manually in our experiment. This part can be done by some human face feature recognition techniques[19]. However, it is a difficult problem and is not the focus of our paper. Moreover, automatic extraction means possible locations and correspondences error in the process, which may cause errors in the calculation of the 3D coordinates.

6.1 Head Shape Modeling Error

There are many factors affecting the correctness of the resulting model. The modeling error of the head shape can be measured by the non-overlapping area between the projection of resulting model and the real head shape in the image. The sizes of those mismatching areas are measured for each of the three original views. They are represented as proportions out of the sizes of the entire head shapes, as shown in Table 1.

We can observe that the margins of errors are always limited under 8%. It shows that the head shaping of our process is quite accurate. The left and right views have

Table 1: Mismatching area proportion in percentage

	Example 1	Example 2	Example 3
Front view	4.70%	6.12%	5.67%
Left view	6.56%	7.15%	7.77%
Right view	6.60%	5.06%	7.87%
Mean	5.95%	6.11%	7.10%

Table 2: Mismatching area proportion in percentage

	Example 1	Example 2	Example 3
Left 45^o	8.23%	7.58%	5.62%
Right 45^o	6.42%	8.27%	8.20%
Mean	7.32%	7.93%	6.91%

slightly larger mismatching areas than the front view. This may be resulted from the more complex shapes of the side views of a human head. For more complex shapes, the contour-based model will be too coarse for details. To improve the quality, more contours have to be used.

Since the above three views are used to generate the models, we have also measured the errors for left 45^o and right 45^o as shown in Table 2. This is done to demonstrate the correctness for those in-between views, which are not used in modeling the head shape. We found that the margins of errors in these two views are also limited. Their values are only slightly larger than those of the previous three original views. This is obviously resulted from the fact that the models are more conformed to the original three views used in modeling than the two in-between views.

The number of contours used in the generic model is an important factor to determine the resolution and correctness of the resulting model. The larger the number of contours, the higher the resolution we will obtain. As shown in Figure 11, the non-overlapping area is smaller when we have more contours in the model. At the same time, the large number of contours for representing a more complicated generic model, require more time to construct and heavier computation load in model scaling. The

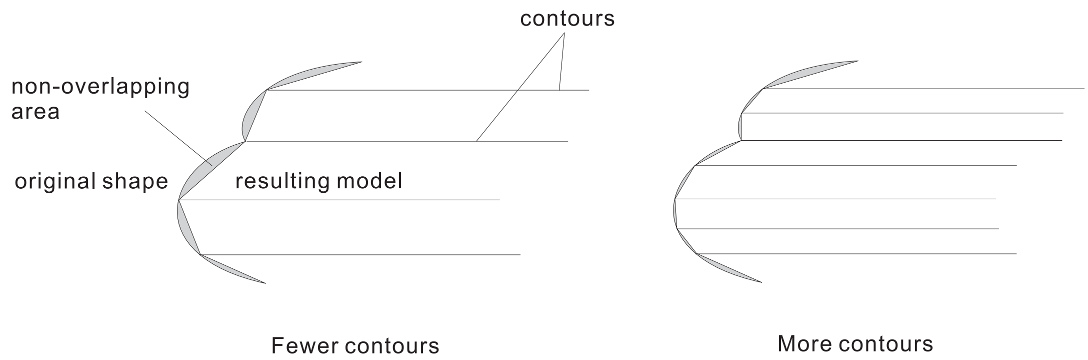

Figure 11: Modeling error from mismatching areas

number of contours used in our experiment is 16, which is a compromise between time and quality.

Another factor affecting the quality of the resulting model is the difference between the shape of contours and the shape of the target head. The mismatch is due to the fact that the shape of every human head will not be exactly the same. There is no "universal" generic model that perfectly fits for everyone. It is possible to introduce more than one type of generic model, for example, using different generic models for man and woman. Moreover, the deformation parameter is another factor affecting the modeling accuracy. The deformation parameters used in our design are the directions and scaling factors of contours in the x-axis and the z-axis, together with those of the entire model in the y-axis. If more images can be used, more deformation parameters can be chosen.

7 Conclusion and Future Direction

As accurate 3D model of a human head has great potential in many applications such as video-conferencing or population database. However, the manual construction of the 3D model of a real object is expensive and thus demands the need for highly automatic model construction. In this paper we present a robust and cost-effective method to construct a 3D human head model of a user from only three perspective camera images. The estimation of head shape is done by the deformation of a generic human head model that is constructed from a set of deformable contours. We also present a method to calculate the focal length of the camera and the 3D coordinates of the facial landmarks, which are used in the shaping and positioning of face organs, as well as the integration of head shape and face organs. We also provide an alternative for the 3D coordinates estimation using epipolar geometry when the extrinsic parameters are not available. The resulting model is texture-mapped by the images captured to increase realism. Our experiment result demonstrates that our system can produce realistic appearance of the generated model when viewed from different angles. Our work differs from the others in using perspective projection rather than some approximated orthographic image projection models. The future directions of our work include the utilization of variable number of images and variable angles between viewpoints, improving the extraction method of facial landmarks and improving the generic model by more detail deformation parameters. The use of multiple generic models will be investigated. The metrics and methods to measure modeling errors are also interesting topics to be studied.

Acknowledgement

This research was supported by the Research Grant council of Hong Kong, under the UGC earmarked grant of Number CUHK4389/99E.

References

[1] R. Koenen, "MPEG4 Overview", International Organisation For Standardisation, Coding Of Moving Pictures And Audio, ISO/IEC JTC1/SC29/WG11 N2995, October 1999, Melbourne

[2] H. Busch, "Automatic Modelling of Rigid 3D Objects Using an Analysis by Synthesis System", SPIE Proceedings, Visual Communication and Image Processing IV, vol. 1199, pp. 356-364, 1989

[3] C. H. Chien and J. K. Aggarwal, "Identification of 3D objects from multiple silhouettes using quadtrees/octrees", Computer Vision, Graphics, and Image Processing, vol. 36, no. 2/3, pp. 256-273, 1986.

[4] W. Niem, "Robust and Fast Modelling of 3D Natural Objects from Multiple Views", SPIE Proceedings, Image and Video Processing II, vol. 2182, pp. 388-397, 1994

[5] M. Potmesil, "Generating octree models of 3D objects from their silhouettes in a sequence of images", Computer Vision, Graphics, and Image Processing, vol. 40, no. 1, pp. 1-29, 1987

[6] R. Szeliski, "Rapid Octree Construction from Image Sequences", Computer Vision, Graphics, and Image Processing. Image Understanding, vol. 58, no. 1, pp. 23-32, 1993

[7] A. Lanitis, C. J. Taylor and T. F. Cootes. "A Unified Approach for Coding and Interpreting Face Images", Proceedings ICCV '95, pp. 368-373, 1995

[8] Y. Moses, D. Reynard, A. Blake, "Robust Real Time Tracking and Classification of Facial Expression", Proceedings ICCV '95, pp. 296-301, 1995

[9] I. Essa, A. Pentland, "Facial Expression Recognition Using a Dynamic Model and Motion Energy", Proceedings ICCV '95, pp. 360-367, 1995

[10] D. DeCarlo, D. Metaxas, "The integration of Optic flow and Deformable models with Applications to Human Face Shape and Motion Estimation", Proceedings CVPR '96, pp. 231-238, 1996

[11] L. G. Farkas, "Anthropometry of the Head and Face", Raven Press, 1994

[12] G. Xu, H.Agawa, Y. Nagashima, F. Kishino, Y. Kobayashi, "Three-Dimensional Face Modeling for Virtual Space Teleconferencing Systems", Transactions of the IEICE, vol. E 73, no. 10, pp. 1753-1761, 1990

[13] T. Akimoto, Y. Suenaga, "3D Facial Model Creation Using Generic Model and Front and Side Views of Face", IEICE Transactions on Information & Systems, vol. E 75, pp. 191-197, 1992

[14] L. A. Tang, T. S. Huang, "Automatic Construction of 3D Human Face Models Based on 2D images", Proceedings ICIP '96, pp. 467-470, 1996

[15] C. J. Kuo, R. S. Huang, T. G. Lin, "Synthesizing Lateral Face From Frontal Image Using Anthropometric Estimation", Proceedings ICIP '97, pp. 133-136, 1997

[16] L. Yin, A. Basu, "MPEG4 Face Modeling Using Fiducial Points", Proceedings ICIP '97, pp. 109-112, 1997

[17] G. Xu, Z. Zhang, "Epipolar Geometry in Stereo, Motion and Object Recognition. A Unified Approach", Kluwer Academic Publishers, 1996

[18] H. C. Longuet-Higgins, “A Computer Algorithm for Reconstructing a Scene from Two projections”, Nature, vol. 293, no. 10, pp. 133-135, 1981

[19] DeSilva, L.C., Aizawa, K., Hatori, M., “Detection and Tracking of Facial Features by Using a Facial Feature Model and Deformable Circular Template”, IEICE(Info. and Sys. E78-D), no. 9, pp. 1195-1207, 1995