THREE DIMENSIONAL FACIAL MODEL ADAPTATION

Nikolaos Sarris and Michael G. Strintzis

Information Processing Laboratory, Electrical and Computer Engineering Department
Aristotle University of Thessaloniki, Thessaloniki 54006, Greece
phone: +30 (31) 996349, fax: +30 (31) 996398, e-mail: nikos@dion.ee.auth.gr

ABSTRACT

This paper addresses the problem of adapting a generic 3D face model to a human face of which the frontal and profile views are given. Assuming that a set of feature points have been detected on both views the adaptation procedure initializes with a rigid transformation of the model aiming to minimize the distances of the 3D model feature nodes from the calculated 3D coordinates of the 2D feature points. Then, a non-rigid transformation ensures that the feature nodes are displaced optimally close to their exact calculated positions, dragging their neighbors in a way that does not deform the facial model in an unnatural way.

1. INTRODUCTION

The adaptation process of a generic three-dimensional mesh onto one or more projected images of a 3D object involves the registration of the mesh on the projections, followed by its rigid and non-rigid 3D deformation. The success of this method relies on deforming the mesh in a way that it retains its basic characteristics while its final shape resembles the object to be modeled. Depending on the nature of the object to be modeled a different generic model can be used thus varying the field of possible applications from 3D modeling of simple geometric objects to complex surgical simulation on the human body.

This work studies an adaptation method of a generic 3D human face model on two orthogonal views of a particular face. The realistic deformation of generic 3D face models has been widely sought by the research community as the potential applications in model based image coding have been remarkably popular the recent years with video telecommunications systems requiring the lowest possible bandwidth and graphics applications demanding immoderate realism [1].

Several approaches have been proposed for the deformation of a generic 3D face model: In both [2] and [3] the required geometry of the face is captured by a 3D laser scanner. In [2] the generic geometric model is deformed by applying physical spring forces to predetermined nodes according to the positions of the corresponding features in the target face geometry. In [3] the generic model is built using triangular B-spline patches and the deformation is performed by displacement of the spline control points which correspond to facial features. The required face geometry in [4] is deduced from two orthogonal photos of the target face and the generic face model is deformed by an extension of the Free Form Deformation (FFD) [5], the Dirichlet FFD [6]. In [7] the required 3D positions of the facial features are estimated from a series of captured image frames of the target face and the generic model is transformed by applying an interpolation function based on radial basis functions. Information from one view of the target face is utilized in [8] to measure the face, eyes, and mouth dimensions which are used to adapt the Candide model by rigidly transforming the whole face and locally correcting the position and orientation of the face, eyes and mouth. Geometric assumptions have to be made however, as the 3D characteristics of the features cannot be totally deduced from only one view.

Our approach differs from the above methods mainly in treating the facial model as a collection of facial parts, which are given the liberty to deform under separate affine transformations. This should be a natural assumption as the physiological differences in characteristics between faces are based on such local variations i.e. a person may have a longer or shorter nose, narrower eyes etc. Thus, in the proposed method after rigidly transforming the whole model so that it is aligned and scaled according to the target face, each facial part is stretched, rotated and translated separately and in the optimum way to fit to the required corresponding target facial part. The model nodes between different parts are interpolated to give a natural smooth transition from one facial part to the other.

2. THREE DIMENSIONAL MODEL ADAPTATION

2.1. Stereo Image Acquisition

In the proposed method the 3D positions of a set of characteristic facial feature points are required to initiate the generic 3D facial model deformation. In order to calculate accurately the coordinates of the feature points in the 3D space, more information is needed than that provided by a single 2D image frame. In the proposed scheme we utilize the combined information from two views of the face at right angles to each other. These may be obtained either by a stereoscopic system of two cameras at right angles to each other (as can be seen in Figure 1), or by a rotating single camera.

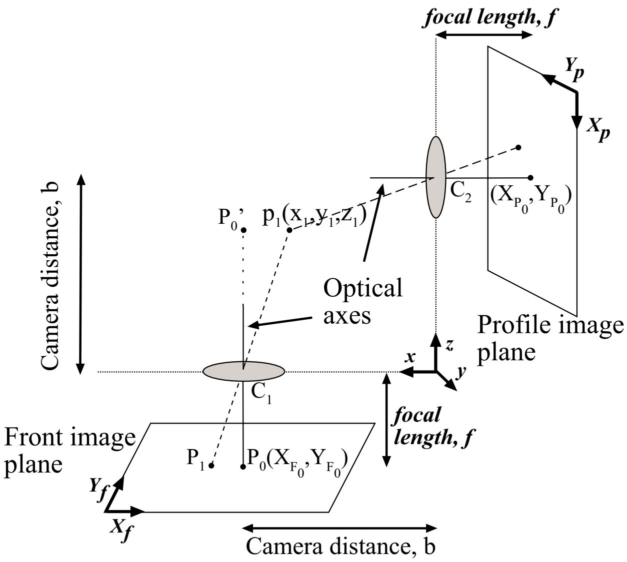

Figure 1: Image acquisition layout

To locate the set of feature points in the 2D image views various methods have been proposed in the literature [9][10][11][12]. Our approach [13] is based on the utilization of chrominance information for the localization of the face and the exploitation of knowledge of the geometrical structure of the face for finding the exact location of the feature points.

Having located the set of characteristic feature points in both views the calculation of their 3D coordinates is carried out in the following way:

We shall assume the use of a perspective projection camera system, which is shown in Figure 1. The 3D points with coordinates *(x, y, z)* are projected on two image planes, the front and the profile, with perspective rays passing through the two corresponding projection centers C_1 and C_2, which lie within the physical camera and are at a distance b from each other. The focal length f is the distance of the image planes from the corresponding centers of projection, while the point where the optical axis of the camera intersects the image plane is called the principal point. For the frontal image plane the principal point coordinates are (X_{Fo}, Y_{Fo}) and for the profile (X_{Po}, Y_{Po}). In our experiments these were fixed to the centers of the corresponding projection images. Finally, X_F and Y_F are the projections on the frontal view, while X_P and Y_P are the projections on the profile view of the person's face.

The projection coordinates for the frontal image can be computed from the similar triangles $P_1C_1P_0$ and $p_1C_1P_0$', shown in Figure 1. In the same way, the projection coordinates for the profile image may also be determined:

$$X_F = f\frac{-b+x}{z} + X_{F_0} \qquad X_P = f\frac{b-z}{x} + X_{P_0}$$
$$Y_F = f\frac{y}{z} + Y_{F_0} \qquad Y_P = f\frac{y}{x} + Y_{P_0} \tag{1}$$

This system gives for every feature point four equations with three unknowns (the 3D position of the feature point *(x, y, z)*). This is solved by least squares methods to provide the 3D position *(x, y, z)* of every feature point, given the projections of the feature points on the frontal (X_F, Y_F) and profile view (X_P, Y_P).

2.2. Rigid Adaptation

Having calculated the positions of the feature points in 3D, a generic 3D facial model needs to de deformed so that its corresponding feature nodes are displaced as close as possible to these required positions, while preserving the natural smoothness and geometry of the human face.

The first part of the deformation involves a rigid transformation of the model. Thus, the model is rotated and translated to match the pose of the real face. The rotation and translation transformations are calculated by a slight modification of the 'spatial resection' problem in photogrammetry [14]. Specifically, the relation between the initial and the transformed model is given by:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathbf{R}\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + \mathbf{T} \tag{2}$$

where (x_0, y_0, z_0) are the coordinates of a model node at its initial position, *(x', y', z')* are the coordinates of the transformed node, **R** is a 3x3 rotation matrix and **T** is a 3D translation vector $[T_x, T_y, T_z]^T$. We shall assume a left-hand coordinate system, and let ω be the angle of rotation around the x-axis of the camera reference frame, φ the angle of rotation around the y-axis, and κ the angle of rotation around the z-axis. The κ, φ and ω angles are often called "Euler angles". Then the rotation matrix **R** will be given by [14]:

$$\mathbf{R} = \mathbf{R}(\kappa)\mathbf{R}(\phi)\mathbf{R}(\omega), \text{ where:}$$
$$\mathbf{R}(\omega) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{bmatrix}$$
$$\mathbf{R}(\phi) = \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{bmatrix} \tag{3}$$
$$\mathbf{R}(\kappa) = \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Our purpose is the minimization of the following sum of squares by appropriate selection of the 6 unknown parameters (3 translation coefficients: T_x, T_y, T_z and 3 rotation coefficients: κ, φ, ω):

$$\sum_{i=1}^{N}\left(\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} - \begin{bmatrix} x_i' \\ y_i' \\ z_i' \end{bmatrix}\right)^2 \tag{4}$$

where N is the number of feature points located in the front and profile image frames, (x_i, y_i, z_i) are their required positions in the 3D space calculated as described in the previous subsection and (x_i', y_i', z_i') are the positions of the model 3D nodes after the transformation (2). This minimization is accomplished by use of the non-linear Levenberg-Marquadt algorithm [15]. A number of iterations of the above method are used to determine the values of the Euler angles corresponding to the rotation matrix **R** and the three components of the translation vector **T.**

2.3. Non-Rigid Adaptation

The rigid transformation may align the generic model with the required face and scale it to meet the total face dimensions. However, the local physiology of the face cannot be altered this way. Thus, in the second step of the adaptation process the model is transformed in a non-rigid way aiming to further displace the feature nodes bringing them as close as possible to their exact calculated positions while the facial parts retain their natural characteristics. To perform this adaptation the model is split into face parts (left eye, right eye, mouth, etc.) and a rigid adaptation is performed on every part separately. This means that every 3D face part is rotated, translated and stretched optimally minimizing the distances of the feature points (belonging to that face part) from their required positions. This is accomplished with the following transformations:

1. Centering at the origin: The center of the face part is found as the 3D center of the contained feature nodes and the whole part is translated towards the origin so that this center falls on the origin. The same is done with the set of required feature positions (i.e. their center is found and they are translated towards the origin in the same way).

2. Alignment: The face part is rotated around the origin so that three lines connecting three pairs of feature nodes are aligned with the corresponding lines connecting the required feature positions. This is accomplished by minimizing the differences in the gradients of these lines. For one pair of nodes at (x_1, y_1, z_1) and (x_2, y_2, z_2), with required positions at (x_1', y_1', z_1') and (x_2', y_2', z_2') this would involve finding the 3x3 rotation matrix **R** which would perform the following minimization:

$$\min\left\|\left(grad_y - grad_y'\right)^2 + \left(grad_z - grad_z'\right)^2\right\| \tag{5}$$

where:

$$grad_y' = \frac{y_2' - y_1'}{x_2' - x_1'},\quad grad_z' = \frac{z_2' - z_1'}{x_2' - x_1'},$$
$$grad_y = \frac{y_2'' - y_1''}{x_2'' - x_1''},\quad grad_z = \frac{z_2'' - z_1''}{x_2'' - x_1''} \tag{6}$$

and:

$$\begin{bmatrix} x_i'' \\ y_i'' \\ z_i'' \end{bmatrix} = \mathbf{R}\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix},\quad i = 1,2 \tag{7}$$

3. Stretching: The face part is scaled around the origin with different scale factors for every axis (s_x, s_y, s_z) so that the distances of the transformed nodes from their required positions are minimized. Thus, if (x_i, y_i, z_i) are the coordinates of the feature nodes after the rotation and (x_i', y_i', z_i') their required positions this would involve finding the 3x3 scaling matrix **S** which would minimize the following set of equations:

$$\sum_{i=1}^{N}\left(\begin{bmatrix} x_i' \\ y_i' \\ z_i' \end{bmatrix} - \begin{bmatrix} x_i'' \\ y_i'' \\ z_i'' \end{bmatrix}\right)^2 \tag{8}$$

where:

$$\begin{bmatrix} x_i'' \\ y_i'' \\ z_i'' \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix}\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix},\quad i = 1,\ldots,N \tag{9}$$

4. Translation: After the stretching transformation the face part is translated back towards its original position by adding the position vector of the face part center calculated and subtracted in step 1.

Results of these steps for two face parts are shown in Figure 2 and Figure 3.

Between all face parts one series of nodes is left as border nodes (i.e. these nodes do not belong to any face part). The positions of these nodes after the deformation are found by linear interpolation of the neighboring nodes belonging to face parts. This is done to assure that a smooth transition is made from one facial part to the other and possible discontinuities are filtered out.

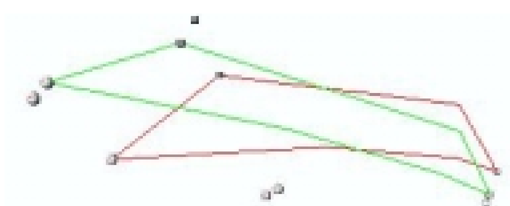

Figure 2: Local Adaptation of the right eyebrow

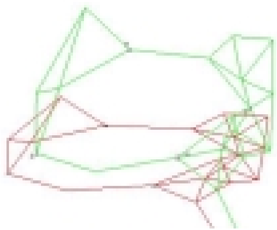

Figure 3: Local Adaptation of the right eye

Thus, the final deformed model adapts to the particular characteristics implied by the feature points (e.g. bigger nose or smaller eyes) keeping the generic characteristics of a human face (e.g. smoothness of the skin and symmetry of the face).

3. RESULTS

Current results (Figure 4) show the rigid adaptation to bring the model as close as possible to the required face, followed by the non-rigid transformation which adapts the model to the characteristics required and retains the natural smoothness and symmetry of a human face.

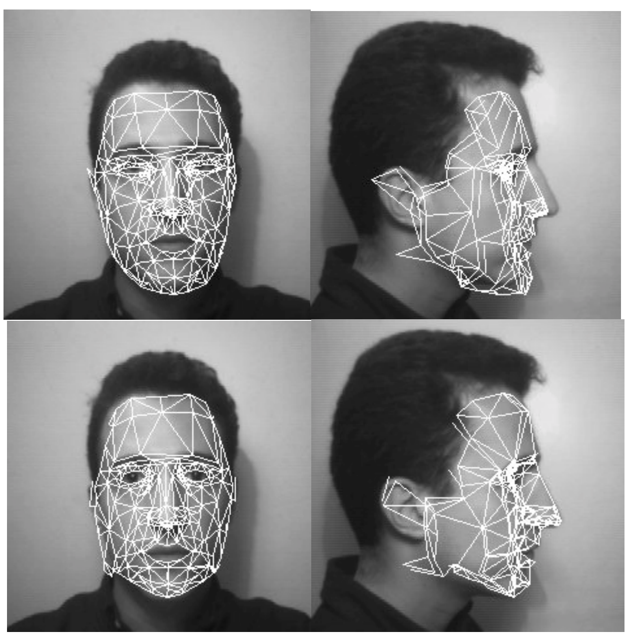

Figure 4: Rigid (upper) and non-rigid (lower) adaptation of the face model shown on the front and profile views

4. ACKNOWLEDGEMENTS

This work was supported by the Greek Secretariat of Research and Technology, AMEA project 'LipTelephone: A Videophone for the Deaf' and by the European Community IST project 'HISCORE'

5. REFERENCES

[1] SIGGRAPH 97 Panel on Facial Animation: Past, Present and Future, Los Angeles, CA, August 1997

[2] Y. Lee, D. Terzopoulos, K. Waters, "Realistic Modeling for Facial Animation", Computer Graphics (Proc. SIGGRAPH'95), pp.55-62, 1995

[3] P. Eisert and B. Girod, "Analyzing Facial Expressions for Virtual Conferencing", IEEE Computer Graphics and Applications, pp. 70-78, September 1998

[4] W.S. Lee, M. Esher, G. Sannier, N. Magnenat-Thalmann, "MPEG-4 Compatible Faces from Orthogonal Photos", International Conference on Computer Animation, Geneva, Switzerland, pp. 186-194, May 1999

[5] T.W. Sederberg, S.R. Parry, "Free Form Deformation of Solid Geometric Models", Computer Graphics (Proc. SIGGRAPH'86), pp.151-160, 1986

[6] L. Moccozet, N. Magnenat-Thalmann, "Dirichlet Free-Form Deformations and their Application to Hand Simulation", International Conference on Computer Animation, IEEE Computer Society, Geneva, Switzerland, pp.93-102, 1997.

[7] F. Pighin, J. Auslander, D. Lischinski, D. Salesin, R. Szeliski, "Realistic Facial Animation Using Image Based 3D Morphing", Technical Report UW-CSE-97-01-03, University of Washington, May 1997.

[8] L. Zhang, "Automatic Adaptation of a Face Model Using Action Units for Semantic Coding of Videophone Sequences", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No 6, pp.781-795, October 1998

[9] D. Chai, K.N. Nghan, "Face Segmentation using Skin Color Map in Videophone Applications", IEEE Transactions on CSVT, Vol. 9, No. 4, June 1999

[10] L.A. Tang, T.S. Huang, "Automatic Construction of 3D Human Face Models Based on 2D Images", International Conference on Image Processing 1996, vol.3, pp467-470, 1996.

[11] T.C. Chang, T.S. Huang, "Facial Feature Extraction from Color Images", 12th International Conferene on Pattern Recognition, vol 2, pp 39-43, 1994

[12] A.A. Qayedi and A.F. Clark, "An algorithm for Face and Facial Feature Location Based on Grey Scale Information and Facial Geometry", 7th International Conference on Image Processing and its Applications, vol. 2, pp. 625-629, 1999

[13] N. Sarris, P. Karagiannis and M.G. Strintzis "Automatic Extraction of Facial Feature Points for MPEG4 Videophone Applications", International Conference on Consumer Electronics, Los Angeles, USA, June 2000

[14] R.M. Haralick and L.G. Shapiro, Computer and Robot Vision, Volume II, pp.125-150, Addison Wesley, 1993.

[15] The minpack library, Netlib repository, [online]: http://www.netlib.org/minpack/