From 3D Data to Animated Face Models through MPEG-4 Model Reshaping

Ludovico Basili, Augusto Sarti, Stefano Tubaro
Dipartimento di Elettronica e Informazione, Politecnico di Milano
Piazza Leonardo da Vinci 32,
20133 Milano, ITALY

ABSTRACT

In this paper we propose a method for the creation of an animation-ready face model based on the reshaping of a given source model onto a reference mask.

1. INTRODUCTION

The multimedia MPEG-4 standard describes the audio-visual scene as made of different "media" objects organized in a hierarchical fashion. One of the primitive objects defined in this standard is the "face model", which can be displayed and animated according to a number of predefined rules. In particular, FAPs (Facial Animation Parameters) are used to model and encode facial movements. In order to do so, the face object is treated as a pseudo-muscular model [1].

There are currently a significant number of systems that are able to collect rich and accurate three-dimensional and textural information relative to a human face. Laser scanners, for example, are widely employed in the creation of virtual actors and avatars in the entertainment industry. Moreover, there are several modeling systems, based on digital cameras or photo-cameras, which are able to generate dense 3D information using images taken from different viewpoints. Such systems, in addition to geometric 3D information, also retrieve the textural information that is needed to obtain a realistic model of the viewed face (see, for example, [2][3] and Fig. 2). These models can be easily saved in VRML format, and their textural information can be obtained from all the available views after compensating problems such as the non-uniform illumination and the non-Lambertian reflectance of the skin (see, for example, [4]).

Unfortunately, such models cannot be animated as they are, mostly because the vertices of their mesh of triangles do not correspond to physical (semantic) features of the face (see Fig. 1). In fact, the distribution of such triangles, which is usually quite regular, depends on the 3D data acquisition/estimation process.

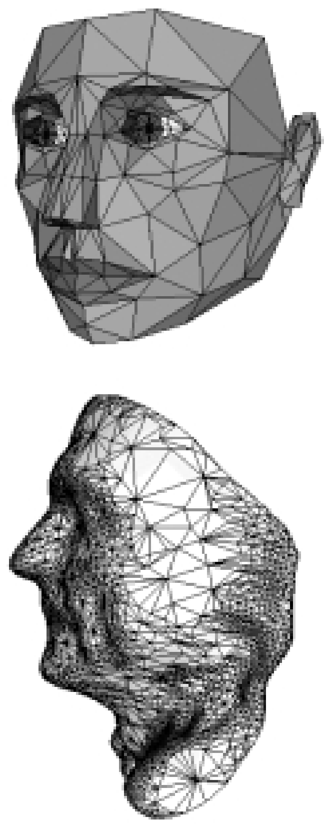

Fig. 1. Two meshes of triangles for face modeling. Above: specialized model for face animation. Bottom: unstructured model obtained through image-based modeling.

One possible solution to this problem could be the adjustment of a sub-set of the vertices, chosen as the closest ones to significant key-points of the face. Examples of relevant face features are the corners of eyes and mouth, the tip of the nose, the "prominent" points of the cheeks, some points that mark the eyebrow's profile, etc. Once all such points are optimally relocated, their animation rules, and the consequent motion of all the other vertices, can be defined based on the displacements of the key features. Indeed, this requires the definition of different animation rules for each model, as the mesh geometry may vary from model to model.

In order to overcome this difficulty, we could reverse the above strategy by reshaping a predefined face model (*source model*) onto the acquired 3D model (*target mask*) (see Fig. 3). The advantage of this approach is that the animation rules can be directly defined on the source model, and re-used with reshaped model (as long as the model deformations are not excessive).

2. MODEL RESHAPING

As described above, our approach to the creation of an animation-ready face model is based on the reshaping of a given *source model* onto a *reference mask*. We developed and tested our solution considering two MPEG-4 source face models created by Lavagetto's research group at DIST (University of Genova, Italy): the "Oscar" and the "Mike" models [5].

The fist step of the reshaping process consists of the identification of the control-points on the reference mask, which correspond to the pre-assigned control points on the source model (see Fig. 4). The number of control points is usually between 30 and 50 or more [6]. In order to do so as quickly as possible, with a certain flexibility in the 3D-model handling, and without having to rely on commercial software, we developed a specific Java-3D application based on the metaphor of the "light pen". A mouse click on a view of the face mesh selects the vertex that lies the closest to the intersection between the surface and the "light beam" (a thick visual ray passing through the optical center of the viewpoint and the arrow tip of the mouse). The thickness of the light beam can be changed in order to increase or reduce the selection accuracy. This preliminary process provides us with a set of one-to-one relationships between control points on the reference mask and on the source model.

Model reshaping is achieved by changing the position of all the vertices of the source mesh by means of an appropriate deformation function, which is defined for all the points of the 3D space, but it is evaluated only at the vertices of the source model.

More specifically, the deformation function can be defined as follows

$$s(P) = \sum_{i=1}^{N} c_i R\left(\left\|P - P_i\right\|\right) \qquad (1)$$

where N is the number of the selected feature-points and P_i are their positional coordinates on the source model. The function $R(\cdot)$ is chosen to belong to the class of Radial Basis Functions [7], as they are known to provide a good interpolation of sparse data sets [8]. In our case, the sparse data set is made of the 3D deformation vectors associated to the key-points of the source model. The coefficients c_i are computed by solving a linear system whose vector of constants is the set of available deformations. Among the possible choices of the function $R(\cdot)$ we obtained the best results using Hardy functions [8]:

$$R(P, P_i) = \left(\left\|P - P_i\right\|^2 + r^2\right)^{1/2} \qquad (2)$$

The r parameter is defined as the "stiffness radius" and it controls the smoothness of the interpolator. Our implementation uses an adaptive version of such radial basis functions, where each P_i is given a different value of r (r_i) in order to take into account the fact that the control-points are not uniformly scattered in the space.

Once the function defined by eq. (1) is available, the deformation of the source model may take place. Our experiments showed that the proposed solution to the reshaping problem performed well in a variety of situations (see Fig. 5).

In order give our approach a better ability to adapt to small details, we added a second deformation phase based on the so-called Discrete Smooth Interpolation [9][10]. The vertices of the source model mesh (after the first deformation phase) are iterately "pulled" toward the reference model in their current normal direction, while taking a global smoothness constraint into account in order to avoid unnatural results.

At the end of the process, textural information is extracted from the images used to create the reference mask [4] and applied to the reshaped source model.

The whole process takes between 20 and 30 minutes on a PC equipped with Pentium III 600 MHz processor. More that half of this time, however, goes in the selection of the key-points on the reference mask.

3. CONCLUSIONS

In this paper we proposed and described a system for the reshaping of a source face model onto a reference face model with the goal of generating animation-ready face models that resemble a real person as much as possible.

The tests that we carried out demonstrated the potential of the proposed approach. Currently, very simple source models (e.g. OSCAR, with about 400 vertices) have been used, therefore the final quality of the reconstructed face is quite limited. Better results are to be expected using more complex source models. We are currently working on the automatization of the key-point detection on the reference model, although a complete automatic process is very difficult and often impossible [6].

4. REFERENCES

[1] F.I. Parke, K. Waters: *Computer Facial Animation.* A.K.Peters, Wellesley, MA, 1996.

[2] P. Pigazzini, F. Pedersini, A. Sarti, S. Tubaro: “3D Area Matching with Arbitrary Multiview Geometry”. EURASIP Signal Processing: Image Communications, special issue on 3D Video Technology, edited by S.A. Benton, B.Choquet, R. ter Horst, K.N. Ngan and M. Tanimoto. Vol. 14, Nos. 1-2, 1998, pp. 71-94.

[3] F. Pedersini, A. Sarti, S. Tubaro: “Multi-resolution 3D reconstruction through texture matching”. EUSIPCO 2000. Tampere, Finland, Sept. 5-8, 2000. Vol. 4, pp. 2093-2096.

[4] F. Pedersini, L. Piccarreta, A. Sarti, S. Tubaro: *Estimation of Radiometric Parameters for a Realistic Rendering of 3D Models*. International Conference on Image Processing, ICIP-99, October 25-28, 1999, Kobe, Japan..

[5] http://www-dsp.com.dist.unige.it/~pok/RESEARCH/index.htm

[6] V. Blanz, T. Vetter: *A Morphable Model for the Synthesis of 3D Faces*. Computer Graphics Proceeding SIGGRAPH-2000, July 2000, New Orleans

[7] N. Arad et al.: *Image Warping by Radial Basis Functions: Applications to Facial Expression*. Computer Vision, Graphics and Image Processing, Vol. 56, pp. 161-172, March 1994.

[8] B.J.C. Baxter: *The Interpolation Theory of Radial Basis Function*. PhD Thesis, Cambridge University, August 1992.

[9] J.L. Mallet: Discrete Smooth Interpolation. ACM Transaction on Graphics, Vol.8, N.2, pp.121-144, April 1989.

[10] F. Pedersini, A. Sarti, S. Tubaro: *Visible Surface Reconstruction with Accurate Localization of Object Boundaries*. IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on Three-Dimensional Video Technology, Vol. 10, N. 2, March 2000, pp. 278-291.

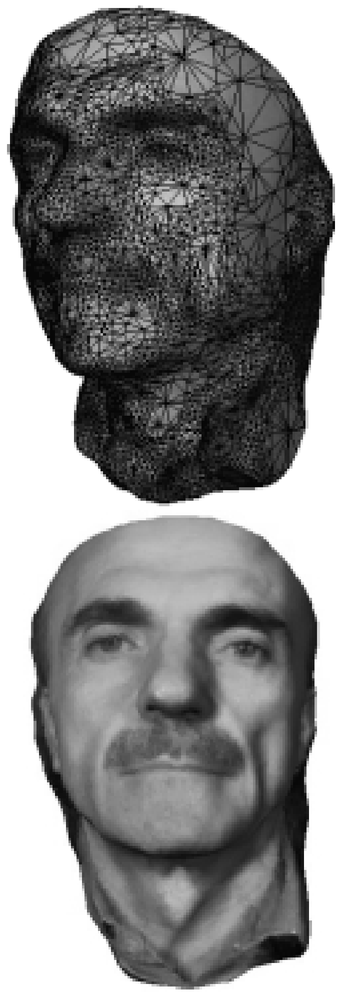

Fig. 2. A 3D model obtained from the system illustrated in [2] and [3]

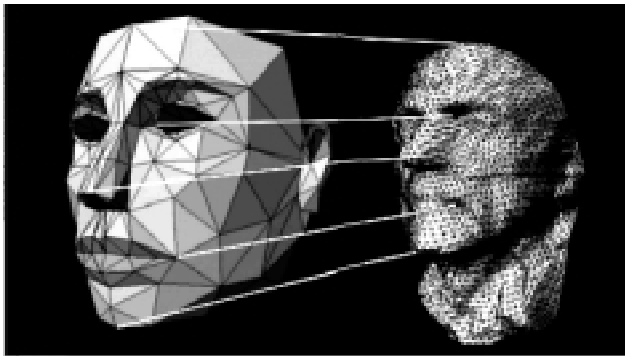

Fig. 3. The source model is adapted on the target model as a deformable mask.

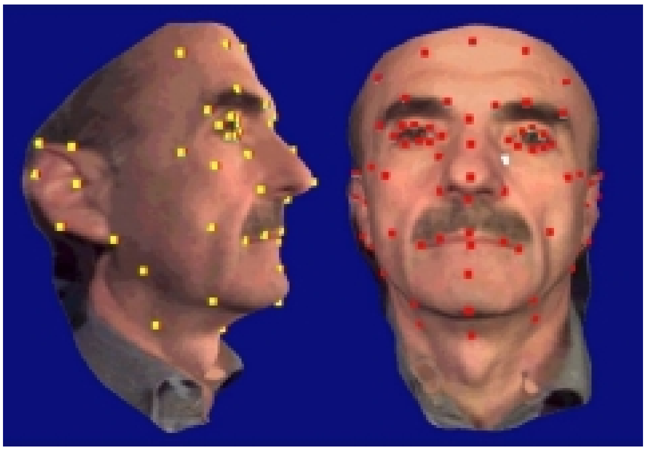

Fig. 4. The selected key-points on the target model.

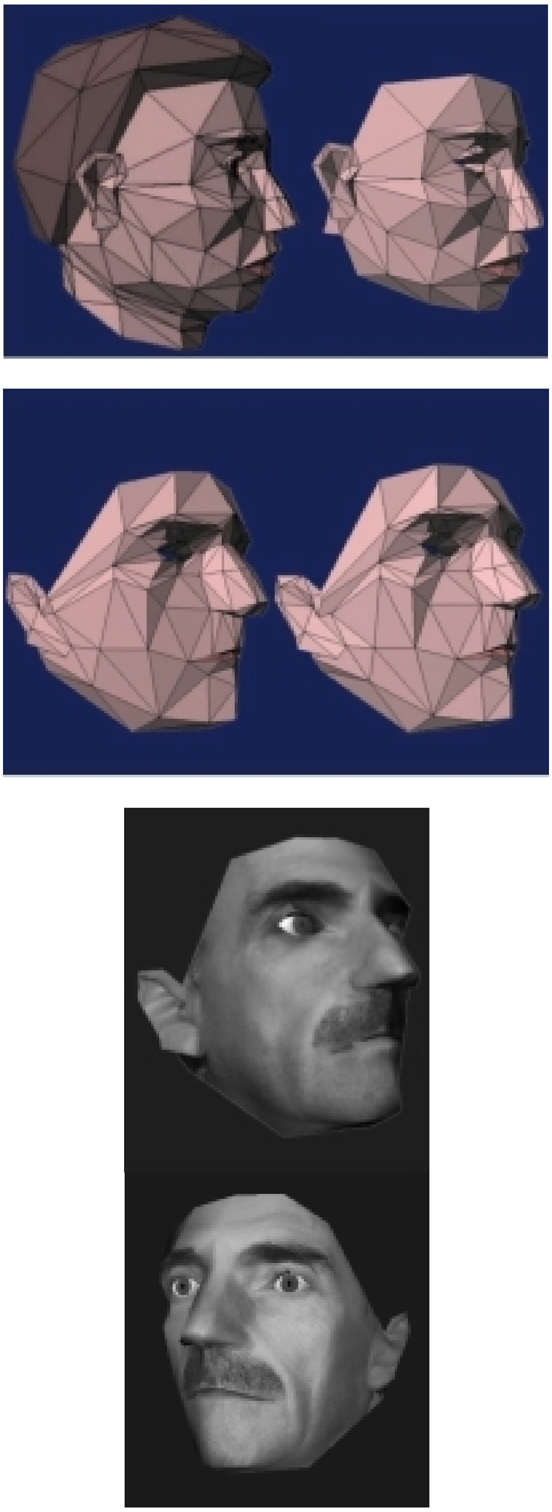

Fig. 5. The various phases of the source model deformation