A SOLUTION FOR MODEL-INDEPENDENT ANIMATION OF MPEG-4 FACES

Roberto Pockaj, Maurizio Costa, Fabio Lavagetto, Carlo Braccini

DIST – University of Genova, Italy
pok@dist.unige.it, maury@dist.unige.it, fabio@dist.unige.it, carlo@dist.unige.it

ABSTRACT

The technique proposed in this paper provides operative indications for implementing 3D head models compliant to the facial parameters recently standardized in MPEG-4, capable of being animated for reproducing the behavior of true human faces. The major contribution of this paper consists of an algorithmic proposal for MPEG-4 compliant model animation, independent of the particular 3D mesh used to model the synthetic face. The methodology is based on the definition of a set of pre-defined movements and their eventual composition, expressed as a function of each specific MPEG-4 animation parameter. Experimental results, as reported in the following, prove the effectiveness of the proposed method showing how a 3D mesh can be animated for reproducing both low-level movements and high-level facial expressions. A variety of applications are envisaged in multimedia products, virtual and augmented reality, very low bitrate videophone and efficient 3D facial data archiving. Ideas of this kind are proposed and discussed in the introduction and in the conclusions of the paper with reference to the work carried on within the European projects VIDAS[1] and INTERFACE[2].

1. INTRODUCTION

During the last three years there has been relevant standardization activity in MPEG-4 for defining an efficient system for encoding Audio-Visual Objects (AVO), either natural or synthetic. Towards the revolutionary objective of going beyond the conventional concept of the audio/visual scene as composed rigidly of a sequence of rectangular video frames with associated audio, MPEG-4 is suitably structured to manage any kind of AVO, with the capability to compose them variously in 2D and 3D scenes. In this framework, synthetic objects like human faces and human bodies have been given particular attention. In fact, among the multitude of possible objects that can be created artificially at the computer, MPEG has chosen to dedicate specific room to this class of objects that are considered to play a role of particular interest in audio-visual scenes. Therefore, the necessary suitable syntax has been standardized in MPEG-4 for the definition and animation of synthetic faces and bodies.

It is up to those who implement MPEG-4 face and body decoders to use the encoded parameter for synthesizing virtual faces and bodies. International research in computer graphics has proposed several solutions in the past decades for modeling and animating realistic 3D meshes reproducing human faces and bodies [1-3]. High level research has also been carried out for the anatomically correct modeling of bones, muscles and tissues [4-5]. A lot of expertise has been consolidated also in psychological interpretation of facial expressions [6] and in bimodal acoustic-visual analysis and modeling [7-8] of speech.

What is missing in this area, is how to merge all this knowledge into an integrated technology for MPEG-4 face and body animation capable of guaranteeing the automatic faithful calibration of any 3D models to the specific face and body geometry encoded by the MPEG-4 bitstream, of assuring similar capability for adapting the original texture from natural images to the virtual face/body, of performing realistic facial expression and lip movements in synchronization with speech.

The algorithms described in this paper offer a concrete contribution to anyone interested in implementing MPEG-4 animated faces for developing such kind of friendly user interfaces, either from the calibration and animation point of view. The recently achieved animation algorithms are described in Section 2 while final conclusions are drawn in Section 3.

[1] VIDAS was a 4FP ACTS project, coordinated by DIST, University of Genoa, Italy. It has been started in September 1995 and has been concluded in February 2000.

[2] INTERFACE is a 5FP IST project, coordinated by DIST, University of Genoa, Italy. It has been started in January 2000 and will be concluded on December 2002.

2. MPEG-4 FACIAL MODEL ANIMATION

MPEG-4 recommends the implementation of a parametric facial model with movements controlled by a limited set of parameters. However, the identification of a complete set of parameters turns out to be very difficult and a rigorous mathematical formalism is missing. For the first time MPEG-4 has proposed a standard set of animation parameters, called Facial Animation Parameters (FAP), leaving to each specific decoder implementation the responsibility of making the best possible use of them. Anyone of the many available facial animation software, which have followed up to now proprietary approaches, is able to animate only the specific facial model it has been tailored on. More than to the use of non standard parameters, which is easily overcome by suitable parameter conversion, this limitation is due to the non flexible relation that determines the effects produced by parameters on the geometry of the facial mesh. In other words, up to now no animation software can offer the high-level functionality for:

- modifying the shape of the facial mesh (with no topology modification)
- modifying the effects determined on the facial mesh by a single specific parameter

For this reason, we have chosen to develop an animation system with such functionality and fully compliant with the FAP animation parameters defined in MPEG-4. The solution that has been adopted consists of associating a semantic file to each model, with all the necessary hi-level information for the automatic generation of the animation rules, parameter by parameter.
In our approach, the structure, colors and texture of a facial model are specified in a data file written in VRML 2.0, format preferred to other alternatives not only because of its world wide diffusion, which has allowed easy exchange of models with other research centers and universities working on similar activities, but also for the possibility to use already existing applications like browsers and authoring tools for the generation, modification and visualization of 3D meshes.

Of the many complex functionalities provided by VRML 2.0, only a limited subset has been used for the description of the facial mesh. For ease of implementation, the facial mesh have been constructed as a patchwork of triangular rather than polygonal tiles, exploiting the planarity of the triangle surface and the easy computation of the normal vector.
The data file also contains a cross-reference table (ignored by standard browsers) to specify the correspondence of the 84 *feature points* defined in MPEG-4 to specific mesh vertices.
Differently from the model geometrical description, based on the existing VRML2.0 language, the semantic description has been specified according to a proprietary file format containing high-level information like:

- the *look-up* table for mapping high-level FAPs (visemes and facial expressions) in terms of a vector of low-level FAPs; each face model will then reproduce viseme and expressions in its unique way;
- correspondences between the *wireframe* vertices and the feature points;
- the displacement domain, defining the mesh region affected by each FAP;
- for rotational FAPs , the corresponding rotation axis;
- the indexes of the vertices affected by the displacement of the feature point associated to each specific FAP;
- the criterion for weighting the amplitude of vertices displacement with respect to the displacement of the *feature points.*

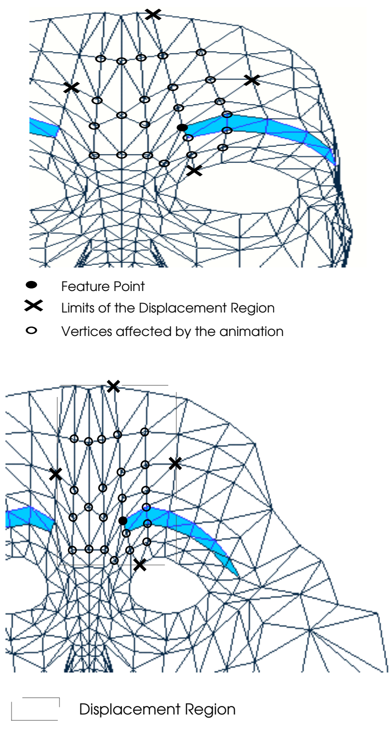
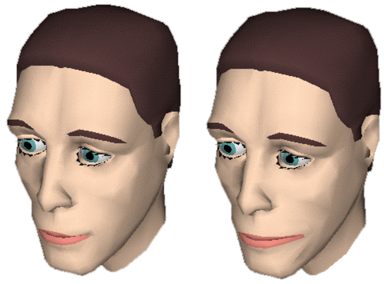


Figure 1 (Top) Graphical representation of the semantic information associated to FAP31 for the model "Oscar". (Bottom) Graphical representation of the semantic information associated to FAP31 with reference to the cylindrical projection of the model "Oscar"

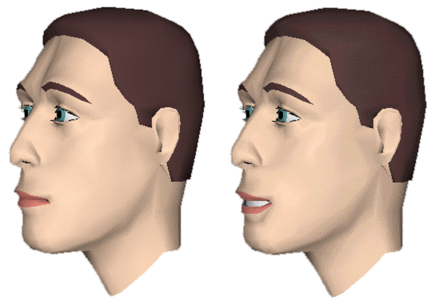
It should be noticed how all this information is expressed in a simple way, as a function of the indexes of the *wireframe* vertices, instead of their spatial coordinates or of their topology (see Figure 1). This fact makes the animation algorithm full compliant with the optional calibration phase that could have been applied beforehand. Whatever reshaping could have occurred on the proprietary model, the semantic information remains interpretable anyway.

The criterion to assign weights to a set of vertices, together with predefined movements, represents a critical key issue for generating the model animation rules. In our implementation, the displacement of each vertex associated to a feature point can be equal or proportional (according to a weight of value equal or less than one) to the displacement of the corresponding feature point. The computation of the weights is done automatically, based on the information contained in the semantic file, through a two-step procedure.
At the first step the cylindrical projection of the model vertices is computed with respect to an axis parallel to the coordinated axis Y aligned vertically in the center of the head model and the displacement domains associated to each FAP are determined.
At the second step, the corresponding weighting function, defined in the semantic file, is applied. The weighting function, defined on the [XY] projection plane, indicates the weight associated to each vertex.
The weighting functions have null value at least along two boundary sides of the displacement domain. In this way, a unitary weight will be always assigned to the *feature point* while weights will be assigned to the remaining vertices with values that decrease as closer they are to the boundary of the displacement domain. The rate to which weights decrease, therefore, depends on the extension of the displacement domain, thus allowing flexible modulation of the movement effects.
With the term “animation rule” we refer to the full set of rules that control the movement of a set of vertices, as a function of the value of each FAP. The software developed for animating the facial mesh exploits the semantic information for generating automatically these animation rules. All the possible displacements have been grouped into 15 classes of predefined movements and then associated individually to each FAP. The number of predefined movements is lower than the number of FAPs since the same action can be associated to different FAPs.

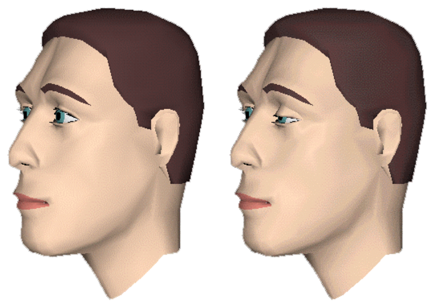
In the following pictures, the effects produced by the predefined movements on the model “Oscar” are presented. For each movement, the picture on the left shows the model in neutral position while the picture on the right shows the effects produced by the animation. The orientation of the head has been chosen appropriately each time to give more evidence to the specific movement applied.



(A)



(B)



(C)

Figure 2 (A) Example of movement applied to FAP 37 “squeeze_l_eyebrow”, squeezing of the right eyebrow, simultaneously with FAP 6 “stretch_l_cornerlip” and FAP 53 “stretch_l_cornerlip_o”, widening of internal and external left mouth corner. (B) Example of movement applied with FAP 31 “raise_l_i_eyebrow”, FAP 33 “raise_l_m_eyebrow” and FAP 35 ”raise_l_o_eyebrow”, raising of the internal, central and external part of the left eyebrow. (C) Example of movement applied to FAPs 5, 10, 11, 52, 57 and 58, opening of the internal and external lower left lip.

3. CONCLUSIONS

The algorithmic solutions proposed in this paper provide a concrete suggestion for the implementation of MPEG-4 facial animation decoders. The stress has been posed in Section 2 on the animation issue, proposing the latest results achieved by the authors' research, since it represents the complementary key tool for integrating the previously developed calibration methodology described in details in [9].

The work described in this paper has been carried out within the European projects VIDAS and INTERFACE through constant critical participation to the MPEG-4 standardization process and relation to the European Research programs ACTS and IST. Live demos of the achieved results have been given recently at IBC-2000 in Amsterdam and at IST-2000 in Nice.

4. REFERENCES

[1] F.I. Parke, "Parameterized models for facial animation", IEEE Computer Graphics Appl. Mag., Vol. 2, n.9, pp. 61-68, 1982.

[2] N.M. Thalmann, D. Thalmann, "Complex models for animating synthetic actors" IEEE Xomputer Graphics Appl. Mag., Vol. 11, n. 5, pp.32-44, 1991.

[3] F.I. Parke, K.Waters, "Computer Facial Animation", Wellesley, MA, Peters, 1996.

[4] K. Waters, "A muscle model for animating three-dimensional facial expression", Computer Graphics, Vol. 22, pp. 17-24, 1987.

[5] K. Waters, D. Terzopoulos, "Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models", IEEE PAMI, Vol. 15, n.6, pp. 569-575, 1993.

[6] P. Ekman e W.E. Friesen, "Facial Action Coding System", Consulting Pychologist Press, Palo Alto (California), 1977.

[7] M.M.Cohen, D.W.Massaro, "Modeling coarticulation in synthetic visual speech.", in N.M.Thalmann & D.Thalmann (Eds.) Modes and Techniques in Computer Animation. Tokyo: Springer-Verlag, pp.139-156, 1993.

[8] F.Lavagetto, "Time-Delay Neural Networks for Estimating Lip Movements from Speech Analysis: A Useful Tool in Audio/Video Synchronization", IEEE Trans. on Circuits and Systems for Video Technology, Vol. 7, n. 5, pp. 786-800 , 1997.

[9] F.Lavagetto, R. Pockaj, "The Face Animation Engine: towards a high-level interface for the design of MPEG-4 compliant animated faces", IEEE Trans. on Circuits and Systems for Video Technology, Vol. 9, n.2, pp.277-289.