

MPEG-4 Compliant Calibration of 3D Head Models

Fabio Lavagetto^{†1}, Roberto Pockaj², Maurizio Costa³

DIST – University of Genova Via Opera Pia 13
16145, Genova, Italy 16154, Genova, Italy

¹ fabio@dist.unige.it, ² pok@dist.unige.it, ³ maury@dist.unige.it

Abstract

The technique proposed in this paper provides operative indications for implementing 3D head models compliant to face parameters recently standardized in MPEG-4 and capable of being calibrated on any specific human face. Since the set of calibration guide points, defined in MPEG-4 as feature points, is very limited with reference to the complexity of the human head geometry, suitable interpolation must be performed to assure smooth and realistic surface rendering. On the other hand, care must be paid in texture adaptation to avoid annoying artifacts in correspondence to deformable face features, which are even more appreciable when the model is animated. The proposed methodology is based on the use of Radial Basis Functions (RBF) for smooth surface interpolation, applied according to a multi-step procedure for progressive detail definition. The set of interpolating functions has been derived empirically exploiting a priori knowledge on the geometry of human heads. The adaptation of the texture information, achieved using the MPEG-4 feature points as anchor points, is also performed according to a multi-step fashion by organizing the head model polygons into homogeneous subsets (i.e. polygons of the hair region define a specific subset). Experimental results, as reported in the following, prove the effectiveness of the proposed method showing how it is possible to adapt the same generic model to different face parameters. A variety of applications is envisaged in multimedia products, virtual and augmented reality, very low bitrate videophone and efficient 3D facial data archiving.

1. Introduction

The achievement of realistic and reliable techniques for the synthetic representation of human faces has been representing an ambitious goal in computer graphics research for more than 25 years. Up to 1993, when MPEG-4 activities were started, a quantity of papers has been published proposing a large choice of proprietary solutions without any serious concern about the need of reaching convergence on some form of reference model or standard approach. It was only with MPEG-4, promoted by ISO-ITU with the main mandate of defining an efficient system for encoding natural/synthetic Audio-Visual Objects (AVO) [1], that a concrete standardization process was activated in synthetic face representation and animation. The specific "ad hoc" groups responsible of carrying on

this task in MPEG-4 were SNHC [2] (Synthetic Natural Hybrid Coding) initially and FBA (Face and Body Animation) later on. Face parameter standardization has interested only aspects affecting the syntax of the MPEG-4 bitstream and, indirectly, the MPEG-4 decoder, while no indications have been given about the implementation of the head model responsible of employing the standard parameters for reproducing accordingly the encoded face expressions.

In the following sections an overview is given on 3D head model definition, head model calibration and head texture mapping, respectively. Experimental results and final conclusions are then reported to provide concrete evidence of the achieved results and of possible exploitation in products and applications.

2. The head model

Feature points, which define relevant somatic points on the face, represent a key-concept. Feature points are subdivided in groups, mainly depending on the particular region of the face they belong to. Each of them is labeled with a number identifying the particular group it belongs to, and with a progressive index identifying them within the group.

Every proprietary model available at any decoder must be in a neutral position and all the FDP used for calibration are referred to a neutral face corresponding to the posture represented in Figure 1.

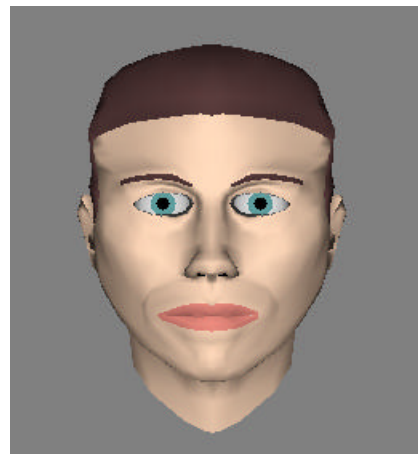


Figure 1: Example of Neutral Face, obtained on the "Oscar" model

MPEG-4 has made some basic assumptions forcing the coordinate system being right-handed, head axes being parallel to the world axes; the gaze direction being aligned to the Z axis, all face muscle being relaxed, eyelids being tangent to the iris, the pupil diameter being 1/3 of the iris diameter, the line of the lips being horizontal and at the same height of lip corners, the mouth being closed and the upper teeth touch the lower ones and the tongue being flat, horizontal with the tip of the tongue touching the boundary between upper and lower teeth.

2. Model calibration

The adaptation of the texture information, achieved using the MPEG-4 feature points as anchor points, is also performed according to a multi-step fashion by organizing the head model polygons into homogeneous subsets.

MPEG-4 defines a set of parameters for the calibration of a synthetic face, called FDP (Facial Definition Parameters), responsible of defining the appearing of the face [3]. These parameters can be used either to modify (though with various levels of fidelity) the geometry and texture of a face model already available at the decoder, or to encode the information necessary to transmit a complete model together with the criteria which must be applied to animate it.

The position of the various features on the face is shown in Figure 2. Considering that the number of feature points is limited to 84, it is evident how critical is the use that the model will make of this information to adapt its shape to that of a specific person. The MPEG-4 specification requires the feature points of the proprietary model to be put in correspondence to the feature points encoded in the FDP, while no constraint is forced on the remaining vertices. Therefore, the quality of the decoder can be measured in terms of its capability to move all the wire-frame vertices without introducing too coarse distortions on the synthetic face.

The algorithm employed in our approach is based on the use of Radial Basis Functions (RBF). The calibration of the model is achieved through a few iterations that reshape the 3D mesh geometry in dependence on the decoded calibration FDP, preserving in the meantime the smoothness and somatic characteristics of the model surface. These calibration points are very few (around 80) with respect to the global number of vertices on the wire-frame which, depending on its complexity, can be as numerous as 1000 or more. Model calibration is solved as a problem of "scattered data interpolation".

The proposed method works correctly also in case of subsets of feature points and it is therefore compliant with the specifications of the Calibration Facial Animation Object Profile [4] which encounter for the possibility of transmitting only a subset of the calibration feature points. To satisfy our specific needs, we have chosen monotone

RBFs decreasing to zero like Gaussian and inverse multi-quadrics RBFs. A significant disadvantage with such functions is due to the global effect produced by the interpolation, which makes it difficult to produce local details. However, this effect becomes negligible imposing that each feature point (on which RBF are centered) influences only a limited region of wire-frame.

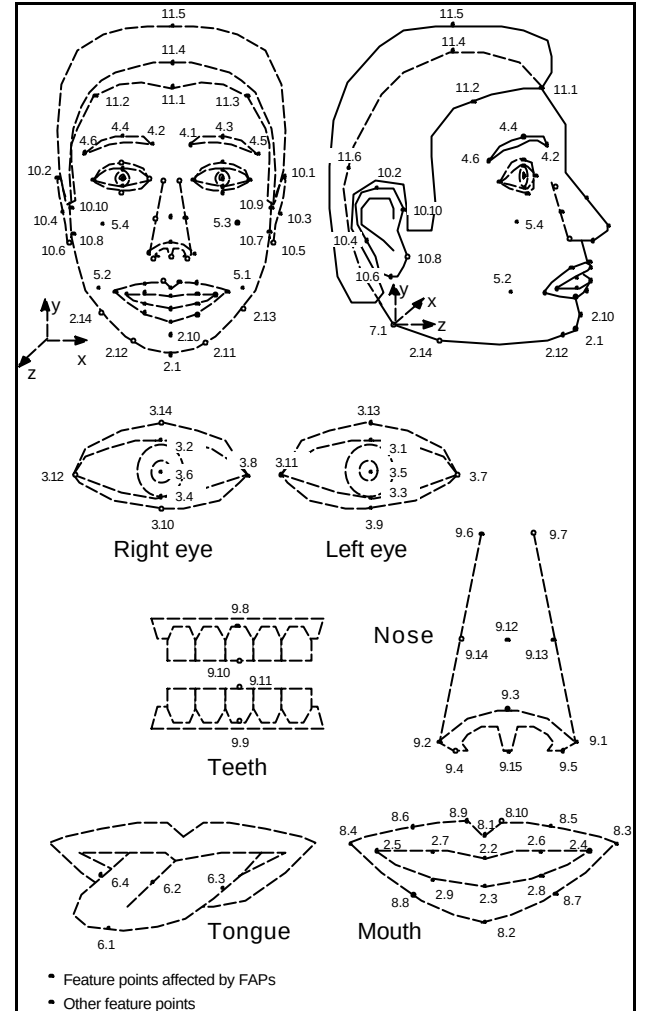


Figure 2: Description of the feature points standardized MPEG-4.

A particular kind of RBF called Compactly Supported (RBFCS) has been employed [5-7]. This type of RBF is widely used in a variety of applications because of the significant computational advantages they offer and the good properties they exhibit like the locality of their domain, characteristic that in our case is of big importance to bound the reshaping action applied on the model.

The non-uniformity of feature point distribution makes it complicated to build an interpolation function capable to satisfy both the requirements of precision and smoothness. The basic idea originated by these requirements, is that of subdividing the ensemble of feature points into a number of subsets such that, in each of them, data are distributed in an as uniform as possible fashion, at least for the smallest

ones. The calibration process consists of in t levels of interpolation. In order to preserve the global surface smoothness, low-order RBF are employed at the first iterations, while high-order RBF are applied in the last iterations to add local somatic details.

3. Experimental results

Experimental results prove the effectiveness of the proposed method showing how it is possible to adapt the same generic model to different face parameters. In the following section we report some results obtained using the data available in the TDS of the FBA “ad hoc” group of MPEG-4 [8]. The proposed algorithm has been employed to animate two different models. The former, “Mike”, is simpler and has been completely developed at DIST, the latter, “Oscar”, is more complex and has been derived from the Geoface model implemented by Keith Waters [9] by adding missing feature points like tongue, palate, teeth, eyes, hair and back of the head [10].

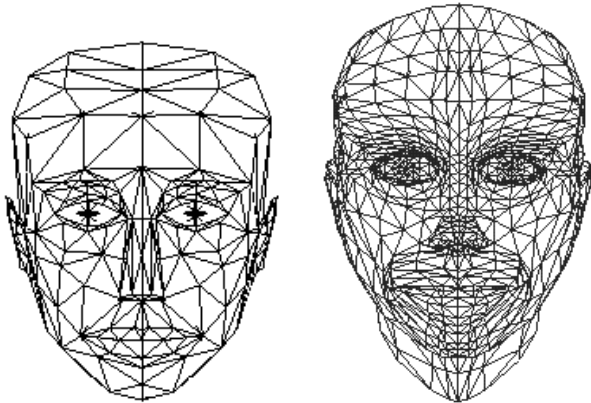


Figure 3: Wireframe of “Mike” (left) and “Oscar” (right).

As shown in Figure 3, “Mike” is composed of 682 polygons including external and internal face structures, while “Oscar” employs 878 polygons for modeling only the frontal face surface and 2.346 polygons for the whole face. The geometric resolution of the model, that is the number of polygons, greatly influences the level of quality achievable through the calibration procedure.

Besides calibrating the model geometry, also the texture information and the texture coordinates for each feature point can be mapped on the model surface. Two different texture data have been employed for running experimental tests, acquired through a 3D scanner or extracted from a 2D frontal picture of the face.

The first problem to be solved for adding the texture information to the calibrated model, is that of mapping the complete set of 3D model vertices onto the 2D-texture domain. The MPEG-4 specifications define the texture coordinates only for the feature points while the decoder

must compute the coordinates of the other wire-frame vertices.

To implement this operation a two-step algorithm has been adopted:

- ?? 2D Projection of the calibrated model onto the texture domain
- ?? Mapping of the 2D projection of the feature points on the texture coordinates

The kind of projection that is employed is evidently dependent on the specific kind of texture available.

In the first case, a cylindrical projection is employed from the 3D space on the plane (u, v) by means of the following equations:

$$u = \arctan(x/z)$$

$$v = y$$

In the second case, a planar projection is employed of the kind:

$$u = x$$

$$v = y$$

For what concerns the mapping of the feature points on the texture coordinates, the same algorithm based on RBF already used for the calibration of the 3D model. The 2D projections of the feature point are forced to coincide with the texture coordinates of the same feature points. In this case, instead of using a multilevel reshaping algorithm, data have been processed one shot. Evidently, in this case RBF in $\mathbb{R}^2$ are used.

One of the major problems concerned with fitting texture information on a 3D deformable mesh is that of being able to map correctly each part of texture on each corresponding 3D model component without incorrect overlapping among object which will be affected by different deformation or motion. This problem is caused, as in many other parts of the model, by the coarseness of the feature points as opposed to the high complexity of a real face geometry. To solve this problem, we have classified the polygons of the model into a few subsets, referred to specific face regions, and we have mapped the texture information separately on each of them inserting synthetic texture or duplicating natural texture where necessary. The quality of the achieved results is shown in Figures 4 and 5.

4. Conclusions

A variety of applications is envisaged in multimedia products, virtual and augmented reality, interfaces to multimedia titles, very low bitrate videophone and efficient 3D facial data archiving, natural and virtual studio fusion in movie production and VR telepresence. Particular reference is made to the results recently achieved within the European

project VIDAS, coordinated by DIST, in applications concerned with virtual telepresence and virtual videophone.



Figure 4: (Top) Frontal and side views of the texture calibration target “Claude” [13]; (center) model “Oscar” reshaped with the feature points of “Claude”; (bottom) model “Oscar” reshaped with feature points and with the texture of “Claude”.

References

- [1] R. Koenen, F. Pereira, L. Chiariglione, “MPEG-4: Context and Objectives”, Image Communication Journal, May 1997.
- [2] P.Doenges, T.Capin, F.Lavagetto, J.Ostermann, I.S.Pandzic, E.Petajan, “MPEG-4: Audio/Video & Synthetic Graphics/Audio for Real-Time, Interactive Media Delivery”, Signal Processing: Image Communication, Vol. 9, n.4, 1997, pp. 433-463.
- [3] MPEG Video and SNHC, “Study of CD 14496-2 (Visual)”, ISO/IEC JTC1/SC29/WG11/N1901
- [4] MPEG SNHC, “Initial thoughts on SNHC visual object profiling and levels”, ISO/IEC JTC1/SC29/WG11/N1972
- [5] R. Schaback, “Creating surfaces from scattered data using radial basis functions”, Mathematical methods in CAGD III, M.Daelhen, T.Lyche, and L. Shumacker (eds.), pp. 1-21, 1995.
- [6] O. Saligon, A. Le Mehaute, C. Roux, “Facial Expression simulation with RBF”, International Workshop on SNHC and 3D Imaging, Rhodos (Greece), September 1997.
- [7] F.J. Narcowich, R. Schaback, J.D.Ward, “Multilevel interpolation and approximation”, revised for ACHA
- [8] FBA Test Data Set: “http://www-dsp.com.dist.unige.it/snhc/fba_ce/”.
- [9] F. I. Parke, K. Waters, “Computer Facial Animation”, A K Peters, Wellesley, 1996.
- [10] 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,

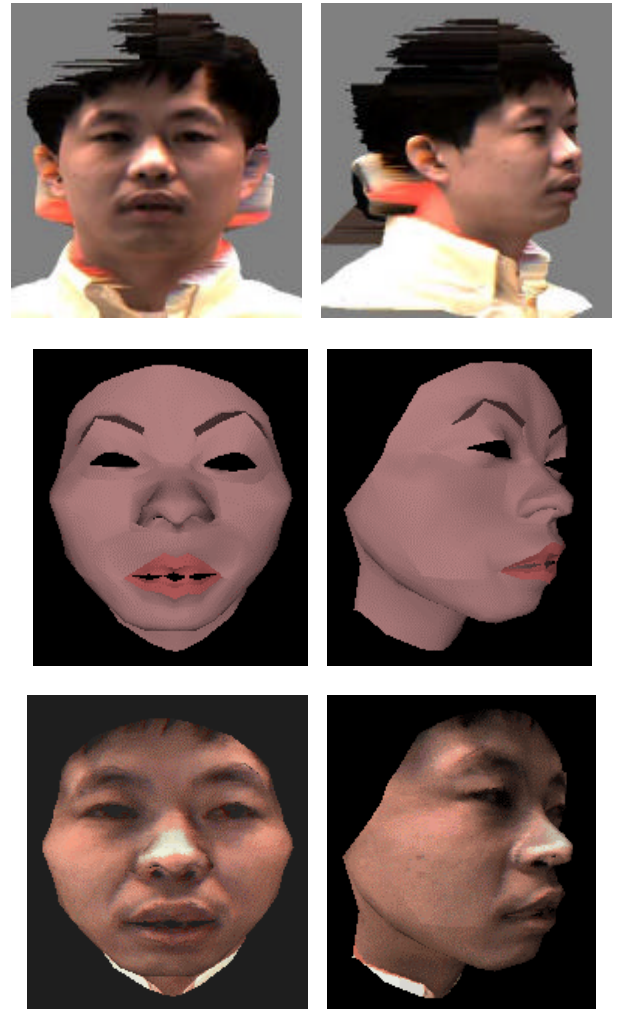


Figure 5: (Top) Frontal and side views of the texture calibration target “Chen” [13]; (center) model “Oscar” reshaped with the feature points of “Chen”; (bottom) model “Oscar” reshaped with feature points and with the texture of “Chen”.