3D TALKING HEAD CUSTOMIZATION BY ADAPTATING A GENERIC MODEL TO ONE UNCALIBRATED PICTURE

Ana C. Andrés del Valle[*], *Jörn Ostermann*[**]

[*]Institut Eurécom
2229, route des Crêtes
Sophia Antipolis, 06904, France

[**]AT&T Labs - Research
200, Laurel Ave. South
Middletown, NJ 07748, USA

ABSTRACT

In this paper we propose a fast and simple interactive method to create a customized 3D talking head by adapting a generic model to a given picture. Our approach uses an interpolation technique, based on Radial Basis Functions with Compact Support, to get the 3D coordinates of the head mesh. We obtain a complete texture-mapped model after selecting only 18 feature points.

1. INTRODUCTION

A great group of applications, most of them developed to be used over the Internet, (teleconferencing, e-commerce, information desktops, etc.) will enrich their human-computer interface by integrating *avatars* or *Talking Heads*. There are several animation techniques to implement *Talking Heads*. One of the common ways is to define a three-dimensional head object, to animate it following some specific rules and to add the person's voice from a natural recorded source or a Text-To-Speech Synthesizer (TTS) [1, 6].

The creation of 3D face models will eventually become a basic need for those Face Animation systems that use 3D polygon meshes. Currently, the existing techniques to build personalized face models are costly and take much time [7]. This kind of expense is worth for applications where high quality is the major constraint. This procedure is not helpful when many models are needed, the individual cannot participate in the creation of the model or we are looking for a fast solution. Since no model creation that starts from scratch is inexpensive, the adaptation of a previously created model becomes necessary. Several approaches to design this model adaptation can be taken. Some are based on anatomy and physics, whereas some others have their basis on mathematical techniques. The first methods are suitable to adapt anatomical head models that simulate facial tissue and organs. They adapt the lattices of the 3D head model to a given picture by rearranging the polygon mesh according to the physical behavior [1]. The second method is suited for models made of vertices that simply define surfaces and whose animation moves these nodes according to predefined facial animation parameters (FAPs) [5]. These adaptation methods take the general model as a collection of points and they apply an interpolation process to adapt it to the given picture(s). The nature of the wireframe models makes scattered data surface interpolation the method commonly used. This interpolation uses specific functions during the process. Radial Basis Functions are the most used from all the families that are available [2, 3].

This paper describes a quick, inexpensive and feasible procedure to obtain new 3D head models from a predefined prototype. It explains the complete process to adapt a generic head model to a given picture applying an interpolation method based on Radial Basis Functions with Compact Support (RBFCS). We show how from just a head model with no more than 1000 vertices and a non-calibrated photograph of the front view of a face we obtain customized head models ready to be used in Internet applications. Our method differs from another similar approach [2] in that all the vertices of the predefined head model are equally treated and that we do not need any iterative process during the adaptation. Sections 2 and 3 contain the theoretical basis of the adaptation process, our *Camera Model* and the mathematical implications of using *RBFCS* are explained. In section 4 we describe the process followed by our adaptation system. Section 5 and Section 6 show the main results and conclusions derived from some tests with the implemented software that performs this adaptation.

2. CAMERA MODEL

To obtain the projection (x', y') of an object onto the camera plane we need the *distance* from the object to the camera (f + z), the *focal length* (f) and the object dimensions (x, y, z), see Fig. 1.

Applying the parameters of a real camera to an adaptation procedure in not feasible; the conditions in which the photograph was taken are usually unknown. To overcome this problem we predefine a camera model. The definition of our camera model and its focal length is based on the following hypothesis:

- All front pictures are taken from more or less the same distance;
- All human heads have similar size.

The first hypothesis allows us to fix the distance between camera and head to a specific value that we always use regardless of the nature of the picture. The second one permits to establish the dimensions of our 3D model. We obtain the focal length by relating the 3D coordinates of two vertices (P1, P2) with their 2D coordinates on the image through the expression $f = (\Delta y' \cdot z)/\Delta y$. The two points in the image are manually located during the system calibration, see Section 4. Once f is

0-7803-6685-9/01/$10.00©2001 IEEE

determined we can apply the projection formulae shown in Fig. 1.

Our face model adaptation uses only one picture therefore only two of the three dimensions can be recovered from the projected points. In the process we use the z-coordinate of the generic mesh we are adapting. The z-coordinate of our new model will not be adapted and the x and y coordinates will just be approximations. For our practical purposes, having the exact information is not crucial.

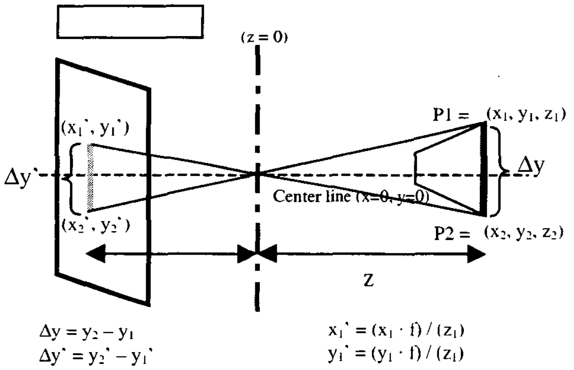


Figure 1 Relationships between 3D coordinates (x, y, z) and 2D projections (x'; y')

3. INTERPOLATING USING RADIAL BASIS FUNCTIONS WITH COMPACT SUPPORT

With the projection relationships we can obtain the 3D location of feature points marked on the image. A different problem is to relocate the other points of the mesh. To solve this issue, we use interpolation of surfaces from scattered data using Radial Basis Functions [3]. This method has provided good results to a similar problem [2]. We built the interpolating function:

$$s(x) = \sum_{i=1}^{N} b_i \Phi(x - x_i) + \sum_{l=1}^{4} c_l p_l(x) \qquad (1)$$

from the following system:

$$\begin{bmatrix} A & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} B \\ C \end{bmatrix} = \begin{bmatrix} Y \\ 0 \end{bmatrix} \quad A = \begin{bmatrix} \Phi(x_1 - x_1) & \Phi(x_1 - x_2) & \cdots & \Phi(x_1 - x_N) \\ \Phi(x_2 - x_1) & \Phi(x_2 - x_2) & \cdots & \Phi(x_2 - x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi(x_N - x_1) & \Phi(x_N - x_2) & \cdots & \Phi(x_N - x_N) \end{bmatrix} \qquad (2)$$

$$P = \begin{bmatrix} 1 & x_{x1} & y_{x1} & z_{x1} \\ 1 & x_{x2} & y_{x2} & z_{x2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{xN} & y_{xN} & z_{xN} \end{bmatrix} \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} \& C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_4 \end{bmatrix} \quad \& Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad (3)$$

N is the number of vertices x_i we use to build the function; we need to know their old and their new location. x is any of the head mesh vertices whose adapted position is unknown and *s(x)* is its interpolated location. 0 represents a matrix or a vector with null values, *b* and *c* are the coefficients of the interpolation, x_{xi}, y_{xi}, z_{xi} are the coordinates of vertex x_i, and y_i is the new position of x_i. $\Phi(x_i - x_j)$ represents a Radial Basis Function centered in x_i.

$p(x)$ is a polynomial of 4th degree ($1{+}x{+}y{+}z$) added to the general RBF interpolation to better adjust global translations. This addition forces the constraint $\sum_{j=1}^{N} b_j p_j(x_i) = 0$ to make the complete system solvable. x_i-x_j is the distance between two vertices of the predefined head model computed with the Euclidean norm in R^3. Calculating the distances among vertices for the entire head model is a compute-intensive task; it is only performed once and the distance values are stored in a file.

From all RBFs we are interested in those which have the property of being monotone decreasing to zero for increasing values of the radius. This kind of functions provides us with the mathematical way to apply the knowledge that each of the feature points (on which the RBFs are centered) influences a limited region of the wireframe. It is physically understandable that feature points located in the chin should not interact with points situated on the upper part of the head. For our process we have been testing and applying a family of RBFCS [4]:

$$\phi_{0,0} = (1 - r)_+ \in C^0 \cap \mathrm{PD}_1$$
$$\phi_{1,0} = (1 - r)^3_+(1 + 3r + r^2) \in C^2 \cap \mathrm{PD}_1$$
$$\phi_{1,1} = \mathrm{D}\phi_{1,0} = (1 - r)^2_+(2 + r) \in C^0 \cap \mathrm{PD}_3$$
$$\phi_{2,0} = (1 - r)^5_+(1 + 5r + 9r^2 + 5r^3 + r^4) \in C^4 \cap \mathrm{PD}_1$$
$$\phi_{2,1} = \mathrm{D}\phi_{2,0} = (1 - r)^4_+(4 + 16r + 12r^2 + 3r^3) \in C^2 \cap \mathrm{PD}_3$$
$$\phi_{2,2} = \mathrm{D}^2\phi_{2,0} = (1 - r)^3_+(8 + 9r + 3r^2) \in C^0 \cap \mathrm{PD}_5$$
$$\phi_{3,0} = (1 - r)^7_+(5 + 35r + 101r^2 + 147r^3 + 101r^4 + 35r^5 + 5r^6) \in C^6 \cap \mathrm{PD}_1$$
$$\phi_{3,1} = \mathrm{D}\phi_{3,0} = (1 - r)^6_+(6 + 36r + 82r^2 + 72r^3 + 30r^4 + 5r^5) \in C^4 \cap \mathrm{PD}_3$$
$$\phi_{3,2} = \mathrm{D}^2\phi_{3,0} = (1 - r)^5_+(8 + 40r + 48r^2 + 25r^3 + 5r^4) \in C^2 \cap \mathrm{PD}_5$$
$$\phi_{3,3} = \mathrm{D}^3\phi_{3,0} = (1 - r)^4_+(16 + 29r + 20r^2 + 5r^3) \in C^0 \cap \mathrm{PD}\text{-}$$

r = distance between 2 points; C^x = x-times continue and x-l-times derivable; PD_y = y-times partially derivable

4. PROCEDURE OVERVIEW

Our system uses a predefined 3D head model and a front picture of a face. The conditions in which the picture was taken, focal length of the camera distance to the individual, etc. are unknown.

The adaptation process follows these four main steps (Fig. 3):

i. ***Calibration:*** During this step the camera parameters are established and we ensure that the vertices will be projected on the face in the image.

We define our own *camera model* (see Section 2) with information taken from the photograph. We automatically select two vertices, one on each ear of the head model, and the user marks the location of those vertices on the picture.

The adaptation has been designed to adjust little changes made to the initial model; therefore the process does not work for great changes of size or large translations. Since the projections of the head model vertices and the face of the picture will neither be located in the same position nor have the same size, we first align the center of the models; the user clicks on the tip of the nose on the picture. To find the correct size for the projections we automatically project the outer eye corners and the user shifts them on the picture. We compute the scale factor between the head model and the face as the ratio between the former distance of the two projections on the image and the distance of the manually located projections.

ii. ***Rearrangement of feature points:*** Several specific vertices of the 3D model are selected (Fig. 2) and projected onto the photograph. They are manually shifted to fit their corresponding position in the face in the picture. The adaptation requires around 18 vertices to obtain a properly reshaped and good quality final model (Fig. 4).

iii. ***Interpolation:*** When the 18 points are properly positioned, we use their (x, y) coordinates, the original 3D coordinates of the head model and the distances between the selected vertices to build the interpolation function that we need to get the adapted position of the rest of the polygon mesh, see Section 3.

iv. ***Texture addition:*** The last step is applying the image from the photograph as the texture for the head model. We assign the 2D coordinates from the projection of all the adapted vertices onto the photograph.

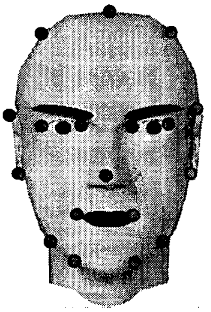


Figure 2 The selected 18 vertices on the generic 3D head.

For an on-line demonstration of the procedure see [8].

5. EXPERIMENTAL RESULTS

Several tests were carried out to find suitable parameters for the process, to evaluate the quality of the obtained 3D head models, and their behavior during animation.

The result of the adaptation depends on the area of influence of the RBFCS centered on the feature points. The area of influence used during the interpolation has to be such that the generated adapted model has the shape of the new face and arrange the parts of the face (eyes, mouth, ears, etc.) in the proper positions. We consider areas ranging from 0.1 to 0.25 (normalized to the largest distance between two vertices of the 3D model), around 2.2 cm and 5.5 cm on a typical real face. The criteria to evaluate the RBFCS are highly subjective. Our goal was to find the function that gives us an adapted model that *looks fairly good.* Generally, a 3D head model looks good when its surface is smooth and the parts that belong to it (eyes, lips, etc.) are well shaped.

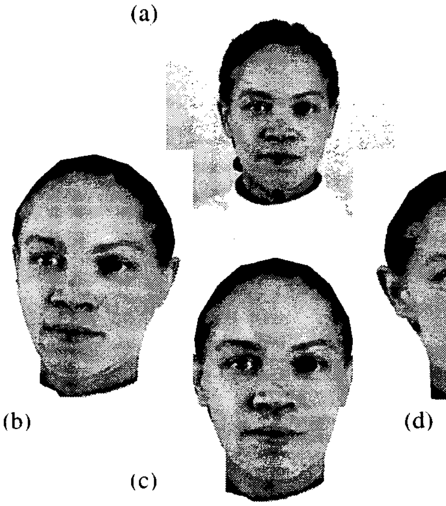


Figure 4 Customized model (b, c, d) obtained from the adaptation of photograph (a)

We have seen two major general behaviors. As the degree of the RBFCS increases the smoothness of the adaptation improves and the natural shape of the parts gets deformed. Both behaviors are due to the fact that that high order RBFCS have more derivatives and therefore they are smoother around their points of discontinuity than low order RBFCS.

Since we want to have good results in smoothness and shape we will choose an RBFCS that responds well enough to both requirements. That could be either $\phi_{2,1}$ or $\phi_{2,2}$

After performing tests to animate our adapted models we have extracted the following remarks:

- The rough adaptation of the area around the lips became too obvious during the animation of some adapted models. To adapt the lips we are using only two feature points, the lip corners, therefore we assume that the lips of the person are straight as in our 3D model. This assumption is not correct when the person has not straight lips. To have a good adaptation in such a case some more processing has to be done after the general adaptation. A possible solution, not

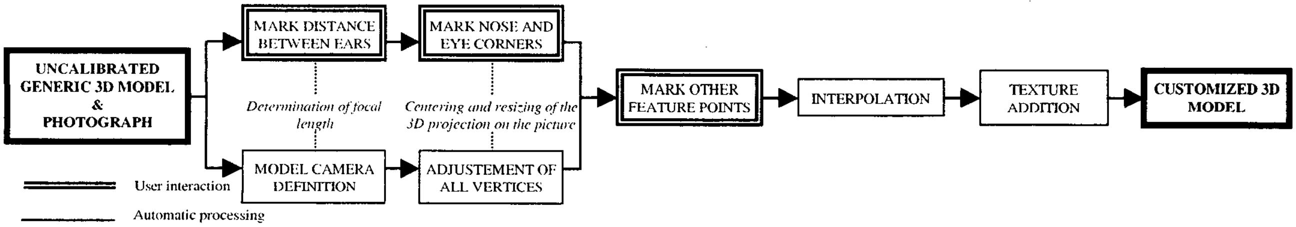
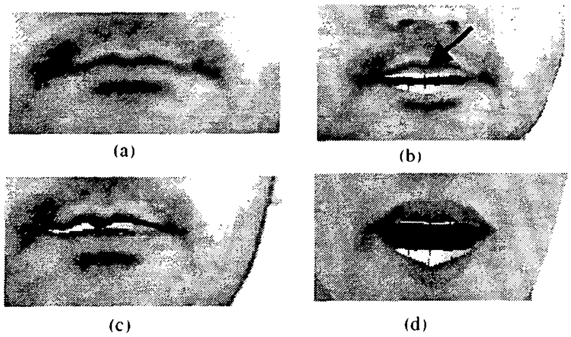

UNCALIBRATED GENERIC 3D MODEL & PHOTOGRAPH
MARK DISTANCE BETWEEN EARS
MARK NOSE AND EYE CORNERS
Determination of focal length
Centering and resizing of the 3D projection on the picture
MODEL CAMERA DEFINITION
ADJUSTEMENT OF ALL VERTICES
MARK OTHER FEATURE POINTS
INTERPOLATION
TEXTURE ADDITION
CUSTOMIZED 3D MODEL
User interaction
Automatic processing

Figure 3 Block diagram of the complete adaptation process



(a) (b) (c) (d)

Figure 5 (a) Original mouth adapted in b and c. (b) Adaptation of crocked lips when selecting only the corners of the mouth. We obtain incorrect texture on the lips. (c) Adaptation of crocked lips when selecting enough points to reshape them. We see undesirable waves on the lips while animating. (d) Adaptation from straight lips. They are correctly mapped and nicely shaped

yet implemented, could be adapting only using the corners so we get the proper mouth shape for the animation (Fig. 5 (d)) but the texture for the lips should be provided giving the points that define the natural lines of the lips on the picture so we do not map part of one lip onto the other (Fig. 5 (b))

- The heads of the adapted 3D models can only be rotated 10 degrees to the left and right and 5 degrees up and down because the texture is taken only from one picture and we give the adapted model the z-coordinate from the prototype. If more freedom is required, we have to use at least one more picture, from a different view.
- The proposed procedure is based on the fact that all human heads are alike. This concept cannot be applied to the hair. Using the procedure but selecting the hairpiece depending on the individual could be a solution to this limitation.

6. CONCLUSIONS

This paper proposes a fast and simple method to create 3D head models by adapting a generic head mesh to a given front picture of a face. Thanks to an interpolation technique based on RBFCS and a precise calibration, very little user interaction is needed. The obtained models behave correctly while being animated and have enough quality to be used in applications developed for the Internet (e-commerce, information desktops, etc.)

7. REFERENCES

[1] D. Terzopoulos and K. Waters, "Physically based facial modeling analysis, and animation", *J. Visualization and animation* 1(2), 1990, pp. 73-80.

[2] F. Lavagetto and R. Pockaj, "The facial animation engine: toward a high-level interface for design of MPEG-4 compliant animated faces", *IEEE Transactions on circuits and systems for video technology*, Vol. 9, No. 2, march 1999, pp. 277-289.

[3] Schaback, R., "Creating surfaces from scattered data using Radial Basis Functions", *Mathematical Methods for curves and Surfaces*, M. Daehlen, T. Lyche, and L. L. Schumaker (eds.), Vanderbilt University Press, 1995, pp. 477-496.

[4] Wu, Z., "Multivariate compactly supported positive definite radial functions", preprint, 1994.

[5] ISO/IEC JTC/SC 29/WG 11 N 2502, "Coding of moving pictures and audio", 14496-2 Visual, Atlantic City meeting, October 1998.

[6] Jörn Ostermann, "Animated Heads" http://www.research.com/~osterman/AnimatedHead

[7] Ostermann, J., Chen, L. And Huang, T., "Adaptation of a generic 3D human face to 3D range date" *J. VLSI Signal Processing Systems for Speech, Image and Video Technology*, vol. 20, 1998, pp. 999-107.

[8] AT&T Labs-Research. "VideoTalks - Playmail": http://www.videotalks.com/playmail