Automated Modelling of Real Human Faces for 3D Animation

B. Nagel, J. Wingbermühle, S. Weik, C.-E. Liedtke
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover - Germany
email: nagel/wingber/weik/liedtke@tnt.uni-hannover.de

Abstract

This paper describes the automated creation of a textured face mask of a natural person suitable for realtime animation. A 3D surface polygon face mask is used. A set of muscles is defined to animate the mask. The mask is adapted to a real person using 3D surface data from a calibrated stereo sensor. Firstly, the facial features of the mask are aligned to the 3D data by scaling the mask. Secondly, a local shape adaptation moves the mask vertices along their surface normals to be coplanar with the measured 3D surface. The adapted mask is textured from one of the stereo images yielding a highly realistic impression. The predefined muscles are used to animate the textured mask.

Index terms - Face Modeling, Shape from Stereo, 3D Adaptation, Facial Animation

1. Introduction

The modelling of realistic looking human faces and facial expressions has been a challenge to computer scientists for many years. Today, the hardware processing power permits real-time modelling of human faces even on a home PC. Research has fostered applications in the area of virtual actors, for the entertainment industry, the planning of facial surgery and telecommunications.

The approaches to facial modelling described in literature differ in many respects, depending on the application, type of sensors, computing efficiency, real time constraints, cost and required accuracy of the model. Active and passive sensors are used such as laser scanners, line projectors and cameras. The modelling strategies range from flat 2D templates over 3D surface models to models which involve detailed simulations of physical properties of the underlying bone structures of the head and the layers of tissue and skin[1][2].

The application area we are aiming at is highly realistic 3D telecommunication [3][4][5] which suggests the virtual presence of the communication partner [6]. Telecommunication is online and involves the real-time transmission of audible words and visual facial and body expressions. The modelling can be restricted to a compute-efficient parametric surface modelling, where only the animation parameters are transmitted to the receiver. Modelling consists of two major stages: a computationally expensive initialization phase, which results in an initial model and an online update phase during the communication process. The following description is restricted to the creation of an initial model.

In our approach a surface model is used. The shape is modelled by a mesh of triangles and the fine structure is modelled by photo texture obtained from a camera which is mapped onto the mesh. The initial surface mesh is a generic facial mask, which resembles an average human face. A muscle based animation scheme has been developed for the generic mask. To enable the animation of the real person's model the mask and its relevant facial features have to be adapted to those of the real head.

In the following, the acquisition of 3D surface data, the generic model and the model adaptation are described in more detail.

2. Shape Reconstruction from Stereo

Figure 1. Left and right input Image of stereo camera and calculated 3D surface points

The 3D shape data of the real person is currently generated using a shape from stereo approach. Two Kodak

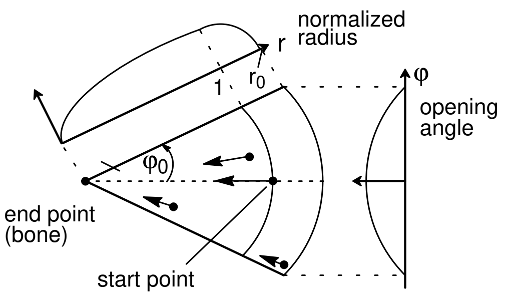

Figure 2. Muscle model - a muscle pulls its start point and also moves surrounding vertices

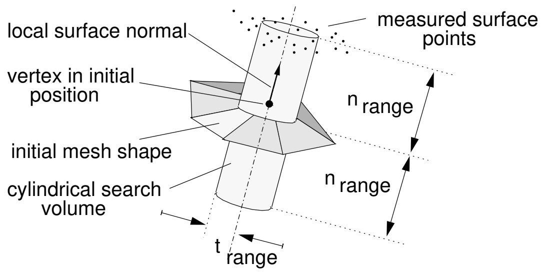

Figure 4. Cylindrical search volume for measured surface points

DCS410 cameras each with a resolution of 1524x1012 pixels build the stereoscopic sensor. The relative position of the two cameras to each other is determined in advance by taking a picture of a calibration pattern and using standard calibration techniques. The viewing directions of the left and right camera are approximately 15 degree rotated to each other.

Figure 1 shows the input images for the stereoscopic analysis. From these two images a so called depth map is calculated describing the distance of the surface points to the camera [7]. This depth map can be transformed into a set of 3D points representing the surface of the person (figure 1). A monochromatic background was used to distinguish the 3D surface of the face from the background.

3. Animated 3D Face Mask

A 3D face mask resembling an average human face and consisting of a polygon mesh is used to model the human face. Muscles are defined in the mask using anatomical data. Definition parameters of the muscle model [8] are the vertices where the muscle attaches to bone and tissue and the region of action (figure 2). A muscle pulls its start point towards its bone attachment. The surrounding elastic tissue, i.e. the vertices within the defined region, is affected. The weighting functions are normalized cosines. It is assumed that corresponding muscles in different faces lie in corresponding areas. Hence, this parametric definition allows the application of the predefined muscles to the face mask after adaptation.

4. Mask Adaptation

4.1. Proportioning and Scaling

It is essential for the quality of the animation results that the particular facial features (eyes, mouth, eyebrows and nose) of the mask are brought to coincide with those of the real person. Automatic approaches for feature detection have been described in [9] and [10], here a manual selection has been used. In the first step, the proportioning, the mask is scaled: the distance between the outer corners of the eyes (the eyeline, $\vec{a}$ in figure 3) and the distance between the center of the mouth and the center of the eyeline (the mouthline, $\vec{m}$) are adapted to the 3D data.

Once the mask is proportioned and scaled properly it is 3D positioned within the measured cloud of surface points. For the extracted 2D features *eyes* and *mouth* the respective 3D positions are taken from the depth map and used to determine the mask's position and orientation in 3D (fig. 3).

4.2. Local Shape Adaptation

In the preceding step the face mask was proportioned, scaled and moved to the correct global 3D position using a set of three predefined feature points. The following section describes an approach to fit the position of all other mask vertices to the measured surface data (local adaptation). Since we use a stereo approach, both shape and texture information is available which permits texturing the shape adapted mask as outlined at the end of this section.

The predefined face mask associates vertices to the facial features (nose, lips, iris contours, etc.). This permits the application of specialized shape adaptation methods for particular sets of vertices. Currently three different adaptation methods are used: one for vertices corresponding to features with clearly distinguishable texture that have been previously located in the image plane, one for vertices belonging to the nose and one default method for all remaining vertices. The proposed default shape adaptation procedure can be understood as an inflation process, where each vertex of the mask is shifted along the respective local surface normal (inwards or outwards) until it fits to the measured data. To avoid erroneus shifts, a robust method has been devised: all measured surface points contained in a cylindrical search volume surrounding the vertex normal (figure 4) are evaluated and the vertex is shifted along the normal to the

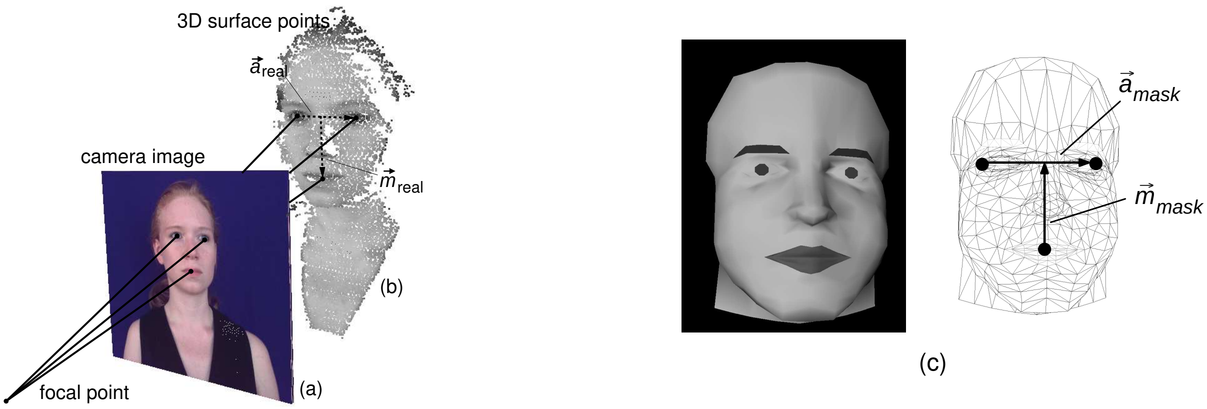

Figure 3. Proportioning of the mask (c) using the eyeline $\vec{a}$ and the mouthline $\vec{m}$ from corresponding features in the 2D image (a), the 3D map of surface points (b) and the mask (c)

determined mean position. If the search volume does not contain any of the measured surface points the amount of shift for a vertex is interpolated from the shifts of its adjacent vertices. Figure 4 depicts the cylindrical search volume around a mask vertex $\vec{v}$. The cylinder axis points in the direction of the local surface normal $\vec{s}_{unit}$, the mean normal of the surrounding triangles. Adjusting the cylinder radius (or tangential search range) t_{range} permits controlling the interpolation area on the surface which implies restricting the possible surface curvature. The cylinder height (or normal search range) n_{range} is chosen equal to the expected maximum distance between the measured surface shape and the inital mask shape. Both parameters can easily be adjusted: a more accurate modelling of the higher curvature on the lips, for example, is achieved by chosing the cylinder radius for adapting a lip vertex smaller than for adapting a cheek vertex. The shift $\vec{s}_{mean}$ applied to a vertex $\vec{v}$ is the mean of the projections δ_i^n of all I measured surface points $\vec{p}_i$ within the search volume on the respective cylinder axis. Additional weighting factors w_i are used to consider the reliablity of each evaluated measured point i (eq. 1).

$$\vec{s}_{mean} = \frac{\sum_{i=1}^{I} w_i \cdot \delta_i^n}{\sum_{i=1}^{I} w_i} \cdot \vec{s}_{unit} \tag{1}$$

The projection δ_i^n of a measured surface point $\vec{p}_i$ on the cylinder axis and its radial distance δ_i^t to the axis can be obtained from eq. 2.

$$\delta_i^n = (\vec{p}_i - \vec{v}_i)^T \cdot \vec{s}_{unit} \qquad \delta_i^t = \sqrt{(\vec{p}_i - \vec{v}_i)^2 - {\delta_i^n}^2} \tag{2}$$

Based on the assumption that the reliability of a measured point $\vec{p}_i$ decreases with the radial distance δ_i^t from the cylinder axis the following weighting factors $w_i^{default}$ are applied if the vertex is adapted using the default method (eq. 3).

$$w_i^{default} = \left(\frac{t_{range} - \delta_i^t}{t_{range}}\right)^2 \tag{3}$$

If the vertex to be adapted belongs to the nose different weights w_i^{nose} are applied. These weights do not only depend on the radial distance δ_i^t but also on the distance δ_i^n in direction of the axis (eq. 4). This relation is introduced since measurement outliers in this area are more likely than in other parts of the face when analyzing frontal stereo views. In addition very narrow cylinders are used because the curvature is significantly higher on the nose than on the face.

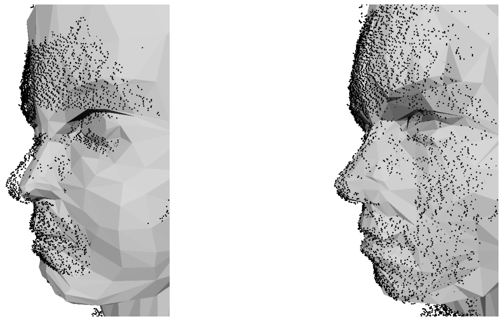

Figure 5. Face mask and measured surface points before and after local shape adaptation

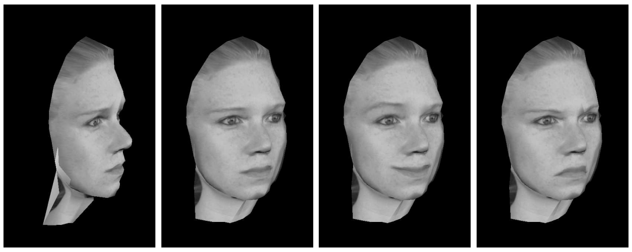

Figure 6. Animated mask with several synthesized facial expressions

$$w_i^{nose} = 0.5 \cdot \left(\left(\frac{t_{range} - \delta_i^t}{t_{range}} \right)^2 + \left(\frac{n_{range} - \delta_i^n}{n_{range}} \right) \right) \quad (4)$$

A third adaptation method is applied to vertices which lie on the contours of accurately locatable features. For animation, features and texture of the real face need to be aligned with those of the mask, e.g. if the model's eyes are moved to the left and part of the iris texture has been assigned to the eyelid this part of the texture will remain in its original position and the model's iris will appear to be torn into two parts. Knowing the exact 2D position of vertices that lie on the boundary of such a feature we can restrict the adaptation method for these vertices to shifts along the respective line of sight. This is achieved by setting the axis direction of the cylindrical search volume to the direction of the line of sight. An example of the local shape adaptation results is given in figure 5. The left image shows a side view of the mask after the global adaptation step and some of the measured surface points, the right image depicts the mask from the same viewpoint after applying the proposed local shape adaptation method.

Once allligned to the real face the mask can be automatically textured by mapping either one or both camera images on the polygonal mesh[11]. The camera parameters that are necessesary for this automatic mapping are available from the camera calibration step which is part of the stereoscopic depth measuring.

5. Animation Results

For the stereoscopic image pair in figure 1 the presented modelling stages have been performed. Figure 6 depicts some animation results achieved with the presented approach. The left two images are rendered views for the neutral expression of the person, i.e. without any mimic synthesis. The right two images are synthetically animated 3D models derived from the neutral adapted mask using the described muscle algorithms.

6. Acknowledgements

The authors would like to thank the Deutsche Telekom Berkom, without whose funding part of this work would have been impossible.

References

[1] Y. Lee, D. Terzopuolos, K. Waters, "Realistic Modeling for Facial Animation", Proc. SIGGRAPH 1995, pp. 55–61, Aug. 6–11, Los Angeles

[2] N. M. Thalmann, S. Carion, M. Courchesne et al.: "Virtual Clothes, Hair and Skin for Beautiful Top Models", in Proc. Computer Graphics International, June 24-28, 1996, Pohang, Korea, IEEE Computer Society Press, June 1996, pp. 132–141.

[3] L.S. Chen, T.S. Huang and J.Ostermann, "Animated Talking Head with Personalized 3D Head Model," Proc. 1997 IEEE 1st Workshop on Multimedia Signal Processing, Princeton NJ, June 1997, pp. 274–279.

[4] R. Boulic, "Human Avatar and Agent Animation for Virtual Reality", 50p, Computer Science Postgraduate Course in Virtual Reality and Animation, Module M4, EPFL DI-LIG, September 1996.

[5] K. Nagao and A. Takeuchi, "Speech Dialogue With Facial Displays: Multimodal Human-Computer Conversation", Proc. of the 32nd Annual Meeting of the Association for Computational Linguistics, pp. 102-109, 1994

[6] J. Wingbermuehle, S. Weik and A. Kopernik, "Highly Realistic Modeling of Persons for 3D Videoconferencing Systems," Proc. 1997 IEEE 1st Workshop on Multimedia Signal Processing, Princeton NJ, June 1997, pp. 286–291.

[7] L. Falkenhagen, "Block-Based Depth Estimation from Image Triples with Unrestricted Camera Setup", IEEE Workshop on Multimedia Signal Processing, Princeton, New Jersey, USA, June 23-25, 1997.

[8] K. Waters, "A muscle model for animating three-dimensional facial expressions", in Computer Graphics (SIGGRAPH 1987), July 1987

[9] M. Kampmann, "Precise face model adaptation for semantic coding of videophone sequences", Picture Coding Symposium, Sept. 1997, Berlin, Germany.

[10] R.S. Wang, Y. Wang, "Facial Feature Extraction and Tracking in Video Sequences," Proc. 1997 IEEE 1st Workshop on Multimedia Signal Processing, Princeton NJ, June 1997, pp. 233–238.

[11] W. Niem, H. Broszio, "Mapping Texture from Multiple Camera Views onto 3D-Object Models for Computer Animation," in *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Greece, 1995.