Adaptation of a Generic Face Model to a 3D Scan

Axel Weissenfeld, Nikolče Stefanoski, Shen Qiuqiong and Joern Ostermann

Institute of Information Technology
University of Hannover
Appelstr. 9A, 30167 Hannover, Germany
aweissen, stefanos, qshen, ostermann@tnt.uni-hannover.de

Abstract— In this paper we propose a fast and simple adaptation of a generic face model to a 3D scan. The adaptation is divided into global and local adaptation. The global adaptation is based on radial basis functions (RBFs), whereas the local adaptation refines the mesh of a face model to achieve better adaptation results. The 3D scan is several times low-pass filtered before the mesh of a generic face model is iteratively adapted first to the most and then to less filtered scans. Finally the generic model is adapted to the original scan. In this way the face model is precisely and efficiently adapted to the scan.

I. INTRODUCTION

In computer vision 3D generic models are frequently used for tackling many tasks, such as 3D motion estimation of a human face, model-based coding or facial animation (FA). A model-based algorithm for tracking a human face is described in [1]. This algorithm is based on adapting a generic face model to the human subject. Another application for adaptation is facial animation. Research in FA started in the early 70's [2]. From that time on different animation techniques [3] [4] were developed, which continuously improved the animation. Finally, in MPEG-4 a Facial Animation specification was standardized [5], that enables the use of a generic face model. In order to synthesize personalized facial animation, the generic face model has to be adapted to the geometric shape of an individual human subject.

For the described tasks a precise adaptation is necessary to achieve good tracking or animation results. Since faces have a highly complex 3D shape, a precise adaptation is a challenging task. Classically a generic face model is adapted to an image in two steps: First facial feature points, such as eye corners and nostrils are detected. Then the generic model is adapted to these features. These approaches range from single view [6] to multiple-views [3]. In single views depth information is obviously not available, so that the model cannot be precisely adapted. Both methods attempt to automatically detect facial features. However, they lack in robustness, so that the reconstructed 3D shape is often not precise. Even manual assistance will not lead to precise 3D shapes. Laser scanners offer the opportunity to automatically capture precise 3D shapes of objects by using a 3D shape acquisition system.

Acquired 3D-shapes are usually represented by triangle meshes with a large number of vertices. Since these mesh vertices are semantically unrelated, facial animation parameters as defined in MPEG-4 cannot be directly extracted from the scan. Furthermore, 3D scans have a high resolution, so
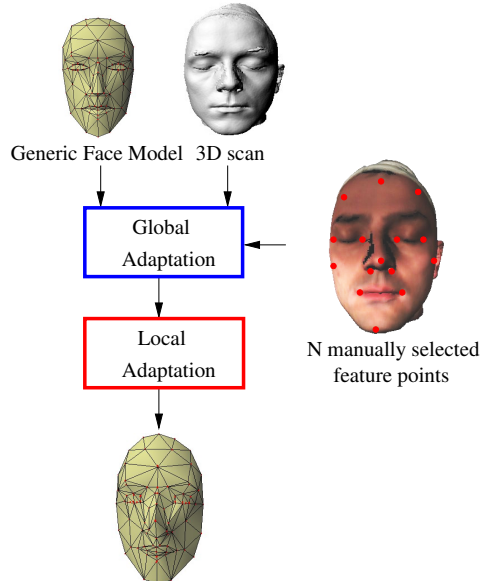

Fig. 1. Pipeline of adapting a generic face model to 3D scan.

that the computational effort is too large for real time applications. Consequently, generic face models are indispensable for tracking and animations.

A precise adaptation of generic models to 3D scans is possible. Earlier approaches [7]–[9] discussed the adaptation of a generic face model to a 3D scan for facial animations. However, these papers did not investigate the adaptation error, whereas we are addressing the problem of adapting generic face models as precise as possible. Our adaptation process is divided into global and local adaptation. The global adaptation roughly adapts the geometric shape of the generic face model to the 3D scan via RBFs. This paper describes a novel approach of locally adapting a generic face model to a 3D scan by low-pass filtering the original scan a few times, before the face model is iteratively adapted. Low-pass filters for 3D objects are simple to implement. They requiring low computational effort and they are very effective for a precise adaptation.

In the following some notations are introduced, which will be used in the remainder of this paper. A mesh $M = (V, E)$ is considered as a tuple of sets of vertices V and edges E. E is the connectivity of mesh M. $p_i \in V$ denotes a vertex of M and $e_{ij} \in E$ denotes an edge connecting p_i and p_j. The set of neighbors $N(i)$ of vertex p_i consists of all vertices p_j which
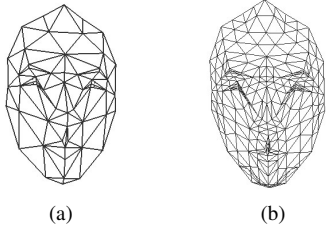
Fig. 2. Mesh of generic face model before (a) and after uniform plain subdivision (b).

have a common edge e_{ij} with p_i. $N(i)$ is also called the 1-neighborhood of vertex p_i. We will denote the face model with $M^m = (V^m, E^m)$ and the 3D scan with $M^s = (V^s, E^s)$. The cardinality of an arbitrary set S will be denoted as $|S|$.

In the remainder of this paper, we give an overview of our system (Section 2). Furthermore are discussed the global (Section 3) and local adaptation (Section 4). The error measure is explained in section 5. In section 6 some adaptation results are presented.

II. SYSTEM OVERVIEW

The overall framework of our face model adaptation system is described in Fig.1. The inputs are a 3D scan acquired by a Cyberware scanner and a generic face model. As face models we selected a simplified Candide-1 model for rigid 3D motion estimation and a Candide-3 model, which is compliant to MPEG-4 Face Animation, for animation. The Candide-1 model consists of 76 vertices and 99 triangles, while the Candide-3 model consists of 113 vertices and 168 triangles. Note that our face model adaptation system is not limited to the two models used.

First the generic face model M^m is globally adapted to the 3D scan M^s by selecting N manually corresponding points in both meshes resulting in mesh $\bar{M}^m$. This adaptation achieves a scaling and position of the face model's facial features to the 3D scan.

Secondly, the globally adapted face model $\bar{M}^m$ is further improved locally by shifting vertices successively within a small domain.

The number of vertices of the scan is much higher than the number of vertices of the face model, i.e. the number of degrees of freedom needed for adaptation is strongly limited. This restricts the overall adaptation to coarse features of the scan, like nose, mouth, eyes, eyebrows. In order to increase the accuracy of the adaptation process to finer details the user has the opportunity to increase the resolution, i.e. to increase the number of vertices to the generic face model. Therefore we use uniform plain subdivision, which divides each mesh triangle into four smaller triangles (Fig. 2).

III. GLOBAL ADAPTATION

The face model is globally adapted to the 3D scan by selecting manually N facial feature points in the 3D scan; e.g. for Candide-1 N=16. This is the only manual interference required by the user. Then the Candide mask is adapted to the 3D scan by interpolation, which is based on radial basis

functions [10]. RBFs are well known in the head animation community for adapting generic head models [3] and even animating facial parts [11]. A function describing the linear transformation of $f(p_i^m) = p_{i,new}^m$, in which the N 3D feature points of the Candide mask are mapped onto the 3D scan. This transformation can be described as follows:

$$f\left(p_i^m\right) = \sum_{j=1}^{N} b_j \Phi\left(p_i^m - p_j^m\right) + \sum_{l=1}^{4} c_l \mathbf{g}_l\left(p_i^m\right) \qquad (1)$$

from which we can determine the control points b_j and c_l.

For geometric applications, as described here, the first term Φ is a radial basis function, which is shift and rotation invariant. In literature a high number of RBFs are proposed [12]. In our work we selected $\phi(r) = (1 - r)_+^3 (8 + 9r + r^2)$ with $\Phi(r) = \phi(\| r \|_2)$ for adapting the Candide mask to a 3D scan. The second term represents a polynomial space with degree 3 representing the x, y, and z coordinate of 3D space coordinates.

Note that after the global adaptation only the N corresponding points of the generic face model are located on the scan. Since the positions of the other vertices are calculated as an interpolation of the corresponding points (1), these vertices may not be located on the 3D scan. Thus, a local adaptation is necessary.

IV. LOCAL ADAPTATION

The process of local adaptation refines the globally adapted model $\bar{M}^m$ to achieve a more precise adaptation of the face model M^m to the 3D scan M^s. This local adaptation is achieved by gradually adapting the face model $\bar{M}^m$ to a series of meshes $M_1^s, \ldots, M_l^s, \ldots, M_L^s = M^s$ generated from the 3D scan. Here the mesh M_1^s captures the low "frequency" part of the scan, while meshes M_l^s with increasing index l possess more and more details leading to the scan $M_L^s = M^s$ itself, the most detailed mesh (Fig. 3). It is assumed that in this series only vertex positions change, while the connectivity E_l^s, which is given by the connectivity of the scan E^s, remains the same for all meshes M_l^s. Local adaptation is performed by iterative application of an operator T, i. e.

$$M_l^m := T(M_{l-1}^m, M_l^s) \text{ for } l = 1, \ldots, L, M_0^m := \bar{M}^m. \quad (2)$$

After a series of gradual face model adaptations the final adapted face model M_L^m is received. In order to obtain an overall accurate adaptation of the face model to the details of the 3D scan gradual adaptations preformed by the operator T in (2) have to be accurate. In our framework we employ the next neighbor operator T_{nn}. $T_{nn}(M_{l-1}^m, M_l^s)$ updates each vertex p_i^m of the face model M_{l-1}^m with the nearest vertex $p_{j_0}^s$ in the 3D scan M_l^s, i.e.

$$p_i^m = p_{j_0}^s \quad \text{with} \quad p_{j_0}^s = \operatorname{argmin}_{p_j^s \in V_k^s} \left\| p_i^m - p_j^s \right\|_2.$$

The series of meshes $(M_l^s)_{l=1\ldots,L}$ induces a hierarchy in the details of the 3D scan. Such a hierarchy can be received by low-pass filtering. A hierarchy is required in order to allow accurate gradual adaptations while repeatedly applying operator T in (2); i.e. the face model is successively adapted to the details of the scan from coarse to fine.
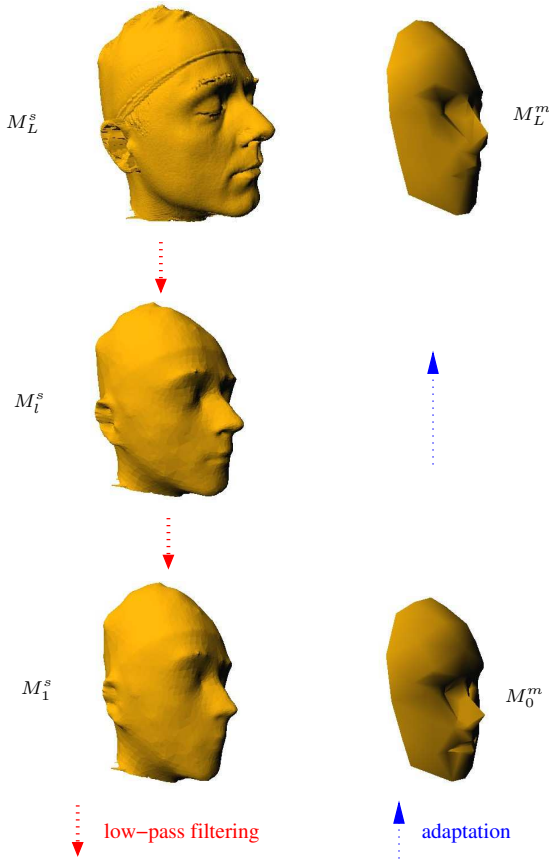
Fig. 3. Local adaptation: First the scan is low-pass filtered, before the generic face model is iteratively adapted.

A. Low-pass filtering

In order to obtain a hierarchy of meshes with increasing details low-pass filtering is applied. There are several filtering techniques for 2-manifold meshes known in literature [13]–[15]. In our mask adaptation framework we employ a filter based on a discretized version of the Laplace-Operator [13] Δ_d which is defined as

$$\Delta_d(p_i) = \frac{1}{|N(i)|} \sum_{p_j \in N(i)} p_j - p_i.$$

A low-pass filtered version $M_l = H_\lambda(M_{l+1})$ of mesh M_{l+1} is obtained by updating all vertices $\tilde{p}_i$ of M_l according to

$$\tilde{p}_i = p_i + \lambda \Delta_d(p_i) \quad \text{with } \tilde{p}_i \in V_l, p_i \in V_{l+1} \qquad (3)$$

for $\lambda \in [0, 2]$. The parameter λ controls the amount of attenuation of the details of the scan, e.g. for $\lambda = 1$ H_1 shifts all vertices p_i into their local barycenter. For values of λ outside of $[0, 2]$ H_λ has no low-pass filtering properties anymore. Thus by successive application of the low-pass filter H_λ to the 3D scan M^s we get a series of meshes $M_L^s, \ldots, M_1^s$ with attenuating high "frequencies":

$$M_l^s := H_\lambda(M_{l+1}^s) \quad \text{for } l = 1, \ldots, L-1, \quad M_L^s := M^s. \quad (4)$$

The number of vertices and the connectivity in all low-pass filtered scans remain the same as in the original. Vertices are

only shifted, as described by equation (3). Hence, the mean Euclidean distance between the vertices of two consecutive low-pass filtered scans M_l^s and M_{l+1}^s can be easily calculated. If this distance is smaller than a threshold d_{th}, which is defined relatively to the length of the bounding box diagonal of the scan, the low-pass filtering is stopped and the number of low-pass filtered scans L in (4) is determined.

Equations (3) and (4) are a generalization of the classical diffusion equation $\partial_t \rho = \lambda \cdot \Delta \rho$, which describes the flow of heat in Euclidean space [14]. In fact the mesh series $M_L^s, \ldots, M_1^s$ represents consecutive states of a discretized diffusion process, where vertices are shifted resp. diffuse in order to reduce local curvature. In this context the local adaptation process in equation (2) can be interpreted as an approximation to the reverse diffusion $M_1^s, \ldots, M_L^s$ which is derived from the 3D scan.

V. Error Measure

For 1D and 2D signals a great number of error or distortion measures are known. These measures range from simple objective such as the mean square error to more elaborate subjective measures regarding human perception [16]. Error measurements for 3D data sets are much more complex, since the comparison between two data sets with a different number of vertices is not straight forward. We choose the Hausdorff distance as an objective measure to determine the distortion between two 3D data sets [17]. This distance is more appropriate as an error measure then a simple vertex to vertex metric.

The distance between a point $p_i^m \in V^m$ belonging to the face model and scan is defined as:

$$d(p_i^m, M^s) = \min_{p_i^s \in V^s} \| p_i^m - p_i^s \|_2 \qquad (5)$$

The root mean square Hausdorff distance between face model and scan is defined as:

$$d_{\text{RMS}}(M^m, M^s) = \sqrt{\frac{1}{|V^m|} \sum_{p_i^m \in V^m} d(p_i^m, M^s)^2} \qquad (6)$$

It is important to note, that d_{RMS} is in general not symmetric. Hence, the symmetrical Hausdorff distance is defined as:

$$d_s(M^m, M^s) = \max[d_{\text{RMS}}(M^m, M^s), d_{\text{RMS}}(M^s, M^m)] \qquad (7)$$

The symmetrical distance $d_s(M^m, M^s)$ describing the error between two data sets is used for the evaluation of our adaptation algorithm. For that we are using the tool 'M.E.S.H.', which is publicly available and can be obtained on the Web at http://mesh.epfl.ch.

VI. Results

We have implemented the complete system described in this paper and adapted both generic face models to different 3D scans. For evaluation of the local adaptation the symmetrical

TABLE I

HAUSDORFF DISTANCE BETWEEN CANDIDE1 AND 3D SCAN.

Face Model	global	$\lambda=0$	$\lambda=0.25$	$\lambda=0.5$	$\lambda=0.75$	$\lambda=1$
original	3.238	1.452	1.362	1.316	1.313	1.338
subdivision 1	3.238	0.837	0.682	0.667	0.655	0.671
subdivision 2	3.238	0.665	0.510	0.489	0.478	0.490

TABLE II

HAUSDORFF DISTANCE BETWEEN CANDIDE3 AND 3D SCAN.

Face Model	global	$\lambda=0$	$\lambda=0.25$	$\lambda=0.5$	$\lambda=0.75$	$\lambda=1$
original	1.066	0.890	0.906	0.898	0.877	0.913
subdivision 1	1.066	0.539	0.504	0.495	0.485	0.493
subdivision 2	1.066	0.445	0.388	0.381	0.382	0.387

Hausdorff distance (7) between the original 3D scan M_L^s and the global $\bar{M}^m$ as well as overall adapted face model M_L^m is calculated. The resulting distances $d_s(M^m, M^s)$ for the different adaptations are presented for both Candide models in Tab. I and II. Not only the original Candide model is investigated, but also the model after applying one (subdivision 1) and two (subdivision 2) uniform plain subdivisions. In the columns of both tables the following adaptations are compared: global adaptation, local adaptation without filtering ($\lambda = 0$), and with filtering using $\lambda = 0.25$, $\lambda = 0.5$, $\lambda = 0.75$ and $\lambda = 1$.

The Hausdorff distance is significantly reduced by the local adaptation with respect to the globally adapted model $\bar{M}^m$. For this reason a precise adaptation requires a local adaptation. Moreover, subdivision also significantly improves the adaptation results; e.g. in Tab. I the Hausdorff distance is almost reduced by half between the original mesh and once subdivided. That is because the original model's number of vertices needed for a more precise adaptation is too low. Hence, a larger number of vertices describing the geometric shape is significantly improving the adaptation.

The best adaptation results are obtained with a low-pass filter using approximately λ=0.75 for both generic models. This filter seems to reduce local curvatures in the most beneficial way for adapting the generic models to a human face. The Hausdorff distance between the adapted Candide-1 model without ($\lambda = 0$) and with filtering using λ=0.75 is unequal zero in areas with small details. There low-pass filtering improves the adaptation, because the generic model is iteratively adapted first to the most and then to less filtered scans.

We tested our adaptation algorithm for other scans and obtained similar results, which lead to the same conclusions.

VII. CONCLUSIONS

For an adaptation system we discussed the global and local adaptation of a generic face model to a 3D scan. The global adaptation is based on RBFs, which scale and orient the face model's mesh to the 3D scan. For the local adaptation different low-pass filters are investigated, which filtered the original scan before the meshes of the generic face model are iteratively adapted to the scan. We explicitly addressed the problem of adapting generic models as precisely as possible, which former proposed algorithms did not. The low-pass filter with λ=0.75 leads to the smallest Hausdorff distances. Moreover, the number of vertices can be increased by uniform plain subdivision. Face models with a higher number of vertices lead to much better adaptation results. Hence, a precise adaptation with the introduced algorithm is achievable. Furthermore, our algorithm has the advantages that it is simple to implement and requires only low-computational effort. The proposed algorithm is not limited to adapt generic face models, but arbitrary generic models can be adapted to 3D scans.

VIII. ACKNOWLEDGEMENTS

REFERENCES

[1] J. Ahlberg, "Extracting mpeg-4 faps from video," in *MPEG-4 Facial Animation: The Standard, Implementation and Applications*, I. S. Pandzic and R. Forchheimer, Eds. Chichester, England: Wiley, 2002, pp. 17–56.

[2] F. I. Parke, "Computer generated animation of faces," *Proc. ACM annual conf.*, 1972.

[3] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin, "Synthesizing realistic facial expressions from photographs," *Computer Graphics*, vol. 32, no. Annual Conference Series, pp. 75–84, 1998.

[4] G. A. Kalberer and L. Van Gool, "Face animation based on observed 3D speech dynamics," in *Proceedings of Computer Animation (CA2001)*, November 2001, pp. 20–27.

[5] J. Ostermann, "Face animation in mpeg-4," in *MPEG-4 Facial Animation: The Standard, Implementation and Applications*, I. S. Pandzic and R. Forchheimer, Eds. Chichester, England: Wiley, 2002, pp. 17–56.

[6] M. Kampmann and R. Farhoud, "Precise face model adaptation for semantic coding of videophone sequences," in *Picture Coding Symposium (PCS 97) Berlin, Germany, 10-12 Sept. 1997*, 1997.

[7] J. Ostermann, L. S. Chen, and T. S. Huang, "Animated talking head with personalized 3d head model," *J. VLSI Signal Process. Syst.*, vol. 20, no. 1-2, pp. 97–105, 1998.

[8] Y. Zhang, T. Sim, and C. L. Tan, "Rapid modeling of 3d faces for animation using an efficient adaptation algorithm," in *GRAPHITE '04: Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and Southe East Asia*. New York, NY, USA: ACM Press, 2004, pp. 173–181.

[9] Q. Wang, H. Zhang, T. Riegel, E. Hundt, and G. Xu, "Creating animatable mpeg4 face," EUROIMAGE ICAV3D, 2001.

[10] R. Schaback, "Comparison of radial basis function interpolants," in *Multivariate Approximation: From CAGD to Wavelets*, K. Jetter and F. I. Utreras, Eds. Singapore: World Scientific, 1993, pp. 293–305.

[11] N. Arad, N. Dyn, D. Reisfeld, and Y. Yeshurun, "Image warping by radial basis functions: applications to facial expressions," *CVGIP: Graphical Models and Image Processing*, vol. 56, pp. 161–172, 1994.

[12] Z. Wu, "*Multivariate Compactly Supported Positive Definite Radial Functions*," Advances in Computational Mathematics, vol. 4, pp. 283–292, 1995.

[13] G. Taubin, "A signal processing approach to fair surface design," in *Proceedings of the Conference on Computer Graphics (SIGGRAPH-95)*, R. Cook, Ed. New York: ACM Press, Aug. 6–11 1995, pp. 351–358.

[14] M. Desbrun, M. Meyer, P. Schröder, and A. H. Barr, "Implicit fairing of irregular meshes using diffusion and curvature flow," *Computer Graphics*, vol. 33, no. Annual Conference Series, pp. 317–324, 1999.

[15] H. Yagou, A. Balyaev, and D. Wei, "Mesh median filter for smoothing 3-D polygonal surfaces," in *Proceedings of the First International Symposium on Cyber Worlds (CW'02)*, 2002, pp. 488–498.

[16] S. Winkler, "Visual fidelity and perceived quality: Towards comprehensive metrics." in *Proc. SPIE Human Vision and Electronic Imaging*, vol. 4299, San Jose, California, 2001, pp. 114–125.

[17] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, "Mesh: Measuring errors between surfaces using the hausdorff distance," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, vol. I, 2002, pp. 705 – 708.