Personalization of Statistical Face Models for Tracking and Animation

Markus Kettern, Anna Hilsmann, Peter Eisert
Fraunhofer HHI, Berlin*

Figure 1: *Mean model shape; model fit by parameters; target shape; mean shape plus personalization offsets; fit of personalized model.*

1 Introduction

Linear and multilinear geometric models of human faces (e.g. blend shapes) are one of the prime representations for facial action in computer vision and graphics. Well-crafted, person-specific models involve lots of manual labour but enable photorealistic animations. On the other hand, simple and generic models are well established for tracking facial action in video since they are robust and allow the estimation of facial geometry even from monocular data, e.g. [Pighin et al. 1999]. Between those extremes, statistical models created from semantically aligned sets of 3D face scans can cover a part of the variety of shapes and expressions that human faces may take on while still looking rather realistic, e.g. [Bolkart and Wuhrer 2013]. However, the number of shapes they can resemble is limited and they often lack detail since they mainly represent lower order subspaces of the data used to create them.

We present an approach to adapt statistical geometry models to a specific person via one or more 3D face scans which are not semantically aligned. Statistical models usually can be controlled by a set of parameters that alter face shape and expression. Our approach adds another layer in which each vertex is individually manipulated in order to match a set of target expressions. In this way, the flexibility of these models can be exploited for tracking a 3D face model in video or for animation with increased level of detail and many facial characteristics of the target person.

2 Geometry-based Personalization

We assume that the statistical model we want to personalize is composed of a mean shape represented by a triangle mesh and sets of vertex offsets that express a change either in shape or expression. The core idea of our approach to personalizing the statistical model is to extend it by a person specific set of vertex offsets that can be added to the mean shape. These offsets are optimized to fit one or more 3D reconstructions yielding different expressions of the traget person. In order to keep the model consistent, we choose an approach that makes use of the model's inherent adaptability towards shape and expression and only adds offsets where needed to better fit the reconstructions.

The mean shape is given by vertices $\mathbf{s} \in S$. For linear models, offsets for shape i and expression j are denoted U_i, V_j and instances of the model are created by $M(\boldsymbol{\theta}, \boldsymbol{\psi}) = S + \sum_i \theta_i U_i + \sum_j \psi_j V_j$. In the multilinear case, the deformation vectors are stored in a tensor $\mathcal{U}$ and instances are given by $M(\boldsymbol{\theta}, \boldsymbol{\psi}) = S + \mathcal{U} \times_2 \boldsymbol{\theta}^T \times_3 \boldsymbol{\psi}^T$ where $\times_n$ denotes the n-th mode product (see [Vlasic et al. 2005] for a detailed explanation). Models that do not use an explicit mean shape can easily be converted to one of these definitions. The target shapes are denoted by T_k. We define the difference between a model instance $M(\boldsymbol{\theta}, \boldsymbol{\psi})$ and a target mesh T_k to be

$$\mathcal{E}_d = \sum_{\mathbf{m} \in M(\boldsymbol{\theta}, \boldsymbol{\psi})} \min \left(\|\mathbf{m} - \mathbf{t}_m\|^2, \|\mathbf{m} - \mathbf{p}_\mathbf{m}\|^2 \right)$$

where $\mathbf{t}_\mathbf{m}$ denotes the closest vertex to $\mathbf{m}$ in T_k and $\mathbf{p}_\mathbf{m}$ the projection of $\mathbf{m}$ onto the closest triangle in T_k. Both are determined efficiently using a suitable extension to a standard acceleration hierarchy based on axis-aligned bounding boxes.

To prevent noise and degeneration of the mesh structure, a regularization term $\mathcal{E}_r = \Phi(LM(\boldsymbol{\theta}, \boldsymbol{\psi}) - LS)$ based on the mesh Laplacian L [Sorkine and Alexa 2007] is added that penalizes deviations in the Laplacian detail vector according to a suitable penalty function Φ, e.g. the Charbonnier Norm. When creating a personalization shape for K target meshes, the resulting objective function $\mathcal{E} = \lambda \mathcal{E}_d + \gamma \mathcal{E}_r$ is minimized in two passes, first over the shape and expression parameters $(\boldsymbol{\theta}, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K)$ and then jointly over these parameters together with the novel vertex offsets U^*. The novel offsets extend the model such that its expression subspace can be moved as close as possible to the geometric subspace spanned by the reconstructions T_k. This extension enhances the resemblance of the target person by the statistical model, thus enabling better performance in applications like tracking and animation.

References

BOLKART, T., AND WUHRER, S. 2013. Statistical analysis of 3d faces in motion. In *Int. Conference on 3D Vision - 3DV*.

PIGHIN, F., SZELISKI, R., AND SALESIN, D. 1999. Resynthesizing facial animation through 3D model-based tracking. *Proc. IEEE International Conference on Computer Vision - ICCV 1*.

SORKINE, O., AND ALEXA, M. 2007. As-rigid-as-possible surface modeling. *Proc. 5th Eurographics symposium on Geometry processing*, 109–116.

VLASIC, D., BRAND, M., PFISTER, H., AND POPOVIĆ, J. 2005. Face transfer with multilinear models. *ACM Trans. Graphics 24*, 3 (July), 426–433.

*{markus.kettern, anna.hilsmann, peter.eisert}@hhi.fraunhofer.de