# Multilinear Models for Face Synthesis

Daniel Vlasic* Matthew Brand† Hanspeter Pfister† Jovan Popović*

*Massachusetts Institute of Technology
†Mitsubishi Electric Research Laboratories

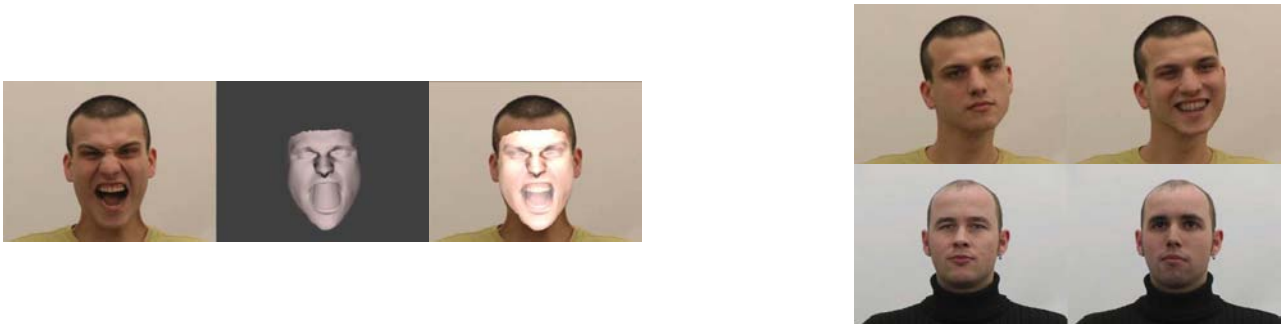

Figure 1: We build a 3D face model with separate identity and expression controls from an incomplete data set of scanned faces. The model parameters can be set via tracking, as shown in left three images: tracked video, the recovered 3D shape, and their overlay. We can then easily modify expression or identity in each frame, as demonstrated on the top- and bottom-right respectively.

## 1 Introduction

Multilinear models offer a natural way of modeling heterogenous sources of variation. We are specifically interested in facial geometry variations due to identity and expression changes. In this setting, the multilinear model is able to capture idiosyncrasies such as style of smiling, i.e. the smile depends on the identity parameters.

Two properties of multilinear models are of particular interest to animators: **Separability** – expression can be varied while identity stays constant, and vice versa; and **Consistency** – expression parameters encoding a smile for one person will encode a smile for every person spanned by the model, *appropriate to their facial geometry and style of smiling.*

We introduce methods that make multilinear models a practical tool for animating faces, addressing two key obstacles. The key obstacle in *constructing* a multilinear model is the vast amount of data (in full correspondence) needed to account for every possible combination of attribute settings. The key problem in *using* a multilinear model is devising an intuitive control interface. For the data-acquisition problem, we show how to estimate a detailed multilinear model from an incomplete set of high-quality face scans. For the control problem, we show how to drive this model with a video performance, extracting identity, expression, and pose parameters.

## 2 Method

After acquiring 3D facial meshes of a number of subjects performing various expressions, we arrange them into a 3-mode data tensor, where the first mode corresponds to identity, second to expression, and the third contains vertex coordinates. Any missing data can be statistically filled with our imputation algorithm. We derive and reduce a generative model by decomposing the data matrix with HOSVD, a higher-order generalization of SVD, formalized by [De Lathauwer et al. 2000], and introduced to vision community by [Vasilescu and Terzopoulos 2002]. This yields the model tensor, which, when multiplied by identity and expression weights, generates the appropriate face geometry.

Using model-based optical flow, we can relate our model to video data and solve for its parameters: per-frame expression and a globally consistent identity. Linear models (e.g. [Blanz and Vetter 1999]) cannot do this without knowing either the identity or the expression in one of the frames. As implementation, we adapt the method by [Brand and Bhotika 2001] to multidimensional models, essentially running a Newton's method (keeping identity constant while solving for expressions, then updating the identity, and repeating until convergence).

## 3 Results

With our system one can lift a performance from video and change its expression and identity parameters, as illustrated in the figure, thus simplifying applications such as video-rewrite or puppetry.

## References

Blanz, V., and Vetter, T. 1999. A morphable model for the synthesis of 3d faces. In *Proceedings of SIGGRAPH 99*, Computer Graphics Proceedings, Annual Conference Series, 187–194.

Brand, M., and Bhotika, R. 2001. Flexible flow for 3d nonrigid tracking and shape recovery. In *2001 Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, vol. 1, 315–324.

De Lathauwer, L., De Moor, B., and Vandewalle, J. 2000. A multilinear singular value decomposition. *Siam Journal on Matrix Analysis and Applications 21*, 1253–1278.

Vasilescu, M. A. O., and Terzopoulos, D. 2002. Multilinear analysis of image ensembles: Tensorfaces. In *7th European Conference on Computer Vision (ECCV 2002) (Part I)*, 447–460.