# Supplementary Material: Learning a model of facial shape and expression from 4D scans

TIANYE LI*†, University of Southern California and Max Planck Institute for Intelligent Systems
TIMO BOLKART*, Max Planck Institute for Intelligent Systems
MICHAEL J. BLACK, Max Planck Institute for Intelligent Systems
HAO LI, Pinscreen, University of Southern California, and USC Institute for Creative Technologies
JAVIER ROMERO†, Body Labs Inc.

Our FLAME model (Faces Learned with an Articulated Model and Expressions) is a statistical head model that combines a linear shape space with an articulated jaw, neck, and eyeballs, pose-dependent corrective blendshapes, and additional global expression blendshapes. This supplemental contains additional figures and results that were omitted in the paper for brevity. We show more results to assess the quality of the registration, the quality of the identity space of FLAME for different number of components, and we give more visual comparisons to FaceWarehouse and Basel Face Model.

CCS Concepts: • **Computing methodologies** → *Mesh models*;

**ACM Reference Format:**
Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero. 2017. Supplementary Material: Learning a model of facial shape and expression from 4D scans. *ACM Trans. Graph.* 36, 6, Article 194 (November 2017), 5 pages. https://doi.org/10.1145/3130800.3130813

*Data:* FLAME is built from three heterogeneous sources, using more than 33, 000 3D scans in total. This comprises shape data ( 3800 shapes), pose data (8000 shapes), and 21, 000 registered expression frames sampled from the 69, 000 registered expression frames.

Figure 1 shows sample head registrations of the CAESAR database [Robinette et al. 2002], showing the large variation in shape present in the database. Figure 2 shows samples of the captured neck rotation (top) and jaw motions (bottom) used as pose data. Figure 3 shows the expression data, namely registrations of D3DFACS [Cosker et al. 2011] (top) and self-captured sequences (bottom).

*Registration quality:* Figure 4 shows further sample registrations of the D3DFACS dataset (top) and our self-captured sequences (bottom). Our registration is able to track subtle motions such as raising the eyebrows (Figure 4 top) or extreme facial expressions such as a wide open mouth (Figure 4 bottom).

*Both authors contributed equally to the paper
†This research was performed while TL and JR were at the MPI for Intelligent Systems.

Authors' email addresses: tianyeli@usc.edu; {tbolkart, black}@tue.mpg.de; hao@hao-li.com; javier.romero@bodylabs.com.


0730-0301/2017/11-ART194
https://doi.org/10.1145/3130800.3130813

*Model quality:* Figure 5 gives further qualitative evaluations on the influence of a varying number of identity components for fitting the neutral BU-3DFE face scans. Increasing the number of components increases the ability of the model to reconstruct localized details. FLAME 300 leads to registrations with an error that is close to zero millimeters in most facial region Figure 6 gives further qualitative comparisons of BFM [Paysan et al. 2009], FW [Cao et al. 2014], and FLAME. Compared to FLAME, BFM has lots o high-frequency details that make the fits look more realistic. Nevertheless, the comparison with the scans reveal that these details are hallucinated and spurious, as they come from people in their original training dataset, rather than from the scans. While lower-resolution and less detailed, FLAME is actually more accurate.

*Shape reconstruction from images:* Figure 7 shows the 2D landmark fitting using FW (top) and FLAME (bottom). FLAME better fits the identity and produces a lower 3D scan distance.

*Expression transfer:* Figure 8 shows the expression transfer between a subject in our test dataset and a high-resolution static scan of Beeler et al. [2011]. The synthetic sequence looks realistic despite the large face shape difference of source and target.

## REFERENCES

T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, R. W. Sumner, and M. Gross. 2011. High-quality passive facial performance capture using anchor frames. *Transactions on Graphics (Proceedings of SIGGRAPH)* 30, 4 (2011), 75:1–75:10.

C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou. 2014. FaceWarehouse: A 3D facial expression database for visual computing. *Transactions on Visualization and Computer Graphics* 20, 3 (2014), 413–425.

P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. 2009. A 3D Face Model for Pose and Illumination Invariant Face Recognition. In *International Conference on Advanced Video and Signal Based Surveillance*. 296–301.

K. Robinette, S. Blackwell, H. Daanen, M. Boehmer, S. Fleming, T. Brill, D. Hoeferlin, and D. Burnsides. 2002. *Civilian American and European Surface Anthropometry Resource (CAESAR) Final Report.* Technical Report AFRL-HE-WP-TR-2002-0169. US Air Force Research Laboratory.

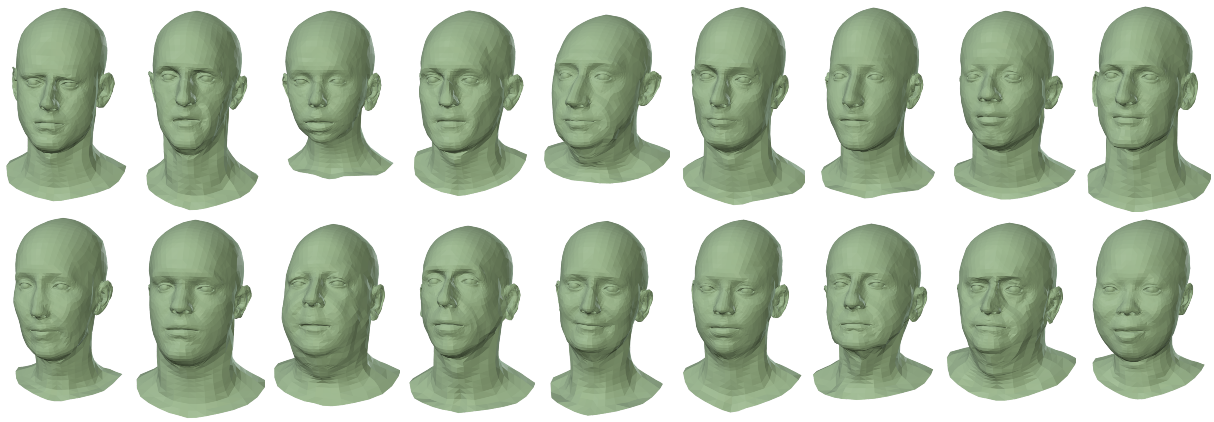

Fig. 1. Sample registrations of the shape data extracted from the CAESAR body database.

Fig. 2. Sample registrations of the self captured pose data. Top: Head rotations around the neck. Bottom: Mouth articulations.

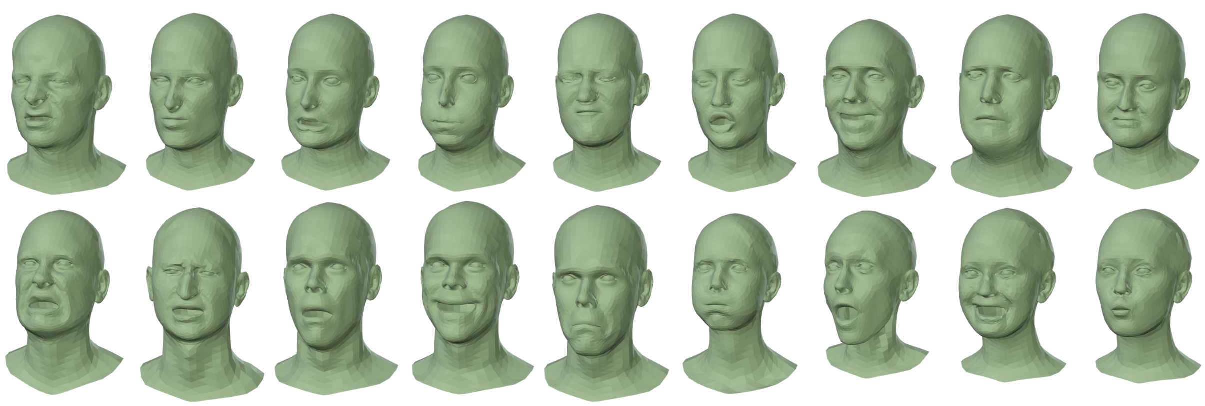

Fig. 3. Samples registrations of the expression data from D3DFACS (top) and self captured sequences (bottom).

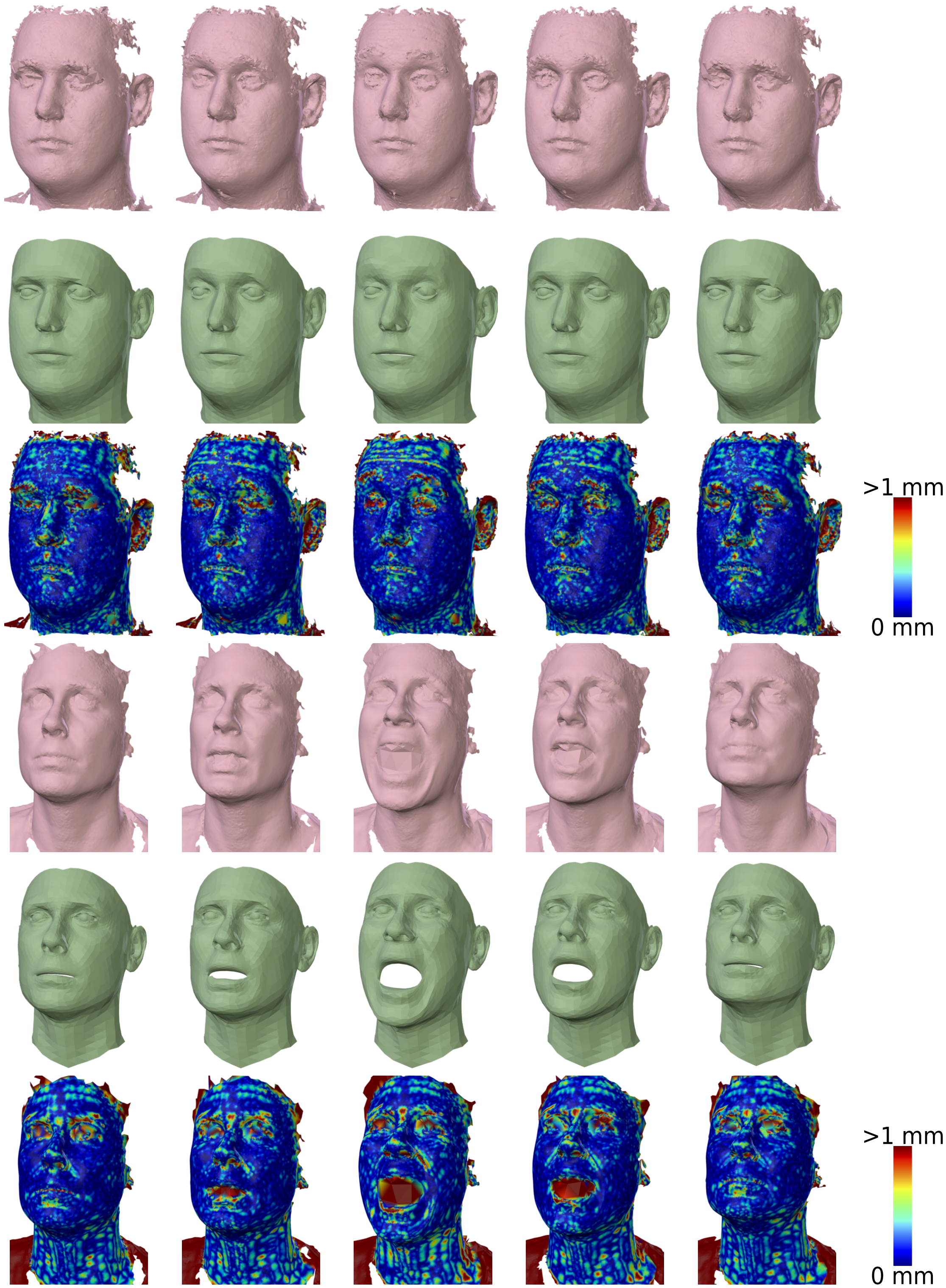


Fig. 4. Sample frames, registrations, and scan-to-mesh distance of one sequences of the D3DFACS database (top) and one sequence of our self-captured sequence (bottom). The texture-based registration allows to track subtle motions such as raising eyebrows (top).

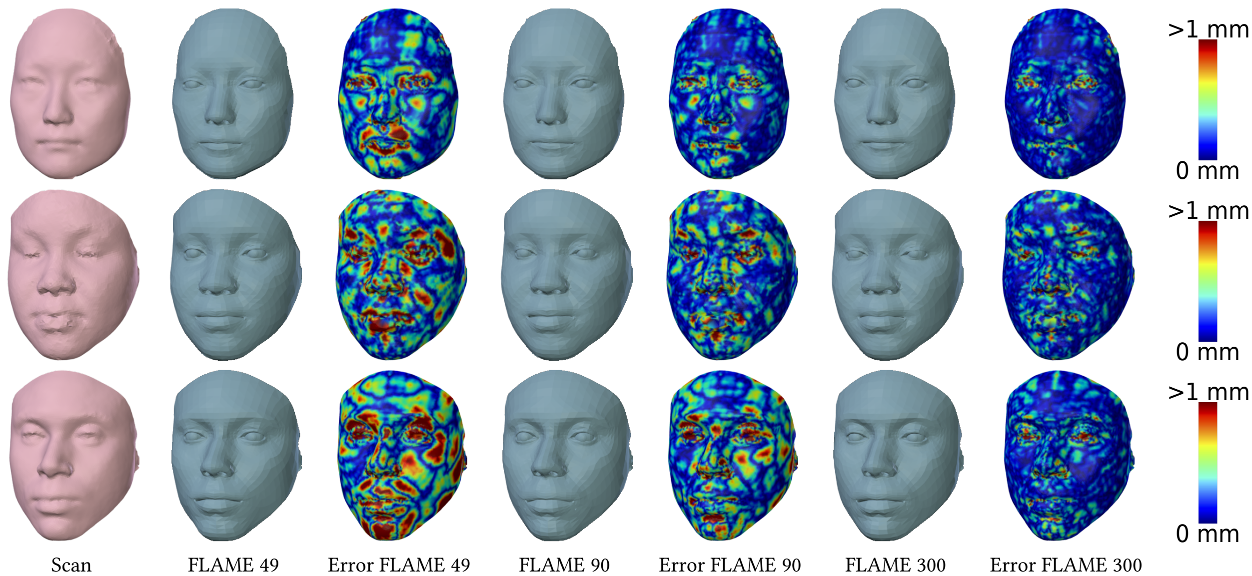


Fig. 5. Expressiveness of the FLAME identity space for fitting neutral scans of the BU-3DFE face database with a varying number of identity components.

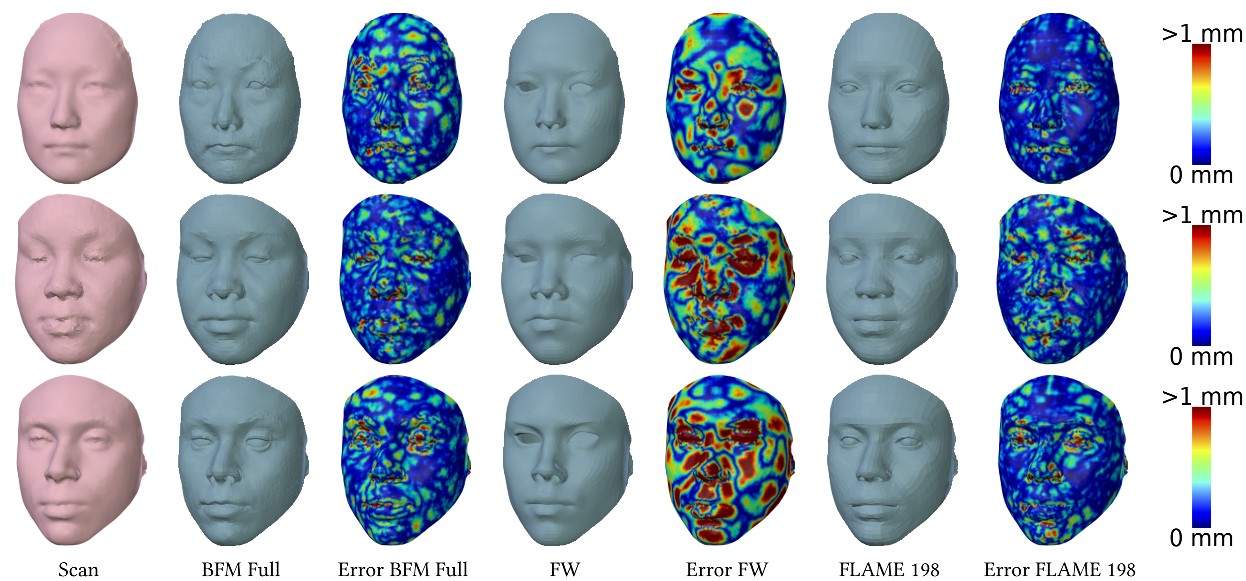


Fig. 6. Comparison of Basel Face Model (BFM) [Paysan et al. 2009], FaceWarehouse model [Cao et al. 2014] and FLAME for fitting neutral scans of the BU-3DFE database.

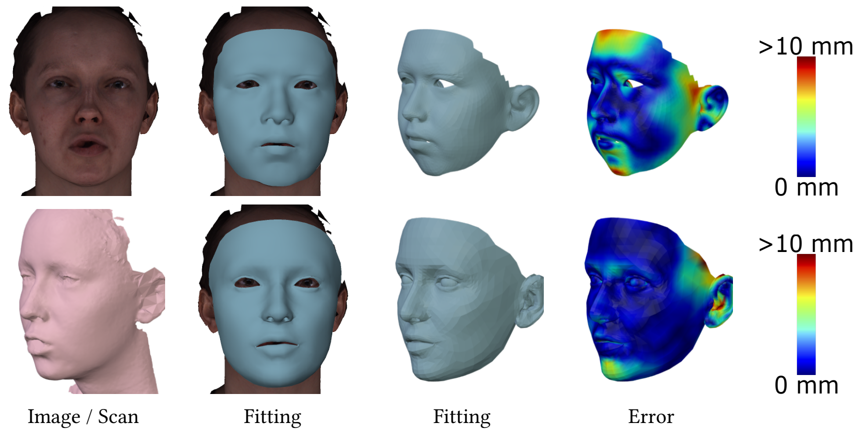


Fig. 7. Comparison of FaceWarehouse model (top) and FLAME (bottom) for 3D face fitting from a single 2D image. Note, that the scan is only used for evaluation.

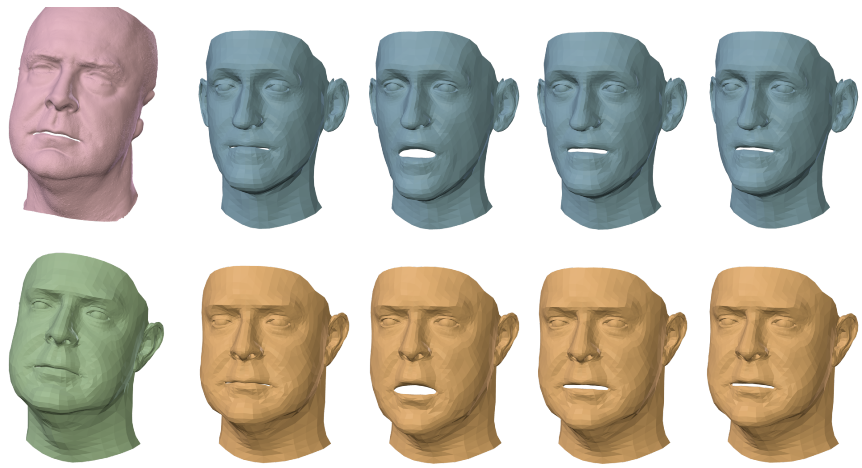

Fig. 8. Expression transfer from a source sequence (blue) to a static target scan (pink). The aligned personalized template for the scan is shown in green, the transferred expression in yellow.