AMASS: Archive of Motion Capture as Surface Shapes
*Supplementary Material**

Naureen Mahmood[1] Nima Ghorbani[2] Nikolaus F. Troje[3]
Gerard Pons-Moll[4] Michael J. Black[2]
[1]Meshcapade [2]MPI for Intelligent Systems [3]York University [4]MPI for Informatik
nmahmood@meshcapade.com, {nghorbani, black}@tue.mpg.de
troje@yorku.ca, gpons@mpi-inf.mpg.de

1. Optimization and Runtime

Section 3.3 of the main paper briefly describes the implementation of MoSh++. Here we present the runtime details of the method on AMASS, also comparing them with MoSh [3]. All results are computed on a 2013 MacPro, 3GHz Intel Xeon E5, 16GB RAM.

MoSh++ is faster than MoSh when fitting a body with the same number of parts (i.e. without hands).

1. Shape stage:
   - MoSh++ (SMPL)=554.12s;
   - MoSh (SCAPE)=654.92s per subject;
2. Pose stage:
   - MoSh++ (SMPL)=1.54s;
   - MoSh (SCAPE)=2.64s per frame

Adding hands increases the number of parameters and, consequently, the runtime for MoSh++ (SMPL-H):

1. Shape stage: 959.03s;
2. Pose stage: 2.23s.

2. Data Collection

Sec. 4.1 of the main paper presents the SSM (Synchronized Scans and Markers) dataset. To record this dataset we use an optical motion capture system synchronized and calibrated together with a high resolution 4D scanning system.

We used an OptiTrack motion capture system (NaturalPoint, Inc. DBA OptiTrack. Corvallis, OR) [4] consisting of 24 Optitrack Prime 17W optical mocap cameras. Each subject was fitted with 67 reflective mocap markers based on the *optimized marker-set* layout proposed in [3]. The subjects wore minimal clothing to avoid artifacts due to sliding of cloth. The markers were placed directly on the skin of the subjects wherever possible.

The motion capture system was synchronized to be triggered with a 3dMD 4D body scanning system (3dMD LLC, Atlanta, GA) [1]. The 4D scanner is capable of capturing high-resolution 3D scans of a person at 60 frames per second. The 4D system uses 22 pairs of stereo cameras, 22 color cameras and 34 speckle projectors and arrays of white-light LED panels.

3. Synthetic DFAUST Mocap

SSM is the first and only effort to evaluate 3D body surface estimation from mocap using *ground truth* synced 4D scans. Creating this dataset was a complex process and for unbiased evaluation we took particular care to partition SSM into different train, test and validation sets over multiple trials for hyper-parameter search as detailed in Sec. 4.2 of the paper.

As an additional evaluation, we performed another experiment using one of the largest publicly available 4D-scan datasets, Dynamic FAUST (DFAUST) [2]. We use this to create a *virtual mocap dataset* which lets us evaluate MoSh++ accuracy on a new, larger dataset. We place virtual mocap markers on the registered meshes of DFAUST and add spherical Gaussian noise to the virtual DFAUST markers to simulate variations due to manual marker placement. This gives us a virtual mocap dataset DFAUST-Mocap, which like SSM, has corresponding synchronized 3D scans to serve as ground truth. We fit DFAUST-Mocap with MoSh and MoSh++ and compute accuracy of the recovered 3D meshes from both methods relative to the ground truth 3D scans. As shown in Fig. A.1, MoSh++ reconstruction errors are comparable to those on SSM in Fig. 6 of the paper. This strengthens our original conclusions as DFAUST is not used in hyperparameter turning. The DFAUST dataset includes 5 male and 5 female subjects with varying body shapes, and motions. We release the DFAUST synthetic mocap dataset and the corresponding code used to generate virtual markers at https://amass.is.tue.mpg.de

4. Model Size

Section 3.1 of the main paper describes the SMPL body model incorporated in the MoSh++ pipeline. We experi-

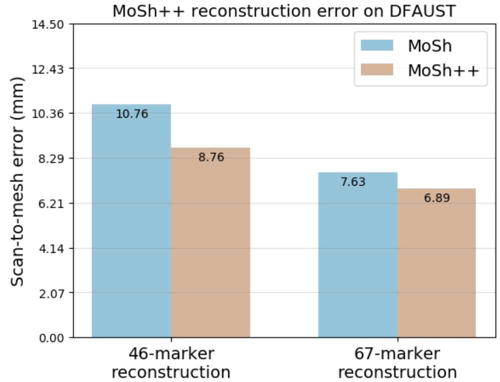

Figure A.1: New DFAUST experiment. Mean absolute distance between ground truth 4D scans and the bodies estimated by MoSh [3] (blue) and MoSh++ (orange). We run both methods twice: using a standard 46-markerset (left), and the specialized 67-markers proposed in MoSh (right).

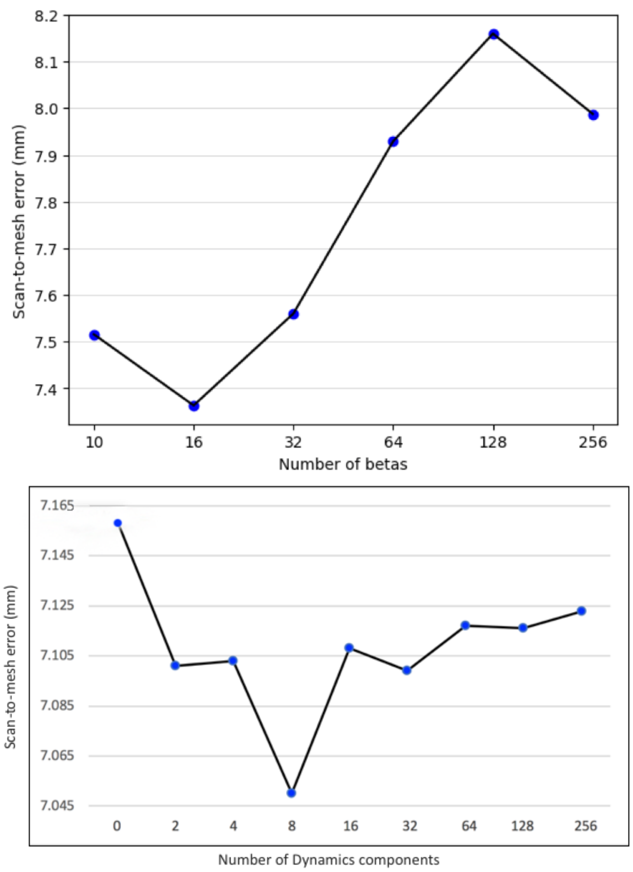

Figure A.2: Optimal number of shape and dynamics components. Mesh reconstruction errors on the SSM dataset using varying numbers of SMPL shape components β (top), and DMPL dynamic components ϕ (bottom) to find the optimal number to use for shape and soft-tissue optimization.

mented with varying numbers of SMPL shape components and DMPL dynamic components to find the optimal number to use to capture shape and soft-tissue motion. We found that $|\beta| = 16$, and $|\phi| = 8$ do the best job of minimizing error on the held-out validation set and also produce natural looking soft-tissue deformations. Given a limited number of mocap markers, allowing more shape variation results in over-fitting (see Fig. A.2).

5. Diversity and Quality

We provide a video to illustrate the variations in the motions in AMASS and the quality of reconstructed body surface deformations from the mocap markers. Please see the video. Note that AMASS captures a wide range of body shapes and motions.

References

[1] 3dMD LLC. 4D Scan. http://www.3dmd.com/. 1

[2] Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Dynamic FAUST: Registering Human Bodies in Motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1

[3] Matthew Loper, Naureen Mahmood, and Michael J. Black. MoSh: Motion and Shape Capture from Sparse Markers. *ACM Trans. Graph.*, 33(6):220:1–220:13, Nov. 2014. 1, 2

[4] NaturalPoint, Inc. Motion Capture Systems. https://optitrack.com/. 1