An Open-Source Articulated Multi-Person Shape Model Training and Inference Pipeline

Samuel ZEITVOGEL, Astrid LAUBENHEIMER
Intelligent Systems Research Group, Karlsruhe, Germany

https://doi.org/10.15221/20.17

Abstract

We propose an open source markerless articulated multi-person shape model training and inference pipeline. Our open source implementation is freely available for non-commercial research purposes.

Keywords: Body shape, skinning, blend-shapes

1. Introduction

Articulated multi-person shape models are crucial for many applications e.g. for shape completion, 3D virtual avatar generation, virtual try-on, gaming, and markerless motion capture. While many trained models are already publicly available, we believe that training custom application-specific models is vital to address many novel use cases. Therefore, we provide an open-source training and inference pipeline for articulated shape models. Our approach neither requires any cost-intensive data registration such as manually set markers or existing correspondences, nor does it depend on existing, preprocessed shape models. Instead, in our approach the 2D pose estimator OpenPose [4,5] is applied on automatically generated synthetic views of the model training data. This allows the direct application to unregistered 3D data. Our approach achieves good results even on comparatively small data sets. The overall pipeline is available for non-commercial research purposes.

2. Our approach

2.1 Related Work

Many state-of-the-art methods for the construction and inference of statistical human shape models are readily available. One of the most important is SMPL [6], which was extended to faces [7], hands [8], and infants [9]. SMPL is compatible with 3D modelling software but relies heavily on high-quality registrations and is thus limited to the application on costly preprocessed data.

Other models can be combined to construct a fully articulated morphable 3D human shape model [8]. Osman et al. [10] extend SMPL by improving the model formulation and by using an additional 10,000 scans for training.

In [1], we propose an open-source training and inference pipeline for articulated shape models. In this work, we focus on the technical aspects of the published framework and extend our method to different object classes.

2.2 Model Formulation

Our model consists of an underlying skeleton and a template base mesh, which defines the resolution and the topology of the expressible surfaces. Both can be adapted for the respective application. As other popular articulated multi-person shape models our model factorizes into subject-specific and pose-specific components.

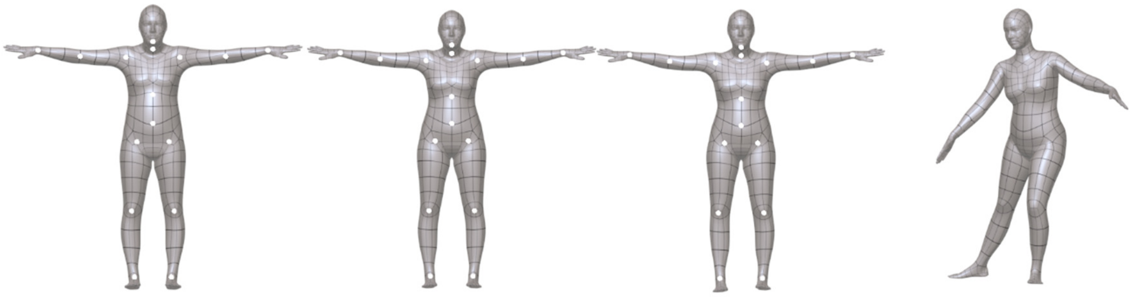

Fig. 1. Following SMPL, from left to right: mean shape with skeleton (white dots) ; subject-specific shape and skeleton; pose-specific shape correction; linear blend skinning. Surface patch borders are shown in black.

Methodically, our approach incorporates aspects of SMPL on the one hand and subdivision surfaces from [11] on the other hand. The latter is advantageous for the optimization task, but also results in good reconstruction quality with a comparatively low number of surface patches. A detailed description of the overall method is given in [1].

2.3 Pre-trained Models

We emphasize the added value of our solution by constructing several articulated multi-person shape models with the same approach: a female, a male, and a unisex articulated multi-person shape model.

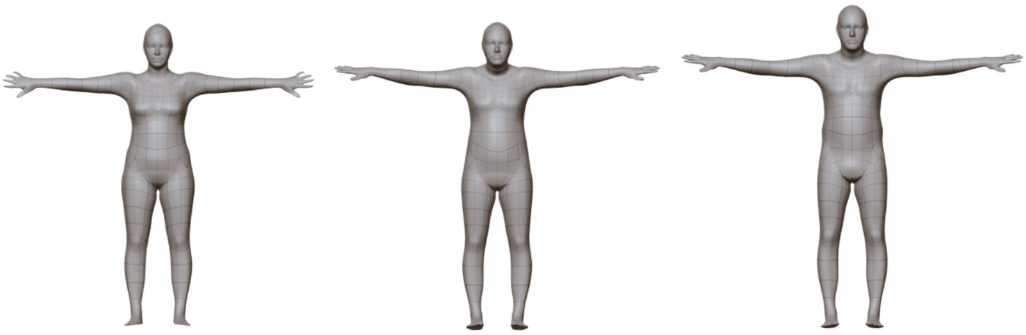

Fig. 2. Mean shape of our female (left), unisex (middle) and male (right) model (see [1] for details).

The multi-person shape models were trained on roughly 1000 raw scans (each), covering a wide range of body shapes and body postures of the D-FAUST and the CAESAR dataset. Training required up to three days on an NVIDIA RTX 2080 TI and the amount of training data was limited by the GPU memory (11 GB).

For quantitative evaluation, we applied our models to the FAUST correspondence challenge dataset [2], where we achieve accuracy comparable to strong human shape models that have higher resolution and rely on (semi)-supervised training schemes. For details see [1].

To test the applicability of our approach to other object classes, we have built an adult and an infant multi-person head model. The latter will be used for plagiocephaly classification in future work. The adult head model was trained on data extracted from the CAESAR data set and the infant head model on data which was acquired for helmet therapies and is characterized by a very low level of detail.

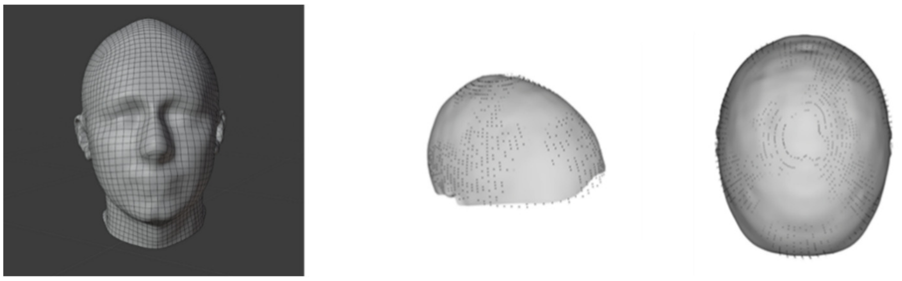

Fig. 3. Adult head model (left) and its fit to an infant scan (middle and right).

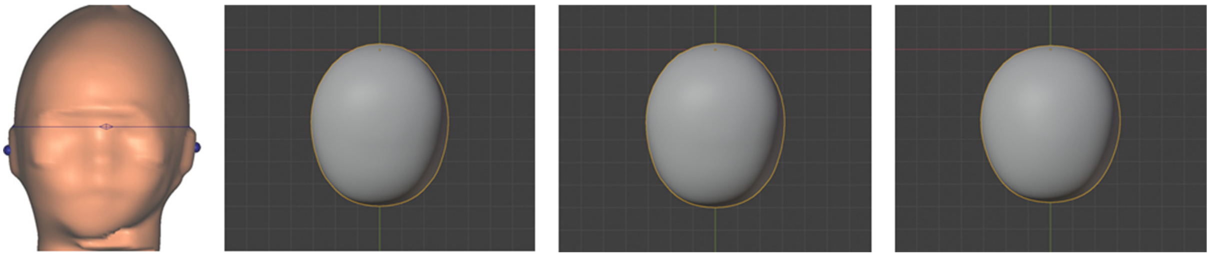

Fig. 4. Input data (left) and three different head contours of our infant head model (top view).

The resulting infant head model shows variations of head contours but due to the lack of features in the training data the infant head model needs to be enhanced, e.g. by bootstrapping with the adult head model, what we will be the next focus on in our future work.

3. Implementation Details

In this section, we describe the necessary steps to train an articulated multi-subject shape model using our training pipeline.

3.1 System and Software Requirements

The implemented model training and inference solution run in a virtualized docker environment making the training and evaluation algorithm easily accessible and deployable. Most of the configuration can be done using the free 3D modelling software blender. A GPU with Cuda 9.2 support is mandatory. The supported dataset size and model complexity is limited by the available GPU memory.

3.2 Model Template and Skeleton Definition

To train the model, a template mesh can be modeled in blender. The template mesh consists of quad faces and has to be watertight. In the next step, the mesh has to be segmented into body parts as depicted in Fig. 5. Body part segmentations are realized in blender using vertex groups. The names of the vertex groups start with “m_” followed by a unique body part identifier. A provided blender script writes the body parts to a text file. The body part text file together with the mesh are fed into the training pipeline.

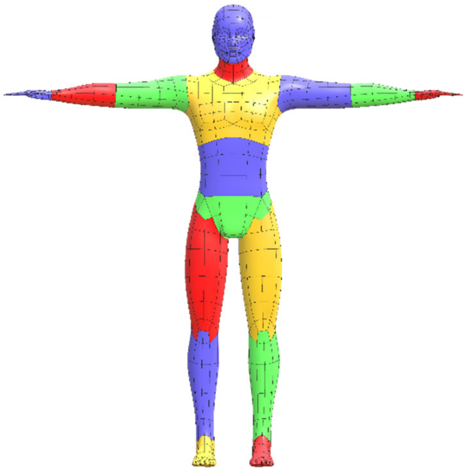

Fig. 5. Template mesh and body part segmentation

The joints of the underlying skeleton will be placed at the center of the intersecting vertices. The skeleton topology can be defined in the skeleton.json file. For our articulated multi-person shaped model training, the skeleton is defined as follows:

```
{
"part_names" :
        ["m_chest", "m_neck","m_head","m_rupperarm",
         "m_rlowerarm","m_rhand","m_lupperarm",
         "m_llowerarm","m_lhand","m_hip","m_pelvis",
         "m_rupperleg","m_rlowerleg",“m_rfoot","m_lupperleg",
         "m_llowerleg","m_lfoot"],
"symmetry" : {
        "m_rupperarm" : "m_lupperarm",”m_rlowerarm" : "m_llowerarm",
        "m_rhand" : "m_lhand",       "m_rupperleg" :  "m_lupperleg",
        "m_rlowerleg" : "m_llowerleg", "m_rfoot" : "m_lfoot",
        "m_lupperarm" : "m_rupperarm","m_llowerarm" : "m_rlowerarm",
        "m_lhand" : "m_rhand", "m_lupperleg" :  "m_rupperleg",
        "m_llowerleg" : "m_rlowerleg", "m_lfoot" : "m_rfoot"     },
"ancestors" : {
        "m_neck" : "m_chest",
        "m_head" : "m_neck",
        "m_rupperarm" : "m_chest"
        etc.           },
"influence" : {
        "m_rlowerarm" : ["m_rupperarm","m_chest","m_rhand","m_neck" ],
        "m_lupperarm" : [ "m_chest","m_hip","m_llowerarm","m_neck" ],
        etc.
}}
```

The part name list contains all the body segment identifiers. The "symmetry" list encodes the symmetry of the body. The "ancestors" list defines the skeleton topology. The "influence" list denotes the linear blend skinning influence (most game engines only support four blend-skinning weights per vertex).

3.3 Model Initialization with OpenPose

For initialization during training we use OpenPose [4,5]. The available keypoints for OpenPose are depicted in Fig. 6.

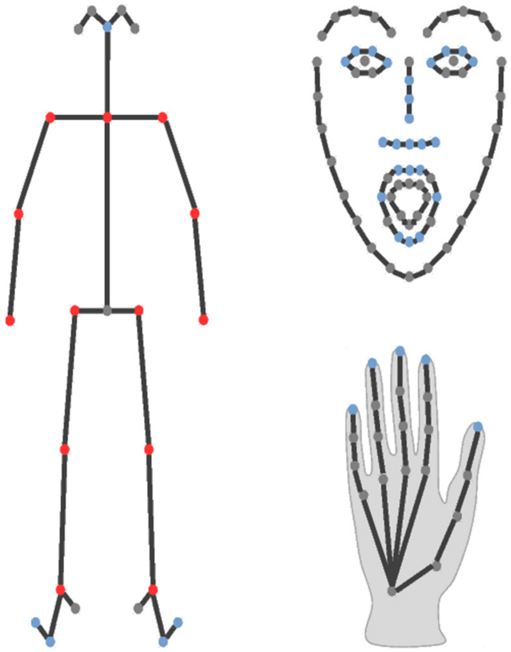

Fig. 6. OpenPose[4,5] keypoints

For each scan, eight images are rendered from different viewpoints. Then, OpenPose is applied on each image. The result on a scan from the Dynamic FAUST dataset [3] is depicted in Fig. 7.

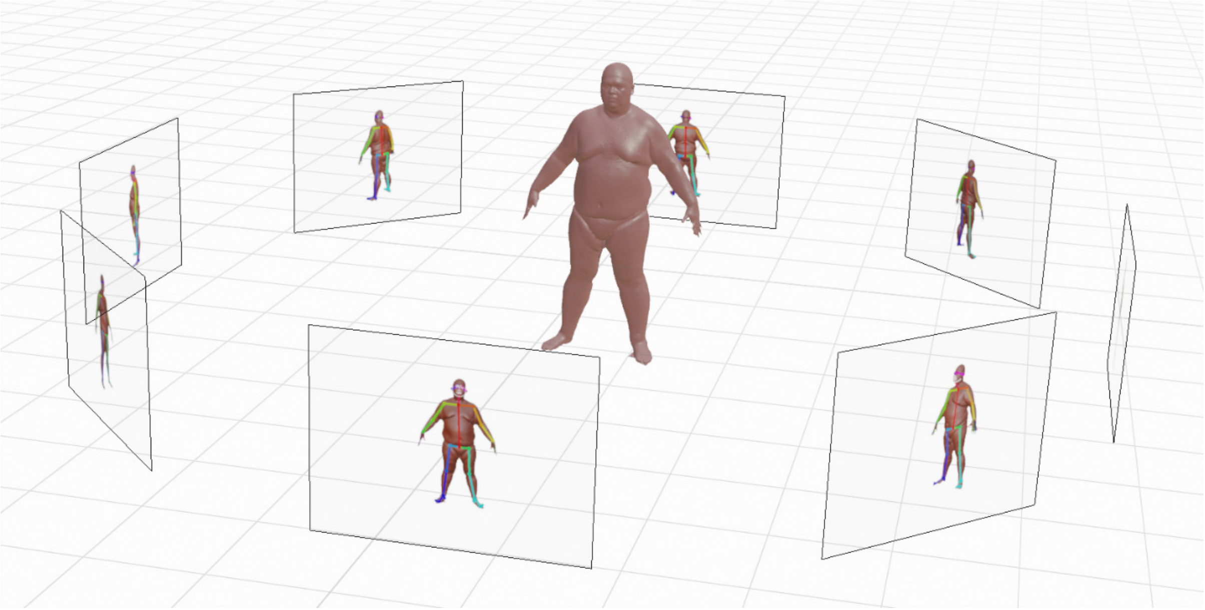

Fig. 7. OpenPose applied on a scan of the Dynamic FAUST dataset [3]

For the OpenPose keypoints, the corresponding vertices on the template mesh and corresponding body joints have to be defined. The correspondences for our articulated multi-person shape model are depicted in Fig. 8.

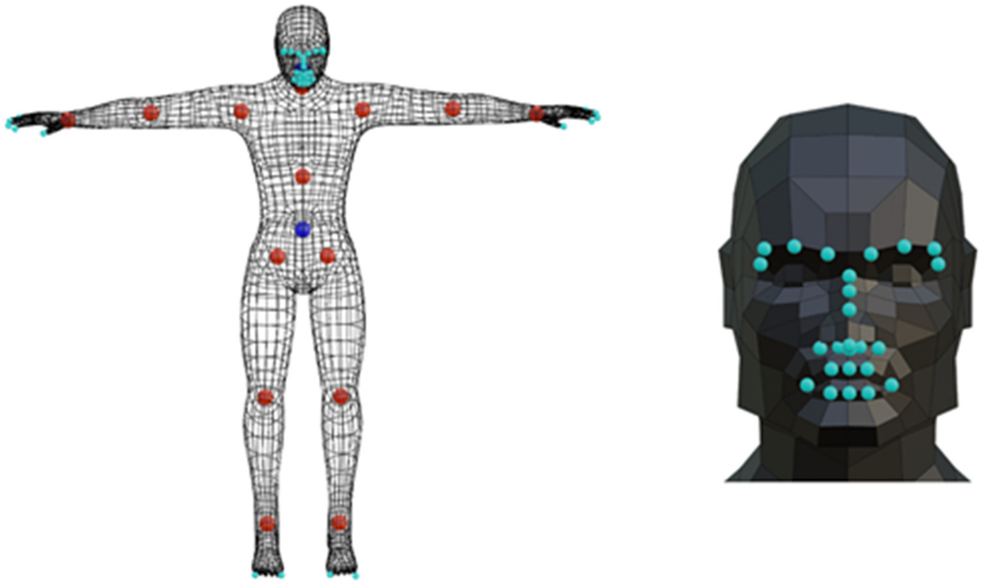

Fig. 8. User-defined OpenPose correspondences to the template mesh and template skeleton

The correspondences are defined in the “surface-to-keyponts.json” file:

```
{
“surface”: [
    "1"    : "Nose",
    "658"  : "RBigToe",
    "637"  : "RSmallToe",
    "222"  : "LBigToe",
    "198"  : "LSmallToe",
    "1141" : "RThumb4FingerTip",
    "876"  : "RIndex4FingerTip",
    "904"  : "RMiddle4FingerTip",
    "874"  : "RRing4FingerTip",
    etc.
]
“skeleton“: [
    {"m_neck","Neck"},{"m_rupperarm","LShoulder"},
    {"m_rlowerarm","LElbow"},{"m_rhand","LWrist"},
    {"m_lupperarm","RShoulder"},{"m_llowerarm","RElbow"},
    {"m_lhand","RWrist"},{"m_rupperleg","LHip"},
    etc.
]“ }
```

Vertex-Id to OpenPose surface keypoint prediction correspondences are defined in the “surface” list (e.g. keypoints on hands, feet and face). Template skeleton joints to OpenPose skeleton keypoint prediction correspondences are defined in the “skeleton” list.

3.4 Hyperparameter Tuning

We expose a configuration file, which enables fast hyperparameter tuning. In particular, the error term weights which may change during the minimization of the error function. As with any training method, finding hyperparameters that are suitable for the respective application might require a deeper understanding of the underlying error function.

3.5 Training and Inference Output

After inference, we output articulated character models that explain the input data. The output is in a standard file format (FBX), which is compatible with many 3D animation and modeling tools. We provide a blender add-on which allows for easy multi-person morphing of the shape and the underlying skeleton.

4. Conclusion

We presented our method for end-to-end construction of articulated multi-subject shape models and presented some results to showcase the transferability of the approach to new object classes. Our method is published as an open-source end-to-end system, which includes the training and the inference step. The system allows customization of the template, the skeleton, the training data and the objective function. In future work, specialized shape models can be applied to novel applications.

Acknowledgment

This work was supported by the German Federal Ministry of Education and Research (BMBF) under Grant 13FH025IX6.

References

[1] S. Zeitvogel et al, “Joint Optimization for Multi-Person Shape Models from Markerless 3D-Scans”, in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[2] F. Bogo et. al, "FAUST: Dataset and evaluation for 3D mesh registration”, in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, https://doi.org/10.1109/CVPR.2014.491.

[3] F. Bogo et. al, "FAUST: Dataset and evaluation for 3D mesh registration”, in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, https://doi.org/10.1109/CVPR.2017.591.

[4] Z. Cao et. al, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[5] T. Simon et. al, "Hand Keypoint Detection in Single Images using Multiview Bootstrapping”, in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, https://doi.org/10.1109/CVPR.2017.494.

[6] M. Loper et. al, "SMPL: A Skinned Multi-Person Linear Model”, in *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 2015, https://doi.org/10.1145/2816795.2818013.

[7] T. Li et. al, "Learning a model of facial shape and expression from 4D scans”, in *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 2017, https://doi.org/10.1145/3130800.3130813.

[8] J. Romero et. al, "Embodied Hands: Modeling and Capturing Hands and Bodies Together”, in *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 2017, https://doi.org/10.1145/3130800.3130883.

[9] N. Hesse et. al, "Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D sequences”, in Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019, https://doi.org/10.1109/TPAMI.2019.2917908.

[10] A. Osman et al, “STAR: Sparse Trained Articulated Human Body Regressor”, in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. https://doi.org/10.1109/CVPR.2014.491.

[11] S. Khamis et al, “Learning an Efficient Model of Hand Shape Variation from Depth Images”, in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, https://doi.org/10.1109/CVPR.2015.7298869.