Generating Realistic and Accurate SMPL Body Shapes from Anthropometric Measurements

Maja Katharina Nikolic[1,2], Sophie Kaltenleithner[1], Ulrich Bodenhofer[2], and Michael Giretzlehner[1]

[1]RISC Software GmbH
[2]University of Applied Sciences Upper Austria
[1]{maja.nikolic, sophie.kaltenleithner, michael.giretzlehner}@risc-software.at
[1]ulrich.bodenhofer@fh-hagenberg.at

Abstract

Accurate human 3D avatars are essential for various medical applications, such as pain visualization and burn size estimation. Generating these avatars from simple anthropometric measurements offers a cheaper and more practical alternative to conventional 3D body scanners and image-based reconstruction methods, especially when patients have limited mobility. In this work, we investigate the reliability of predicting Skinned Multi-Person Linear model (SMPL) shape parameters from anthropometric measurements in the presence of real-world noise. We further introduce β-likelihood to quantify the anatomical plausibility of generated shapes against a learned distribution. Multiple regression models are evaluated on two external datasets, revealing a clear trade-off between metric accuracy and shape plausibility. The results indicate that regularized regression models are best suited to balance this trade-off when dealing with real-world measurement noise.

1 Introduction

Human 3D avatars are gaining more relevance in medical contexts for various applications. For instance, in burn care, human 3D models can be used to visualize burn injuries and calculate the percentage of total body surface area (TBSA) burned, which is a key metric for determining appropriate burn treatment [8]. Another application is pain visualization, where patient-specific 3D body models can help patients to communicate the location and extent of their pain to medical professionals. Since pain is still often documented on generic, frequently male, 2D templates [12], patient-specific 3D avatars can improve anatomical correspondence of pain markings, especially when perceived pain depth relates to tissue volume.

As 3D body scanners are costly, image-based reconstruction methods have been proposed [4, 11, 13]. However, patient photos may be impractical in clinical settings, due to patient's limited mobility or privacy concerns. In such cases, fitting a parametric body model from anthropometric measurements alone can serve as a practical alternative. While Ludwig et al. [6] demonstrate that machine learning can predict body shape from such measurements, their evaluation does not consider measurement errors, which are common in manual anthropometry due to variation in landmark placement, soft-tissue compression and posture. To better understand the reliability of anthropometric measurements the the authors of The Virtual Caliper [10] further investigate which body measurements can be measured most reliably. Their results show that individual measurements can be accurately reconstructed on a 3D model but as the number of approximated measurements increases, the generated body shapes

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

tend to appear less realistic. However, these observations are based solely on perceptual evaluation, leaving the quantitative assessment of shape plausibility unexplored.

To address this gap, this study investigates the reliability of Skinned Multi-Person Linear Model (SMPL) shape parameters predicted from anthropometric measurements under real-world measurement noise. Instead of relying on perceptual judgment to assess plausibility, we propose β-likelihood as an objective metric that measures how likely the predicted shape parameters are under the learned shape distribution. To investigate the trade-off between measurement accuracy and anatomical plausibility across model classes, multiple regression models are trained to predict SMPL shape parameters from a set of anthropometric measurements. The models are then evaluated on external validation datasets containing measurement noise with respect to both metric accuracy and shape plausibility.

2 Data Preparation

Skinned Multi-Person Linear Model (SMPL) [5] is the state-of-the-art parametric human body model, developed and published by the Max Planck Institute for Intelligent Systems. The body shape is controlled by β-coefficients, which are referred to as **shape parameters** throughout this paper. These coefficients encode the principal components of variation derived from a large dataset of aligned 3D human meshes, and capture differences in height, limb proportions, torso shape and overall body composition. Each β-coefficient represents a distinct dimension of shape variation, for example controlling leg length, hip circumference or other morphological features. SMPL provides gender-specific models, which capture typical shape differences between male and female bodies, as well as a gender-neutral model.

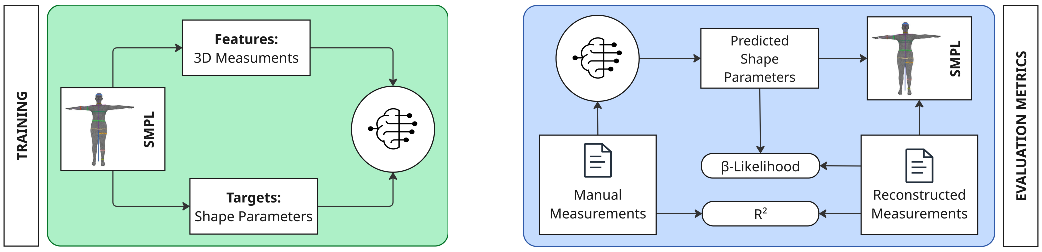
Anthropometric measurements of SMPL body models are calculated using the SMPL-Anthropometry repository [1]. The repository defines two types of measurements, lengths and circumferences, that are computed differently. Both approaches use predefined landmarks, whose exact spatial positions are determined using the SMPL shape parameters. Length measurements are computed by summing the Euclidean distances between defined landmarks, while circumference measurements are obtained by slicing the mesh at a defined landmark and summing the distances along the convex hull of the resulting intersection points.

Training Dataset This study utilizes the Avatars in Geography Optimized for Regression Analysis (AGORA) dataset [9], which provides SMPL shape parameters based on a neutral body model, fitted to high-quality 3D human scans. The dataset includes 3187 subjects and captures a diverse range of body shapes, each represented by 10 shape parameters. Using the SMPL shape from the AGORA dataset, 3D body models are generated, and anthropometric measurements are computed. The resulting dataset, consisting of SMPL shape parameters paired with their derived anthropometric measurements, serves as the foundation for model training (see Figure 1a). As the measurements are derived directly from the parametric model, they are free from manual measurement noise, ensuring a consistent and deterministic relationship between measurements and shape parameters.

Validation For external validation, the anthropometric datasets Anthropometric Survey of US Army Personnel (ANSUR) [3] and Study of Health in Pomerania (SHIP) [2] are used. ANSUR contains 93 anthropometric measurements collected from over 6000 adult US military personnel, collected under real-world measurement conditions, including potential measurement errors. As most participants are subject to military recruitment standards, the dataset does not fully represent the diversity of body types in the general population. The SHIP dataset originates from a study in which 2313 full-body scans were collected and processed to derive an anthropometric dataset representing the German working-age population, containing 39 anthropometric measurements.

For the present study a set of 16 commonly used anthropometric measurements was identified as a suitable starting point. Future work will focus on determining which subset of measurements is truly essential and most effective for this task. Since the real-world anthropometric data available in SHIP and ANSUR datasets do not perfectly match the predefined set in terms of landmark positions, closely related measurements were selected where necessary. Accordingly, the landmarks used for measurement computation on the SMPL model were adapted to match the ANSUR measurement definitions. Note that the SHIP dataset includes only 11 of the 16 measurements, as corresponding measurements for the remaining five are not available.

3 Experimental Setup and Evaluation



(a) The regression models are trained using SMPL body measurements as input features and the corresponding SMPL shape parameters as target outputs.

(b) The predicted shape parameters are directly used to assess plausibility, while for metric accuracy they are first used to reconstruct the corresponding body measurements, from which R^2 is computed.

Figure 1: Overview of the model training and evaluation process, showing how SMPL body measurements are used to predict shape parameters and how these predictions are evaluated in terms of plausibility and metric accuracy.

Evaluation Metrics Model performance is evaluated using two complementary metrics. The coefficient of determination (R^2) measures the accuracy of reconstructed body measurements, while β-likelihood quantifies the plausibility of generated body shapes. As depicted in Figure 1b, R^2 is computed by reconstructing body measurements from predicted shape parameters and calculating a weighted average across measurements, with weights proportional to each measurement's variance. More precisely, if R_j^2 denotes the coefficient of determination for measurement j and σ_j^2 its variance, the reported score is computed as

$$R_w^2 = \frac{\sum_{j=1}^{M} \sigma_j^2 R_j^2}{\sum_{j=1}^{M} \sigma_j^2},$$

where M is the number of evaluated measurements. This ensures that errors in measurements with greater natural variability, such as waist circumference, contribute more to the overall metric than measurements with lower variability, like wrist circumference. To address the problem of generating unrealistic body shapes, we propose using β-likelihood to quantify the plausibility of predicted shapes. Using the SMPL shape parameters in the AGORA dataset, a multivariate distribution is estimated by computing the mean vector and covariance matrix of the β-coefficients. The likelihood of a given shape is computed relative to this distribution using multivariate normal log-density. For a predicted shape-parameter vector $\hat{\beta}$ we report the log-likelihood

$$\log p(\hat{\beta}) = -\frac{1}{2}\left[(\hat{\beta}-\mu)^\top \Sigma^{-1}(\hat{\beta}-\mu) + d\log(2\pi) + \log|\Sigma|\right],$$

where μ and Σ are the mean vector and covariance matrix estimated from AGORA, and d is the dimensionality of the shape space.

Model Training Multiple regression models, namely Linear Regression, Ridge Regression, Elastic Net, K-Nearest-Neighbors (KNN), Random Forest and a Multilayer Perceptron (MLP), were trained on the AGORA dataset, using anthropometric measurements as input features and the corresponding SMPL shape parameters as target outputs (see Figure 1a). Hyperparameter tuning was performed to jointly optimize both the coefficient of determination (R^2) and log-β-likelihood on the validation datasets. The hyperparameters were selected from predefined search spaces, using the Optuna Multi-Objective TPE Sampler [7]. For each model class, optimization was performed over 20 trials. As the measurement sets of SHIP and ANSUR differ, hyperparameters were optimized separately for each dataset.

4 Results and Conclusion

Results A Pareto front analysis was conducted to compare the different model classes, based on their performance on the SHIP and ANSUR validation datasets. In Figure 2, Pareto-optimal points

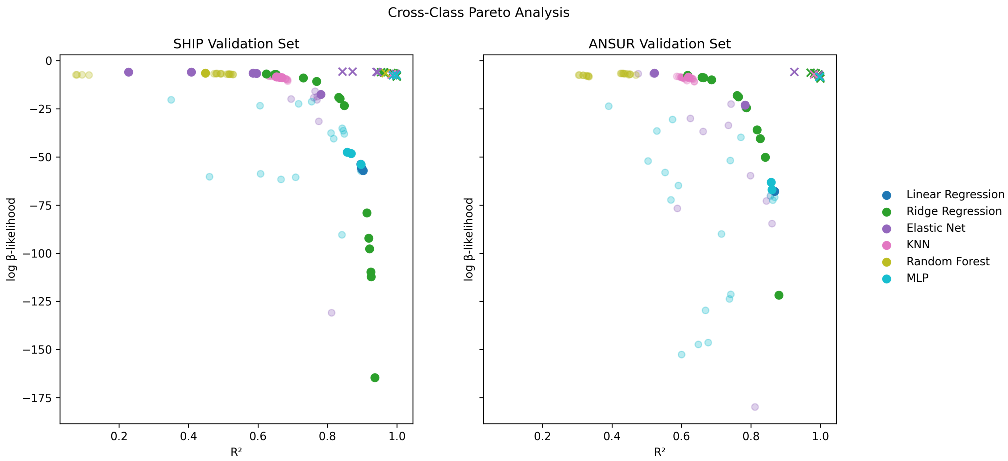


Figure 2: Pareto front analysis of different model classes, evaluated on the SHIP and ANSUR validation datasets. Each point corresponds to one hyperparameter configuration of the respective model class. Pareto-optimal points are highlighted, while the remaining validation results are shown in faded colors. Crosses indicate the corresponding results on a validation split of the AGORA training dataset. The figure illustrates the trade-off between R^2 and β-likelihood on the validation datasets under real-world measurement conditions.

are highlighted, while the remaining validation results are shown in faded colors. Crosses indicate the corresponding results on an AGORA validation split. The analysis reveals a clear trade-off between the two evaluation metrics for the validation datasets. Higher metric accuracy is associated with less plausible body shapes, and vice versa. However, this pattern is not observed for the validation split of the AGORA training dataset, suggesting that it is mainly caused by dataset shift between deterministic SMPL-derived measurements and real-world anthropometric measurements. KNN and Random Forest models produce shapes with better β-likelihood, indicating greater plausibility, but at the expense of metric accuracy. In contrast, linear regression-based models like Ridge Regression and Elastic Net achieve the highest metric accuracy but tend to generate less plausible body shapes. Notably, Ridge Regression demonstrates that the trade-off between metric accuracy and shape plausibility can be controlled by tuning the regularization parameter, which pulls the predicted shape parameters closer to zero.

Conclusion In this work, the AGORA dataset, which provides SMPL shape parameters, was used to generate a dataset consisting of paired SMPL shape parameters and corresponding measurements. This dataset served to train multiple regression models, where anthropometric measurements were used as input features and the associated SMPL shape parameters were predicted as outputs. In addition, the dataset was used to estimate a multivariate distribution that forms the basis of the proposed β-likelihood metric for quantifying the plausibility of generated body shapes. For validation, additional anthropometric datasets were incorporated to assess model performance under real-world measurement conditions including potential measurement errors. While both high metric accuracy and plausible body shapes can be achieved on a validation split of the training dataset, evaluation on the external validation datasets reveals a clear trade-off between metric accuracy and the plausibility of the generated body shapes. In this context, linear regression models with regularization are particularly suitable, as the trade-off can be controlled by tuning the regularization parameter. Future work will focus on training models whose loss function optimizes the error of reconstructed measurements, rather than the shape parameters, addressing a key limitation of the current study. To simplify the measurement process in a clinical setting, essential measurements will be identified to define a minimal set of necessary inputs, and strategies will be explored to handle missing measurements.

Acknowledgments This work was funded by the FFG (Austrian Research Promotion Agency) under the grant 913010 (Embodied Perceptions). This project is also financed by research subsidies granted by the government of Upper Austria. RISC Software GmbH is Member of UAR (Upper Austrian Research) Innovation Network. This work is part of a Master's thesis at the University of Applied Sciences Upper Austria.

References

[1] David Bojanić. SMPL-Anthropometry. https://github.com/DavidBoja/SMPL-Anthropometry, February 2026.

[2] Dominik Bonin, Alexander Ackermann, Dörte Radke, Markus Peters, and Sascha Wischniewski. Anthropometric dataset for the German working-age population using 3D body scans from a regional epidemiological health study and a weighting algorithm. *Ergonomics*, 66(8):1057–1071, August 2023.

[3] C. C. Gordon, C. L. Blackwell, B. Bradtmiller, Joseph L. Parham, Patricia Barrientos, S. Paquette, B. Corner, J. Carson, Joseph Venezia, Belva M. Rockwell, M. Mucher, and S. Kristensen. 2012 Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics, December 2014.

[4] Muhammed Kocabas, Nikos Athanasiou, and Michael J. Black. VIBE: Video Inference for Human Body Pose and Shape Estimation, April 2020.

[5] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A Skinned Multi-Person Linear Model. In Mary C. Whitton, editor, *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 851–866. ACM, New York, NY, USA, 1 edition, August 2023.

[6] Katja Ludwig, Julian Lorenz, Daniel Kienzle, Tuan Bui, and Rainer Lienhart. Leveraging Anthropometric Measurements to Improve Human Mesh Estimation and Ensure Consistent Body Shapes, April 2025.

[7] Yoshihiko Ozaki, Yuki Tanigaki, Shuhei Watanabe, Masahiro Nomura, and Masaki Onishi. Multiobjective Tree-Structured Parzen Estimator. *Journal of Artificial Intelligence Research*, 73:1209–1250, April 2022.

[8] Daryousch Parvizi, Lars-Peter Kamolz, Michael Giretzlehner, Herbert L. Haller, Maria Trop, Harald Selig, Peter Nagele, and David B. Lumenta. The potential impact of wrong TBSA estimations on fluid resuscitation in patients suffering from burns: Things to keep in mind. *Burns*, 40(2):241–245, March 2014.

[9] Priyanka Patel, Chun-Hao P. Huang, Joachim Tesch, David T. Hoffmann, Shashank Tripathi, and Michael J. Black. AGORA: Avatars in Geography Optimized for Regression Analysis, April 2021.

[10] Sergi Pujades, Betty Mohler, Anne Thaler, Joachim Tesch, Naureen Mahmood, Nikolas Hesse, Heinrich H. Bülthoff, and Michael J. Black. The Virtual Caliper: Rapid Creation of Metrically Accurate Avatars from 3D Measurements. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1887–1897, May 2019.

[11] Akash Sengupta, Ignas Budvytis, and Roberto Cipolla. Probabilistic Estimation of 3D Human Shape and Pose with a Semantic Local Parametric Model, November 2021.

[12] K.V. Stein, T.E. Dorner, W. Ilias, and A. Rieder. Schmerzpatienten und ihre Erwartungen an die ärztliche Versorgung. *Der Schmerz*, 24(5):468–473, September 2010.

[13] Moyu Wang and Qingping Yang. From prediction to measurement, an efficient method for digital human model obtainment. *International Journal of Metrology and Quality Engineering*, 15:1, 2024.