ISSN 1213-6972 (print)
ISSN 1213-6964 (online)
Journal of WSCG
http://www.wscg.eu
Vol.33, No-1-2, 2025

Metrically Accurate 3D Human Avatars from Silhouette Images

Martin Halaj
Comenius University
Faculty of Mathematics, Physics and Informatics
Bratislava, Slovakia
martin.halaj@fmph.uniba.sk

Dana Skorvankova
Skeletex Research
Bratislava, Slovakia
skorvankova@skeletex.xyz

Martin Madaras
Comenius University
Faculty of Mathematics, Physics and Informatics
Bratislava, Slovakia
martin.madaras@fmph.uniba.sk

ABSTRACT

In recent years, the demand for realistic human avatars has escalated across diverse industries, ranging from gaming and virtual or augmented reality to fashion and healthcare. Creating an accurate, lifelike virtual duplicate of a human subject requires a precise reconstruction of anthropometric measurements from the real body to the rendered body model. In this paper, we propose a novel pipeline for 3D human body model reconstruction from a set of two input images. Our method generates metrically accurate virtual human avatars, based on body measurements extracted from an input image. Our approach is capable of parameterizing the generated mesh with body measurements and yields accurate results. To augment the metrically accurate body meshes, we introduce a pipeline for generating textured clothes to enhance user virtual experience. As part of our work, we generated a synthetic dataset that serves as a foundation for further enhancing the training process by extracting anthropometric measurements from an input photo. This dataset contains over one million files derived from 120,000 virtual avatars that can extend existing real datasets.

Keywords

3D Human Avatars, Mesh Generation, Body Measurements

1 INTRODUCTION

Deep learning excels at solving diverse tasks in computer science, but its effectiveness is based on large amounts of training data. Although some domains have abundant data, others, such as X-rays or human images, face challenges in collection and distribution. This issue can be partially addressed by generating artificial data to replace missing real samples.

One compelling application of synthetic data lies in the creation of realistic digital humans. As virtual environments become more immersive, there is an increasing demand for lifelike avatars, especially in fields such as gaming, virtual reality, and film production. These domains require not only photorealistic graphics, but also anatomically accurate human representations that can reflect individual user characteristics.

The popularity of movie effects, animations, and video games continues to grow steadily. Over the past decade, significant advancements have been made in terms of visual realism, largely driven by the enhanced performance of graphics cards and computers overall. Today, video games and animations boast synthetic environments that are nearly indistinguishable from reality, complete with sophisticated weather simulations and lifelike portrayals of natural phenomena.

Central to these immersive experiences are the characters within them. Thanks to improvements in computer graphics, there are new ways to deepen the user's experience in a virtual environment. One such method involves integrating a lifelike replica of the user directly into the game or animation, fostering stronger emotional connections with the virtual experience. The cornerstone of creating these realistic virtual personas lies in accurately visualizing the human body, including precise body measurements. Anthropometry, as a scientific discipline, offers invaluable insight to tackle this challenge. However, despite the potential benefits, there are currently only a few tools available to generate human body replicas based on anthropometric measurements, and those available often suffer from limited versatility and accuracy [Nguyen2021, Zeng2018]. While an accurate representation of the body is crucial, a bare form lacking clothes, texture, and other details would fail to enhance the user experience. Incorporating these elements is

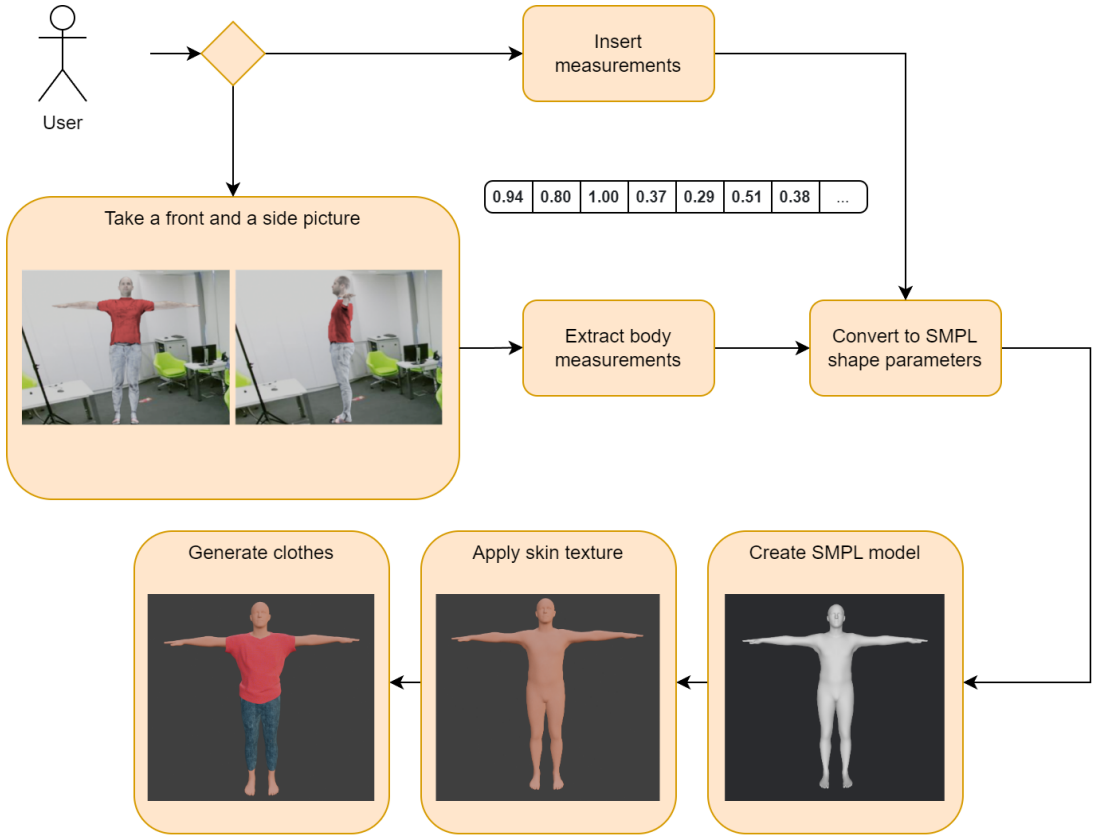

Figure 1: The diagram of our proposed pipeline for creating a virtual avatar of a real person.

essential to enrich the final experience. Therefore, we introduce a pipeline to create a virtual avatar based on a real user, as illustrated in Figure 1.

The main contributions of our paper can be summarized as follows:

1. We address the growing demand for realistic human avatars across diverse industries by establishing a novel pipeline for generating antropometrically accurate 3D human avatars, based on a set of two input images.
2. Our method is capable of precisely reconstructing a 3D body model from a set of 16 body measurements that collectively define the overall shape of the body. This is achieved by learning the mapping between body measurements and shape coefficients of the parameterized human body model.
3. We propose an approach to generate textured clothing for virtual avatars, thus enhancing the visual realism of body models.
4. We created and shared for research purposes an annotated synthetic dataset containing over one million files, serving as a valuable resource for augmenting real datasets in anthropometric measurement estimation.

2 RELATED WORK

In this section, we closely look at existing works related to our research. We divided them into several parts, considering the corresponding areas.

2.1 Parametrized 3D Avatars

Over the last decade, several different models have been proposed to parameterize the human body. One of the most popular human body models available is the Skinned Multi-Person Linear (SMPL) model [Loper2015]. The base SMPL model is the first one in the SMPL family. The more advanced models based on SMPL are SMPL+H [Romero2017], which includes a fully articulated hand model; SMPL-X [Pavlakos2019], which further extends the model by introducing face expressions; and STAR model [Osman2020], which incorporates pose-corrective deformations dependent on body shape. The SMPL models are parameterized with T, β, θ and δ, which correspond to the template mesh, shape parameters, pose parameters and the soft tissue control parameter, respectively. The goal of parametric human body models is to create realistic human bodies, including the natural deformation of the body or simulating the plausible motion of soft tissue. Another motivation was to create a model that can be quickly rendered without the need for manual intervention and is fully compatible with existing rendering

engines or graphics tools. The resulting model is easy to animate or control. For our purposes, we focus on high-level model definition with an emphasis on the practical usability for our task.

2.2 Human Mesh Reconstruction

In recent years, there has been a surge in research focused on 3D human mesh reconstruction from images [Choutas2020, Feng2021, Lin2023], driven by the increasing demand for realistic human avatars in various applications such as virtual reality, games, and healthcare.

Deep learning-based approaches have gained significant attention in the field of 3D human mesh reconstruction due to their ability to learn complex mappings directly from data. These methods leverage large-scale datasets of 3D human scans or images to train neural networks to directly predict 3D mesh geometry. Convolutional Neural Networks (CNNs) have been widely employed for 3D mesh reconstruction from 2D images [Huang2017, Kanazawa2018, Pavlakos2018, Varol2018]. These methods typically consist of two stages: feature extraction and mesh generation. In the feature extraction stage, the CNN processes the input image to extract high-level features that capture important visual cues related to human anatomy. These features are then used as input to the mesh generation stage, where a decoder network generates the corresponding 3D mesh geometry. Nonetheless, the meshes constructed in most of these methods either lack the underlying skeleton structure of a human body, therefore it is not trivial to further use them for animation; or do not use body measurements to reproduce anthropometrically accurate body models. The Virtual Caliper [Pujades2019] generates an SMPL model from input anthropometric measurements; however, the measurements need to be inserted manually into the program or calibrated using virtual reality headset with controllers.

2.3 Body Measurements

Creating a virtual metrically accurate copy of a person is not a trivial task. However, there is an increasing demand for this type of technology. One of the many applications is the virtual dressing room, where a accurately reproduced body shape is crucial. Anthropometry is a scientific field that studies various physical attributes of the human body. While several analytical methods have been proposed over the years [Albances2019, Gan2020, Guillo2020, Song2017], they often lack robustness or suffer from low precision.

In recent years, numerous learning-based approaches have been proposed to infer anthropometric body measurements automatically from a visual input [Gonzalez2021, Yan2020, Yan2021]. The recent research in this field shows that body measurements can be effectively estimated from an input image even using simple CNN-based models.

2.4 Synthetic Human Data

Incorporating synthetic data into neural network training is a commonly followed strategy to enhance the available real data and improve the generalizability of the model. The utilization of synthetic data presents several advantages in this context. Recognizing the importance of abundant training data for ensuring model accuracy and reliability, the integration of synthetic data becomes particularly valuable in scenarios where data scarcity or logistical challenges impede the acquisition of sufficiently large datasets.

SURREAL [Varol2017] is a wide-ranging dataset that consists of synthetic data. It includes more than six million images, including ground truth positions, depth maps, and segmentation information for each frame. To represent the person in the image, the dataset uses the base SMPL model. The generation pipeline consists of several steps. The first step is to render an SMPL avatar using random pose and shape parameters. To obtain a random shape, a human subject is randomly selected from the real CAESAR dataset [Robinette2002] and the corresponding body shape is approximated using SMPL, while also performing a similar process with the pose parameters chosen from the motion capture dataset. Next, a random texture is applied to the avatar. Finally, images from two viewpoints are rendered with and without a realistic background.

Evaluation of a CNN model on these datasets showed that the best results were achieved when the real training data were enhanced by the synthetically generated samples. The results suggest that augmenting the training process using synthetic data improves the resulting accuracy of the neural network.

SURREACT [Varol2021] shifts the focus toward action recognition, specifically targeting the challenge of recognizing human actions from perspectives or points of view that are not represented in training data. This approach involves synthesizing human action videos from various viewpoints to enhance the robustness and generalization capability of action recognition models.

2.5 Texturing Avatars

Incorporating textures into 3D human avatars is essential to enhance their visual realism and perceptual fidelity in virtual environments. Texturing plays a crucial role in simulating surface details such as skin tone, clothing patterns, and other intricate features, which are fundamental for creating lifelike representations of human characters.

SMPLitex [Cassas2023] is a tool that creates a texture map for the SMPL model from an input photo. It extends generative models for 2D images into the 3D domain by establishing pixel-to-surface correspondences computed from the input image. Initially, a generative model for complete 3D human appearance is trained, then fitted into the input image by conditioning the model to the visible subject parts. In addition to generating texture from the input image, the framework is capable of generating texture maps from text prompts. Nevertheless, the model demands significant hardware resources, particularly in terms of computational power. Moreover, the quality of the produced textures may be lacking in specific instances.

In comparison to SMPLitex, HOOD [Grigorev2023] is focused on generating dynamic clothes for 3D avatars. The core functionality of the approach is based on graph neural networks. A primary limitation of the approach is its restriction to a single garment per 3D avatar. For example, while the avatar can wear a shirt, it cannot simultaneously wear trousers.

3 OUR APPROACH

In this section, we introduce our proposed pipeline for generating realistic and metrically accurate 3D human avatars from a set of two input images. Each stage of the pipeline is described in detail.

3.1 Data Acquisition

As already mentioned, synthetic data have proven to be beneficial for training neural networks [Varol2017, Varol2021]. In our work, we also focused on generating artificial data to improve the training process of learning-based methods.

For this purpose, we have established a data generator. Since strategies such as SURREAL [Varol2017] or SURREACT [Varol2021] come with prepared data used to generate synthetic outputs, such as textures and image backgrounds; we followed their work by adjusting the existing SURREACT framework.

Figure 2 depicts an example set for one avatar. The set consists of twelve images: frontal and lateral images with a background, frontal and lateral images without a background, a mesh of the corresponding avatar, and a set of joints in 3D space. The same pictures are available for both the T-pose and the A-pose. Moreover, anthropometric measurements and beta parameters employed in the construction of each avatar are available.

Subjects used for generation were instanced using random parameters for the SMPL model. Therefore, we have subjects of various shapes that cover the variability of plausible human body shapes.

Figure 3 presents a histogram that provides an overview of the BMI distribution of the avatars generated and the

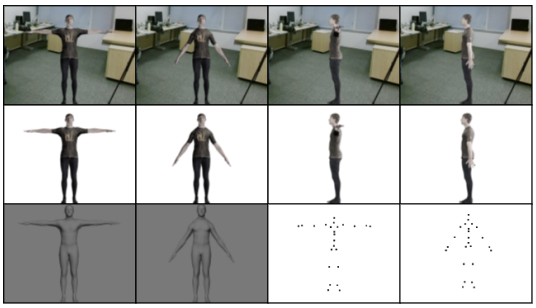

Figure 2: Example of a set generated with the referenced modified generator.

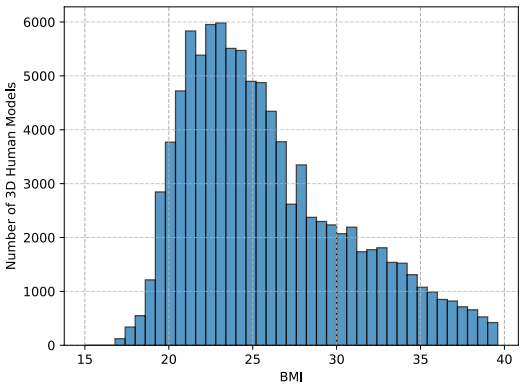

Figure 3: Overview of BMI values between 15 and 40 in the training portion of the generated dataset.

number of samples for each BMI value. The BMI was calculated based on the volume of avatar mesh, multiplied by the average human body density of 985 kg/m^3.

In summary, we have generated over one million files with a total size of more than 100 GB corresponding to 120,000 different virtual subjects. The examples of randomly generated subjects are shown in Figure 4.

3.2 Body Measurements Extraction

To train deep learning models, annotated data is essential. In our case, this means providing ground truth labels of anthropometric body measurements for the generated synthetic human data. To obtain these labels, we use the skeleton-driven annotation pipeline [Skorvankova2022] on the generated meshes. It uses the skeleton joint locations to measure specific anthropometric data on the human mesh in a T-pose. Using this tool, we generate 16 anthropometric measurements for each subject within the generated synthetic dataset. Later, we used them as ground truth to evaluate the synthetic data.

We used the generated silhouette images and the corresponding measurement labels to train a neural model to automatically predict body measurements from a set of two binary images, capturing a human subject from a frontal and lateral view. Silhouette images were created

Figure 4: Example of dataset generated with the referenced modified generator, using varied lighting.

from background-free generated images. The proposed model is described in detail in Section 4.3.

3.3 Anthropometric Avatar Generator

As illustrated in Figure 1, after acquiring anthropometric measurements by extracting them from an input image or inserting them manually, we want to use them to parameterize the SMPL model. Therefore, we need to convert them to SMPL shape parameters. As suggested in previous research [Pujades2019], there could be a linear relation between body measurements and shape parameters used by the SMPL model. Therefore, we have experimented with various types of regression models, such as linear and polynomial regression.

Firstly, it is vital to understand that the foundation for training any model is an extensive training dataset. Accordingly, the input parameters for each avatar have been generated together with other previously referenced output data. Throughout our work, we use SMPL models with 10 shape parameters, since they are sufficient to create a plausible human model.

In this way, we created a dataset with a total of 120,000 data pairs. One pair consists of a set of anthropometric measurements and the ground truth SMPL shape parameters.

3.4 Augmenting Avatar Meshes

After successfully obtaining the user's body measurements and translating them into SMPL parameters, allowing us to generate a body mesh using the SMPL model, we still need to enhance the model by adding colors, textures, and clothing to fully engage users in a virtual environment.

3.4.1 Clothes Generator

Considering the limitations highlighted in SMPLitex and HOOD, we opted to introduce an alternative approach for straightforward avatar texturing and garment creation, utilizing Blender [1] as the cornerstone of our pipeline.

We employ procedural materials supported by Blender for skin texture generation. This method offers several benefits, including the ability to easily adjust the skin tone, thereby enabling us to maintain a wide range of variability.

We explored two approaches to creating clothes for 3D avatars, one of which is the *sewing* method. This technique simulates sewing by creating a plane adjusted to the avatar's dimensions, cutting out sleeves and a head opening, and extruding it to cover the body. The front and back planes remain connected by edges at the seams, acting as springs that pull them together. Blender's built-in cloth simulation then enables dynamic interactions. Although effective, this method struggles to generalize to avatars with varying waist circumferences, potentially leaving gaps at the sides. Scaling the planes to match body volume helps, but does not fully ensure a perfect fit.

Our second approach, the *extruding* method, addresses the limitations of the sewing technique. Here, specific face indices of the SMPL model are extruded to form clothing. Since all SMPL models share the same face indices regardless of shape or gender, we can pre-define selections for different types of clothing. The selected faces are duplicated, upscaled, and simulated as cloth,

[1] www.blender.org

Figure 5: Result of textured avatars with clothes created using extruding method.

ensuring that the clothes conform to the avatar's body shape. This method simplifies extending the range of supported clothing. Figure 5 illustrates avatars dressed using this approach.

3.4.2 Adding Textures

We described the process of creating clothes for virtual avatars. However, clothes require the application of textures to make them more realistic. Many different databases store various materials that we can use for clothes generation in Blender. We have decided to use BlenderKit [2], specifically some of the included procedural materials, as it is straightforward to further adjust their color and properties.

4 EXPERIMENTS AND EVALUATION

Here, we describe the conducted experiments and evaluate the proposed approaches.

4.1 Real Data

While there is a lack of available labeled datasets suitable for evaluating our pipeline, a real dataset annotated with ground truth body measurements has been published in a recent study [Ruiz2022]. The BodyM dataset consists of frontal and lateral silhouette images of real human subjects. A total of 8,978 samples were created by capturing 2,505 subjects, with 16 body measurements annotated for each subject. Although the image data captures real humans, the measurement annotations were obtained not manually, but by fitting 3D body scans to a mesh and measuring it virtually.

Unfortunately, this dataset has several disadvantages. Most importantly, the authors do not provide any definitions of how the measurements were calculated or the exact locations of the landmarks used to extract the measurements on the human body. They provide only labels for individual measurements. Additionally, the original RGB images of participants are not available due to privacy reasons.

Another drawback is the size of the dataset. It includes 6,134 training samples, consisting of frontal and lateral silhouette images, which may not represent a sufficient number of samples to train a neural network to generalize well on unseen examples. The remaining images in the dataset are divided into test set A ($Test_A$) and test set B ($Test_B$). Test set A contains images captured in a laboratory environment, avoiding issues such as various camera angles and inaccurate foreground segmentation. In contrast, test set B is captured in a home-like environment with various camera settings.

Based on these shortcomings, we decided to further extend this dataset for training purposes using synthetically generated data.

[2] www.blenderkit.com

Training data	Test data	MAE (cm) ↓	mAP@5 (%) ↑	Batch size
Real	$Test_B$	3.379	78.68	32
Synthetic + Real	$Test_B$	**3.114**	**80.75**	32
Real	$Test_B$	3.325	79.26	64
Synthetic + Real	$Test_B$	3.176	80.28	64

Table 1: Test results of the base model showing the mean absolute error (MAE) in cm, mean average precision (mAP) at the 5 cm threshold, and the size of the used batch.

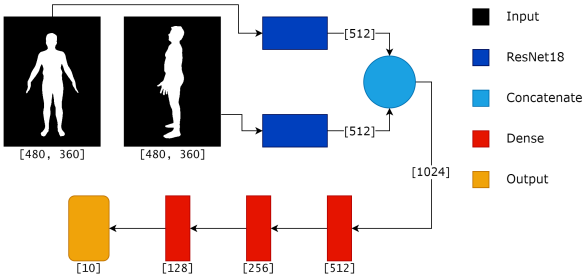

Figure 6: Structure of our base CNN model for anthropometric measurements estimation.

4.2 Synthetic Data

As mentioned before, we extended the BodyM dataset with synthetic data to improve the capabilities of a convolutional neural network that estimates anthropometric measurements from input images. To evaluate the benefit of the synthetic data, we provide test results using exclusively real training data vs. adding synthetic data to the training set.

In order to mix the synthetic data with the real training samples, we need to unify them. We resize all images to the same dimensions. Both the synthetic images and the images from the BodyM dataset have 16 measurements each; however, the measurements differ between the two sets. The intersection of these two sets includes the following 10 parameters: 1. chest circumference, 2. waist circumference, 3. bicep circumference, 4. thigh circumference, 5. arm length, 6. leg length, 7. calf length, 8. wrist circumference, 9. shoulder width, 10. height.

4.3 Body Measurements Estimator

Figure 6 illustrates the structure of the convolutional neural network that we propose for the estimation of body measurements. The network is based on two ResNet18 blocks [He2016] to maintain a general structure and avoid any concerns about modifications for demonstration purposes. We aim to keep the structure of the neural network as general as possible. The network processes two silhouette images (front and side views) as inputs. These images are processed by ResNet18, concatenated, and then passed through three dense layers. Finally, the network produces a set of 10 output measurements.

We used real data and a mixture of real and synthetic data as training datasets. The mixed training dataset contains 6,134 real images, which we augmented with an equal amount of synthetic data to preserve balance between the two sets. For evaluation, we exclusively used the test set B ($Test_B$) from the BodyM dataset, considering its higher image variability compared to the test set A. Test set B also contains various camera settings while capturing human subjects.

We conducted experiments with two distinct networks. While the first model (marked as *base model*) takes only two silhouette images as input, the second model (marked as *height-driven model*) also includes a person's overall height as an additional parameter (as illustrated in Figure 7). This approach sets the scale of the input images, considering that the test images may be captured by various cameras with different parameters and positioned at varying distances from the human subject. We consider a person's height a basic parameter that most people are aware of; thus, it should not require any additional measuring. Both models were implemented in the PyTorch framework. The mean absolute error loss function was used in both cases.

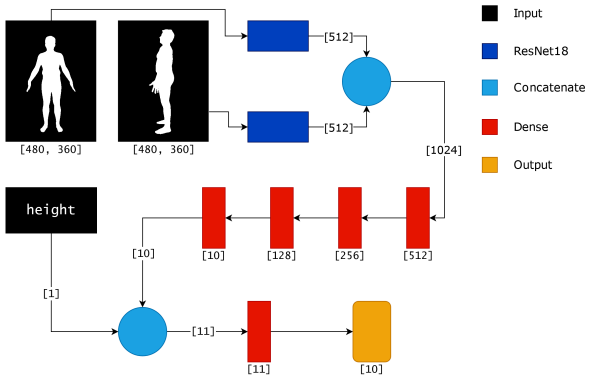

Figure 7: Structure of our height-driven CNN model for anthropometric measurements estimation.

In Table 1, we observe that adding synthetic data to the training process enhanced the capabilities of the body measurement estimator. Additionally, we provide per-measurement metrics (Table 2), including the mean absolute error (MAE) in centimeters and the mean absolute percentage error (MAPE). Moreover, incorporating information on the height of a person further improved accuracy substantially, as demonstrated in Table 3. Fi-

Body measurements	Real data		Synthetic + Real data	
	MAE (cm) ↓	MAPE (%) ↓	MAE (cm) ↓	MAPE (%) ↓
chest circumference	4.876	4.799	**4.432**	**4.340**
waist circumference	5.586	6.223	**5.035**	**5.819**
bicep circumference	1.888	6.119	**1.723**	**5.596**
thigh circumference	3.276	6.106	**2.786**	**5.063**
arm length	**2.434**	**5.116**	2.508	5.272
leg length	3.804	5.045	**3.774**	**5.011**
calf length	**1.993**	**5.263**	2.119	5.613
wrist circumference	1.006	6.277	**0.836**	**4.991**
shoulder width	1.642	4.629	**1.455**	**4.129**
height	6.743	4.056	**6.468**	**3.883**
mean	3.325	5.363	**3.114**	**4.972**

Table 2: Performance comparison of the base model trained on real data only versus a model trained on a mixture of real and synthetic data. For consistency, the real-data model used a batch size of 64, and the mixed-data model used a batch size of 32. Mean absolute error (MAE) and mean absolute percentage error (MAPE) are reported.

nally, Table 4 compares the per-measurement errors of both proposed models trained on a mixture of real and synthetic data. The evaluation metrics are shown for the model trained for 40 epochs, since additional training iterations did not substantially improve the final error.

Training data	MAE (cm) ↓	mAP@5 (%) ↑
Real	2.085	90.64
Synthetic + Real	**1.888**	**92.86**

Table 3: Test results of the updated height-driven model showing the mean absolute error (MAE) in cm and mean average precision (mAP) at the 5 cm threshold. Batch size was set to 32 samples.

4.4 Avatar Generation

We tested several machine learning models for the purpose of learning the mapping between the set of explicit body measurements and the SMPL shape parameters.

As seen in Table 6, we managed to achieve the best results using fourth-degree polynomial regression. Aside from polynomial regression, good results were also achieved using a random forest regressor with 200 estimators and a k-nearest neighbors model with 10 neighbors. We have performed experiments with different neural network benchmark architectures, but the accuracy did not outperform the models mentioned.

Our final model is based on second-degree polynomial regression and can take up to 16 input anthropometric measurements and produce the corresponding 10 SMPL shape parameters.

We performed two experiments. In the first (as shown in Table 5), we predict the SMPL shape parameters using five basic anthropometric measurements as follows: 1. height, 2. chest circumference, 3. inside leg height, 4. shoulder breadth, 5. waist circumference.

In the second experiment (shown in Table 6), we utilized a total of 16 available measurements. As depicted in the tables, polynomial regression outperforms other models when using all measurements, while random forest performs better when using only the basic subset of measurements.

5 CONCLUSIONS

In this paper, we develop a pipeline to obtain a metrically accurate 3D human avatar from frontal and lateral input images to address the demand for realistic virtual avatars in various applications. We have shown that we are capable of creating highly accurate virtual bodies based on input silhouette images.

As part of our pipeline, we designed a tool to effectively map explicit body measurements extracted from an input image to SMPL shape parameters to define the body shape of an SMPL mesh. Based on our experiments, we assume that the relationship between the SMPL shape parameters and the anthropometric body measurements is quadratic. Additionally, we propose two distinct approaches to augment the virtual bodies with clothing and textures to enhance the visual experience. To improve visual realism, we extract the original color of the clothes by segmenting the input RGB image.

Our contributions also include the acquisition of a large-scale dataset of synthetic human bodies, including rendered images and 3D meshes annotated with a set of the most commonly used body measurements.

We have experimented with estimating the body measurements from exclusively visual inputs and incorporating explicit prior knowledge of the person's height along with the input images. Although the accuracy of our system is already relatively high using just the visual inputs, this can be an issue when capturing human subjects in different scenes, using various camera settings, and varying the distance of the camera from the

Body measurements	Base model		Height-driven model	
	MAE (cm) ↓	MAPE (%) ↓	MAE (cm) ↓	MAPE (%) ↓
chest circumference	4.432	4.340	3.627	3.666
waist circumference	5.035	5.819	3.812	4.420
bicep circumference	1.723	5.596	1.367	4.520
thigh circumference	2.786	5.063	2.213	3.993
arm length	2.508	5.272	1.079	2.246
leg length	3.774	5.011	1.518	1.962
calf length	2.119	5.613	1.549	4.065
wrist circumference	0.836	4.991	0.660	3.955
shoulder width	1.455	4.129	1.167	3.399
MAE (cm)	2.741		**1.888**	
mAP@5 (%)	84.56		**92.86**	

Table 4: Test results of the base model and the height-driven model. Both models are trained on a mixture of real and synthetic data. Mean absolute error (MAE) and mean absolute percentage error (MAPE) are reported.

Metric	Pol. regression (4th)	Random Forest	K-Neighbors
Mean squared error (betas)	0.5352	**0.0930**	0.1015
Mean absolute error (betas)	0.5340	**0.1178**	0.1184

Table 5: Test results of estimating SMPL shape parameters from basic body measurements.

Metric	Pol. regression (4th)	Random Forest	K-Neighbors
Mean squared error (betas)	**0.0004**	0.0539	0.0618
Mean absolute error (betas)	**0.0088**	0.0873	0.0922

Table 6: Test results of estimating SMPL shape parameters from all body measurements.

captured subject. To include an indication of the scale, we suggest explicitly passing the person's height as an additional input to the network, since it is an easily acquirerable parameter, usually known to the subject.

In deployment, another potential workaround could be to determine the scale of the synthetically generated images by measuring the ratio between the ground truth height of the person and the pixel distance within the image area corresponding to the same stature. The obtained scale could then be utilized to rescale real input images based on the person's height to address the issues related to varying camera distances by matching the scale of the synthetic training images.

6 REFERENCES

[Nguyen2021] Nguyen, D. and Hoang, T. 3D Reconstruction Human Body From Anthropometric Measurements Using Diversity Control Oriented Genetic Algorithm. Mendel, vol. 27, pp.49-57, 2021.

[Zeng2018] Zeng, Y., Fu, J. and Chao, H. 3D Human Body Reshaping with Anthropometric Modeling. Springer Singapore, Internet Multimedia Computing and Service, pp.96-107, 2018.

[Choutas2020] Choutas, V., Pavlakos, G., Bolkart, T., Tzionas, D. and Black, M. J. Monocular Expressive Body Regression Through Body-Driven Attention. Springer-Verlag, ECCV 2020: 16th European Conference, Glasgow, UK, Proceedings, Part X, pp.20-40, 2020.

[Feng2021] Feng, Y., Choutas, V., Bolkart, T., Tzionas, D. and Black, M. J. Collaborative Regression of Expressive Bodies using Moderation. International Conference on 3D Vision (3DV), pp.792-804, 2021.

[Lin2023] Lin, J., Zeng, A., Wang, H., Zhang, L. and Li, Y. One-Stage 3D Whole-Body Mesh Recovery With Component Aware Transformer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.21159-21168, 2023.

[Huang2017] Huang, H., Chao, W. and Guibas, L. J. Beyond Silhouettes: Surface Reconstruction Using Multi-View Consistency. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

[Kanazawa2018] Kanazawa, A., Black, M. J., Jacobs, D. W. and Malik, J. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[Pavlakos2018] Pavlakos, G., Zhu, L., Zhou, X. and Daniilidis, K. Learning to Estimate 3D Human

Pose and Shape from a Single Color Image. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[Varol2018] Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M. J., Laptev, I. and Schmid, C. BodyNet: Volumetric Inference of 3D Human Body Shapes. Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[Pujades2019] Pujades, S., Mohler, B., Thaler, A., Tesch, J., Mahmood, N., Hesse, N., Bülthoff, H. H. and Black, Michael J. The Virtual Caliper: Rapid Creation of Metrically Accurate Avatars from 3D Measurements. IEEE Transactions on Visualization and Computer Graphics, vol. 25, pp.1887–1897, 2019.

[Loper2015] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G. and Black, M. J. SMPL: A Skinned Multi-Person Linear Model. ACM, Transactions on Graphics, vol. 34, pp.248:1-248:16, 2015.

[Romero2017] Romero, J., Tzionas, D. and Black, M. J. Embodied Hands: Modeling and Capturing Hands and Bodies Together. ACM, Transactions on Graphics, vol. 36, pp.245:1-245:17, 2017.

[Pavlakos2019] Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A A, Tzionas, D. and Black, M. J. Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.10975-10985, 2019.

[Osman2020] Osman, A. A A, and Bolkart, T. and Black, M. J. STAR: A Sparse Trained Articulated Human Body Regressor. European Conference on Computer Vision (ECCV), pp.598-613, 2020.

[Yan2020] Yan, S., Wirta, J., and Kämäräinen, J. Anthropometric clothing measurements from 3D body scans. Machine Vision and Applications, vol. 31, 2020.

[Yan2021] Yan S. and Kämäräinen J. Learning Anthropometry from Rendered Humans. ArXiv, 2021.

[Gonzalez2021] Gonzalez Tejeda Y. and Mayer H. A. A Neural Anthropometer Learning from Body Dimensions Computed on Human 3D Meshes. IEEE Symposium Series on Computational Intelligence (SSCI), pp.1-8, 2021.

[Albances2019] Albances, X., Binungcal, D., Cabula, J. N., Cajayon, C., and Cabatuan, M. RGB-D Camera Based Anthropometric Measuring System for Barong Tagalog Tailoring. IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), 2019.

[Gan2020] Gan, X. Y., Ibrahim, H., and Ramli, D. A. A simple vision based anthropometric estimation system using webcam. IOP Publishing, Journal of Physics: Conference Series, vol. 1529, 2020.

[Guillo2020] Guillo, A., Azorin-Lopez, J., Saval-Calvo, M., Castillo-Zaragoza, J., Garcia-D'Urso, N., and Fisher, R. RGB-D-Based Framework to Acquire, Visualize and Measure the Human Body for Dietetic Treatments. Sensors, vol. 20, 2020.

[Song2017] Song, D., Tong, R., Chang, J., Wang, T., Du, J., Tang, M., and Zhang, J. Clothes Size Prediction from Dressed-Human Silhouettes. Lecture Notes in Computer Science, vol. 10582 LNCS, pp.86-98, 2017.

[Varol2017] Varol, G., Romero, J., Martin. X., Mahmood, N., Black, J. M., Laptev I. and Schmid, C. Learning from Synthetic Humans. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.4627-4635, 2017.

[Robinette2002] Robinette, K.M., Blackwell, S., Daanen, H., Boehmer, M., Fleming, S., Brill, T., Hoeferlin, D. and Burnsides, D. Civilian American and European surface anthropometry resource (CAESAR), final report, volume I: Summary. Sytronics Inc Dayton Oh, 2002.

[Varol2021] Varol, G., Laptev, I., Schmid, C. and Zisserman, A. Synthetic Humans for Action Recognition from Unseen Viewpoints. IJCV, pp.2264-2287, 2021.

[Cassas2023] Casas, D. and Comino-Trinidad, M. SMPLitex: A Generative Model and Dataset for 3D Human Texture Estimation from Single Image. ArXiv, 2023.

[Grigorev2023] Grigorev, A., Thomaszewski, B., Black, M.J. and Hilliges, O. HOOD: Hierarchical Graphs for Generalized Modelling of Clothing Dynamics. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.16965-16974, 2023.

[Skorvankova2022] Skorvankova, D., Riecicky, A. and Madaras, M. Automatic Estimation of Anthropometric Human Body Measurements. SciTePress, Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 4: VISAPP, pp.537-544, 2022.

[Ruiz2022] Ruiz, N., Bellver, M., Bolkart, T., Arora, A., Lin, M.C., Romero, J. and Bala, R. Human Body Measurement Estimation with Adversarial Augmentation. ArXiv, 2022.

[He2016] He, K., Zhang, X., Ren, S. and Sun, J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770-778, 2016.