IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 4, 2019, accepted December 24, 2019, date of publication December 30, 2019, date of current version January 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962833

Non-Parametric Anthropometric Graph Convolutional Network for Virtual Mannequin Reconstruction

HAOYANG XIE[1], YUEQI ZHONG[1,2], ZHICAI YU[1], AND AZMAT HUSSAIN[3]

[1]College of Textiles, Donghua University, Shanghai 201620, China
[2]Key Laboratory of Textile Science and Technology, Ministry of Education, Donghua University, Shanghai 201620, China
[3]Department of Textile Engineering, Bahuddine Zakariya University, Multan 54000, Pakistan

Corresponding author: Yueqi Zhong (zhyq@dhu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572124, and in part by the Fundamental Research Funds for the Central Universities under GrantCUSF-DH-D-2017006.

ABSTRACT In this paper, we present a novel non-parametric method for precisely reconstructing a three dimensional (3D) virtual mannequin from anthropometric measurements and mask image(s) based on Graph Convolution Network (GCN). The proposed method avoids heavy dependence on a particular parametric body model such as SMPL or SCPAE and can predict mesh vertices directly, which is significantly more comfortable using a GCN than a typical Convolutional Neural Network (CNN). To further improve the accuracy of the reconstruction and make the reconstruction more controllable, we incorporate the anthropometric measurements into the developed GCN. Our non-parametric reconstruction results distinctly outperform the previous graph convolution method, both visually and in terms of anthropometric accuracy. We also demonstrate that the proposed network possesses the capability to reconstruct a plausible 3D mannequin from a single-view mask. The proposed method can be effortless extended to a parametric method by appending a Multilayer Perception (MLP) to regress the parametric space of the Principal Component Analysis (PCA) model to achieve 3D reconstruction as well. Extensive experimental results demonstrate that our anthropometric GCN itself is very useful in improving the reconstruction accuracy, and the proposed method is effective and robust for 3D mannequin reconstruction.

INDEX TERMS Graph convolution network, non-parametric mannequin reconstruction, anthropometric mannequin design, parametric reconstruction.

I. INTRODUCTION

Three dimensional (3D) virtual mannequin plays an essential role in many applications such as virtual try-on, made-to-measure, sports science, movie industry, personalized entertainment (e.g., augmented reality games and virtual reality), etc. [1]. Recently, the overwhelming success of Convolutional Neural Network (CNN) has achieved for 3D human body estimation from image(s). By taking advantage of convolutional layers on Multilayer Perception (MLP) or regular girds, 3D tasks involved in deep learning usually present 3D shapes as point clouds [2]–[4] or voxels [5], [6] to accommodate convolution in the underlying Euclidean space. However, point cloud and voxel are non-trivial in reconstructing a surface model. Contrarily, the mesh is more desirable for many real scenarios. There have been several parametric approaches [7]–[14] that can reconstruct 3D human mesh from image(s) by various CNNs accompany with some parametric body models such as SCAPE [15] and SMPL [16]. However, two major problems exist in these current approaches. 1) Most image-based 3D human reconstruction methods only utilize CNNs as feature extractors to regress the parametric space corresponding to a particular parametric body model rather than a real 3D shape, i.e., the 3D body is not reconstructed explicitly and directly. Also, committing to a specifically parametric space is rather limited by itself. For example, SMPL cannot model facial expressions [17], [18]. 2) Although the existing approaches can estimate the complete shape and pose from image(s), the regression results are challenging to be used in some applications that have requirements on accuracy, such as made-to-measure, since they lack constraints

The associate editor coordinating the review of this manuscript and approving it for publication was Ah Hwee Tan.

on anthropometry. In this paper, we propose an end-to-end network based on Graph Convolutional Network (GCN), which can predict the 3D vertices directly rather than a specific parametric space, and the network also explicitly integrates anthropometric measurements to improve the accuracy and make the reconstruction more controllable.

Our goal is to reconstruct an accurate 3D mannequin mesh from the anthropometric measurements and mask image(s) through GCN. Since the texture is unavailable in our dataset and a single-view mask or silhouette cannot provide enough shape information [14], [19], we therefore primarily focus on the reconstruction from two-view mask images, i.e., the front view mask and the lateral view mask. However, we still confirm the capability to reconstruct a complete 3D shape from a single-view mask in our experiments. The proposed approach requires the integration of knowledge learned from three modalities, i.e., geometry, image, and anthropometric measurements. For 3D geometry, we apply graph convolution [20] on a human mesh, where the vertices and edges of the mesh are represented as nodes and edges of a graph. The graph convolutional layers encode the features across neighboring nodes (vertices) and predict the 3D vertices eventually. For the mask image(s), we use ResNet-18 to extract image features since 1) ResNet has achieved significant success in many visual tasks [21], and 2) the mask image is more straightforward than regular RGB image, so there is no need to utilize more complex architectures. Meanwhile, ResNet-18 consumes less computing and memory resources than other deeper architectures. For anthropometry, we adopt five measurements, including stature, weight, chest circumference, waist circumference, and hip circumference by default. Notice that more measurements can be involved in our network, but for end-users, these five anthropometric measurements are the most widely used and readily accessible. We also compare the reconstructions with more anthropometric measurements, and the results with five measurements are well enough. Since the anthropometric measurements are related to each other in practice and the scales of the measurements are also different, which may result in the optimization being stuck at a local minimum. Therefore, we map the original measurements into another space to obtain an abstract anthropometric feature using a fully connected network. The anthropometric feature and image feature extracted from ResNet-18 are attached to the mesh vertices, and GCN is employed to process them on the mesh topology. Furthermore, our non-parametric method preserves the topology of a mesh, and the proposed approach can thus be easily extended to regress a specific parametric space by appending a Multilayer Perceptron (MLP), although our goal is not to investigate the parametric method.

The major contributions of this paper are summarized as follows:

- The proposed method utilizes the GCN to predict the 3D location of each vertex on a mesh directly, which avoids the dependency on a particular parametric space.
- By explicitly incorporating the anthropometric measurements, the reconstruction accuracy can be improved further, and the result is controllable.
- Extensive experimental results demonstrate that our anthropometric GCN itself is very useful in improving the reconstruction accuracy, and the proposed method is effective and robust for 3D mannequin reconstruction.
- We exhibit that the proposed approach can be used for anthropometric body shape design.

II. RELATED WORK

3D body reconstruction has always been an active field in the past decades. Since the literature on 3D human reconstruction is fairly broad, we only review the work closely related to our approach and refer the interested readers to recent surveys [22]–[25]. To make readers better understand our approach, we divide the previous 3D mannequin reconstruction approaches into two categories: the parametric approach and the non-parametric approach.

A. PARAMETRIC APPROACH

A parametric body model is the heart of the parametric reconstruction approaches. SCAPE [15] is a widely used model in parametric human reconstruction. Most prior parametric methods use traditional optimization algorithms to fit the parametric space of the SCAPE from various modalities, such as scans [26] and depth frames [27]–[31]. The approach of Guan *et al.* [32] is one of the few works to reconstruct 3D shape from a single image before the success of deep learning. They use labeled 2D landmarks and optimize for the SCAPE parameters to generate a 3D object. Although the SCAPE can simulate the human shape and pose, even muscles, naturally using non-rigid deformation, it is rather time-consuming. There are also several methods that use the SMPL model [16], which represents the 3D shape as a function of pose and shape parameters. For example, Zanfir *et al.* [33] formulate shape estimation as optimization with SMPL.

Although the optimization-based parametric approaches can obtain reliable results, there are several problems like the reliance on a proper initialization, the slow running-time, and the typical failures due to poor local minima, which have recently shifted the focus of 3D reconstruction to deep learning. The learning-based method provides an alternative approach that can be expected to learn appropriate priors from training data automatically, and then reconstruct a complete 3D human according to a particular parametric body model. Dibra *et al.* [9] propose one of the first methods using deep learning and train CNN to estimate the shape parameters of SCAPE from 2D silhouettes. Subsequently, they improve their performance by designing a more advanced architecture to regress a latent space consists of Heat Kernel Signature (HKS) [19]. Furthermore, their improved method is one of the few works that focus on reconstruction accuracy. They do not explicitly employ the anthropometric measurements. Instead, all the training mask images are calibrated implicitly. Ji *et al.* [14] predict the shape parameter of a simplified

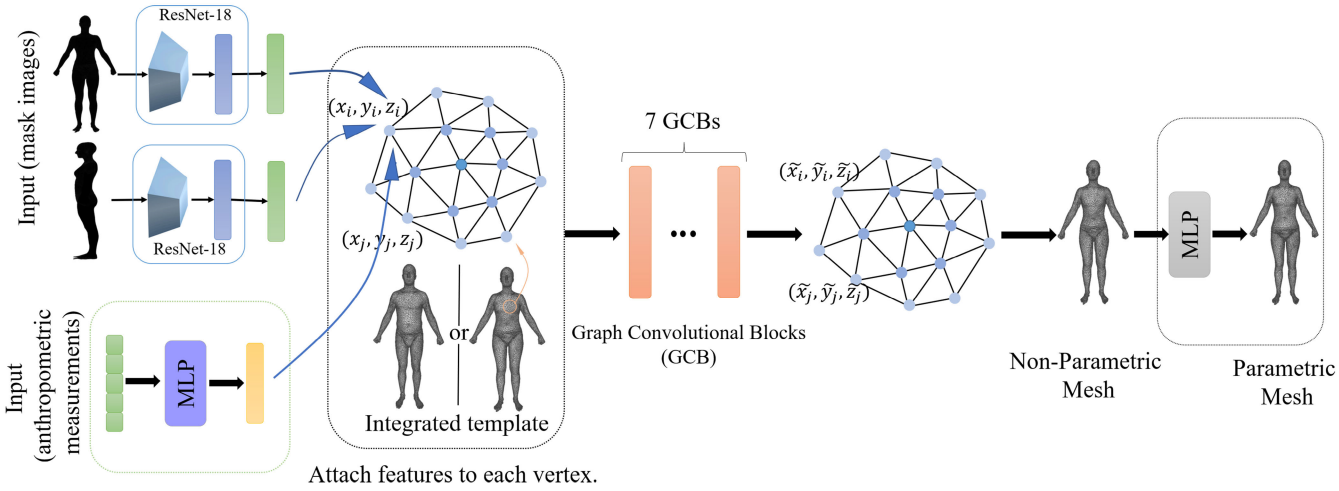


FIGURE 1. **Overview of the proposed approach. Two mask images and five measurements are employed as the inputs to our method. A template is also integrated into our network. For the end-user, the template is transparent. All the features are attached to the template's vertices, and a serial of graph convolutions are implemented on the mesh to generate the final vertices. The network can also be extended to generate a parametric body mesh by appending an MLP.**

SCAPE model from mask images by designing a double-branch DenseNet architecture. Recently, Huang *et al.* [34] employ a fully connected network to regress the PCA space of the feature curves that decompose a 3D human mesh into various patches from semantic parameters. These feature curves keep the underlying spatial relations of various patches, and a full 3D human mesh can thus be generated by combining the patches. However, their dataset only contains 77 female models, which severely limits the generalization ability of their model, and the generated human shapes share a similar geometry. Their method is completely different from ours in design philosophy. Other works [7], [10], [12], [13], [35], [36] train neural networks to predict the SMPL parameters from the input image(s). Kanzawa *et al.* [7] predict the SMPL parameters with a re-projection loss on the 2D joints and introduce a discriminator to distinguish whether the generated 3D human is realistic. Similarly, Tung *et al.* [13] learn the parameters by combining the re-projection loss on the 2D joints, optical flow, and silhouettes. Tan *et al.* [12] train an encoder-decoder architecture to regress SMPL shape parameters at the bottleneck layer. Omran *et al.* [10] integrate the SMPL model into a deep CNN architecture and introduce an intermediate region-based 2D representation to predict the SMPL parameters. They also demonstrate that only a small fraction of training data needs to be paired with 3D annotations. Overall, the parametric method aims to regress the parameters of a particular parametric mode, rather than generating a mesh directly.

B. NON-PARAMETRIC APPROACH

Recently, learning-based non-parametric methods have been proposed. BodyNet [35] is an end-to-end network that infers volumetric body shape from a single image. By extending a face reconstruction network [37], Jackson *et al.* [38] also propose a volume-based human shape reconstruction method. These two methods present 3D objects as voxel representation rather than mesh. Wang *et al.* [39] propose one of the first approaches to reconstruct 3D mesh with GCN. They reconstruct 3D mesh from a single image by deforming an initial ellipsoid. However, they do not demonstrate that their model can generate a body shape very well, and the network is designed for a generic object rather than human shape. Lately, several human-related kinds of literature based on GCN are proposed as well. For example, Litany *et al.* [40] build a Variational Autoencoder (VAE) using graph convolution for shape completion. Our method is inspired by a recent approach proposed by Kolotouros *et al.* [41]. They estimate the body shape and posture from an RGB image using GCN. However, similar to many image-based methods, their method fails to generate the precise shape, and the 3D meshes predicted by their non-parametric method is rather coarse. Instead, we pay more attention to the accuracy and propose a novel loss function as well as a capable architecture. The reconstruction results significantly outperform the previous approaches, both visually and in terms of anthropometric accuracy.

III. TECHNICAL APPROACH

A. IMAGE-BASED FEATURE EXTRACTION

The pipeline of the proposed approach starts with a CNN based image feature extractor, as shown in Fig.1. Theoretically, many state-of-the-art CNN architectures can be used as the feature extractor. In this paper, we utilize transfer learning technology and employ the ResNet-18 that has been trained on ImageNet [42] to achieve image-based feature extraction. Specifically, we replace the last fully connected layer with a new layer consists of 512 neurons and only retrain the last layer rather than training the entire ResNet-18 from scratch, which makes our network more suitable for the human mask image. For single mask training, we ignore the ResNet-18 for the lateral image.

B. ANTHROPOMETRIC MEASUREMENTS

Given a 3D human mesh, we extract five anthropometric measurements by default, including stature, weight, chest circumference, waist circumference, and hip circumference. These measurements are essential for clothing and are also readily accessible to consumers. Since all bodies are standing in our dataset, we use the height of the bounding box to approximate the statue. The weight is estimated by multiplying the body volume by body density, and the body density is 0.985 kg/cm^3. We use a recent algorithm proposed by Zhong *et al.* [43] to calculate the chest circumference, chest circumference, and hip circumference automatically. These five measurements are fed into an MLP that includes two hidden layers and outputs a 5-D feature, where each hidden layer contains eights neurons with ReLU activation. This simple architecture works well in our experiments. We implement the algorithm with trimesh [44] so that this implementation can be integrated into our network to calculate the anthropometric measurements of the reconstructed mannequin directly.

C. GCN ARCHITECTURE

We aim to predict the vertices of a 3D mannequin mesh by GCN. A 3D mesh $\mathcal{M}$ can be naturally converted into an undirected graph, i.e., $\mathcal{M} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$, where $\mathcal{V} = \{v_i\}_{i=1}^N$ is the set of N vertices, $\mathcal{E} = \{e_i\}_{i=1}^N$ is the set of edges that connecting adjacent vertices, and $\mathcal{X} = \{x_i\}_{i=1}^N$ are the feature vectors attached to vertices. For graph convolution, we employ a simple but effective formulation proposed by Kipf and Welling [20], which is defined as:

$$Y = \tilde{A}\mathcal{X}W \tag{1}$$

where $\mathcal{X} \in \mathbb{R}^{N\times k}$ is the input feature matrix, i.e., the features attached on vertices, and each row x_i denotes a *k-dim* feature vector. $\tilde{A} = I + D^{-\frac{1}{2}}AD^{\frac{1}{2}}$ is the row-normalized adjacency matrix of a graph, where $A \in \mathbb{R}^{N\times N}$is the adjacency matrix, and $D \in \mathbb{R}^{N\times N}$ is the degree matrix. $W \in \mathbb{R}^{k\times l}$ is the weight matrix, and $I \in \mathbb{R}^{N\times N}$ is the identity. Notice that $\tilde{A}$ is substantially constant in the process of a series of graph convolutions since the topology of the graph is unchanged. In our case, the attached feature vector x_i is the concatenation of 3D vertices, anthropometric features, and features extracted from ResNet-18. Individually, for the two masks training, the dimension of x_i is 1032, and 520 for single mask training.

A template (female or male) mesh with N vertices is integrated into the architecture, as shown in Fig.1. Both the female and male networks are trained using the same architecture but with different templates. The entire GCN architecture consists of a series of Residual Graph Convolution Blocks (RGCBs), as shown in Fig.2, which is derived from the residual block [45], where the 3×3 convolution in residual block is replaced by (1), the 1×1 convolution is replaced by a linear layer. Due to the limitation of GPU memory, we use a small batch size for training. Therefore, we replace the Batch Normalization (BN) with Group Normalization (GN) since BN does not perform well for relatively small batch size [46]. For the case where the input and output channels are inequal, an extra linear layer is added to the shortcut to ensure the correct element-wise addition on the output. Furthermore, we find that seven RGCBs are the optimal trade-off between reconstruction accuracy and training consumption. Less than seven blocks cause a drastic drop in accuracy, more than seven blocks consume expensive memory and time but without significant improvement in precise.

To produce precise and reliable reconstruction, we experiment with a well-designed loss. Let $\mathcal{V} \in \mathbb{R}^{N\times3}$ denotes the ground-truth mesh vertices, and $\tilde{\mathcal{V}} \in \mathbb{R}^{N\times3}$ is the predicted vertices. Since our network predicts the 3D coordinates directly, an intuitive approach is to minimize the per-vertex loss between $\mathcal{V}$ and $\tilde{\mathcal{V}}$, i.e.,

$$\mathcal{L}_{\mathcal{V}} = \sum_{i=1}^{N} ||\mathcal{V}_i - \tilde{\mathcal{V}}_i||^2 \tag{2}$$

Additionally, we explicitly incorporate the anthropometric error into the loss. Let $M \in \mathbb{R}^5$ denotes the ground-truth anthropometric measurements, and $\tilde{M} \in \mathbb{R}^5$ is the corresponding measurements of the predicted shape. The measurement loss can be defined as:

$$\mathcal{L}_M = \|M - \tilde{M}\|^2 \tag{3}$$

Even with $\mathcal{L}_{\mathcal{V}}$ and $\mathcal{L}_M$, the optimization is easily stuck at a local minimum. Visually, the model will generate sharp vertices that make the surface unsmooth. To handle this problem, we propose a Laplacian smoothing term $\mathcal{L}_\delta$ to encourage the neighboring vertices to have a similar movement. Let δ_i and $\tilde{\delta}_i$ denote the Laplacian coordinates of the *i-th* vertex on the ground-truth mesh and the reconstructed mesh, respectively. Then, $\mathcal{L}_\delta$ is defined as:

$$\mathcal{L}_\delta = \sum_{i=1}^{N} ||\delta_i - \tilde{\delta}_i||^2 \tag{4}$$

where $\delta_i = v_i - \sum_{j\in\mathcal{N}(v_i)} \frac{j}{||\mathcal{N}(v_i)||}$. $\mathcal{N}(v_i)$ denotes the neighboring vertices of v_i. Finally, the complete training loss for our non-parametric anthropometric GCN is:

$$\mathcal{L}_{Non-param} = \mathcal{L}_{\mathcal{V}} + \mathcal{L}_M + \mathcal{L}_\delta \tag{5}$$

D. PCA SPACE AND APPENDING MLP

The low-dimensional parametric representation is not entirely worthless [7], [47]. In this section, we present that the proposed non-parametric anthropometric GCN can be efficiently extended to a parametric method by appending an MLP network. Our parametric model is inspired by SCAPE [15]. Since this paper focuses on reconstructing a mannequin in a standard A-pose, therefore, we ignore the pose space introduced by SPCAE and utilize the shape scape only, i.e., PCA space. A PCA model represents each body shape with a parameter vector $\phi_s \in \mathbb{R}^k$:

$$B_s = \bar{B} + \mathcal{Q}\phi_s \tag{6}$$

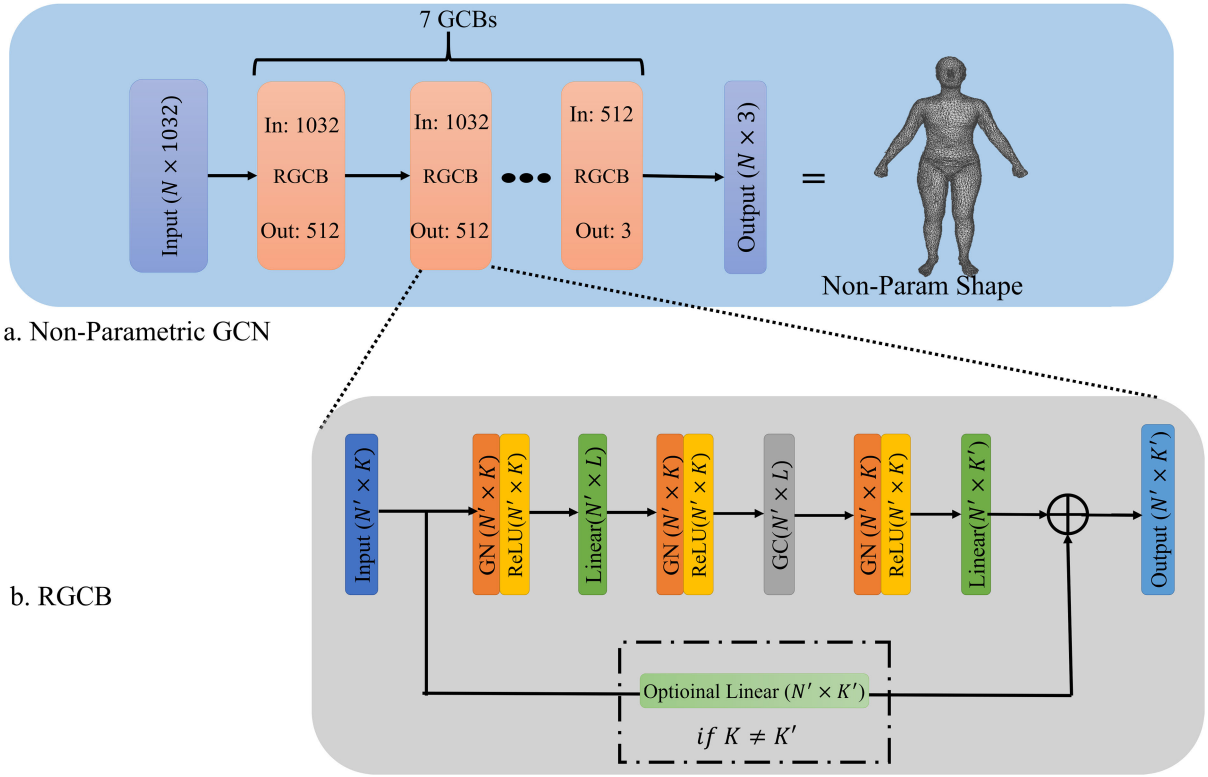


FIGURE 2. The non-parametric anthropometric GCN architecture and Residual Graph Convolution Block (RGCB): (a) illustrates our anthroprometric GCN architecture, where N is the number of vertices, and $1032 = 512 \times 2 + 5 + 3$ denotes the dimension of each feature on a vertex.The non-parametric GCN consists of seven RGCBs, and the number marked on each RGCB denotes the dimensions of input and output feature. The output dimension is 3 since the proposed netowrk predicts the 3D location; (b) demonstrates a Residual Graph Convolution Block (RGCB), where N' is the dimension of the output of the last layer, K is the number of output feature of this layer, and K' is the output dimension of a block. In our experiment, $L = \frac{K}{2}$. If $K \neq K'$, we add a linear layer to the shortcut to enable the element-wise addition. GN denotes the group normalization, and GC denotes the graph convolution.

where $B_s \in \mathbb{R}^{3N}$ is a body shape with N vertices, and $\mathcal{Q} \in \mathbb{R}^{3N \times k}$ is the principal matrix composed of k principal vectors $\{\mu_1, \mu_2, \ldots, \mu_k\}$. $\bar{B}$ is the mean body shape of the dataset. We append an MPL to the non-parametric anthropometric GCN to regress the parametric space of the PCA model, and then to achieve 3D reconstruction, as shown in Fig.1. By fixing the previously trained GCN, we train the MLP to regress the parameter vector ϕ_s. To enable a broad range of variations, we set $k = 64$ in our implementation, which captures 98.93% of the energy and performs well in our experiments. Since the output of parametric regression is a parameter vector rather than the 3D location of vertices, the 3D coordinate loss $\mathcal{L}_\mathcal{V}$ can be ignored, and the loss function for the parametric method is:

$$\mathcal{L}_{Param} = \mathcal{L}_M + \mathcal{L}_\phi \tag{7}$$

where $\mathcal{L}_M$ is the loss of anthropometric measurements defined by (3), and $\mathcal{L}_\phi$ denotes the loss of parametric vectors. The appending MLP network is a fully connected architecture consist of three Residual Fully Connected Blocks (RFCBs), as shown in Fig.3. Since the dimension of the parametric space is 64, the output dimension of the last linear layer in MLP is 64 as well.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASET

Although a full 3D human mesh can be estimated using "in-the-wild" images that only have ground-truth 2D annotations [7], [10], [48], [49], ground-truth 2D-to-3D supervision is still necessary to explicitly restore the size of a 3D body [19]. In this paper, we use the SPRING dataset proposed by Yang *et al.* [50]. SPRING is derived from the CAESAR project [51], and all the human meshes in SPRING share identical topology and similar A-pose, which facilitates to calculate the per-vertex loss. SPRING contains 1517 male meshes and 1529 female meshes. We randomly select 300 meshes, 150 per gender, as the test set, and the rest as training data.

B. TRAINING DETAILS

We perform all the experiments on a laptop with an Intel i7-7700HQ at 2.8GHz CPU, 16GB of memory, and NVIDIA GeForce GTX 1070 graphics with a GPU memory of 8GB. The algorithm proposed in this paper is based on PyTorch [52] to verify the performance and computational efficiency.

We employ the same GCN architecture but different template mesh to train our network for males and females separately. The mean male and female shapes are used as

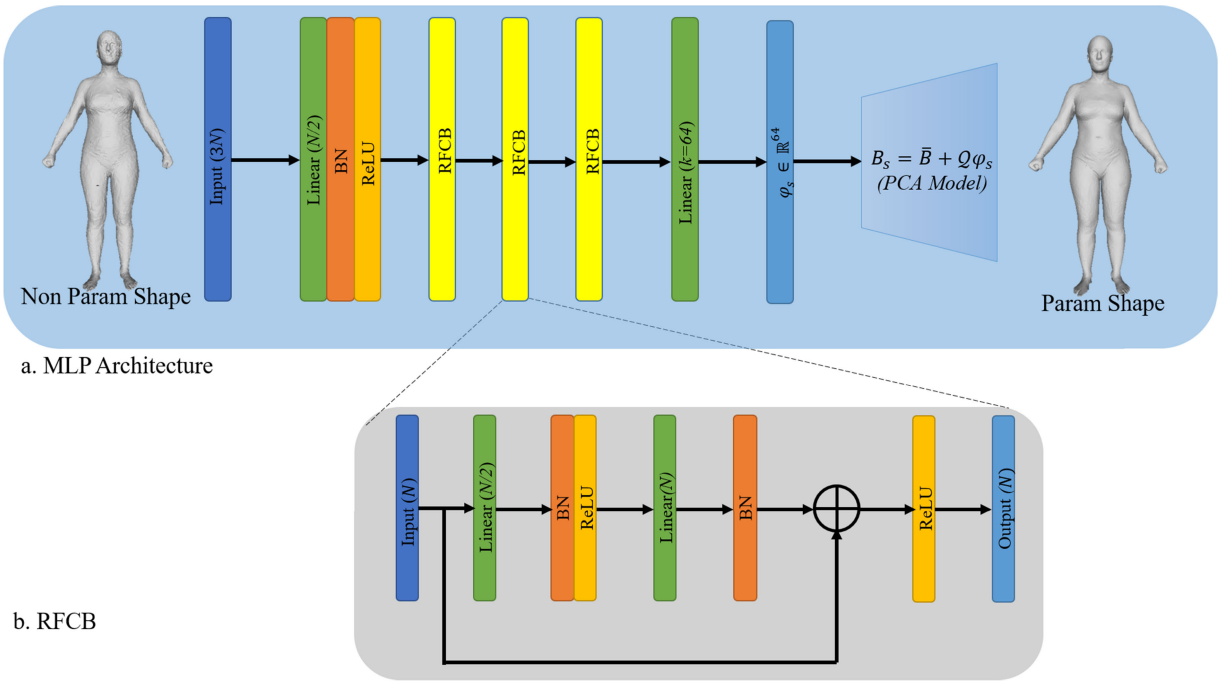


FIGURE 3. The MLP architecture and Residual Fully Connected Block (RFCB). BN denotes the batch normalization.

our templates. We also experiment with other arbitrary templates, but we find that the mean shape can slightly accelerate network training. The dataset utilized in our experiments contains only 3D meshes without any texture and corresponding 2D rendering. For each mesh, we render two masks from the front view and lateral view using the render-to-texture technique in OpenGL. Since the pre-trained ResNet-18 is employed as the feature extractor, all the mask images are rescaled to 224 $\times$ 224 before feeding them to the network. The non-parametric anthropometric GCN training lasts 17 epochs with the Adam optimizer, and the batch size is set to be 2. The initial learning rate is 0.001 and is multiplied by 0.5 at every 5 epochs. Also, we fix the trained GCN when we train the appending MLP. The training for MLP lasts 4 epochs with Adam optimizer, and the learning rate is 0.0001. The non-parametric anthropometric GCN contains 160M parameters totally, and the training takes approximately 12.5 hours. However, the forward prediction (reconstruction) is rather fast and can be completed in less than one second to reconstruct a complete 3D human mesh with 12500 vertices.

C. RESULTS AND ANALYSIS

We qualitatively and quantitatively evaluate the proposed approach. To simplify the description, all experiments demonstrated in this paper are the non-parametric female reconstructions based on two mask images by default, unless state otherwise. Our first ablation analysis is to investigate the effect of the Laplacian smooth term $\mathcal{L}_\delta$ in (5). To this end, we retrain the GCN by ignoring $\mathcal{L}_\delta$, and a reconstructed mannequin is demonstrated in Fig.4. Fig.4(a) is a ground-truth mesh in our test set. Fig.4(b) and Fig.4(c) demonstrate

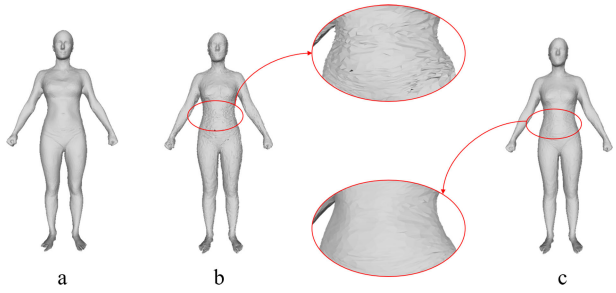


FIGURE 4. Comparison of the non-parametric reconstructions with and without Lapalcian smooth item $\mathcal{L}_\delta$: (a) is the ground-truth; (b) illustrates the reconstruction without $\mathcal{L}_\delta$; and (c) shows our result with $\mathcal{L}_\delta$.

the results without and with $\mathcal{L}_\delta$, respectively. The difference between the results is apparent. For the case where $\mathcal{L}_\delta$ is missing, all vertices only pursue the minimum vertex-to-vertex error during the optimization, which results in sharp vertices and unsmooth surface. Whereas, the movement of the vertices will be affected by their neighbor vertices after involving $\mathcal{L}_\delta$ in, which makes the adjacent vertices share a similar movement and result in a smoother surface. These experiments have also proved that utilizing only the vertex-to-vertex error is not enough to reconstruct a plausible shape.

Similarly, we retrain our GCN by ignoring $\mathcal{L}_M$ in (5) to investigate its impact on the reconstructed results. Fig.5 illustrates the difference between the results with and without $\mathcal{L}_M$. Fig.5(a) is the ground-truth in our test set, Fig.5(b) and Fig.5(c) are the non-parametric results without and with anthropometric measurements, respectively. We can see that the statures of the reconstructed shapes are different. Other

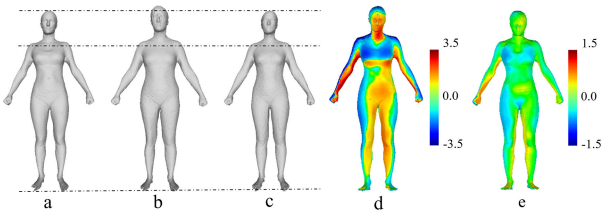


FIGURE 5. Comparison of the non-parametric reconstructions with and without $\mathcal{L}_M$: (a) is the ground-truth; (b) illustrates the reconstruction without $\mathcal{L}_M$; (c) shows our result with $\mathcal{L}_M$; (d) is the heat map of (a) and (b); and (e) demonstrates the heat map of (a) and (c).

TABLE 1. Numerical comparison with and without anthropometric measurements. Errors are express as Mean±Std. Dev in centimeters. Our best achieving is highlighted.

Measurements	*Female*		*Male*	
	WA	*WOA*	*WA*	*WOA*
Stature	***0.54±0.11***	*3.27±1.10*	***0.41±0.15***	*4.82±2.33*
Weight (kg)	***0.98±0.57***	*3.89±2.93*	***1.03±0.62***	*4.61±2.52*
Chest	***0.76±0.42***	*2.12±2.03*	***0.53±0.27***	*1.98±1.18*
Waist	***0.85±0.49***	*2.35±1.87*	***0.81±0.45***	*3.87±2.16*
Hip	***0.68±0.51***	*3.59±2.71*	***0.49±0.33***	*3.71±2.31*

differences can be observed in Fig.5(d) and Fig.5(e), which demonstrate the heat maps of (a) and (b), and (a) and (c), respectively. Although we have employed the vertex-to-vertex loss, and if each vertex ideally regresses to the correct position, we indeed do not need $\mathcal{L}_M$. However, it is impossible to minimize the error to zero without overfitting in practice. Thus, we explicitly give the anthropometric loss to improve the accuracy, which is also consistent with the idea of Conditional GAN [53]. To further verify the effectiveness of incorporating measurements, we calculate the anthropometric errors of the five measurements on the female and male test set. The statistical result is demonstrated in Table 1, where *WA* denotes the results with measurements, and *WOA* denotes the reconstructions without measurements. Errors are expressed as Mean±Std. Dev in centimeters. The result with measurements outperforms the reconstruction without measurements in each size, which indicates that it is useful to integrate anthropometric measurements into the network.

To further verify the contribution of the proposed approach, we experiment to compare our method with three related state-of-the-art body shape reconstruction networks proposed by Kolotouros *et al.* [41], Ji *et al.* [14], and Dibra *et al.* [19]. For Kolotouros et al.'s method, we use the trained model published by the authors. For the methods of Ji et al. and Dibra et al., we reimplement their networks according to the suggested hyper-parameters configuration, and train 27 and 35 epochs on the same dataset, respectively, to get the best results. Kolotouros et al. present a non-parametric method to estimate the shape from a single RGB image. Contrarily, Ji et al.'s work is a parametric approach with two mask images as input. They reconstruct the body shape by regressing the shape space of the SCAPE model, i.e., the PCA model. Similarly, Dibra et al. also use two mask images as input to generate body shape. For fairness of comparison, we use a real body and employ the data format they required as respective inputs.

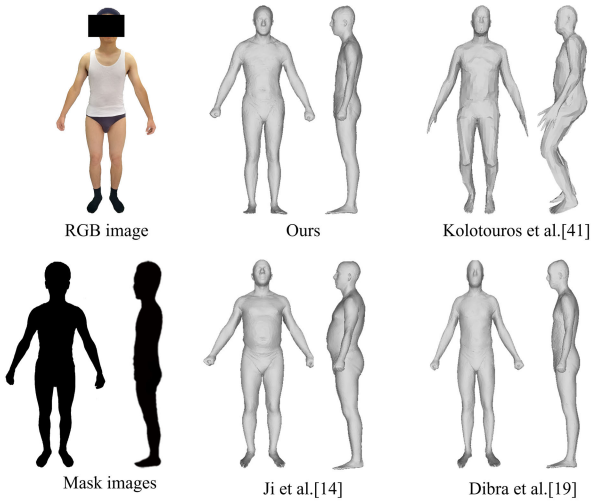


FIGURE 6. Comparison with other state-of-the-art methods proposed by Kolotouros et al. [41], Ji et al. [14], and Dibra et al. [19] using a real person.

TABLE 2. Numerical comparison for the real case. Dev in centimeters.

Measurement error	*Ours*	*Kolotouros et al. [41]*	*Ji et al. [14]*	*Dibra et al. [19]*
Stature	*0.11*	*13.57*	*9.3*	*1.52*
Weight (kg)	*1.28*	*11.13*	*8.76*	*2.78*
Chest	*0.83*	*9.42*	*8.56*	*1.13*
Waist	*0.52*	*7.79*	*9.15*	*1.69*
Hip	*0.65*	*12.66*	*6.55*	*1.46*

The first column of Fig.6 demonstrates input representations. The second and third columns of Fig.6 illustrate the comparisons. Visually, our results significantly outperform the other approaches. The method of Kolotouros et al. almost fails to complete reconstruction, which contains too much noise, and the resulting body shape is inaccurate in all sizes. The comparison with Kolotouros et al.'s method also confirms that integrating the anthropometric measurements in GCN can significantly improve the reconstruction accuracy. Ji et al. regress the parametric space, and the resulting body mesh is smoother than Kolotouros et al. However, their result still lacks anthropometric constraints on the size. Dibra et al. utilize Heat Kernel Signature (HKS) and scaled mask images as the inputs and train a more complex network to generate the 3D shape. Their input images contain calibration information implicitly, and the result is much better than Kolotourous et al. and Ji et al. However, the waist shape is significantly different from the input front mask. Table 2 demonstrates the specific numerical comparison for the real case, and our approach outperforms others on all anthropometric sizes. Table 3 shows a more comprehensive comparison. Specifically, we calculate 13 anthropometric measurements widely utilized in tailoring fitting and compare the anthropometric error (Mean±Std) of the meshes generated by the state-of-the-art methods [14], [19] that require two-view at test time on our test set. Our best achieving method is highlighted. Last but not least, we need to pay special attention to the surface details generated by these methods, including ours. Almost all of them could generate

TABLE 3. Body measurement errors comparison with shapes reconstructed from two-view mask images. Errors are expressed as Mean±Std. Dev in centimeters. Our best achieving is highlighted.

Anthropometric	*Female*			*Male*		
Measurement	*Ours*	*Ji et al. [14]*	*Dibra et al. [19]*	*Ours*	*Ji et al. [14]*	*Dibra et al. [19]*
Stature	***0.54±0.11***	*10.13±4.15*	*1.18±0.78*	***0.41±0.15***	*13.44±5.15*	*1.37±0.69*
Weight (kg)	***0.98±0.58***	*8.68±5.56*	*2.23±1.01*	***1.03±0.62***	*9.23±6.17*	*2.21±1.17*
Chest circumference	***0.76±0.42***	*9.03±4.87*	*1.67±0.88*	***0.53±0.27***	*7.35±4.72*	*1.59±0.75*
Waist circumference	***0.85±0.49***	*7.57±4.43*	*2.34±1.14*	***0.81±0.45***	*7.51±4.33*	*2.33±1.20*
Hip circumference	***0.68±0.51***	*10.12±7.54*	*1.86±0.98*	***0.49±0.33***	*10.89±7.11*	*1.69±0.87*
Neck circumference	***0.13±0.08***	*5.13±3.66*	*0.87±0.55*	***0.15±0.04***	*5.03±3.25*	*0.44±0.31*
Shoulder breadth	***0.22±0.15***	*6.12±4.33*	*0.93±0.47*	***0.17±0.02***	*6.85±4.28*	*0.93±0.54*
Arm length	***0.33±0.11***	*11.45±6.87*	*1.66±1.04*	***0.55±0.32***	*10.32±6.18*	*1.10±0.38*
Forearm length	***0.21±0.09***	*7.81±4.55*	*1.83±1.32*	***0.34±0.28***	*6.89±4.11*	*0.78±0.29*
Thigh circumference	***0.48±0.23***	*5.54±4.17*	*1.05±0.88*	***0.41±0.19***	*5.18±3.82*	*1.27±0.62*
Calf circumference	***0.27±0.14***	*4.19±3.22*	*1.21±0.43*	***0.24±0.15***	*3.89±3.14*	*1.26±0.89*
Inside leg length	***0.45±0.30***	*9.38±5.16*	*1.22±0.79*	***0.67±0.41***	*7.56±4.21*	*1.38±0.66*
Neck-to-crotch length	***0.32±0.12***	*8.27±6.31*	*0.89±0.68*	***0.31±0.13***	*5.47±4.66*	*0.78±0.56*

TABLE 4. Measurement error comparison among our non-parametric method, non-parametric+parametric method, and parametric only method.

Anthropometric	*Female*			*Male*		
Measurements	*Non-Param*	*Non-Param+Param*	*Param only*	*Non-Param*	*Non-Param+Param*	*Param only*
Stature	*0.54±0.11*	***0.51±0.14***	*6.26±3.16*	*0.41±0.15*	***0.40±0.21***	*5.87±3.23*
Weight (kg)	*0.98±0.57*	***0.74±0.55***	*4.37±2.15*	*1.03±0.62*	***0.95±0.62***	*4.32±2.25*
Chest circumference	*0.76±0.42*	***0.71±0.40***	*5.12±3.78*	*0.53±0.27*	***0.50±0.25***	*5.27±3.56*
Waist circumference	*0.85±0.49*	***0.79±0.43***	*4.76±2.66*	*0.81±0.46*	***0.75±0.39***	*4.32±2.21*
Hip circumference	*0.68±0.51*	***0.67±0.49***	*7.22±3.11*	*0.49±0.33*	***0.47±0.34***	*6.78±2.89*

plausible details such as the face, as shown in Fig.6 but those details are not necessarily real. The inputs to these methods do not contain any surface detail information, and the four methods generate four different faces. The details seem to be generated from the training data randomly rather than the truth.

Intuitively, more input anthropometric measurements would characterize the human shape more precisely. However, too many parameters may influence the reconstruction [54], i.e., the reconstruction method maybe not robust to the number of input measurements, and increasing the input may sway the error seriously. To verify the robustness of our approach to the number of parameters and the effect of the number of parameters on the reconstruction results, we conduct an experiment using a different number of input measurements to train our network. Notice our purpose is to investigate the effect of the number of measurements on reconstruction, not determining the optimal measurements for reconstruction. Specifically, besides the previous five measurements, we collect the other eight measurements, including neck circumference, shoulder breadth, arm length, forearm girth, thigh circumference, calf circumference, inside leg length, and neck-to-crotch length. These eight measurements are randomly divided into four groups, and then the four groups are sequentially accumulated to the network for training (need to adjust the size of the input layer). Besides, we also demonstrate the reconstruction with less anthropometric measurements, i.e., stature, weight, and waist. The performance can be observed in Fig.7, where Fig.7(a) is the ground-truth, Fig.7(b), Fig.7(c), Fig.7(d) and Fig.7(e) illustrate the results with three, five, nine, and thirteen input measurements. The results of the reconstruction error against

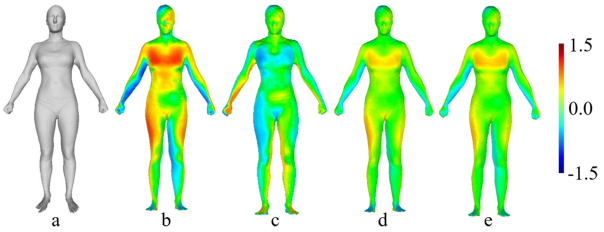


FIGURE 7. Non-parametric reconstructions with different number of measurements: (a) is the ground-turth; (b), (c), (d), and (e) illustrate the results with three, five, nine, and thirteen measurements.

the number of input sizes are plotted in Fig.8. By comparing the results generated by a different number of input measurements, we can find that 1) the reconstruction error decreases with more anthropometric measurements involved. 2) Involving too many measurements does not bring an apparent improvement since anthropometric measurements are usually related to others in practice. Therefore, the proposed approach is insensitive to the number of input parameters, and an increase in the number of anthropometric measurements does not introduce a severe vibration on the error, which makes our system more practical and repeatable.

The proposed method can be effortless extended to a parametric method by appending an MLP to regress the PCA space of our training data. Fig.9 and Fig.10 demonstrate the female and male non-parametric reconstructions and parametric reconstructions from two mask images. Since our parametric reconstruction is based on the trained non-parametric network, the parametric method is denoted as non-param+param in Fig.9 and Fig.10. All the meshes are from our test set. Generally, both the non-parametric method

TABLE 5. Body measurement errors comparison with shapes reconstructed from single-view image. Errors are expressed as Mean±Std. Dev in centimeters. Our best achieving is highlighted.

Anthropometric Measurements	*Female*				*Male*			
	Ours	*Kolotouros et al. [41]*	*Ji et al. [14]*	*Dibra et al. [19]*	*Ours*	*Kolotouros et al. [41]*	*Ji et al. [14]*	*Dibra et al. [19]*
Stature	**0.63±0.12**	18.46±9.78	11.34±4.44	1.48±0.65	**0.59±0.15**	17.55±9.65	14.32±5.13	1.57±0.35
Weight (kg)	***1.12±0.57***	*13.87±7.38*	*9.96±6.51*	*2.73±1.21*	***1.17±0.62***	*12.69±7.21*	*9.86±6.06*	*2.61±1.26*
Chest circumference	***1.01±0.42***	*17.99±10.11*	*10.13±3.59*	*1.96±0.75*	***1.26±0.27***	*16.82±10.15*	*7.91±4.24*	*1.89±0.68*
Waist circumference	***1.22±0.49***	*11.29±8.34*	*9.22±5.27*	*2.66±1.11*	***1.19±0.45***	*10.56±7.32*	*8.17±4.22*	*2.57±1.13*
Hip circumference	***0.89±0.51***	*12.87±9.25*	*11.34±7.25*	*1.97±0.79*	***0.95±0.33***	*13.69±8.85*	*11.86±7.41*	*1.75±0.58*
Neck circumference	***0.25±0.08***	*7.58±5.12*	*6.52±4.13*	*1.13±0.72*	***0.27±0.04***	*7.98±5.62*	*5.83±3.11*	*0.82±0.47*
Shoulder breadth	***0.37±0.15***	*8.46±4.87*	*7.12±4.98*	*1.56±0.51*	***0.49±0.12***	*8.99±4.68*	*7.56±4.15*	*1.26±0.52*
Arm length	***0.53±0.11***	*13.98±7.69*	*12.67±7.04*	*2.03±1.23*	***0.64±0.32***	*12.76±7.34*	*10.98±5.87*	*1.37±0.41*
Forearm length	***0.32±0.09***	*12.56±5.27*	*9.12±5.51*	*1.95±1.46*	***0.58±0.28***	*13.51±6.78*	*7.37±4.26*	*1.12±0.33*
Thigh circumference	***0.58±0.23***	*7.38±5.15*	*6.92±4.48*	*1.57±0.87*	***0.41±0.19***	*7.89±5.03*	*5.79±3.48*	*1.44±0.34*
Calf circumference	***0.41±0.12***	*6.75±4.66*	*5.59±3.46*	*1.84±0.47*	***0.55±0.15***	*5.58±4.21*	*4.67±3.51*	*1.58±0.73*
Inside leg length	***0.67±0.30***	*16.81±11.37*	*10.43±5.98*	*1.65±1.03*	***0.89±0.41***	*15.75±11.24*	*8.29±4.17*	*1.75±0.45*
Neck-to-crotch length	***0.55±0.12***	*11.56±8.62*	*9.95±6.19*	*1.98±0.36*	***0.67±0.13***	*12.79±8.83*	*6.38±4.42*	*1.11±0.39*

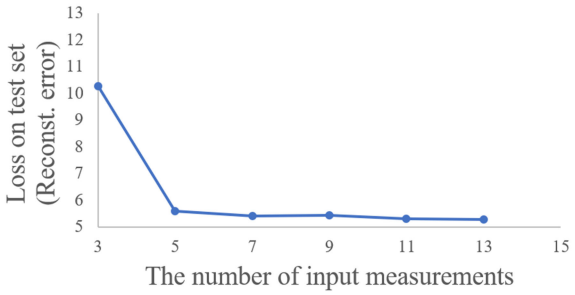


FIGURE 8. Loss on test set with various size of input measurements.

and the parametric method have achieved excellent results in terms of accuracy, and the parametric results seem to have more details than the non-parametric method. However, as we mentioned above, these details are not necessarily correct, which heavily depends on the training data. For the obese body shapes, both the non-parametric method and the parametric method introduce a slightly larger error in anthropometric measurements due to the lack of corresponding samples in the training set. The specific numerical error for these two methods on the female and male test set is demonstrated in Table 4. Since our non-parametric+parametric model is trained based on the non-parametric GCN, the results are slightly better than the non-parametric results. Additionally, Table 4 also demonstrates the reconstruction results using the parametric MLP only, where we ignore the non-parametric GCN and feed the image features and anthropometric features to the appending MLP directly. In this case, the reconstruction performance drastically drops, which indicates that the anthropometric GCN itself is very useful in improving the reconstruction accuracy.

We also demonstrate that the proposed anthropometric GCN can reconstruct a complete 3D mannequin from a single-view mask, in which we ignore the ResNet-18 for the lateral mask and retrain our network. The numerical error for the single-view reconstruction is shown in Table 5. Table 5 also illustrates other state-of-the-art methods that can complete reconstruction from single-view image [14], [19], [41]. Our best achieving method is highlighted.

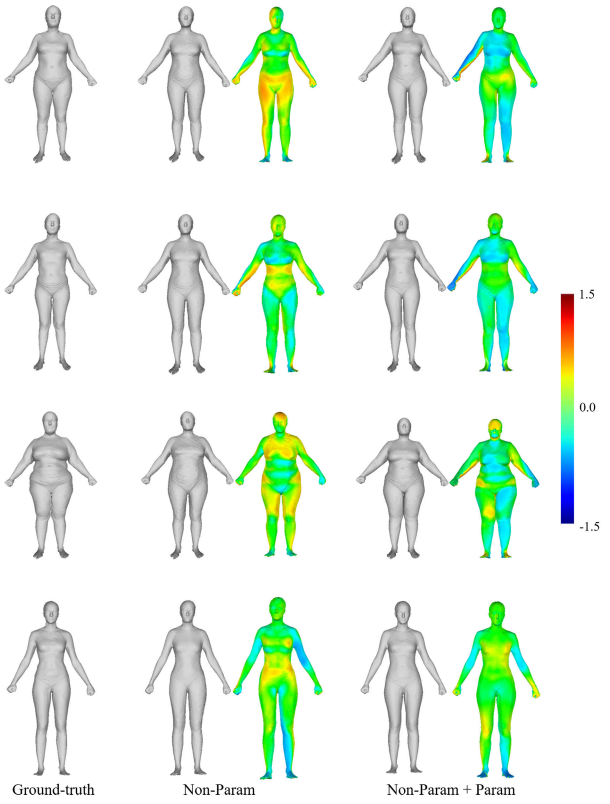


FIGURE 9. Female reconstructions. All the meshes are from our test set. The first column illustrates the ground-truth, the second and third columns show the non-parametric results and corresponding heat maps. And the last two columns demonstrate the non-parametric+parametric reconstructions and corresponding heat maps.

Comparing with Table 3, the single-view reconstruction introduces a slightly larger error, which is also consistent with the previous conclusion (Sect.I). However, the single-view reconstruction is not entirely useless. For example, the single-view image is easier to capture for an end-user. Since our network explicitly requires anthropometric measurements, we conduct an interesting experiment to reconstruct a 3D shape from the frontal mask image of a dressed person, as shown in Fig.11. Our result is also very well in this case.

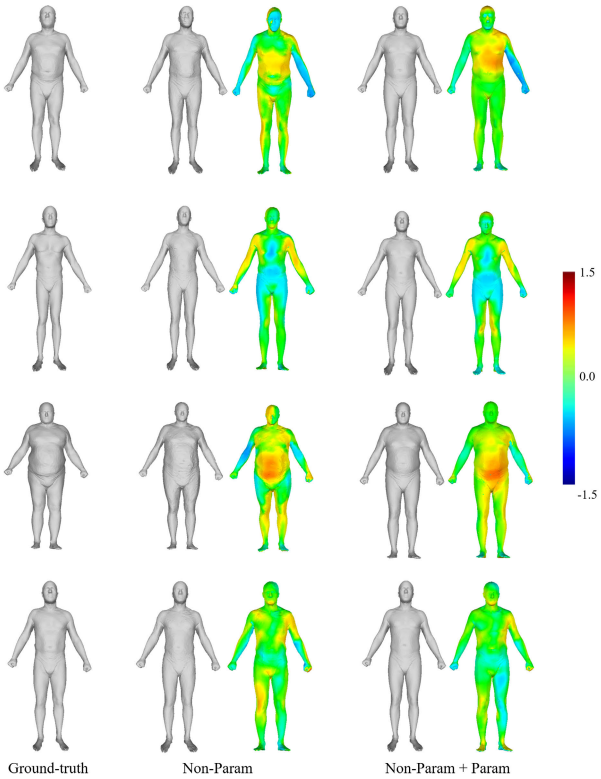


FIGURE 10. Male reconstructions. All the meshes are from our test set. The first column illustrates the ground-truth, the second and third columns show the non-parametric results and corresponding heat maps. And the last two columns demonstrate the non-parametric+parametric reconstructions and corresponding heat maps.

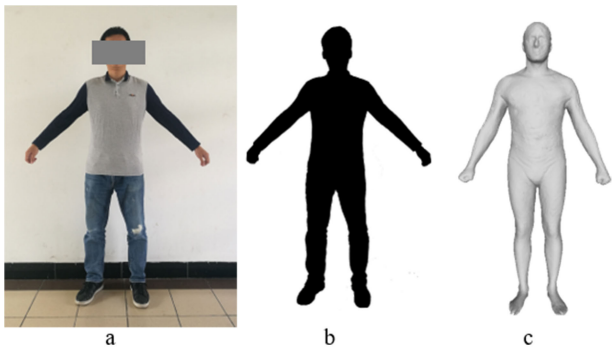


FIGURE 11. Result for a dressed person. From left to right: real photo, mask image, predicted 3D mesh.

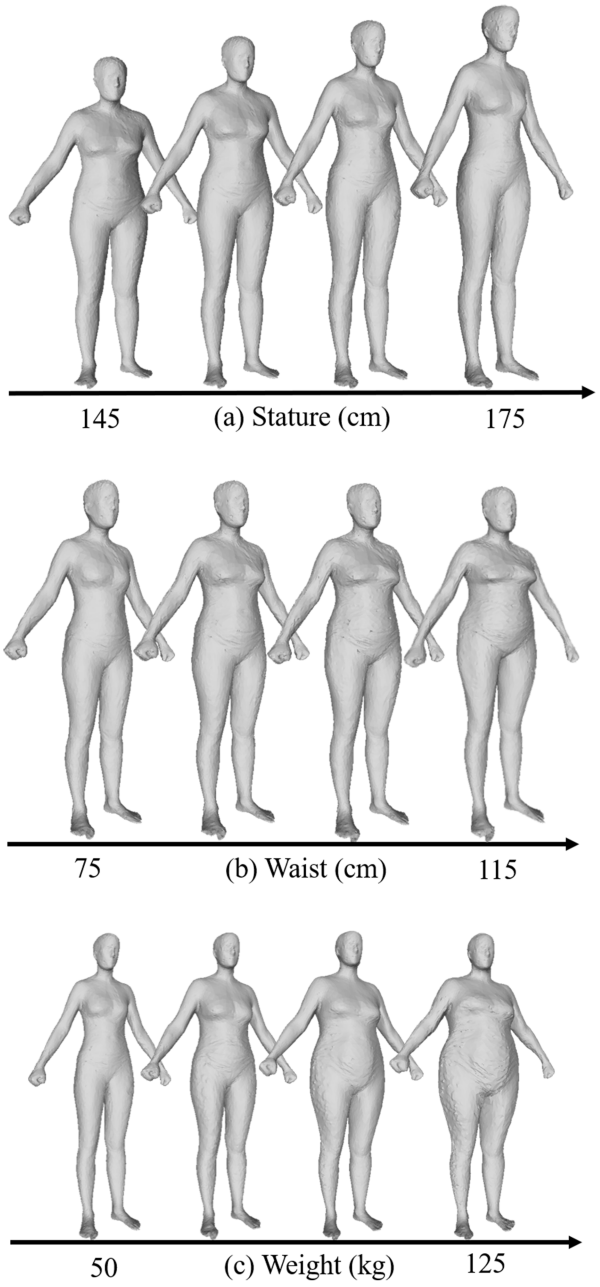


FIGURE 12. Anthropometric body design by varying the input measurements.

D. APPLICATION: ANTHROPOMETRIC BODY DESIGN

Through previous experiments, we can see that the proposed non-parametric method has a significant improvement in mannequin reconstruction from the previous experiments. In this subsection, we provide a potential application for anthropometric body design, which can be applied to facilitate product design and satisfy various individual preferences. A major challenge of anthropometric body design is that the measurements are usually related to each other, e.g., weight is related to the waist, waist is related to hip and chest, and stature may be related to weight as well. Our model has implicitly learned the complicated correlation among human models using deep learning, which can generate a plausible body shape by adjusting one or a few measurements. In our experiments, three anthropometric parameters: stature, weight, and waist are varied with different values to study the effect on the reconstructed body and design capability of the proposed approach. Fig.12 illustrates the experimental results based on the female mean shape. We find that the resulting body shape can reflect the variance of anthropometric measurements. The body gets taller as the input stature increases, and the increase in waist circumference also brings significant changes in shape. This indicates that our model has learned the complicated relationship among various

measurements, and the proposed approach is robust and effective for anthropometric body design. In addition, we only provide one potential application in this paper, and other applications related to 3D human body can try our method as well, which will generate more accurate body shape.

V. CONCLUSION

In this paper, we present a non-parametric approach for precise 3D virtual mannequin reconstruction based on graph convolution, which directly predicts the 3D location of each vertex on a mesh rather than the parameters of a particular parametric body model. The anthropometric measures are explicitly integrated into the network to improve the reconstruction accuracy further and make the reconstruction more controllable. On par with previous parametric methods, the non-parametric method can be extended to a parametric method easily by appending an MLP to regress the parametric space of the PCA model, and then to achieve 3D reconstruction. The proposed method can reconstruct a complete 3D shape from a single-view mask image as well. Extensive experimental results demonstrate that the proposed method significantly outperforms the previous state-of-the-art methods, both visually and in terms of anthropometric accuracy, and our method is effective and robust for 3D mannequin reconstruction. However, there are still some limitations to our approach. All the body shapes in our training set are standard A-pose, and our method thus cannot generate mannequins with various postures. This problem could be addressed by expanding and diversifying our training set further, which may improve the reconstruction accuracy of the obese shapes as well. Moreover, to reduce the noise on the non-parametric reconstructed surface, we introduce a smooth item to our loss function, which also leads the proposed non-parametric method not being able to generate rich surface details. Reconstructing precise and richly detailed 3D mannequin using graph convolution is our future work.

REFERENCES

[1] Y.-W. Cha, T. Price, Z. Wei, X. Lu, N. Rewkowski, R. Chabra, Z. Qin, H. Kim, Z. Su, Y. Liu, A. Ilie, A. State, Z. Xu, J.-M. Frahm, and H. Fuchs, "Towards fully mobile 3D face, body, and environment capture using only head-worn cameras," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 11, pp. 2993–3004, Nov. 2018.

[2] H. Fan, H. Su, and L. Guibas, "A point set generation network for 3D object reconstruction from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 605–613.

[3] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.

[4] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on X-transformed points," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 820–830.

[5] C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese, "3D-R2N2: A unified approach for single and multi-view 3D object reconstruction," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 628–644.

[6] P.-S. Wang, Y. Liu, Y.-X. Guo, C.-Y. Sun, and X. Tong, "O–CNN: Octree-based convolutional neural networks for 3D shape analysis," *TOGACM Trans. Graph.*, vol. 36, no. 4, pp. 1–11, Jul. 2017.

[7] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik, "End-to-end recovery of human shape and pose," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7122–7131.

[8] C. Lassner, J. Romero, M. Kiefel, F. Bogo, M. J. Black, and P. V. Gehler, "Unite the people: Closing the loop between 3D and 2D human representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6050–6059.

[9] E. Dibra, H. Jain, C. Oztireli, R. Ziegler, and M. Gross, "HS–Nets: Estimating human body shape from silhouettes with convolutional neural networks," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 108–117.

[10] M. Omran, C. Lassner, G. Pons-Moll, P. Gehler, and B. Schiele, "Neural body fitting: Unifying deep learning and model based human pose and shape estimation," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2018, pp. 484–494.

[11] G. Pavlakos, L. Zhu, X. Zhou, and K. Daniilidis, "Learning to estimate 3D human pose and shape from a single color image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 459–468.

[12] V. Tan, I. Budvytis, and R. Cipolla, "Indirect deep structured learning for 3D human body shape and pose prediction," in *Proc. Brit. Mach. Vis. Conf.*, 2017.

[13] H.-Y. Tung, H.-W. Tung, E. Yumer, and K. Fragkiadaki, "Self-supervised learning of motion capture," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5236–5246.

[14] Z. Ji, X. Qi, Y. Wang, G. Xu, P. Du, and Q. Wu, "Shape-from-mask: A deep learning based human body shape reconstruction from binary mask images," Jun. 2018, *arXiv:1806.08485*. [Online]. Available: https://arxiv.org/abs/1806.08485

[15] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: Shape completion and animation of people," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 408–416, 2005.

[16] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "SMPL: A skinned multi-person linear model," *TOGACM Trans. Graph.*, vol. 34, no. 6, pp. 1–16, Oct. 2015.

[17] J. Romero, D. Tzionas, and M. J. Black, "Embodied hands: Modeling and capturing hands and bodies together," *TOGACM Trans. Graph.*, vol. 36, no. 6, pp. 1–17, Nov. 2017.

[18] H. Joo, T. Simon, and Y. Sheikh, "Total capture: A 3D deformation model for tracking faces, hands, and bodies," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8320–8329.

[19] E. Dibra, H. Jain, C. Oztireli, R. Ziegler, and M. Gross, "Human shape from silhouettes using generative HKS descriptors and cross–modal neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4826–4836.

[20] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," Sep. 2016, *arXiv:1609.02907*. [Online]. Available: https://arxiv.org/abs/1609.02907

[21] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[22] N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris, "3D human pose estimation: A review of the literature and analysis of covariates," *Comput. Vis. Image Understand.*, vol. 152, pp. 1–20, Nov. 2016.

[23] M. Zollhöfer, P. Stotko, A. Görlitz, C. Theobalt, M. Nieβner, R. Klein, and A. Kolb, "State of the art on 3D reconstruction with RGB-D cameras," *Comput. Graph. Forum*, vol. 37, no. 2, pp. 625–652, May 2018.

[24] Z.-Q. Cheng, Y. Chen, R. R. Martin, T. Wu, and Z. Song, "Parametric modeling of 3D human body shape—A survey," *Comput. Graph.*, vol. 71, pp. 88–100, Apr. 2018.

[25] T. Lescoat, M. Ovsjanikov, P. Memari, J.-M. Thiery, and T. Boubekeur, "A survey on data-driven dictionary-based methods for 3D modeling," *Comput. Graph. Forum*, vol. 37, no. 2, pp. 577–601, May 2018.

[26] A. Weiss, D. Hirshberg, and M. J. Black, "Home 3D body scans from noisy image and range data," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1951–1958.

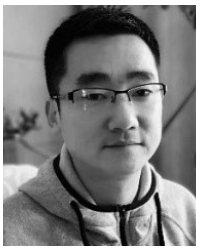
[27] Q. Zhang, B. Fu, M. Ye, and R. Yang, "Quality dynamic human body modeling using a single low–cost depth camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 676–683.

[28] T. Helten, A. Baak, G. Bharaj, M. Müller, H.-P. Seidel, and C. Theobalt, "Personalization and evaluation of a real–time depth–based full body tracker," in *Proc. Int. Conf. 3D Vis.*, Jun. 2013, pp. 279–286.

[29] F. Perbet, S. Johnson, M.-T. Pham, and B. Stenger, "Human body shape estimation using a multi-resolution manifold forest," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 668–675.

[30] F. Bogo, M. J. Black, M. Loper, and J. Romero, "Detailed full–body reconstructions of moving people from monocular RGB–D sequences," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2300–2308.

[31] Z. Liu, J. Huang, S. Bu, J. Han, X. Tang, and X. Li, "Template deformation–based 3-D reconstruction of full human body scans from low–cost depth cameras," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 695–708, Mar. 2017.
[32] P. Guan, A. Weiss, A. O. Balan, and M. J. Black, "Estimating human shape and pose from a single image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1381–1388.
[33] A. Zanfir, E. Marinoiu, and C. Sminchisescu, "Monocular 3D pose and shape estimation of multiple people in natural scenes: The importance of multiple scene constraints," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2148–2157.
[34] J. Huang, T.-H. Kwok, and C. Zhou, "Parametric design for human body modeling by wireframe-assisted deep learning," *Comput.-Aided Des.*, vol. 108, pp. 19–29, Mar. 2019.
[35] G. Varol, "BodyNet: Volumetric inference of 3D human body shapes," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 20–36.
[36] Z. Zheng, T. Yu, Y. Wei, Q. Dai, and Y. Liu, "DeepHuman: 3D Human Reconstruction from a Single Image," Mar. 2019, *arXiv:1903.06473*. [Online]. Available: https://arxiv.org/abs/1903.06473
[37] A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos, "Large pose 3D face reconstruction from a single image via direct, volumetric CNN regression," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1031–1039.
[38] A. S. Jackson, C. Manafas, and G. Tzimiropoulos, "3D human body reconstruction from a single image via volumetric regression," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018.
[39] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, "Pixel2mesh: Generating 3D mesh models from single RGB images," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 52–67.
[40] O. Litany, A. Bronstein, M. Bronstein, and A. Makadia, "Deformable shape completion with graph convolutional autoencoders," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1886–1895.
[41] N. Kolotouros, G. Pavlakos, and K. Daniilidis, "Convolutional mesh regression for single-image human shape reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4501–4510.
[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.
[43] Y. Zhong, D. Li, G. Wu, and P. Hu, "Automatic body measurement based on slicing loops," *Int. J. Cloth. Sci. Technol.*, vol. 30, no. 3, pp. 380–397, Apr. 2018.
[44] Trimesh. (2019). *Dawson-Haggerty*. [Online]. Available: https://trimsh.org/
[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[46] Y. Wu and K. He, "Group normalization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jul. 2019, pp. 3–19.
[47] T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll, "Video based reconstruction of 3D people models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8387–8397.
[48] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: New benchmark and state of the art analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3686–3693.
[49] S. Johnson and M. Everingham, "Clustered pose and nonlinear appearance models for human pose estimation," in *Proc. Brit. Mach. Vis. Conf.*, 2010, pp. 12.1–12.11.
[50] Y. Yang, Y. Yu, Y. Zhou, S. Du, J. Davis, and R. Yang, "Semantic parametric reshaping of human body models," in *Proc. 2nd Int. Conf. 3D Vis.*, vol. 2, Dec. 2014, pp. 41–48.
[51] K. Robinette, H. Daanen, and E. Paquet, "The CAESAR project: A 3-D surface anthropometry survey," in *Proc.2nd Int. Conf. 3-D Digit. Imag. Modeling*, Jan. 2003, pp. 380–386.
[52] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *Proc. NeurIPS Workshop*, 2017, p. 2.
[53] M. Mirza and S. Osindero, "Conditional generative adversarial nets," Nov. 2014, *arXiv:1411.1784*. [Online]. Available: https://arxiv.org/abs/1411.1784
[54] C.-H. Chu, Y.-T. Tsai, C. C. Wang, and T.-H. Kwok, "Exemplar-based statistical model for semantic parametric design of human body," *Comput. Ind.*, vol. 61, no. 6, pp. 541–549, Aug. 2010.



HAOYANG XIE received the B.S. degree in computer science from Zhengzhou University, Henan, China, in 2009, and the M.S. degree in information assurance from Fort Hays State University, USA, in 2011. He is currently pursuing the Ph.D. degree with Donghua University, Shanghai, China. His research interest includes 3D body scanning technology, digital geometry processing, machine learning, and deep learning in 3D.

YUEQI ZHONG received the master's degree in textile engineering from Xi'an Polytechnic University, in 1982, and the Ph.D. degree in textile production design from Donghua University, in 2001. He was a Postdoctoral Fellow with the University of Texas at Austin, from 2001 to 2005. He is currently a Professor with the College of Textiles, Donghua University. His research interests include virtual garments, physically based cloth simulation, geometry processing, interactive techniques, and deep learning.

ZHICAI YU received the master's degree in textile material and textile design from the Zhongyuan University of Technology, Zhengzhou, China, in 2015. He is currently pursuing the Ph.D. degree with Donghua University, Shanghai, China. His major is digital textile engineering. His work is related to 3D scanning technology, image processing, machine learning, deep learning, and fabric inquiry and analysis.

AZMAT HUSSAIN received the master's degree in fiber technology from the University of Agriculture, Faisalabad, Pakistan, in 2004, and the Ph.D. degree in textile engineering from Donghua University, Shanghai, China, in 2019. He is currently a Lecturer with Bahuddine Zakariya University, Multan, Pakistan. His research interests include 3D body scanning, digital geometry processing, and deep learning.