Anthropometric Parameter Measurement from Equivariant Multi-view Images

Lijiao Deng¹,ᵃ, Tingman Yan¹,ᵇ and Qunfei Zhao¹,ᶜ
¹Department of Automation, Shanghai Jiaotong University

E-mails: ᵃdenglijiao@sjtu.edu.cn, ᵇmeander@sjtu.edu.cn, ᶜzhaoqf@sjtu.edu.cn

Abstract. In this paper, we propose an anthropometric parameter measurement method that any customized parameter can be measured online by the pre-selected endpoints on the reconstructed 3D body models of equivariant multi-view images. The method includes 3D body model reconstruction, anthropometric parameter measurement, and parameter modification. In 3D body model reconstruction, we detect and segment the human body from its background and reconstruct a generative 3D body model from the segmented image with deep learning. And then we measure anthropometric parameter on the reconstructed 3D body model of each view. Before parameter measurement, we manually pre-select endpoints associated with all anthropometric parameters on the reconstructed 3D body model since all vertices of the reconstructed body model are ordered. However, the information of a single-view image is insufficient and the measurement result is varied regularly by the view changes. To improve the measurement accuracy, we design a convolutional neural network in the last step which can regress more accurate anthropometric parameters from equivariant multi-view measurements. Experimental results on the representative dataset demonstrate that the proposed method can measure planar and spatial anthropometric parameters automatically with comparable performance.

1. Introduction
The anthropometric parameters measurement has a wide range of applications on virtual try-on and health monitor. Traditional tape-based methods are not only time-consuming but infeasible for some occasions. With the growth of technology, the computer-based anthropometric parameters measurement attracts the attention of many researchers. The computer-based measurement methods can be divided into the 2D (image-based) measurement where images of the person are taken, and 3D (scan-based) measurement where a complete person’s body is scanned to reconstruct exact model. Compared to the image-based measurement system, the scan-based measurement is more accurate, but it is expensive and not portable in actual production [1].

The image-based measurement system usually has four steps: image acquisition, image segmentation, feature points extraction, and parameter measurement. Image acquisition is the primary step which employs a camera to take photos of the person standing in front of the backboard. Image segmentation is to extract the person from its background and common methods are exploiting edge detection or HSV color space [2], [3]. Recently, the powerful convolution neural networks (CNN) are applied to detect and extract the person from its background. Feature points determine the start and end points of the related anthropometric parameter and the distance between the feature points are employed to measure anthropometric parameter [4]. Common methods of feature point extraction are
manually extraction or automatically extraction. The first category is to label feature points manually on new input images [5] or new reconstructed 3D body models [6]. Another category is to extract feature points automatically from new input images [7] or new reconstructed 3D body models [8], [9]. Compared to these feature points extraction methods, our method selects all endpoints manually in advance which will be imported automatically in the measurement process.

Parameter measurement usually includes spatial and planar parameter measurement and the spatial parameter is more complex to measure than planar parameter in the image-based measurement system. Common methods of spatial parameter measurements are to estimate circumference length from the reconstructed body model which includes the geometric body model [10] and surface body model [6]. Compared to these methods, our method employs a generative body model called Skinned Multi-Person Linear model (SMPL) [11] to represent the human body in the image. Spatial anthropometric parameters can be measured when 3D body models are reconstructed from images and common methods including geodesic distance [9], ellipse girth [3], and intersection curves integral [6]. Compared to these methods, we measure spatial parameters as the geodesic distance between the pre-selected endpoints in the reconstructed generative 3D body model.

In this paper, we propose an image-based anthropometric parameter measurement method that any customized parameter can be measured by the pre-selected endpoints on the reconstructed body models of multi-view. The main steps of our method include 3D body model reconstruction, anthropometric parameter measurement, and parameter modification. Additionally, the whole process of our work is fully automatic without any landmarks pasted on the body and the generative body models SMPL are employed to represent the person in the image.

2. Measurement from images
To obtain accurate anthropometric parameter of human body in the image, we propose a systematic strategy in this part. Before parameters measurement, we reconstruct 3D body models from multi-view segmented images. Then anthropometric parameters will be measured by the Euclidean distance and geodesic distance according to the pre-selected endpoint on the reconstructed 3D body models. Moreover, a convolutional neural network is employed to modify multi-view measurement results since the information of a single-view image is insufficient. Figure 1 shows an overview of the proposed framework.

![Figure 1](image.png)

**Figure 1.** The overview of the proposed framework where we segment multi-view images $I$ and reconstruct 3D body models $M$ from the segmented images $S$. Then $M$ and the pre-selected endpoints $F$ are imported to the measurement step to get multi-view anthropometric parameters $P$. Finally, the modification step will modify $P$ to obtain more accurate anthropometric parameters $P^*$ of the person in the image.

2.1. 3D body model reconstruction
The goal of this section is to obtain multi-view images $I$ in an indoor scene and reconstruct 3D body models $M$ from segmented images $S$. To capture multi-view images and control the rotation angles, we design an image acquisition equipment as shown in figure 2. The image acquisition equipment mainly includes the turntable, mobile, and computer. The turntable can rotate every fixed degree and a green backboard is set behind the turntable and a moderate distance from the turntable is a mobile mounted by a tripod. The multi-view images taken by the mobile are defined as follows:

$$I = \{i_k | k = 1,2,\ldots,N\} ,$$

where $I$ is the set of $N$ views images and $i_k$ is the $k$th view image.

**Figure 2.** The image acquisition equipment including the turntable, mobile, and computer.

Image segmentation and 3D body reconstruction are two key steps that can avoid the background interference and reconstruct generative 3D body models. Firstly, multi-view images are segmented by DeepLab V3 [12]. The DeepLab V3+ combines the spatial pyramid pooling module and encoder-decoder structure to encode multiple-scale contextual information and capture sharper object boundaries. The segmentation process is defined as:

$$S = \{s_k = D(i_k) | k = 1,2,\ldots,N\} ,$$

where $s_k$ is the segmented image of each view, $D(.)$ is the DeepLab V3+ network, and $S$ is the set of multi-view segmented images. Another key step is the 3D body reconstruction that HMR [13] is exploited. HMR is an end to end network that can reconstruct a full 3D mesh of a human body from a single RGB image. The reconstruction process is defined as:

$$M = \{m_k = H(s_k) | k = 1,2,\ldots,N\} ,$$

where $m_k$ is the reconstructed 3D body model in the $k$th view, $H(.)$ is HMR network and $M$ is the set of multi-view reconstructed body models.

### 2.2. Anthropometric parameters measurement

The measurement step reads in 3D body models of multi-view images and pre-selected endpoints and output multi-view anthropometric parameters. The step includes end-points selection and anthropometric parameters computation. The pre-selected endpoints will be imported into the measurement process and all anthropometric parameters measured in our work are defined according to [14]. We only measure a total of 17 representative anthropometric parameters in the experiment that can meet the needs of general clothing manufacturing and they can be customized by the user if necessary. The definition and name of all anthropometric parameters is shown in figure 3(a) and table 1. All endpoints are selected in advance since the vertices of the reconstructed 3D body model are ordered. To determine the position of endpoints, we design a graphical interface that can show all vertices on the reconstructed 3D body model as shown in figure 4(a). The endpoints are defined as follows:

$$F = \{f_j | j = 1,2,\ldots,V\} ,$$

where $V$ is the total number of all selected endpoints ($V=32$), and the position and index of each endpoint $f_j$ is shown in figure 3(b).
Figure 3. (a) The definition of all anthropometric parameters. (b) The location of all endpoints.

Table 1. Names of all anthropometric parameters in figure 3

| Index | Name          | Index | Name          | Index | Name          |
|-------|---------------|-------|---------------|-------|---------------|
| 1     | stature       | 7     | waist height  | 13    | head circumference |
| 2     | head length   | 8     | crotch height | 14    | neck circumference |
| 3     | cervical height | 9    | knee height   | 15    | chest circumference |
| 4     | acromial height | 10   | foot length   | 16    | waist circumference |
| 5     | axilla height | 11    | arm length    | 17    | hip circumference |
| 6     | chest breadth | 12    | acromial breadth |      |                |

Anthropometric parameter $p_k$ of the $k$th view body model can be obtained when the $k$th view 3D model $m_k$ and pre-selected endpoints $F$ are determined and the anthropometric parameters $P$ are defined as follows:

$$P = \{ p_k = L(m_k, F) | k = 1, 2, ..., N \} ,$$

where $L(.)$ is the distance function. The anthropometric parameters $p_k$ of each view include the linear length $l_k$ and circumference length $c_k$:

$$p_k = \{ l_k^i, c_k^i \} ,$$

The reconstructed 3D body model $m_k$ of the $k$th view consists of the vertex set $\Omega_k$ and edge set $T_k$:

$$m_k = \{ \Omega_k, T_k | \Omega_k \in R^{2n}, T_k \in R^{2n \times (n-2)} \} ,$$

where $n$ is the number of vertices in $m_k$ and $n=6890$ in the SMPL model. For linear length $l_k^i$, we employ the Euclidean distance to compute the length between endpoints. Compared to the linear length $l_k^i$, the measurement of circumference length $c_k^i$ is more complicated and we employ the geodesic distance to compute all circumference lengths. Geodesic distance is a common method to obtain the shortest path between points on the surface of the 3D model and what we exploit in this paper is the Subdivision Dijkstra algorithm which is improved according to the Dijkstra algorithm. The process of the Subdivision Dijkstra algorithm mainly includes two steps which are the vertex insertion and the shortest path searching.

The first step of circumference length measurement is vertex insertion which is to insert $\lambda$ vertices equally in each edge of the edge set $T_k$:

$$\Omega_k^* = [ \Omega_k, \lambda T_k ] , \Omega_k^* \in R^{3n \times (n-2)+\lambda} ,$$

where $\Omega_k$ and $T_k$ are defined in (7), and $\Omega_k^*$ is the vertex set after vertex insertion as shown in figure 4(c). And then we use the Dijkstra algorithm to search the shortest path between endpoints as shown in figure 4(d). For each anthropometric parameter of $p_k^c$, the circumference length is the sum of the
geodesic distance of the three paired endpoints associated with the parameter. The three paired endpoints $\Gamma_j$ are defined as follows:

$$\Gamma_j = \{\tau_{ji}, \tau_{ij}, \tau_{ji}\} = \{(f_j^1, f_j^2, f_j^3), (f_j^2, f_j^3, f_j^1), (f_j^3, f_j^1, f_j^2) \mid f_j^1, f_j^2, f_j^3 \in F\},$$

(9)

where $f_j^1, f_j^2, f_j^3$ are three endpoints associated with the circumference length which is shown in figure 4(e). Additionally, the circumference length of $p_k^*\Gamma$ is defined as follows:

$$p_k^*\Gamma = \{\ell_j = \sum_{i=1}^3 Dijkstra(\Omega_i^\tau, \tau_{ji}) \mid j \in 13, 14, \ldots 17, \tau_{ji} \in \Gamma_j \},$$

(10)

where $Dijkstra(.)$ is the shortest path searched by the Dijkstra algorithm of each paired endpoints $\tau_{ji}$ on the new vertex set $\Omega_k^\tau$.

2.3. Anthropometric parameter modification

Even though the anthropometric parameters of a single-view image have been obtained in above, however, we find that the information of a single-view image is insufficient and the measurement results are regularly changed by the change of view. Therefore, we design a modification step to combine the measurement results $P$ of multi-view images to get a more reliable result $P'$ of the human body. We propose a convolutional neural network with several layers of convolutional kernels to extract feature maps from multi-view measurement results $P$.

The network consists of six convolutional layers with 256 neurons each with a batch normalization layer in between, followed by a ReLU layer, and a fully-connected layer with 256 neurons is connected behind all convolutional layers as shown in figure 5. Moreover, to reduce the number of parameters and decrease the over-fitting problem in the training process, the output of the second, fourth and sixth convolutional layers are max pooled. The process of each layer is defined as:

$$o_i = \begin{cases} c(o_{i-1}), & i = 1, 3, 5 \\ g(c(o_{i-1})), & i = 2, 4, 6 \\ h(o_{i-1}), & i = 7 \end{cases},$$

(11)

where $c(.)$ is the function of the composite layer including the convolutional layer, batch normalization layer and ReLU activation layer. $g(.)$ is the function of the max-pooling layer and $h(.)$ is the function of a fully connected layer. Additionally, $o_0 = P, o_7 = P', P \in R^{N*17}, P' \in R^{N*17}$ are the input and output of the convolutional neural network.

**Figure 4.** Steps of circumference length measurement. (a) The point cloud of reconstructed model. (b-e) The process of circumference computation.

**Figure 5.** The architecture of the CNN. The input of the is a map of $N*17$ dimensions and the output of the network is a vector of 17 variables.
3. Results

3.1. Experiment setup

1) Datasets: We collect data of a dozen people who are representative of height and shape including 4 females and 11 males, and the height ranges from 161cm to 187cm. Moreover, we set N=24 in the experiment which means that the turntable rotates every 15 degrees and a total of 24 images will be taken every turn. The person in tights should stand statically and keep their arms slightly away from the torso which helps us determine the circumference length. The ratio between the training dataset and validation dataset is 4:1.

2) Parameter Settings: To balance the accuracy and cost of computation in the measurement process, we set \( \lambda = 3 \) in formula (8). The learning rate of the CNN is 0.001 and the optimizer is RMSProp [15]. The train epoch is 2000 and the network will be validated every 10 epochs in the training process, besides, the batch size is 5 in the training and validation process. The step size of the max pooling function is 2, which means that the size of the output of each convolutional layer will drop to half of the original size. Besides, the kernel size of the convolutional layer is 3 and the padding is 1.

3.2. Method of evaluation

The evaluation methods in this paper are accuracy and mean difference that the accuracy is to judge the distribution of the result and the mean difference is to evaluate the deviation between the result and the true value measured manually. The accuracy is defined as follows:

\[
\text{Accuracy} = \frac{1}{17 \cdot N_v} \sum_{i=1}^{N_v} \sum_{j=1}^{17} r_{yt} = \begin{cases} 
1 & |y_j - t_j| \leq \alpha \\
0 & \text{otherwise} 
\end{cases},
\]

where \( \alpha \) is a threshold to determine the scope of the ground truth. Moreover, \( N_v \) is the number of validation datasets and the number of anthropometric parameters measured in our work is 17. Furthermore, \( t_j \) and \( y_j \) are the true anthropometric parameter measured manually and output of our work. Another evaluation method is mean difference including the Mean Absolute Difference (MAD), and Mean Relative Difference (MRD) which can evaluate the error of each anthropometric parameter and the definition are as follows:

\[
\text{e}_{\text{MAD}}(j) = \frac{1}{N_v} \sum_{i=1}^{N_v} |y_j - t_j| , j = 1,2,\ldots,17,
\]

\[
\text{e}_{\text{MRD}}(j) = \frac{1}{N_v} \sum_{i=1}^{N_v} \frac{|y_j - t_j|}{t_j} , j = 1,2,\ldots,17.
\]

3.3. Discussion and result

1) Performance on Test Samples: To demonstrate the whole performance of our method, we choose three samples from the validation dataset and compare the measurement results with the values measured manually. The result is shown in Table 2 where \( t_i \) is the true anthropometric parameter of the \( i \)th person and \( y_i \) is the output of our method. The experimental results show that the method proposed by us has higher accuracy in height, head length, and knee height whose \( y_i \) and \( t_i \) have little difference less than 1cm.

| Name         | t1(m) | y1(m) | t2(m) | y2(m) | t3(m) | y3(m) |
|--------------|-------|-------|-------|-------|-------|-------|
| stature      | 1.66  | 1.65  | 1.74  | 1.74  | 1.70  | 1.71  |
| head length  | 0.23  | 0.23  | 0.23  | 0.23  | 0.23  | 0.23  |
| cervical height | 1.40  | 1.38  | 1.45  | 1.46  | 1.45  | 1.43  |
| acromial height | 1.36  | 1.35  | 1.39  | 1.42  | 1.42  | 1.39  |
2) Performance of Each Anthropometric Parameter: To demonstrate the mean error of each anthropometric parameter, we compare the MAD and MRD of each anthropometric parameter both in the train and test dataset as shown in table 3. We can find that the MAD of head length and knee height are smallest both in the train and test dataset and the MRD of them is also acceptable. To sum up, the average MAD is 0.996cm in the train dataset and 1.94cm in the test dataset which is very small in the image-based measurement system.

Table 3. The MAD and MRD of the anthropometric parameters.

| Name           | MAD(cm) (training) | MRD (training) | MAD(cm) (validation) | MRD (validation) |
|----------------|--------------------|----------------|----------------------|------------------|
| stature        | 1.46               | 0.84%          | 2.22                 | 1.31%            |
| head length    | 0.38               | 1.65%          | 0.34                 | 1.47%            |
| cervical height| 1.29               | 0.88%          | 1.83                 | 1.28%            |
| acromial height| 1.26               | 0.89%          | 2.16                 | 1.55%            |
| axilla height  | 1.04               | 0.80%          | 1.68                 | 1.32%            |
| chest breadth  | 0.58               | 1.78%          | 1.34                 | 4.43%            |
| waist height   | 1.13               | 1.09%          | 2.06                 | 2.04%            |
| crotch height  | 0.76               | 1.00%          | 1.09                 | 1.41%            |
| knee height    | 0.49               | 1.04%          | 0.38                 | 0.83%            |
| foot length    | 0.39               | 1.61%          | 1.26                 | 5.37%            |
| arm length     | 0.59               | 0.83%          | 1.43                 | 2.06%            |
| acromial breadth| 0.73              | 1.73%          | 1.50                 | 3.70%            |
| head circumference| 0.76          | 1.34%          | 2.23                 | 4.17%            |
| neck circumference| 0.73            | 2.05%          | 2.59                 | 7.25%            |
| chest circumference| 1.95         | 2.14%          | 3.49                 | 3.79%            |
| waist circumference| 2.07        | 2.43%          | 4.14                 | 4.84%            |
| hip circumference| 1.34            | 1.34%          | 3.26                 | 3.18%            |
| average difference| 0.996      | 1.38%          | 1.94                 | 2.94%            |

3) Accuracy Analysis: To demonstrate the distribution of the anthropometric parameters measured by our method, we show the accuracy under different threshold \( \alpha \) both in the training and validation dataset. In table 4, the accuracy is 97% in the training dataset and 94% in the validation dataset when \( \alpha = 5cm \). When \( \alpha > 5cm \), the accuracy of the validation dataset and training dataset are almost 100% which means that the difference between the results measured by our method with ground truth is less than 10cm.
Table 4. Relationship between the threshold and the accuracy

| α        | 2cm | 3cm | 5cm | 8cm | 10cm |
|----------|-----|-----|-----|-----|------|
| training dataset | 85% | 90% | 97% | 100%| 100% |
| validation dataset | 62% | 81% | 94% | 97% | 98% |

4) Discussion: The method proposed by us can measure spatial anthropometric parameters from images without any image calibration and the result is comparable with other methods. The whole performance of our method is acceptable according to the experimental result even though there are some parameters such as chest circumference are not ideal. There are some reasons to explain the average MAD is 2cm and most measurement result falls within the ground truth of 3cm in our work. (1)The ground truth of anthropometric parameters in our work is measured manually which will bring artificial error. (2) The training dataset of the Modification network is insufficient.

4. Conclusions

We have proposed an anthropometric parameter measurement method that any customized parameter can be measured online by the pre-selected endpoints on the reconstructed 3D body models of equivariant multi-view images. There are two key components of the method which are the measurement and modification of anthropometric parameters. In the measurement step, we pre-select endpoints associated with all anthropometric parameters before the measurement. Moreover, we design a convolutional neural network in the modification step to regress a more precise result according to measurement results of multi-view images since the information of single-view image is insufficient. Experimental results demonstrate that the method proposed by us can measure spatial parameters from multi-view images with comparable performance.

References

[1] S Luo 2017 Application Research on 2D Image-based Non-contact Anthropometric Technology in Clothing E-commerce 7th International Conference on Applied Science, Engineering and Technology, ICASET, Advances in Engineering Research 122 pp 112-128.
[2] Stancic, I, Supuk, T, Cecic and M 2009 Computer vision system for human anthropometric parameters estimation WSEAS Transactions on Systems 8 pp 430-439.
[3] KOHLSCHÜTTER T and HEROUT P 2012 Automatic human body parts detection in a 2d anthropometric system Lecture Notes in Computer Science 7424 pp 536-544.
[4] P R M Jones 1997 Three-Dimensional Surface Anthropometry: Applications to the Human Body Optics and Lasers in Engineering 28 pp 89-117.
[5] C Barron and I A Kakadiaris 2000 Estimating anthropometry and pose from a single image Computer Vision and Pattern Recognition pp 13-15.
[6] Li et al. 2013 Anthropometric body measurements based on multi-view stereo image reconstruction Thirty Fifth Annual Int. Conf. On Engineering in Medicine and Biology Society pp 366-369.
[7] Jheanel E Estrada, Alexies A, Benedito, John Lester D and Belicano 2017 Digital anthropometry for human body measurement on android platform 2017 IEEE 2nd International Conference on Signal and Image Processing pp 262-265.
[8] T Xiaohui, P Xiaoyu, L Liwen and X Qing 2018 Automatic human body feature extraction and personal size measurement Journal of Visual Languages and Computing 46 pp 9-18.
[9] Tsoli, A, Loper, M, Black and M J 2014 Model-based anthropometry: Predicting measurements from 3D human scans in multiple poses Proc. Winter Conference on Applications of Computer Vision pp 83-90.
[10] C Y Hung, P Witana, and S Goonetilleke 2004 Anthropometric measurements from photographic images 7th Int. Proc. on Work with computing system pp 104-109.
[11] M Loper, N Mahmood, J Romero, G Pons-Moll and M J Black 2015 SMPL: A skinned multi-person linear model ACM Transactions on Graphics 6 pp 248:1-16.
[12] Liang Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff and Hartwig Adam 2018 Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation *the European Conference on Computer Vision.*

[13] A Kanazawa, M J Black, D W Jacobs and J Malik 2018 End-to-end recovery of human shape and pose *Computer Vision and Pattern Recognition* pp 7122-31.

[14] Centers for Disease Control and Prevention https://www.cdc.gov/niosh/topics/anthropometry/default.html.

[15] G Hinton, N Srivastava and K Swersky 2012 Lecture 6a overview of mini-batch gradient descent Coursera Lecture slides https://class.coursera.org/neuralnets-2012-001/lecture.