An Anthropometric Dimensions Measurement Method Using Multi-pose Human Images with Complex Background

Yidong Chen and Yigang Wang

1Hangzhou Dianzi University, Hangzhou 310018, China
*yigang.wang@hdu.edu.cn

Abstract. Automatically acquire anthropometric dimensions using two-dimensional images, providing a convenient, effective and low-cost way for measuring anthropometric parameters. In the existing methods, the background and posture of the user are highly restricted, and the anthropometric dimensions obtained by using the ellipsoid model is not accurate enough. On this observation, we propose a new method for measuring the anthropometric dimensions. This method integrates the deep learning method with the deeplabv3 and openpose frameworks, and introduces the contour matching method to get anthropometric dimensions instead of using ellipsoid model. Experiments show our method can effectively dealing with the complex background and posture problems, and improving the accuracy.

1. Introduction
Anthropometry focuses on anthropometric measurements and observation methods, and explores the characteristics, types, variations, and development of the human body through overall and local measurements. Anthropometric measurement not only plays an important role in the fields of medicine, sports, monitoring, etc. It is also the most critical technology in the field of virtual fitting, so an accurate and convenient anthropometric method has become one of the current research hotspots.

In the past, traditional manual measurement methods used mainly tools such as line-of-sight meters, anthropometers, height gauges and tape measures to collect anthropometric data. This measurement method is simple and easy to operate but the accuracy depends on the operator's expertise and subjective [1].

In order to avoid the influence of subjective factors, Cui Y et al[2] uses multiple depth image registration through three-dimensional scanning equipment such as Kinect, and performs measurement after human body modelling. However, limited by related technologies and equipment, the method is costly and cumbersome, and often faces problems such as insufficient equipment accuracy and lack of depth information.

Some systems[3] use two-dimensional images to obtain planar data, and the ellipsoid model is used to calculate the secondary dimensions. This method is cheaper, more efficient, and more convenient than a three-dimensional scanning system. However, the method of obtaining the secondary dimensions is very limited by posture and background of the subject in the images, and the anthropometric dimensions obtained by the ellipsoid model is only applicable to a relatively standard body type.

Aiming at solving these problems, this paper proposes a multi-pose anthropometric dimensions measurement method based on contour matching in complex background.
The method first extracts the contour of the human body in complex background through deeplabv3[4], then divides the human body points through the openpose[5], eliminates the influence of the subject's posture, and finally retrieves the most similar human body model from the database through silhouette matching. Thereby the subject parameters are estimated.

2. Related Work
In recent years, there have been many studies on methods for measuring human body dimensions for two-dimensional images. Most of the research uses the front and side silhouette of the human body to extract feature points and obtain partial linear dimensions, and then use the ellipsoid model to approximate the human body dimensions.

In Lin and Wang's research[6], they proposed an automatic feature extraction algorithm to analyse the direction difference between two connected edge pixels of the human contour, which calculates the difference in direction between two connected edge pixels. The feature is characterized by the absolute difference ek, and if ek=2, the edge pixels whose direction changes by 90° are used as the feature points.

\[ e_k = |d_j + 1 - d_j| = 2 \]  

In the process of feature point detection, since the detection of each FP (Feature Point) depends on the former one, it is difficult to ensure the accuracy of the extraction dimensions.

In response to the lack of precision and dimensions in the algorithm proposed by Lin & Wang, Murtaza Aslam proposed a technique for automatically measuring anthropometric dimensions.

Based on the Lin & Wang algorithm, the technology divides the human body and performs regionalized FP detection. This technology can obtain more than three times anthropometric dimensions and significantly reduce the MAD value for most dimensions.

However, this method still has some shortcomings. Firstly, the measurement background must be pure blue, and the subject is required to open 45° with both hands. At the same time, the dimensions obtained by the ellipsoid model is only suitable for the standard human body.

3. Our Method
3.1. Silhouette detection and extraction

This experiment uses the deeplabv3 network to extract the contours of the human body. The network structure is shown in Figure 1.

Deeplabv3 is a kind of neural network that uses the cavity convolution and full connection condition random field and proposes a pyramid-shaped atrous spatial pyramid pooling (ASPP) in the spatial dimension, which greatly improves the effect of segmentation network [7].

Compared with the previous Murtaza Aslam method, which needs to be photographed under a strict blue background, the method greatly improves the experimental conditions, reduces the
disturbance of complex background, thereby enhances the efficiency of extracting human silhouettes in complex background.

The effect of deeplabv3 in complex background is shown in Figure 2:

![Figure 2 Silhouette detection and extraction in complex background by Deeplabv3](image)

3.2. Feature extraction and segment

![Figure 3 backbone](image)

![Figure 4 Network structure (Brand1 and Brand2)](image)

In the experiments of Lin & Wang and Murtaza Aslam, there are strict requirements for posture.

We performed human feature point extraction through openpose, which greatly reduced the impact of the posture factor.

OpenPose is a real-time multi-person keypoint detection library open sourced by CMU. It implements a real-time system for detecting human, hand and face key points (130 key points).

The specific process is to first input a picture, the picture through a backbone (such as vgg, resnet, mobilenet) as shown in Figure 3 to get F(F is the feature maps output through the first 12 layers of the
network), and then through the stage shown in Figure 4, each stage has two branches One for detecting heatmaps and one for detecting vectmaps.

With heatmap and vectmap, you can know all the key points in the picture, and then point the points to everyone through PAFs (Part Affinity Fields, Part Affinity Fields refers to the position and orientation information stored in the limb area, represented by the 2D vector field).

Branch1 and Branch2 network Loss layer:

Brand1 Loss:

\[ f_S^t = \sum_{j=1}^{J} \sum_{p} W(p) \| S^t_j(p) - S^*_j(p) \|_2^2 \]  \hspace{1cm} (2)

Brand1 Loss:

\[ f_L^t = \sum_{c=1}^{C} \sum_{p} W(p) \| L^t_c(p) - L^*_c(p) \|_2^2 \]  \hspace{1cm} (3)

W§ is the weight, W§ = 0 when position p is not marked, avoiding prediction errors during training.

\( S^t_j(p) \) is the confidence value of point p on the j-th body part map of Branch1 outputted in the t-th network.

\( L^t_c(p) \) is the vector of point p on the c-th part affinity vector field that Branch2 outputs in the t-th network.

\( L^*_c(p) \) is the vector of the p-point of the c-th body part of the ground truth on the affinity vector field.

In order to solve the gradient disappearance, periodically supplement the gradient \( f' \) :

\[ f = \sum_{i=1}^{T} \left( f_S^i + f_L^i \right) \]  \hspace{1cm} (4)

The human body feature points recognized by openpose are shown in Figure 5 below. After the recognition is completed, more human body feature points are obtained according to the basic feature points of the human body, thereby performing human silhouette segmentation.

![Image](image.png)

**Figure. 5** Identify anthropometric features

### 3.3. Silhouette matching

The ellipsoid model has a good effect on the standard human body, but there is a large deviation in the fat or slightly deformed human body. This experiment used Makehuman to create a human body database with 20,000/cm, so that we can match the front and side contours of the human body to the contours of the human body in the database. In order to find the most similar human data from the database, the difference between the contour of the human body model and the contour of the original 2D image was analyzed by using the contour change comparison method [8]. The contour change Ea
representing the non-overlapping regions of the two contours is used to indicate the similarity level between the contour of the human body model and the original 2D image [9].

Ea can be calculated by the following equation:

\[
E_a = \frac{\sum (T(i,j) - D(i,j))}{\sum T(i,j)} + \frac{\sum (\bar{T}(i,j) - \bar{D}(i,j))}{\sum \bar{D}(i,j)}
\]  

\(T(i, j)\) is the value of the pixel if the pixel at \((i, j)\) is within the human body model.

\(\bar{T}(i, j)\) is the value of the pixel if the pixel at \((i, j)\) is within the human body model.

\(D(i, j)\) is the value of the pixel if the pixel at \((i, j)\) is the foreground pixel.

\(\bar{D}(i, j)\) is the value of the pixel if the pixel at \((i, j)\) is the background pixel.

The human contours of the subject is calculated with the contours of the human body in the database, and the optimal model is selected based on Ea to estimate the human body dimensions.

4. Analysis

The whole experimental process includes: human silhouette extraction, human feature point extraction, and human body contour matching in the database. The corresponding flow chart is as Figure 6.

![Experimental flow chart](image)

First of all, this experiment uses two cameras to shoot the human body at the same time. The two cameras are placed at 90°, the camera angle is parallel to the ground, and the subject is required to wear tights. In order to remove the influence of the arm factor, there is no arm extension in the side image required to be photographed.

We used the deeplabv3 to process the front and side of the subject to obtain the contour of the human body, thereby extracting the basic feature points of the human body using the open pose. From these basic feature points, we will complete the regionalization of the silhouette.

This experiment uses Makehuman human body modeling software to make a large number of models based on human body height, the order of magnitude is 20000/cm, the height range is 160cm-180cm, and there are corresponding frontal photos, side photos, and corresponding 15 human body dimensions. For each subject, we used the contour matching algorithm mentioned in 3, and finally selected the Ea minimum model as the matching result, and compared its corresponding data with the subject data.

We also performed traditional manual measurements to verify the effectiveness of our approach. Twenty subjects, aged 18-27 years, were tested and each subject was measured by a professionally trained staff to obtain 15 specific dimensions (including Height, shoulder width, sleeve length, Chest height, front waist section, back waist section, chest width, bp distance, back width, bust, collar, hip circumference, abdominal circumference, sleeve cage, arm circumference, thigh root circumference). The measured dimensions are accurate to the millimeter.

The data obtained is as Table 1:
Table 1. Comparison of the proposed method and the previous methods of anthropometric dimensions

| Dimension         | code | Proposed method PA | Murtaza Aslam method PA | Lin&Wang method PA |
|-------------------|------|--------------------|-------------------------|--------------------|
| Shoulder width    | C1   | 94.76%             | 95.00%                  | 82.51%             |
| Sleeve length     | C2   | 95.17%             | 83.80%                  | 89.65%             |
| Chest height      | C3   | 99.02%             | 97.68%                  | 92.21%             |
| Front waist section | C4  | 93.70%             | 95.88%                  | 85.14%             |
| Back waist section | C5  | 95.19%             | 89.79%                  | 88.02%             |
| Chest width       | C6   | 96.84%             | 94.36%                  | 86.38%             |
| Bp distance       | C7   | 95.44%             | 93.56%                  | 90.18%             |
| Back width        | C8   | 96.90%             | 98.04%                  | 91.93%             |
| Bust              | C9   | 92.91%             | 98.15%                  | 91.54%             |
| Collar            | C10  | 96.56%             | 94.62%                  | 84.50%             |
| Hip circumference | C11  | 98.08%             | 94.33%                  | 88.67%             |
| Waistline         | C12  | 95.65%             | 90.48%                  | 95.13%             |
| Sleeve cage       | C13  | 94.28%             | 82.27%                  | 84.76%             |
| Arm circumference | C14  | 95.71%             | 93.70%                  | 88.33%             |
| Comprehensive result | CR | 95.59%             | 92.98%                  | 88.50%             |

The PA of jth dimension \(\left(P_{A_j}\right)\) [10] can be calculated by the following equation:

\[
P_{A_j} = \left(\frac{k}{n}\right) \times 100 (6)
\]

n represents the total number of subjects
k represents the number of subjects whose absolute difference is less than MAE [10].

As shown in Table 1, the proposed method has 10 higher precision in the 14 anthropometric dimensions, the average PA is 95.59%, the Murtaza Aslam method is 92.98%, and the Lin&Wang method is 88.50%.

Therefore, the experimental results show that the proposed method has higher accuracy as compared with Murtaza Aslam method and Lin&Wang method for the measurement of anthropometric dimensions.

5. Conclusion
In this paper, we proposes a new method for automatically acquiring anthropometric dimensions using two-dimensional images. Our method first solved the complex background problem with deeplabv3, so that it is no longer limited by the background. In the process of feature point extraction, this method improves the openpose open source library, effectively reducing the user's posture requirements. We also introduce contour matching to retrieve similar human body data and improve the accuracy of the acquired anthropometric dimensions.

In the future, these detected data can be used not only for 3D reconstruction of the human body, but also for custom clothing, clothing simulation, virtual fitting, etc.

Acknowledgment
This research was supported by Zhejiang Provincial Science and Technology Program in China (2018C01030).

References
[1] Leong I F, Fang J J, Tsai M J 2007 Automatic body feature extraction from a marker-less
scanned human body Computer-Aided Design 39(7):568-82

[2] Cui Y, Chang W, Nöll T, Stricker D 2013 KinectAvatar: Fully Automatic Body Capture Using a Single Kinect C. ACCV Workshops (2) pp 133–47

[3] Aslam M, Rajbdad F, Khattak S. et al. 2017 Automatic measurement of anthropometric dimensions using frontal and lateral silhouettes J. IET Computer Vision 11(6):434-47

[4] Chen L C, Papandreou G, Kokkinos I, et al. 2018 DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs J. IEEE Transactions on Pattern Analysis & Machine Intelligence 40(4):834-48

[5] Cao Z, Simon T, Wei S E, et al. 2016 Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields J.

[6] Lin Y L, Wang M J J 2012 Constructing 3D human model from front and side images M. Pergamon Press, Inc.

[7] Chen L C, Papandreou G, Schroff F, et al. 2017 Rethinking Atrous Convolution for Semantic Image Segmentation.

[8] Xi, P, Lee, W, & Shu, C 2007 A data-driven approach to human-body cloning using a segmented body database J. Pacific Graphics 139–47

[9] Seo, H, Yeo, Y, & Wohn, K 2006 3D body reconstruction from photos based on range scan Lecture Notes in Computer Science 39(42) 849–60

[10] Lin Y L, Wang M J J 2011 Automated body feature extraction from 2D images J. Expert Systems with Applications 38(3):2585-91