Girth Measurement Based on Multi-View Stereo Images for Garment Design

LEI YANG[1], (Member, IEEE), QIANQIAN HUANG[1], XIAOWEI SONG[1], (Member, IEEE), MENGLONG LI[1], CHUNPING HOU[2], AND ZIXIANG XIONG[3], (Fellow, IEEE)
[1]School of Electronic and Information, Zhongyuan University of Technology, Zhengzhou 450007, China
[2]School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China
[3]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA

Corresponding authors: Lei Yang (annieyanglei@163.com) and Xiaowei Song (sxw-tju@163.com)

ABSTRACT In this paper, we propose a novel girth measurement system based on multi-view stereo images for garment design. Our system is set in a fixed location to capture three pairs of stereo images for the subject by six calibrated and synchronously triggered cameras. An important feature of this system is the use of an optimized semantic segmentation network that can efficiently segment the girth region in the captured six-view stereo images. Another important feature of this system is the use of color subspace classification and coordinate clustering that can effectively constrain the stereo matching within the scope of markers. Then, the system performs only on the corresponding clusters to extract stereo matching point pairs of markers correctly. The space coordinates of 3D point corresponding to each stereo matching point pair are calculated in each coordinate system of stereo cameras. The unified coordinates of these 3D markers are transformed from three different coordinate systems into one unified coordinate system. Girth is measured by curve fitting of these markers and calculating the length of the fitting curve. Our proposed system performs passive and intelligent girth measurement in garment design, and overcomes the problem of too many invalid stereo matching point pairs in girth measurement. Experimental results demonstrate its accuracy. Our system has a maximum bust measurement error of 1.28cm for woman and 1.31cm for man and a maximum waist measurement error of 1.18cm for woman and 0.99cm for man, which are within the error limit regulated by China national standards GB/T 2664-2017, 2665-2017, 2666-2017 and textile industry standard FZ/T 81004-2012.

INDEX TERMS Girth measurement, semantic segmentation, PSPNet, multi-view, stereo matching.

I. INTRODUCTION

With the rapid progress of information technology, garment design is developing towards automation and digitalization [1]–[3]. Nowadays, simple ready-to-wear design can no longer satisfy people's pursuit of individuality. Garment design is bound to be customized [4]–[8]. Clothing fit is the primary factor affecting the quality of customization. Therefore, precise anthropometric measurements of different individuals must be made first. Traditional anthropometry is to measure the human body size manually with a tape, which has strong controllability and flexibility [9]. However, the manual measurement process is complex and time-consuming, and its measurement error depends on the experience of

different tailors. Modern anthropometry is to measure the human body size automatically with a device. It can be divided into two main categories according to whether the device emits light source: active and passive.

The most common active devices are 3D laser scanner [10]–[12] and 3D structured light scanner [13]–[15]. The 3D laser scanner emits lasers to the subject, receives lasers bounced off the subject's surface, and perform precise anthropometry according to the positions, time intervals and optical axis angles of the received lasers [11]. The 3D structured light scanner emits structured light to the subject, records the interference fringes formed on the subject's surface, and performs accurate anthropometry according to the recorded distortion [14]. However, these devices are expensive and bulky high-end devices [16]. What's more, laser and structured light scanning often make the subject feel uncomfortable worrying

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan-Li Sun.

about eye damage during the data acquisition process [17]. The recent developed active devices are Kinect [18], [19] and RGB-D cameras [20], [21], focusing on low-cost and portability. These devices emit infrared light to the subject, obtain the depth map, and generate a 3D model of the subject [20]. However, the resolution of the depth map is low, the image registration is difficult due to the lack of geometric features, and the anthropometry is rough [22]. For both high-end and low-cost active devices, the subject is required to stand still without any movement for at least a few seconds to complete the scan. Otherwise, the scanned image will be deformed, thus affecting the precision of anthropometry.

The passive device can effectively solve this problem [22]–[29]. It uses common cameras to capture images of the subject in one shot. With technological advancement in vision system, cameras become cheaper and the captured images are of higher resolution, with rich geometry and texture details. These images are used for body reconstruction and anthropometry. References [23] and [24] are single-view anthropometric systems. They are scale-based. They estimate the anthropometric length based on the observed known value of the reference which may be a fixed line segment or a grid. These single-view systems are simple and cost effective. However, they can only measure the length, but not the girth. Moreover, the measurement points need to be selected manually, lack of automation. References [25] and [26] are dual-view anthropometric systems. They capture the front and side views of the subject, calculate the width and thickness of the measured girth based on the scale, and obtain the girth data by linear regression. These dual-view systems are also simple and cost effective. They can measure both the length and the girth. But it is difficult to completely recover the complicated curved surface such as bust and hip, and the measurement error of girth is large. References [27], [28] and [22] are multi-view anthropometric systems. References [27] and [28]are both single camera multi-view systems. They capture the multi-view images of the rotating subject with a single camera. Reference [27] reconstructs 3D human body model from the captured multi-view images, and selects the exact cross section of the model to measure the girth. Reference [28] calculates the length of the same girth in different views based on the scale, and estimates the girth data by model fitting. Nevertheless, the multi-view images are not captured synchronously, and the subject's movement may result in decreased accuracy. Reference [22] is a multi-camera multi-view system. It comprises 60 cameras and a carefully designed synchronization trigger, and captures 30 pairs of multi-view stereo images in a single shot. For each paired stereo cameras, camera calibration is performed. For each paired stereo images, stereo matching is performed, resulting in 30 dense and uniformly distributed point clouds. Multi-view registration and surface meshing are carried out to reconstruct the 3D human body model from the 30 point clouds, and thus the anthropometry is completed. This system can achieve high anthropometric accuracy. However, it's expensive and bulky, with complex matching and fusion for

60-view images. Hundreds of thousands of feature points have to be identified and matched, and complex algorithms must be used to remove as many mismatches as possible, which is a difficult and time-consuming task and not suitable for garment design. Furthermore, in order to improve the matching accuracy and reduce the reconstruction error, the image should be preprocessed to remove the background.
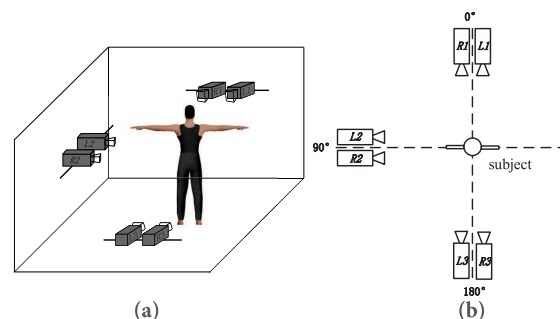

FIGURE 1. Overall view of the proposed girth measurement system based on multi-view stereo images. (a) 3D spatial view of the system. (b) Top view of the system.

In this paper, we propose a novel girth measurement system based on multi-view stereo images for garment design. Fig. 1 shows the overall view of our proposed system.The proposed system integrates semantic segmentation with stereo matching for simple, accurate and intelligent girth measurement. Three pairs of stereo images are captured by six synchronously triggered cameras (R1, L1; R2, L2; R3, L3) from the front (0°), side (90°) and back (180°) of the subject. An optimized Pyramid Scene Parsing Network(PSPNet) is used to segment the specific regions of the measured girths from each pair of the captured stereo images. Stereo matching is only carried out within the segmented regions of the specific girths in each stereo image pair with cluster constraints to extract the markers. The actual spatial coordinates of the markers in the space of each stereo camera pair are calculated in accordance with the respective calibrated parameters. The Euclidean spatial coordinates transformation is performed to transform the coordinates of three groups of markers from three different spaces into one space. The spatial coordinates of these markers are finally fitted to achieve girth measurement. Our proposed system is small in volume and simple in structure. It can effectively eliminate measurement errors caused by the subject's movement. It can simplify the stereo matching process and improve the matching accuracy as well. However, since bust and waist are the most significantly affected girths by the respiratory movement of the subject, these two girths are selected for experiment and verification in our system.

The rest of the paper is divided into four parts. In Section II, we discuss the related works.In Section III, we present our proposed system, which includes system configuration, semantic segmentation, stereo matching and spatial coordinates calculation, coordinates transformation and girth fitting. In Section IV, we report experimental procedures and results. In Section V, we make a conclusion.

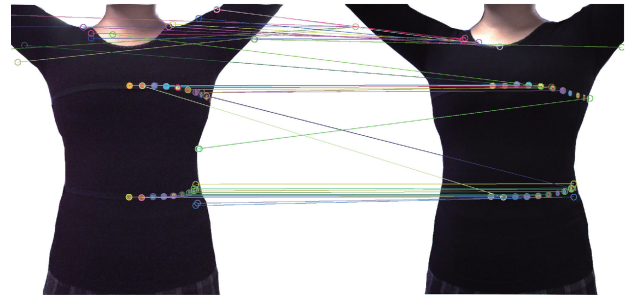**TABLE 1.** Statistical values of the matching results in Fig. 2 and Fig. 3.

Number of matching pairs	Total	Invalid	Valid but mismatching	Valid and matching	Invalid and Valid but mismatching (%)
Complex background	175	157	7	11	93.7%
Simple background	43	23	8	12	72.1%

II. RELATED WORKS

Precise matching in the stereo image pair is a prerequisite for the girth measurement in our system. Survey on the image matching methods can be found in reference [30]. The common matching methods can be broadly divided into three categories: grayscale correlation based method, feature based method and transform domain based method. The grayscale correlation based methods calculate the correlation between the template and the image to be matched to search the best matching position. The possible correlations include Mean Absolute Difference (MAD) [31], Sum of Absolute Differences (SAD) [32], Sum of Squared Differences (SSD) [33], etc. The larger the template, the higher the computation cost. The feature based methods extract the feature descriptors of two images and match the features according to the similarity of descriptor. Potential image features include points [34], edges [35], surfaces [36], etc. The most commonly used point matching algorithms include Scale Invariant Feature Transform (SIFT) [37], Speeded-Up Robust Features (SURF) [38], etc. They have good translation and rotation invariance and anti noise performance. Usually, the mismatch removal is conducted to achieve accurate feature matching results [39]–[41]. The transform domain based methods transform the rotation in time domain into translation in frequency domain by Fourier Transform [42], Walsh Transform [43], Wavelet Transform [44], etc. They have fast algorithms with easy implementation, as long as all points in the two images are shifted by the same direction and amount.
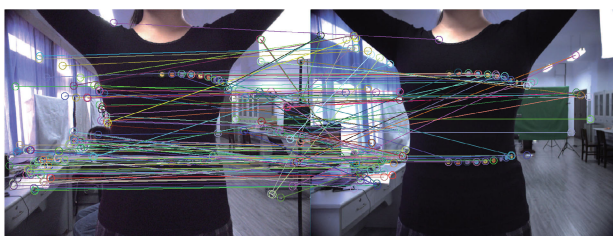
FIGURE 2. Matching result of complex background images with SURF.

In our system, the stereo images to be matched are of the same size, so the method based on grayscale correlation is not suitable. There is not only translation but also slight rotation between the two images. Hence, the method based on transform domain is also inappropriate. In order to complete the girth measurement, color markers in the two images need to be well matched. Therefore, the feature based method SURF is first selected to match the captured stereo image pair. Fig. 2 shows the matching result of the complex background images

FIGURE 3. Matching result of simple background images with SURF.

with SURF. Fig. 3 shows the matching result of the simple background images of the same subject with SURF. It can be observed that in both cases, there are a large number of invalid matching points in the matching results, which are useless for girth measurement. Table 1 shows the statistical values of the matching results in Fig. 2 and Fig. 3. For complex background, there are a total of 175 matching point pairs, of which, 157 are invalid, 7 are valid but mismatching, only 11 are valid and matching. Invalid and valid but mismatching point pairs account for up to 93.7%. Even for simple background, there are a total of 43 matching point pairs, of which, 23 are invalid, 8 are valid but mismatching, only 12 are valid and matching. Invalid and valid but mismatching point pairs also account for up to 72.1%. As shown in Fig. 4, invalid matching point pairs (a) and valid but mismatching point pairs (b) are useless for girth measurement. Only valid and matching point pairs (c) are required. Hence, invalid and valid but mismatching point pairs must be removed as many as possible [45], [46]. This is a difficult task considering the large proportion of these pairs. Therefore, before matching, it is necessary to properly segment the regions where the girths are located (in our system, bust and waist) to reduce the matching areas and background interference.

Traditional image segmentation methods include threshold based segmentation, region based segmentation and graph theory based segmentation. The threshold based segmentation methods [47]–[49] divide the grayscale histogram of an image into several classes by several thresholds. The pixels in the same class belong to the same object. They are simple, fast and efficient, but sensitive to noise, and not suitable for images with complex background. The region based segmentation method [50] connects pixels with similarity, thus forming the final segmented region. It is suitable for images with complex background, but is complicated and slow. The graph theory based segmentation methods [51]–[53] map an image
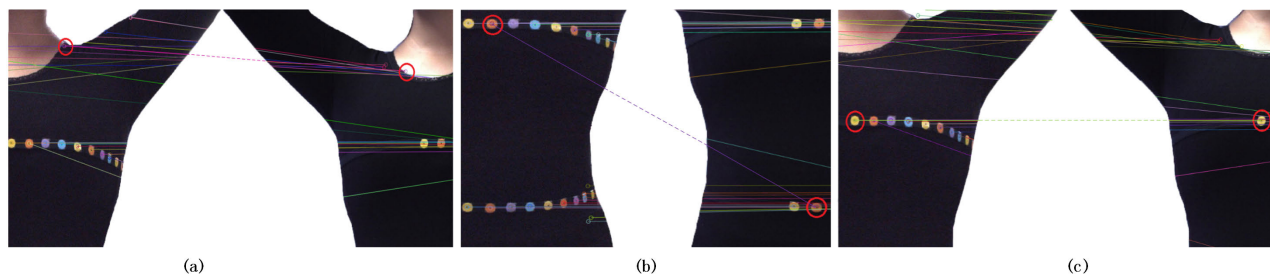
FIGURE 4. Examples of different types of matching point pairs. (a) Invalid matching point pair. (b) Valid but mismatching point pair. (c) Valid and matching point pair.

TABLE 2. Results of methods tested in the PASCAL VOC 2012 data set [63].

Method	FCN [56]	DeepLab [62]	CRFasRNN [60]	DeconvNet [57]	DPN [58]	Piecewise [61]	PSPNet [63]
MIoU	62.2%	71.6%	72.0%	72.5%	74.1%	75.3%	82.6%

into a weighted graph, and the complex image segmentation problem is simplified into an optimization problem by the optimal partition theory of graphs. They are fast, but only suitable for binary segmentation, and require manual intervention. All these traditional image segmentation methods do not meet the segmentation requirements in our system: fast and intelligent, in complex background, and for multiple segmentation results.

Modern image segmentation methods have developed with the progress of Convolutional Neural Networks (CNN) [54], [55]. A Full Convolutional Network was proposed in 2015 [56]. Since then, many semantic segmentation methods based on deep learning have emerged, including Deconvolution Network (DeconvNet) [57], Deep Parsing Network (DPN) [58], RefineNet [59], Conditional Random Fields as Recurrent Neural Networks (CRFasRNN) [60], Piecewise [61], DeepLab [62], PSPNet [63], etc. FCN selects CNN network as the basic framework and introduces the full convolution layer, but it does not make full use of the context information, and the segmentation precision is low. DeconvNet improves FCN by introducing deep deconvolution network, while the segmentation effect is not as good as FCN in the scene with strong illumination contrast. DPN makes use of group convolution to reduce computation complexity, while it ignores the finer details of the image. RefineNet improves the decoder structure and fuses low-level and high-level semantic features by up-sampling. However, the network capacity is large and the training time is long. CRFasRNN combines CRF and RNN into an end-to-end network to improve the segmentation accuracy of FCN, nevertheless, it lacks the utilization of context information. Piecewise combines CNN and CRF to effectively improve performance, but the model takes up much memory and the training time is long. DeepLab utilizes the empty convolutional layer instead of up-sampling, while fails to capture fine object boundaries. PSPNet fuses the features of different scales, so as to learn the features of the subject more effectively, increase the multi-resolution

receptive field, and further improve the segmentation accuracy. Table 2 shows the Mean Intersection over Unions (MIoUs) of various methods tested in the PASCAL VOC 2012 data set [63]. The MIoU of PSPNet is 82.6%, which is the highest among all the methods. Therefore, we choose PSPNet based on the spatial pyramid structure to segment the regions where the girths are located.
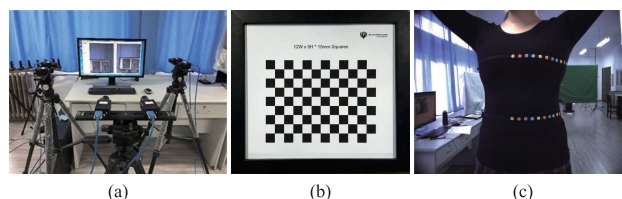

FIGURE 5. Multi-view stereovision system. (a) System configuration. (b) Calibration board. (c) Tights.

III. PROPOSED SYSTEM
In the proposed girth measurement system, we use six POINT GREY GS3-U3-28S4C-C industrial cameras to build the multi-view stereovision system, as shown in Fig. 5(a). This camera has a Sony ICX687 CCD chip, with a 1/1.8 inch size, a 1928 × 1448 maximum resolution, 2.8M effective pixels, a 128 MB onboard buffer, and a 2 MB data flash memory. The multi-view stereovision system also includes a PC, with an Intel Xeon E5-2620 CPU, 32 G RAM and a 8 G Nvidia Geforce GTX 1080 discrete graphics card. The PC controls the cameras and communicates with the cameras via a USB3.0 interface. Zhengyou Zhang's camera calibration method [64] is used to calibrate our multi-view stereovision system. The calibration board is shown in Fig. 5(b) with a 15mm×15mm size. As shown in Fig. 5(c), the subject wears tights during measurement. On the surface of the measurement position of the tights, there are circular markers in a repeated order of yellow, orange, purple and blue, with 1.5 cm spacing.
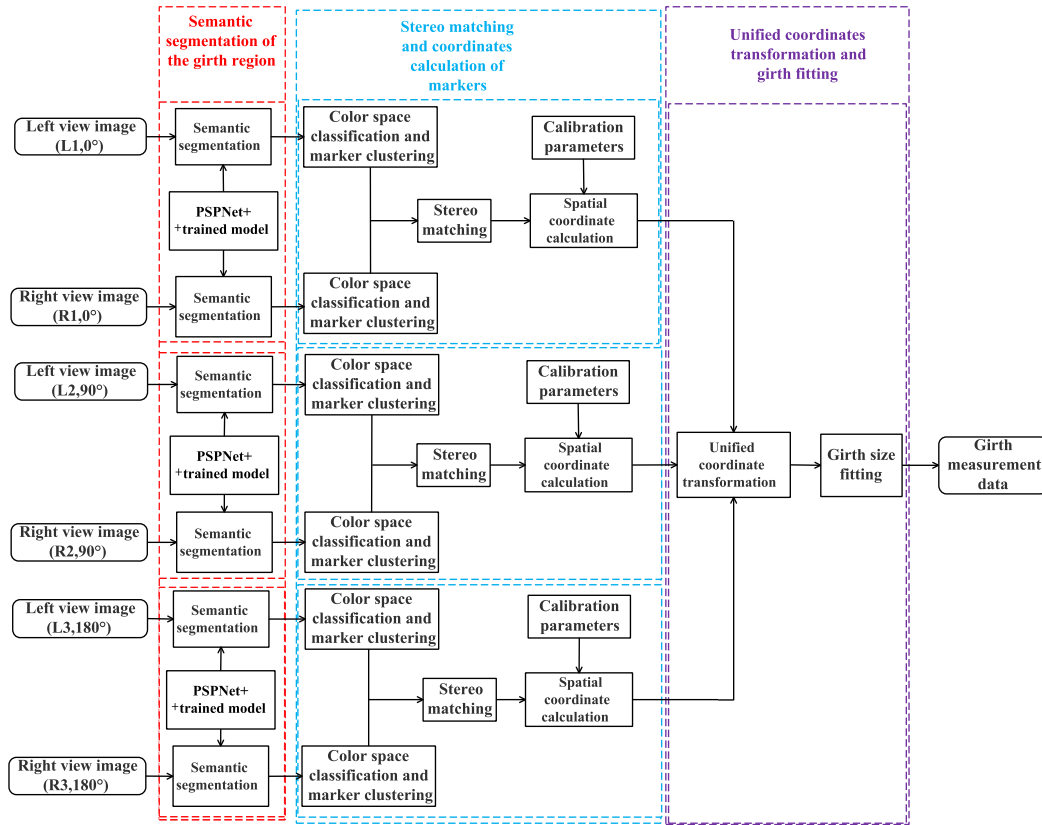
FIGURE 6. The whole procedure of the proposed girth measurement system.

The whole procedure of the proposed girth measurement system is shown in Fig. 6. System inputs include three pairs of left and right view images (0°, 90°, 180°) synchronously captured by the calibrated multi-view stereovision cameras. System output is the girth measurement data. This system consists of three main parts: semantic segmentation of the girth region, stereo matching and coordinates calculation of markers, unified coordinates transformation and girth fitting. In the semantic segmentation part, an optimized PSPNet is trained. The trained network is then used to segment the bust and waist regions respectively from each captured stereo image pair. In the stereo matching and coordinates calculation part, stereo matching is only performed with cluster constraints within the bust and waist regions in each stereo image pair to extract the markers. The actual spatial coordinates of the markers in the space of each stereo camera pair are calculated. In the coordinates transformation and girth fitting part, transformation is carried out to transform the coordinates of markers into a unified space. The unified spatial coordinates of these markers are finally fitted to achieve girth measurement.

A. SEMANTIC SEGMENTATION OF THE GIRTH REGION

We first perform semantic segmentation of the girth region. As described in the introduction, we choose bust and waist as the girths to be measured. We analyze 2700 anthropometric images captured by our system, with image resolution
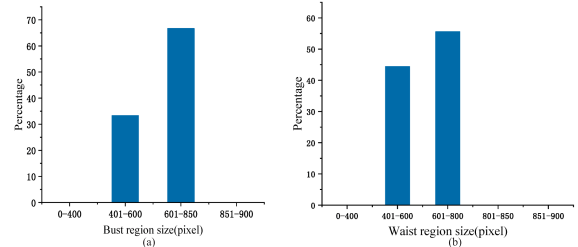

FIGURE 7. Statistical distribution of the girth region. (a) Bust region. (b) Waist region.

1928×1448, to further verify the necessity of segmentation. The statistical distribution bar graph in Fig. 7(a) indicates that the bust regions are all in the resolution range between 400×400 pixels and 850×850 pixels. Specifically, the number of bust region sizes below 400×400 pixels accounts for 0%, the number of bust region sizes between 400×400 pixels and 600×600 pixels accounts for 33.3%, the number of bust region sizes between 600×600 pixels and 850×850 pixels accounts for 66.7%, and the number of bust region sizes above 850×850 pixels accounts for 0%. The statistical distribution bar graph in Fig. 7(b) indicates that the waist regions are all in the resolution range between 400×400 pixels and 800×800 pixels. Specifically, the number of waist region sizes below 400×400 pixels accounts for 0%, the number of waist region sizes between 400×400 pixels and 600×600 pixels accounts for 44.4%, the number of waist region
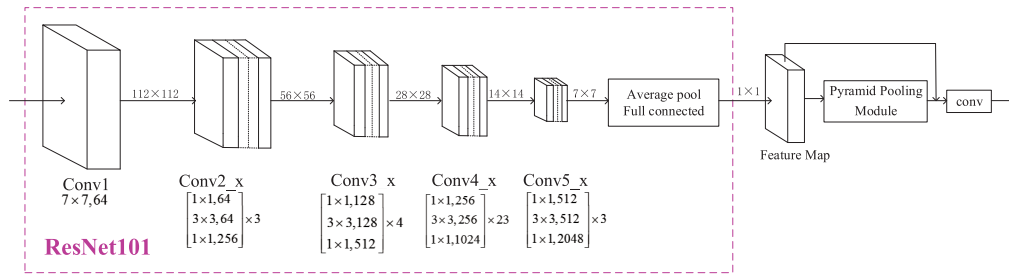
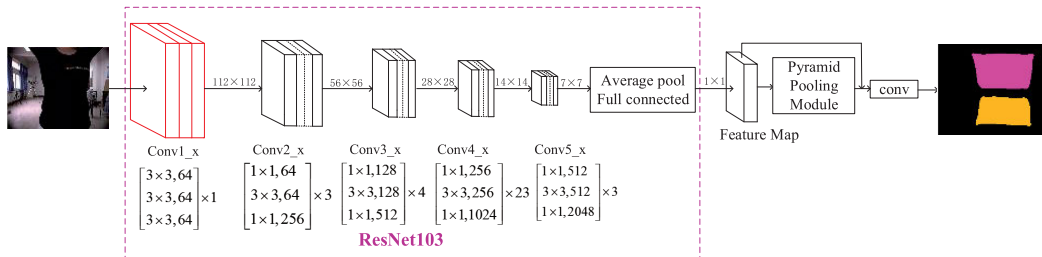**FIGURE 8.** PSPNet network structure [63].

FIGURE 9. PSPNet+ network structure.

sizes between 600×600 pixels and 800×800 pixels accounts for 55.6%, and the number of waist region sizes above 800×800 pixels accounts for 0%. In conclusion, the bust region only accounts for less than 25.9% of the total image size, and the waist region only accounts for less than 22.9% of the total image size. Therefore, semantic segmentation before stereo matching can effectively reduce the matching region and improve the matching accuracy.

TABLE 3. Segmentation performance of ResNet with different depth layers.

Layers	PA	MPA	MIoU
ResNet18	88.78%	37.53%	33.47%
ResNet34	88.96%	37.78%	33.71%
ResNet50	89.24%	38.01%	33.92%
ResNet101	89.40%	39.71%	34.34%

As described in the related works, we select PSPNet as the network structure to segment the girth region. Fig. 8 shows the existing PSPNet network structure [63]. As an important part of the PSPNet, ResNet is used to extract the feature map of the input image. ResNet consists of conv1, conv2_x, conv3_x, conv4_x and conv5_x, with various depth layers [65]. The more layers, the deeper the network, the more adequate the feature extraction, but the more complex the model. Hence, it is necessary to choose an appropriate network depth for our girth measurement system, which not only ensures accurate feature extraction, but also simplifies the model as much as possible. As shown in Table 3, the Pixel Accuracy (PA), Mean pixel Accuracy (MPA) and MIoU of ResNet18, ResNet34, ResNet50 and ResNet101 increase with the number of network layers. For ResNet101, MPA and MIoU show significant increases (1.7% and 0.42%, respectively) compared to other networks (0.24% and 0.22% in average, respectively).

Hence, we select ResNet101 for feature extraction of PSPNet. Although the girth region has no obvious difference with other parts of a human body in color, shape and texture, it has significant difference in spatial and proportional relation. With adequate training, PSPNet can extract the features associated with the spatial and proportional relation of the girth region, and use the features to correctly segment the girth region from the human body.

We also improve the existing PSPNet network structure by replacing the 7×7 convolution kernel of conv1 with three 3×3 convolution kernels in series, as shown in Fig. 9. For the sake of distinction, we use PSPNet+ to refer to this improved version of PSPNet. As shown in Fig. 10, the receptive field and output size are the same for the 7×7 convolution kernel and the three 3×3 convolution kernels in series. The PSPNet+ can increase the network capacity and reduce the number of parameters without significantly increasing the network complexity and can thus enhance the network performance.

We select 2500 images containing bust and waist from the Look Into Person (LIP) open source dataset (50462 images with various resolutions) established by Sun Yat-Sen University [66]. In addition, 1500 images (resolution 1928×1448) are taken by the POINT GREY GS3-U3-28S4C-C industrial camera in our lab at a distance of 2-4 m. In summary, 4000 images are obtained. With random image clipping and scaling, the dataset size is expanded by a factor of three, that is, 12000 images. Fig. 11 shows some image examples of our bust and waist dataset.

We train the girth semantic segmentation model with 7200 images as the training set in the PSPNet and the PSPNet+. We set the model training parameter batch_size to 8 and epochs to 50. After training, we obtain a trained
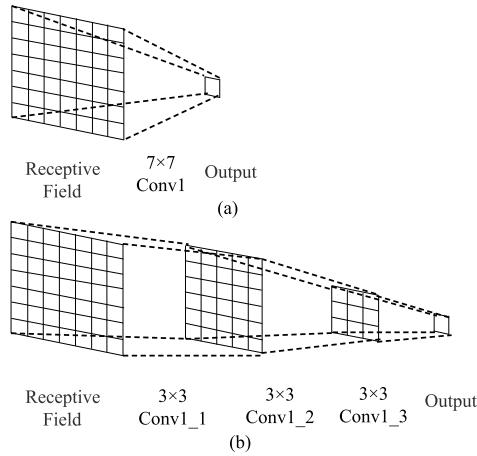
FIGURE 10. Receptive fields of different convolution kernels. (a) 7 × 7 kernel. (b) Three 3 × 3 kernels in series.

FIGURE 11. Image examples of our dataset.

TABLE 4. Segmentation performance of PSPNet and PSPNet+.

Network	PA	MPA	MIoU
PSPNet	89.40%	39.71%	34.34%
PSPNet+	89.69%	44.52%	35.95%

PSP-based model and a trained PSP+-based model. The loss function of the PSP-based model converges to 0.1. The loss function of the PSP+-based model converges to 0.08, which is 0.02 lower than that of PSP. We test the trained PSP-based model and the trained PSP+-based model with the other 4800 images as the testing set. The testing set contains bust and waist shot from various perspectives. As shown in Table 4, the PSP-based model achieves a PA of 89.40%. The PSP+-based model achieves a PA of 89.69%, which is 0.29% higher than that of the PSP-based model. The PSP-based model achieves an MPA of 39.71%. The PSP+-based model achieves an MPA of 44.52%, which is 4.81% higher than that of the PSP-based model. The PSP-based model achieves an MIoU of 34.34%. The PSP+-based model achieves an MIoU of 35.95%, which is 1.61% higher than
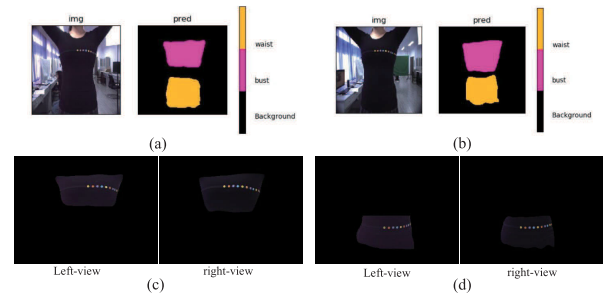

FIGURE 12. Semantic segmentation effect of anthropometric stereo image pair. (a) The semantic segmentation of left-view image. (b) The semantic segmentation of right-view image. (c) The segmented image pair of bust region. (d) The segmented image pair of waist region.
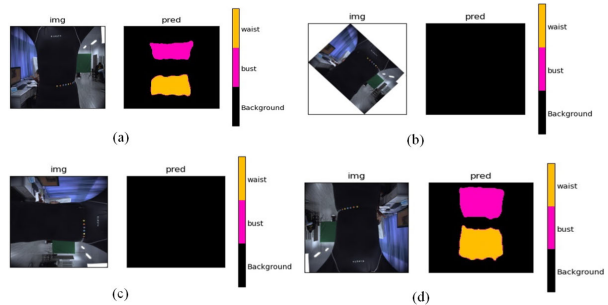

FIGURE 13. Semantic segmentation results of bust and waist regions from images with different rotations. (a) 0°. (b)45°. (c)90°. (d) 180°.

that of the PSP-based model. In summary, the PSPNet+ has better performance. Since there is no exact boundary between the girth region and other regions, there may be irregular upper and lower edges in the segmentation results, and the segmented region is not exactly the same with the ground truth, which leads to relatively low MIoU. However, the measurement results of our system will not be affected because the segmented regions have included the exact girths to be measured. The semantic segmentation results of the anthropometric stereo image pair captured by our girth measurement system with PSPNet+ are shown in Figs. 12(a) and 12(b). The image is divided into three parts, yellow for the waist region, magenta for the bust region, and black for the background. The segmented image pairs of bust region and waist region can be obtained by respective masks, as shown in Figs. 12(c) and 12(d). In this way, the matching areas are intelligently reduced from the whole images to the smaller girth regions. Nevertheless, the spatial related features are sensitive to image rotations. Fig. 13 shows the semantic segmentation results of bust and waist regions from the captured human body images corresponding to clockwise rotation of 0°, 45°, 90°and 180°, respectively. Only the semantic segmentation result in Fig. 13(a) is correct. Figs. 13(b) and 13(c) have no semantic segmentation results, while the semantic segmentation result of Fig. 13(d) is upside down. Therefore, in order to segment the girth region from the human body correctly, there are constraints in the image acquisition. The subject is required to stand upright, keeping feet together and breathing normally to ensure the correct spatial and proportional relation in the captured images.

TABLE 5. Statistical values of the matching results in Fig. 2 and Fig. 3.

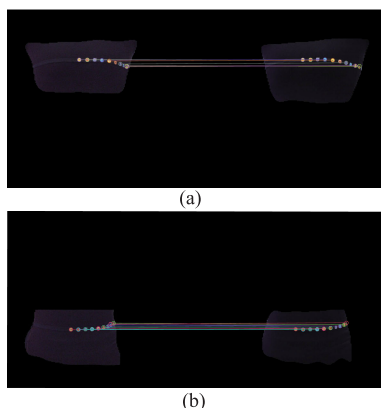| Number of matching point pairs	Total	Invalid	Valid but mismatching	Valid and matching	Mismatching (%)
Bust	9	0	4	5	44.4%
Waist	12	0	4	8	33.3%
Bust and waist	21	0	8	13	38.1%

FIGURE 14. Matching results of the bust and waist segmented image pairs with SURF. (a) Bust. (b) Waist.

B. STEREO MATCHING AND COORDINATES CALCULATION

Fig. 14 shows the matching results of the bust and waist segmented image pairs with SURF, in which the subject is the same as in Fig. 2 and Fig. 3.

Table 5 shows the statistical values of the matching results in Fig. 14. For bust segmented image pair, there are a total of 9 matching point pairs, of which, 0 are invalid, 4 are valid but mismatching, 5 are valid and matching. Mismatching point pairs account for 44.4%. For waist segmented image pair, there are a total of 12 matching point pairs, of which, 0 are invalid, 4 are valid but mismatching, 8 are valid and matching. Mismatching point pairs account for 33.3%. In summary, there are a total of 21 matching point pairs, of which, 0 are invalid, 8 are valid but mismatching, 13 are valid and matching. Mismatching point pairs account for 38.1%, which is much lower than the 93.7% of complex background and 72.1% of simple background in Table 1. Furthermore, the total number of matching point pairs is reduced to 21, which is much lower than the 175 of complex background and 43 of simple background in Table 1 and the computation cost of matching can be greatly reduced. There are no invalid match point pairs, therefore, no more effort is required to remove the invalid matching point pairs. However, there are also about a third of mismatching point pairs. If not removed correctly, they will lead to incorrect girth measurement. Even if removed correctly, the number of matching point pairs is less than the total number of markers, which will lead to a decrease in girth measurement accuracy. Hence, it is necessary to correctly match as many markers as possible.

In our girth measurement system, there are four colors of circular markers on the measured girth. These markers are repeated in yellow, orange, purple and blue pattern.
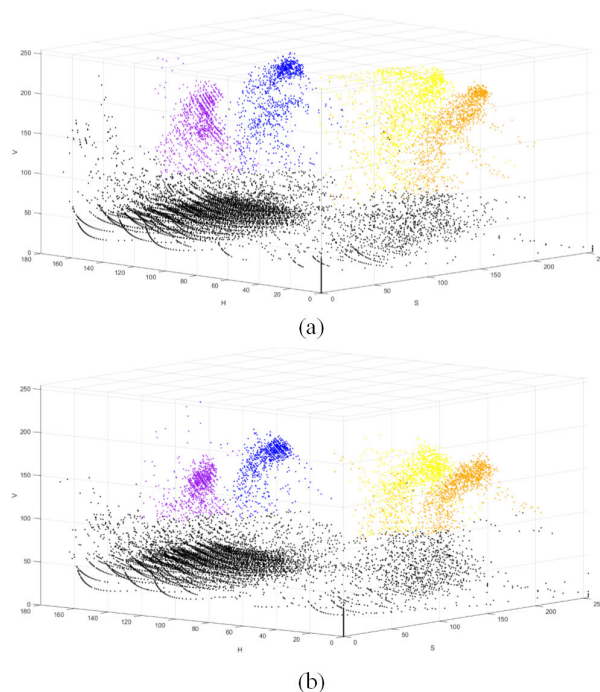

FIGURE 15. Scatter diagrams in HSV space of the segmented images. (a) Bust region. (b) Waist region.

Fig. 15 shows the scatter diagrams in HSV space of the segmented images of bust region and waist region shown in Figs. 12(c) and 12(d). It can be observed that the four colors are located in separate spatial areas of the HSV color space. Therefore, these separate spatial areas can be used to distinguish the markers of different colors. Table 6 shows the HSV range corresponding to the four colors.

TABLE 6. HSV range corresponding to the four colors.

	Yellow	Orange	Purple	Blue	Black
H_{min}	26	11	125	100	0
H_{max}	34	25	155	124	180
S_{min}	43	43	43	43	0
S_{max}	255	255	255	255	255
V_{min}	46	46	46	46	0
V_{max}	255	255	255	255	46

According to this HSV range, markers in the segmented image are classified into four different color categories denoted as $\mathbf{Y}$, $\mathbf{O}$, $\mathbf{P}$ and $\mathbf{B}$. All pixels in the segmented image constitute a data set $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \cdots \mathbf{z}_i, \cdots\}$, $i = 1, 2, \ldots, n$, $n = 1928 \times 1448$. The segmented image is converted from RGB color space to HSV color space. Then the pixel $\mathbf{z}_i$ has three components, i.e., $\mathbf{z}_i (H_i, S_i, V_i)$. If V_i is less than 46, the three components H_i, S_i and V_i of $\mathbf{z}_i$ are all set to 0. If V_i is greater than 46, the pixel $\mathbf{z}_i$ is classified according to its H_i,

and a fourth component C_i representing the color category is added to that pixel $\mathbf{z}_i$. If V_i is greater than 46 and H_i is greater than 11 and less than 25, that pixel $\mathbf{z}_i$ belongs to the color category $\mathbf{O}$, $C_i=orange$. If V_i is greater than 46 and H_i is greater than 26 and less than 34, that pixel $\mathbf{z}_i$ belongs to the color category $\mathbf{Y}$, $C_i=yellow$. If V_i is greater than 46 and H_i is greater than 100 and less than 124, that pixel $\mathbf{z}_i$ belongs to the color category $\mathbf{B}$, $C_i=blue$. If V_i is greater than 46 and H_i is greater than 125 and less than 155, that pixel $\mathbf{z}_i$ belongs to the color category $\mathbf{P}$, $C_i=purple$. According to statistical analysis of the segmented image, the color pixels account for less than 1% of the total pixels of the whole image. In this way, the color pixels in the segmented image constitute a much smaller data set $\mathbf{Z}_C = \{\mathbf{z}_{c1}, \mathbf{z}_{c2}, \cdots \mathbf{z}_{ci}, \cdots\}$. Each element $\mathbf{z}_{ci}$ has five components, i.e., $\mathbf{z}_{ci}$ $(H_i, S_i, V_i, C_i, x_i)$. The dataset $\mathbf{Z}_C$ can be further divided into four subdatasets $\mathbf{Z}_{Cj}, j=1,2,3,4$. Fig. 16 shows the four subdatasets obtained by classification of a left-view segmented image and a right-view segmented image.

$$\mathbf{Z}_{C1} = \{\mathbf{z}_{c1k} \in \mathbf{Z}_C | C_{1k} = yellow\} \quad (1)$$
$$\mathbf{Z}_{C2} = \{\mathbf{z}_{c2k} \in \mathbf{Z}_C | C_{2k} = orange\} \quad (2)$$
$$\mathbf{Z}_{C3} = \{\mathbf{z}_{c3k} \in \mathbf{Z}_C | C_{3k} = purple\} \quad (3)$$
$$\mathbf{Z}_{C4} = \{\mathbf{z}_{c4k} \in \mathbf{Z}_C | C_{4k} = blue\} \quad (4)$$
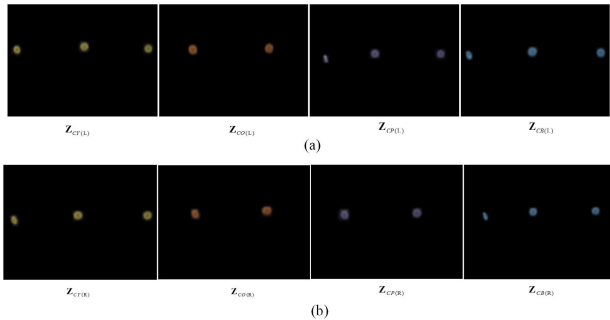

FIGURE 16. Four subdatasets obtained by classification. (a) Left-view. (b) Right-view.

Each of the four subdatasets belongs to a different color subspace. In each color subspace, the spacing between adjacent markers of the same color increases by a factor of 4, while the size of markers remains the same, which provides a guarantee for accurate clustering of pixels belonging to different markers. According to statistical analysis, the maximum horizontal distance of the pixels in the same marker is no more than 40 pixels, while the minimum horizontal distance between pixels in adjacent markers of the same color is no less than 100 pixels. Hence, all pixels in a subdataset can be further clustered into several clusters, each cluster corresponding to a marker. For subdataset $\mathbf{Z}_{Cj}$, each element $\mathbf{z}_{cjk}$ $(H_{jk}, S_{jk}, V_{jk}, C_{jk}, x_{jk})$ has five components. The distance between any two elements is defined as:

$$d(\mathbf{z}_{cjk}, \mathbf{z}_{cjk'}) = |x_{jk'} - x_{jk}| \quad (5)$$

Algorithm 1 Color-Subspace-Classifier

Input: Segmented image
Output: Category C_i of Color
1: **function** Classification(*Orange, Yellow, Blue, Purple*)
2:     $HSV \leftarrow RGB$
3:     Get H_i, S_i, V_i components of $\mathbf{z}_i$
4:     **if** $V_i < 46$ **then**
5:       $H_i \leftarrow 0$
6:       $S_i \leftarrow 0$
7:       $V_i \leftarrow 0$
8:     **end if**
9:     **if** $V_i > 46$ & $11 < H_i < 25$ **then**
10:       $C_i \leftarrow Orange$
11:     **end if**
12:     **if** $V_i > 46$ & $26 < H_i < 34$ **then**
13:       $C_i \leftarrow Yellow$
14:     **end if**
15:     **if** $V_i > 46$ & $100 < H_i < 124$ **then**
16:       $C_i \leftarrow Blue$
17:     **end if**
18:     **if** $V_i > 46$ & $125 < H_i < 155$ **then**
19:       $C_i \leftarrow Purple$
20:     **end if**
21:     **return** C_i
22: **end function**

The $\mathcal{E}$ neighborhood of an element is defined as:

$$\mathbf{N}_\varepsilon(\mathbf{z}_{cjk}) = \{\mathbf{z}_{cjk'} \in \mathbf{Z}_{Cj} : d(\mathbf{z}_{cjk}, \mathbf{z}_{cjk'}) \le \varepsilon\} \quad (6)$$

We set $\mathcal{E}$ to 40 and select the first element $\mathbf{z}_{cj1}$ as the initial pixel. The $\mathcal{E}$ neighborhood of $\mathbf{z}_{cj1}$ is denoted as $\mathbf{M}_{j1}$, and the pixel number in $\mathbf{M}_{j1}$ is denoted as $N_{j1}, N_{j1} = 1$. If the second element $\mathbf{z}_{cj2} \in \mathbf{N}_\varepsilon(\mathbf{z}_{cj1})$, then $\mathbf{z}_{cj2} \in \mathbf{M}_{j1}$, and $N_{j1} = N_{j1} + 1$. If $\mathbf{Z}_{cj2} \notin \mathbf{N}_\varepsilon(\mathbf{z}_{cj1})$, the $\mathcal{E}$ neighborhood $\mathbf{N}_\varepsilon(\mathbf{z}_{cj2})$ is denoted as $\mathbf{M}_{j2}$, and the pixel number in $\mathbf{M}_{j2}$ is denoted as $N_{j2}, N_{j2} = 1$. If the third element $\mathbf{z}_{cj3} \in \mathbf{N}_\varepsilon(\mathbf{z}_{cj1})$ or $\mathbf{z}_{cj3} \in \mathbf{N}_\varepsilon(\mathbf{z}_{cj2})$, then $\mathbf{z}_{cj3} \in \mathbf{M}_{j1}$ and $N_{j1} = N_{j1}+1$ or $\mathbf{z}_{cj3} \in \mathbf{M}_{j2}$ and $N_{j2} = N_{j2}+1$. If $\mathbf{z}_{cj3} \notin \mathbf{N}_\varepsilon(\mathbf{z}_{cj1}) \cup \mathbf{N}_\varepsilon(\mathbf{z}_{cj2})$, the $\mathcal{E}$ neighborhood $\mathbf{N}_\varepsilon(\mathbf{z}_{cj3})$ is denoted as $\mathbf{M}_{j3}$, and the pixel number in $\mathbf{M}_{j3}$ is denoted as $N_{j3}, N_{j3} = 1$. The process continues until all the elements in the subdataset $\mathbf{Z}_{Cj}$ are visited. In this way, we get several clusters $\mathbf{M}_{jm}$ and several $N_{jm}, m = 1, 2, \ldots, M$, where M is the number of markers in $\mathbf{Z}_{Cj}$. We calculate the average horizontal coordinates $\bar{x}_{jm}$ of the pixels in $\mathbf{M}_{jm}$

$$\bar{x}_{jm} = \sum_{n=1}^{N_{jm}} x_{jmn}/N_{jm} \quad (7)$$

We reorder $\mathbf{M}_{jm}$, so that $\bar{x}_{j1} < \bar{x}_{j2} < \ldots < \bar{x}_{jM}$.

In the above way, the subdatasets of the left-view and right-view shown in Fig. 16 are clustered respectively to obtain clusters $\mathbf{M}_{jm_L}$ and $\mathbf{M}_{jm_R}, m_L = 1, 2, \ldots, M_L$ and $m_R = 1, 2, \ldots, M_R$. If $M_L = M_R$, the marker number is the same in the left and right views. If $M_L \neq M_R$, the marker number is different due to parallax and occlusion.

If $M_L = M_R + 1$, the cluster $\mathbf{M}_{jm_L}$ with the smallest N_{jm_L} is removed and other clusters $\mathbf{M}_{jm_L}$ are reordered according to $\bar{x}_{jm_L}$. If $M_L + 1 = M_R$, the cluster $\mathbf{M}_{jm_R}$ with the smallest N_{jm_R} is removed and other clusters $\mathbf{M}_{jm_R}$ are reordered according to $\bar{x}_{jm_R}$. Finally, the number of clusters in the subdatasets $\mathbf{Z}_{Cj_L}$ and $\mathbf{Z}_{Cj_R}$, i.e., the number of markers that can be stereo matched in the left and right views, is the same. Each $\mathbf{M}_{jm_L}$ corresponds to one special $\mathbf{M}_{jm_R}$: $\mathbf{M}_{jm_L} \rightarrow \mathbf{M}_{jm_R}$. Fig. 17 shows the matching results of the bust and waist segmented image pairs with SURF taking $\mathbf{M}_{jm_L} \rightarrow \mathbf{M}_{jm_R}$ as the matching constraint. What's more, only one pair of matching points closest to the center are kept in one pair of $\mathbf{M}_{jm_L}$ and $\mathbf{M}_{jm_R}$. Table 7 shows the improved stereo matching results. The x and y coordinates of the matching pixels in the corresponding markers are output. All the markers are well matched by 100%. However, as shown in Table 8, 2D errors still exist between the 2D matching results and their corresponding 2D ground truths, which will cause 3D errors in the subsequent coordinate calculation.
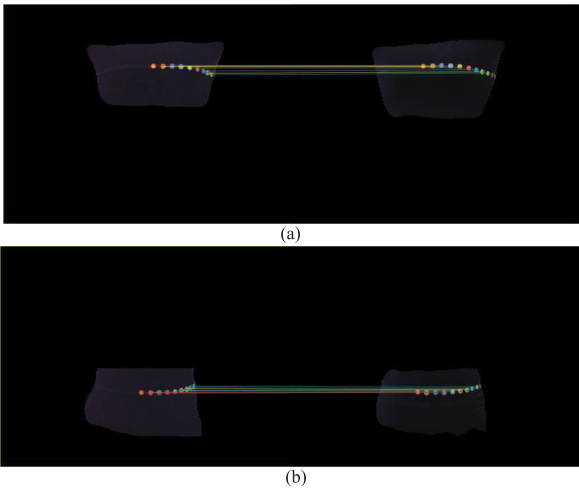

(a)

(b)

FIGURE 17. Matching results of the bust and waist segmented image pairs with constrained SURF. (a) Bust region. (b) Waist region.

TABLE 7. The improved stereo matching results.

Left-view (x_L, y_L)	Right-view (x_R, y_R)
(991.6, 405.7)	(874.8, 406.6)
(1057.2, 405.3)	(936.4, 407.5)
(1115.6, 406.4)	(994.3, 402.7)
(1174.3, 411.1)	(1055.6, 406.8)
(1233.4, 415.3)	(1117.4, 410.1)
(1284.5, 426.3)	(1176.0, 418.4)
(1323.3, 443.5)	(1227.3, 432.5)
(1351.5, 454.8)	(1263.4, 444.6)
(1375.9, 465.3)	(1301.0, 457.5)

After stereo matching of markers, we calculate the space coordinates of the real 3D points corresponding to stereo matching marker pairs with the calibration parameters of the stereovision camera pairs. Fig. 18 shows the convergent stereovision model [67] used in our system. The internal parameters of camera include the focal length of the left-view camera and the right-view camera, f_L and f_R. The external
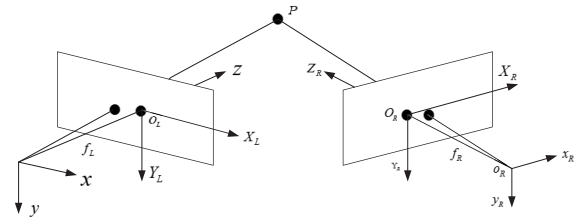

FIGURE 18. Convergent stereovision model.

parameters of camera include a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{T}$, which can be obtained by using Zhengyou Zhang's calibration method [64].

$$\mathbf{R} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T$$

With (x_L, y_L), (x_R, y_R), f_L, f_R, $\mathbf{R}$ and $\mathbf{T}$, the space coordinates (x, y, z) of markers can be calculated as [68]:

$$\begin{cases} x = z x_L / f_L \\ y = z y_L / f_L \\ z = \dfrac{f_L \left(f_R t_x - x_R t_z \right)}{x_R \left(r_7 x_L + r_8 y_L + f_L r_9 \right) - f_R \left(r_1 x_L + r_2 y_L + f_L r_3 \right)} \\ \quad = \dfrac{f_L \left(f_R t_y - y_R t_z \right)}{y_R \left(r_7 x_L + r_8 y_L + f_L r_9 \right) - f_R \left(r_4 x_L + r_5 y_L + f_L r_6 \right)} \end{cases} \tag{8}$$

The stereo calibration method will introduce a relative depth error of 1/20000 in the case of a long focal length (>25mm) and a fixed baseline length [68], which is exactly the case in our system. In our system, the measuring distance is 60cm, so the depth error is about 0.03mm, which is relatively small and can be ignored. Table 9 shows the 3D errors of coordinate calculation for the stereo matching results in Table 8. As shown in Tables 8 and 9, the 2D coordinate deviation of the matching points in one pixel will bring a 3D coordinate deviation of at least 1mm in the subsequent coordinate calculation. The more the 2D coordinate deviation, the greater the 3D error of coordinate calculation.

C. UNIFIED COORDINATES TRANSFORMATION AND GIRTH FITTING

In our system, three pairs of left and right view images (0°, 90°, 180°) are synchronously captured by three stereovision camera pairs. Each pair of cameras corresponds to a special coordinate system. After stereo matching and spatial coordinates calculation, three sets of 3D coordinates of markers in three different coordinate systems are calculated, denoted as $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$.

$$\mathbf{S}_1 = \left\{ \mathbf{s}_{1\,i} \mid \mathbf{s}_{1\,i} = (x_{1\,i}, y_{1\,i}, z_{1\,i})^T, 1 \le i \le I \right\} \tag{9}$$

$$\mathbf{S}_2 = \left\{ \mathbf{s}_{2\,j} \mid \mathbf{s}_{2\,j} = (x_{2\,j}, y_{2\,j}, z_{2\,j})^T, 1 \le j \le J \right\} \tag{10}$$

$$\mathbf{S}_3 = \left\{ \mathbf{s}_{3\,k} \mid \mathbf{s}_{3k} = (x_{3k}, y_{3k}, z_{3k})^T, 1 \le k \le K \right\} \tag{11}$$

TABLE 8. 2D errors of stereo matching results.

	2D matching result (left/right view) (pixel)	2D ground truth (left/right view) (pixel)	2D error (left/right view) (pixel)	2D coordinate deviation (pixel)
1	(910.5,332.5)/(834,333.5)	(910,328)/(833,332)	(0.5,4.5)/(1.0,1.5)	3.04
2	(967.6,332.5)/(890.6,332.7)	(969,330)/(891,333)	(-1.4,2.5)/(-0.4,-0.3)	2.94
3	(1021,334.5)/(944.3,330.8)	(1024,332)/(944,331)	(-3.0,2.5)/(0.3,-0.2)	4.32
4	(1073.8,337.5)/(998,334.8)	(1075,337)/(999,334)	(-1.2,0.5)/(-1.0,0.8)	0.32
5	(1124,343)/(1051.5,339)	(1126,341)/(1053,337)	(-2.0,2.0)/(-1.5,2.0)	0.50
6	(1172,346.9)/(1102.6,340.8)	(1172,347)/(1105,340)	(0,-0.1)/(-2.4,0.8)	2.58
7	(1215,359.5)/(1160.5,345)	(1215,355)/(1156,347)	(0,4.5)/(4.5,-2.0)	7.91
8	(1243.5,359.9)/(1193.9,352.6)	(1246,363)/(1194,352)	(-2.5,-3.1)/(-0.1,0.6)	4.36
9	(1274.5,371.5)/(1231.5,359.2)	(1276,369)/(1234,359)	(-1.5,2.5)/(-2.5,0.2)	2.51

TABLE 9. 3D errors of coordinate calculation for the stereo matching results in Table 8.

	3D coordinate calculation result (mm)	3D coordinate ground truth (mm)	3D error (mm)	3D coordinate deviation (mm)
1	(-40.07,-142.49,620.83)	(-40.20,-143.97,619.99)	(0.13,1.47,0.84)	1.70
2	(-18.95,-143.03,623.12)	(-18.39,-143.60,621.61)	(-0.56,0.57,1.52)	1.71
3	(0.92,-142.83,625.56)	(2.03,-142.53,620.17)	(-1.10,-0.30,5.39)	5.51
4	(20.73,-142.18,627.66)	(20.73,-142.18,627.66)	(-0.44,0.11,0.33)	0.56
5	(39.87,-141.25,632.72)	(40.57,-141.81,631.84)	(-0.70,0.56,0.87)	1.26
6	(58.36,-140.60,636.46)	(58.73,-141.46,640.52)	(-0.37,0.86,-4.05)	4.17
7	(77.47,-140.83,659.99)	(76.54,-140.89,652.05)	(0.93,0.06,7.94)	7.99
8	(89.60,-142.09,666.63)	(90.00,-139.93,662.23)	(-0.39,-2.17,4.40)	4.92
9	(103.39,-139.42,676.10)	(104.26,-140.78,677.78)	(-0.86,1.36,-1.67)	2.33

We take the coordinate system of $\mathbf{S}_2$ as the unified coordinate system. We still use Zhengyou Zhang's calibration method [64] to obtain the relative external parameters $\boldsymbol{R}_{12}$ and $\boldsymbol{T}_{12}$ from the coordinate system of $\mathbf{S}_1$ to the coordinate system of $\mathbf{S}_2$.

$$\boldsymbol{R}_{12} = \begin{bmatrix} 0.043 & 0.033 & -0.999 \\ -0.036 & 0.998 & -0.031 \\ 0.998 & 0.035 & 0.044 \end{bmatrix}$$

$$\boldsymbol{T}_{12} = \begin{bmatrix} -536.79 & -25.63 & 506.18 \end{bmatrix}^T$$

And we also obtain the relative external parameters $\boldsymbol{R}_{32}$ and $\boldsymbol{T}_{32}$ from the coordinate system of $\mathbf{S}_2$ to the coordinate system of $\mathbf{S}_3$.

$$\boldsymbol{R}_{23} = \begin{bmatrix} 0.133 & -0.001 & -0.991 \\ -0.040 & 0.999 & -0.005 \\ 0.990 & -0.040 & 0.133 \end{bmatrix}$$

$$\boldsymbol{T}_{23} = \begin{bmatrix} -519.35 & 0.589 & 481.25 \end{bmatrix}^T$$

The unified coordinate transformation formulas are as follows [69].

$$\mathbf{s}_{2\,i} = \boldsymbol{R}_{12}\mathbf{s}_{1\,i} + \boldsymbol{T}_{12}, \quad i = 1, 2, \ldots, I \tag{12}$$

$$\mathbf{s}_{2\,k} = \boldsymbol{R}_{23}^{-1}\left(\mathbf{s}_{3k} - \boldsymbol{T}_{23}\right), \quad k = 1, 2, \ldots, K \tag{13}$$

Table 10 shows an exemplary unified coordinate transformation result from $\mathbf{s}_{1\,i}$ to $\mathbf{s}_{2\,i}$. As a consequence, we have one set of uniform 3D coordinates of markers in the coordinate system of $\mathbf{S}_2$, denoted as $\mathbf{S}_2'$.

$$\mathbf{S}_2' = \left\{\mathbf{s}_{2\,i} \cup \mathbf{s}_{2\,j} \cup \mathbf{s}_{2k} | 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K\right\} \tag{14}$$

As can be seen from Table 10, the y coordinates of markers are almost the same, i.e., the markers are almost on

TABLE 10. An exemplary unified coordinate transformation result from $\mathbf{s}_{1i}$ to $\mathbf{s}_{2i}$.

	$\mathbf{s}_{1i}\ (x_{1i}, y_{1i}, z_{1i})\ (mm)$	$\mathbf{s}_{2i}(x_{2i}, y_{2i}, z_{2i})\ (mm)$
1	(19.68, 86.05, 400.07)	(-139.59, 74.77, 506.78)
2	(34.17, 85.22, 400.80)	(-138.22, 74.46, 492.32)
3	(45.77, 84.39, 401.54)	(-136.95, 74.05, 480.74)
4	(56.80, 84.38, 404.96)	(-133.06, 74.51, 469.87)
5	(68.34, 84.43, 408.82)	(-128.73, 75.08, 458.52)
6	(78.82, 83.98, 414.00)	(-123.08, 75.16, 448.27)
7	(88.49, 83.21, 419.05)	(-117.60, 74.89, 438.81)
8	(99.67, 83.87, 426.31)	(-109.90, 76.17, 427.99)
9	(109.14, 83.05, 432.87)	(-102.92, 75.89, 418.79)
10	(117.04, 83.01, 439.69)	(-95.78, 76.35, 411.20)

a same horizontal cross section. Therefore, the markers of the measured girth are projected onto the XOZ plane for fitting. Polynomial curve fitting (PCF) [70], Cubic Bezier curve fitting (CBCF) [71] and Polynomial with Intermediate Variable curve fitting (PIVCF) [72] are adopted to fit the markers respectively. Fig. 19 shows the exemplary curve fitting results for four typical cases, that is, woman's bust, woman's waist, man's bust and man's waist. The blue circle represents the original data of markers projected onto the XOZ plane. The green line represents the polyline obtained by directly connecting the original data points. The blue line represents the fitting curve obtained by PCF method. The orange line represents the fitting curve obtained by CBCF method. The red line represents the fitting curve obtained by PIVCF method. For Figs. 19(b), 19(c) and 19(d), these three methods show similar fitting results. However, for Fig. 19(a), the blue line shows obvious under-fitting, that is, the PCF method is not suitable for woman's bust and, thus is not suitable for our system. The other two methods show similar fitting results.

Table 11 shows the RMSE of the other two methods (CBCF and PIVCF) in 10 randomly selected samples from
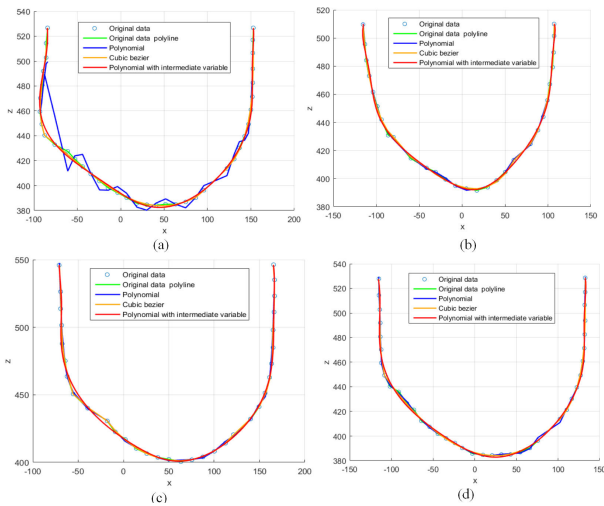
FIGURE 19. The exemplary curve fitting results for four typical cases. (a) Woman's bust. (b) Woman's waist. (c) Man's bust. (d) Man's waist.

TABLE 11. RMSE of the CBCF and PIVCF.

No.	Girth	CBCF (cm)	PIVCF (cm)
1	bust (W)	0.0787	0.1956
2	waist (W)	0.0724	0.1638
3	bust (W)	0.0655	0.1561
4	waist (W)	0.0972	0.1332
5	bust (W)	0.0447	0.2457
6	waist(M)	0.0822	0.1882
7	bust (M)	0.0747	0.1806
8	waist (M)	0.0872	0.1217
9	bust (M)	0.0563	0.1481
10	waist (M)	0.0710	0.1847

140 measurements. The RMSE of CBCF method is lower than that of PIVCF method. Therefore, the CBCF method shows a better fitting for the markers. However, the main purpose of our system is to measure the girth for garment design, while the markers are only a small part of the girth. We should double check the accuracy of these two methods for girth measurement. The fitting curve is divided into many short segments, and the length of the curve L is approximated by the sum of the lengths of these short segments ΔL_i.

$$L = \sum_{i=1}^{N} \Delta L_i = \sum_{i=1}^{N} \sqrt{(x_i - x_{i-1})^2 + (z_i - z_{i-1})^2} \quad (15)$$

In our experiment, $N = 100$. For garment design, the human body is considered to be symmetrical. Hence, the circumference of the measured girth should be $2L$. Table 12 shows the exemplary girth measurement results of these two methods. Table 13 shows the error of the results in Table 12. The maximum absolute error of the circumference obtained by CBCF is 1.36cm, while that by PIVCF is 1.03cm, of which the latter is less than the former. The MAD of the errors obtained by CBCF is 1.07cm, while that by PIVCF is 0.81cm, of which the latter is less than the former. In each measurement, the circumference obtained by PIVCF is always closer to the ground truth than that obtained by CBCF. Therefore,

TABLE 12. The exemplary girth measurement results.

No.	Girth	Ground truth (cm)	Original data polyline (cm)	CBCF (cm)	PIVCF (cm)
1	bust (W)	86.73	88.23	88.03	87.71
2	waist (W)	71.16	72.32	72.19	72.04
3	bust (W)	87.12	88.52	88.27	88.04
4	waist (W)	72.71	71.68	71.59	71.80
5	bust (W)	86.68	88.22	88.04	87.71
6	waist (M)	94.01	93.05	93.08	93.92
7	bust (M)	87.02	87.90	87.89	87.81
8	waist (M)	73.98	75.11	75.02	74.97
9	bust (M)	85.72	86.99	86.91	86.58
10	waist (M)	87.04	88.20	87.79	87.73

TABLE 13. The error of the exemplary girth measurement results.

No.	Girth	Original data polyline (cm)	CBCF (cm)	PIVCF (cm)
1	bust (W)	1.50	1.30	0.98
2	waist (W)	1.16	1.03	0.88
3	bust (W)	1.40	1.15	0.92
4	waist (W)	-1.03	-1.12	-0.91
5	bust (W)	1.54	1.36	1.03
6	waist (M)	-0.96	-0.93	-0.09
7	bust (M)	0.88	0.87	0.79
8	waist (M)	1.13	1.04	0.99
9	bust (M)	1.27	1.19	0.86
10	waist (M)	1.16	0.75	0.69
	MAD	1.20	1.07	0.81

the PIVCF method is more accurate for girth measurement than the CBCF method. We choose the PIVCF method to perform girth fitting in our system. We can not only measure the circumference of the girth, but also draw the contour of the girth, which is very helpful for customized garment design.

IV. EXPERIMENTAL RESULTS
A. EXPERIMENT SETUP
In practical girth measurement test, we adjust the focal length of the six POINT GREY GS3-U3-28S4C-C industrial cameras, so that they can capture the images of the subject with clear markers at a distance of 60cm. The normal breathing rate for an adult is 12 to 20 breaths per minute, so the cameras shoot at a low frame rate of 10 FPS. We choose the size measured manually with a tape as the ground truth of our test, in which anthropometric size definition and measurement method are strictly in accordance with China national standard GB/T 16160-2017 "Anthropometric definitions and methods for garment" [9].

A total of 70 subjects are tested in this experiment, including 33 women and 37 men, aged from 20 to 30 years old, with the height from 150cm to 185cm. Table14 shows the statistical characteristics of these subjects. To avoid occlusion, the subjects stand with their arms outstretched during the measurement. To avoid the error caused by breathing, 6 frame images shot at the same time by the 6 synchronous cameras are selected. Fig. 20 shows three pairs of stereo

images captured synchronously from the front (0°), side (90°) and back (180°) of the subject. To avoid random errors, each subject is measured manually and by our system 5 times, respectively. The average value of the 5 measurements is taken as the final measurement result.

TABLE 14. Statistical characteristics of subjects.

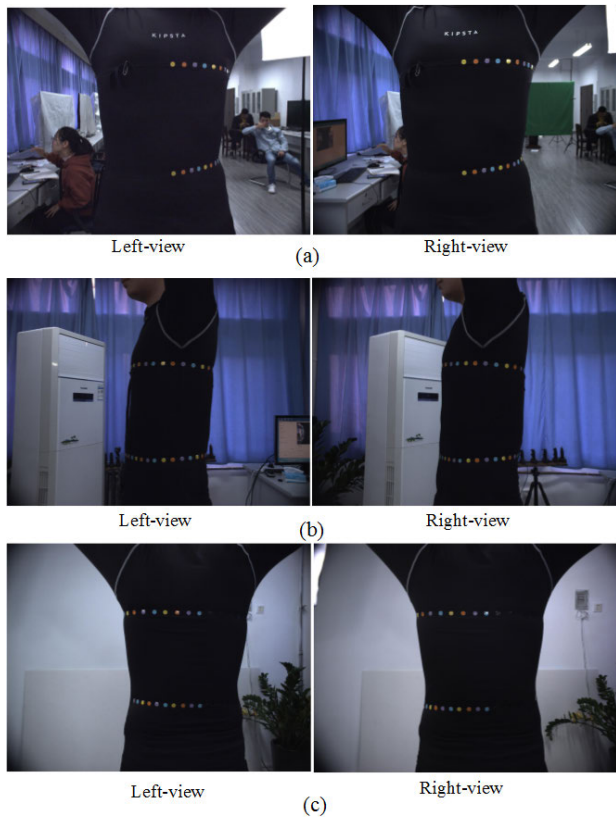|	Woman ($n_1 = 33$)	Man ($n_2 = 37$)	Total ($n=70$)
Age (years)	24.94 ± 4.85	25.72 ± 4.05	25.34 ± 4.46
Height (cm)	163.51 ± 8.35	174.91 ± 9.12	169.50 ± 8.66
Weight (kg)	57.71 ± 7.61	74.77 ± 9.99	66.93 ± 12.36

FIGURE 20. The exemplary three pairs of stereo images captured synchronously. (a) 0°. (b) 90°. (c) 180°.

Our girth measurement experiment is divided into two groups by gender: man and woman. For simplicity, only some of the measurement results are shown, including those with the maximum absolute errors.

B. GIRTH MEASUREMENT EXPERIMENT FOR WOMAN

Table 15 shows the girth measurement results of 12 subjects selected from a total of 33 female subjects, including 2 subjects with the maximum absolute error of bust and the maximum absolute error of waist. The remaining 10 subjects are selected randomly. Subject No.9 has the maximum absolute error of bust, i.e., 1.28cm. It conforms to China national standard GB/T 2665-2017 "Women's suits and coats" in which the tolerance for bust is ± 2.0cm [73]. Subject No.4 has the

TABLE 15. The exemplary girth measurement results of 12 female subjects.

No.	Girth	Proposed (cm)	Manual (cm)	Error (cm)	Error rate (%)
1	bust	83.32	84.54	-1.22	1.44
	waist	74.77	73.78	0.99	1.34
2	bust	85.33	86.29	-0.96	1.11
	waist	75.26	76.15	-0.89	1.17
3	bust	82.43	81.78	0.65	0.79
	waist	71.10	70.01	1.09	1.56
4	bust	90.88	90.03	0.85	0.94
	waist	82.73	81.55	1.18	1.45
5	bust	85.46	86.41	-0.95	1.10
	waist	73.46	72.63	0.83	1.14
6	bust	85.07	84.12	0.95	1.13
	waist	74.07	74.92	-0.85	1.13
7	bust	88.61	89.67	-1.06	1.18
	waist	77.97	78.90	-0.93	1.18
8	bust	90.22	89.31	0.91	1.02
	waist	81.22	82.16	-0.94	1.14
9	bust	84.30	83.02	1.28	1.54
	waist	73.13	73.77	-0.64	0.87
10	bust	87.87	86.9	0.97	1.12
	waist	77.16	78.18	-1.02	1.30
11	bust	90.68	91.53	-0.85	0.93
	waist	80.06	79.68	0.38	0.48
12	bust	89.61	88.45	1.16	1.31
	waist	77.20	76.27	0.93	1.22

maximum absolute error of waist, i.e., 1.18cm. It conforms to China textile industry standard FZ/T 81004-2012 "Dress and lady suit" in which the tolerance for waist is ± 1.5cm [74]. Fig. 21 shows the comparison of the bust and waist measurement results of these 12 subjects between our proposed method and the manual method. The blue line with square represents the measurement results by our proposed method, while the red line with circle represents the measurement results by manual method. The two lines are very close and almost overlapping.
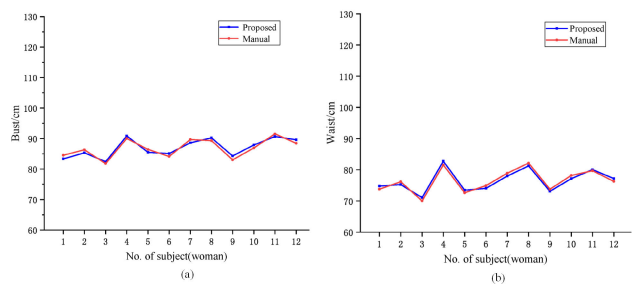

FIGURE 21. Girth measurement results comparison for woman. (a) Bust. (b) Waist.

Table 16 shows the statistical analysis of the girth measurement results of the total of 33 female subjects. It can be seen that the mean value μ and standard deviation σ of the measurement results by our proposed method and the manual method are almost the same, which indicates that the proposed method can replace the manual method. The MAD of the measurement error for bust is 0.98cm, and the corresponding data tolerance is $100\% \leq |\pm 1.5|$cm. The MAD of the measurement error for waist is 0.87cm, and the corresponding data tolerance is $100\% \leq |\pm 1.5|$cm.

TABLE 16. Statistical analysis of the girth measurement results of the total of 33 female subjects.

Woman		Sample Size	μ (cm)	σ (cm)	MAD (cm)	Data Tolerance
bust	proposed	33	86.27	4.36	0.98	$100\% \leq \mid \pm 1.5\mid$cm
	manual	33	86.10	4.11		
waist	proposed	33	75.87	4.26	0.87	$100\% \leq \mid \pm 1.5\mid$cm
	manual	33	75.78	4.25		

TABLE 17. The exemplary girth measurement results of 13 male subjects.

No.	Girth	Proposed (cm)	Manual (cm)	Error (cm)	Error rate (%)
13	bust	91.01	90.44	0.57	0.63
	waist	83.74	82.85	0.89	1.07
14	bust	92.05	92.98	-0.93	1.00
	waist	84.42	85.18	-0.76	0.89
15	bust	88.16	87.12	1.04	1.19
	waist	80.28	81.09	-0.81	1.00
16	bust	82.04	80.73	1.31	1.62
	waist	74.97	73.98	0.99	1.34
17	bust	84.59	85.56	-0.97	1.13
	waist	80.11	79.38	0.73	0.92
18	bust	89.56	88.63	0.93	1.05
	waist	83.35	84.24	-0.89	1.06
19	bust	93.92	94.70	-0.78	0.82
	waist	90.73	89.88	0.85	0.95
20	bust	88.40	87.23	1.17	1.34
	waist	87.73	87.04	0.69	0.79
21	bust	90.18	91.10	-0.92	1.01
	waist	86.2	85.21	0.99	1.16
22	bust	92.41	91.34	1.07	1.17
	waist	89.54	90.43	-0.89	0.98
23	bust	95.41	94.24	1.17	1.24
	waist	90.95	91.39	-0.44	0.48
24	bust	93.69	92.73	0.96	1.04
	waist	93.92	94.01	-0.09	0.10
25	bust	88.65	87.59	1.06	1.21
	waist	83.43	82.50	0.93	1.13

FIGURE 22. Girth measurement results comparison for man. (a) Bust. (b) Waist.

C. GIRTH MEASUREMENT EXPERIMENT FOR MAN

Table 17 shows the girth measurement results of 13 subjects selected from a total of 37 male subjects, including 3 subjects with the maximum absolute error of bust and the maximum absolute error of waist. The remaining 10 subjects are selected randomly. Subject No.16 has the maximum absolute error of bust, i.e., 1.31cm. It conforms to China national standard GB/T 2664-2017 "Men's suits and coats" in which the tolerance for bust is ±2.0cm [75]. Subject No.16 and No.21 have the maximum absolute error of waist, i.e., 0.99cm. It conforms to China national standard GB/T 2666-2017 "Trousers" in which the tolerance for waist is ±1.0cm [76]. Fig. 22 shows the comparison of the bust and waist measurement results of these 13 subjects between our proposed method and the manual method. The blue line with square represents the measurement results by our proposed method, while the red line with circle represents the measurement results by manual method. The two lines are very close and almost overlapping.

Table 18 shows the statistical analysis of the girth measurement results of the total of 37 male subjects. It can be seen that the mean value μ and standard deviation σ of the
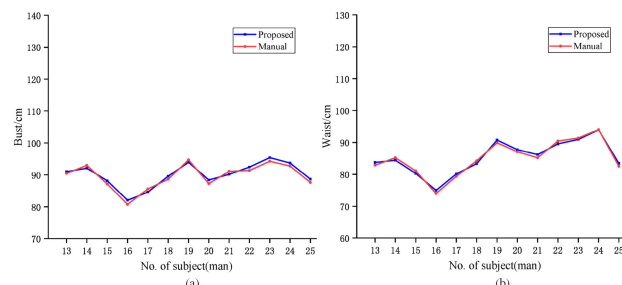
measurement results by our proposed method and the manual method are almost the same, which indicates that the proposed method can replace the manual method. The MAD of the measurement error for bust is 0.99cm, and the corresponding data tolerance is $100\% \leq \mid \pm 1.5\mid$cm. The MAD of the measurement error for waist is 0.83cm, and the corresponding data tolerance is $100\% \leq \mid \pm 1.0\mid$cm.

Overall, the maximum measurement error of bust is 1.28cm for woman and 1.31cm for man, which are within the ±2.0cm tolerance of bust for woman and ±2.0cm tolerance of bust for man regulated by national standards. The maximum measurement error of waist is 1.18cm for woman and 0.99cm for man, which are also within the ±1.5cm tolerance of waist for woman and ±1.0cm tolerance of waist for man regulated by textile industry standard and national standard. In summary, the error mainly comes from four steps: semantic segmentation, stereo matching, coordinate calculation and girth fitting. As mentioned above, the contributions of semantic segmentation and coordinate calculation to the error are relatively small and can be ignored. The contribution of girth fitting to the error is almost constant. However, the contribution of stereo matching to the error decreases with the increase of matching accuracy, which is the main contribution to the final girth measurement error and should be further improved in the future.

We compare the girth measurement error with five other cost-effective and portable anthropometric methods, namely, Sara *et al.*'s method [19], Han *et al.*'s method [77], Lu *et al.*'s method [78], Dekker *et al.*'s method [79] and Gu *et al.*'s method [25], as shown in Table 19. The bust MAD of our proposed system is 0.99cm for man and 0.98cm for woman, which is less than the bust MAD of [77], [78] and [79] with 1.97cm, 1.45cm and 1.60cm, respectively. The waist MAD of our proposed system is 0.83cm for man and 0.87cm for woman, which is less than the waist MAD of [19], [77], [78] and [79] with 2.57cm, 2.03cm, 1.47cm and 2.50cm,

TABLE 18. Statistical analysis of the girth measurement results of the total of 37 male subjects.

Man		Sample Size	μ (cm)	σ (cm)	MAD (cm)	Data Tolerance
bust	proposed	37	89.59	3.71	0.99	$100\% \leq \| \pm 1.5\|$cm
	manual	37	89.27	3.92		
waist	proposed	37	84.73	4.66	0.83	$100\% \leq \| \pm 1.0\|$cm
	manual	37	84.67	4.87		

TABLE 19. Error comparison for girth measurement.

	MAD (cm)		Data Tolerance	
	bust	waist	bust	waist
Sara et al. [19]	NA	2.57	NA	NA
Han et al. [77]	1.97	2.03	NA	NA
Lu et al. [78]	1.45	1.47	NA	NA
Dekker et al. [79]	1.60	2.50	NA	NA
Gu et al. [25]	NA	NA	$86\% \leq \| \pm 2.0\|$cm (W)	$98\% \leq \| \pm 1.5\|$cm (W)
PROPOSED SYSTEM	0.98 (W)/ 0.99 (M)	0.87 (W)/ 0.83 (M)	$100\% \leq \| \pm 1.5\|$cm (W)	$100\% \leq \| \pm 1.5\|$cm (W)

respectively. The bust data tolerance of our proposed system is $100\% \leq \| \pm 1.5\|$cm for woman, which is less than that of [25] with $86\% \leq \| \pm 2.0\|$cm for woman. The waist data tolerance of our proposed system is $100\% \leq \| \pm 1.5\|$cm for woman, which is less than that of [25] with $98\% \leq \|\pm 1.5\|$cm for woman. In summary, our system not only can measure the girth simply and intelligently with low cost and portability, but also can achieve better measurement accuracy than other methods.

V. CONCLUSION

In this study, we solved the problem of intelligently measuring girth on the basis of images captured by the multi-view stereovision system. We presented a system composed of girth region semantic segmentation, marker stereo matching and coordinates calculation, unified coordinates transformation and girth fitting. We integrated girth semantic segmentation within the PSPNet+ network structure particularly for accurate and intelligent girth semantic segmentation. We classified the segmented images into different color subspaces and clustered the pixels in each color subspace into several clusters corresponding to markers. We performed stereo matching only on the corresponding clusters to obtain the matching marker pairs. We calculated the space coordinates of the real 3D points corresponding to stereo matching marker pairs with the calibration parameters of the stereovision camera pairs. We transformed the space coordinates of markers into one unified coordinates. We did curve fitting on the markers with unified coordinates and calculated the length of the fitting curve. The girth was measured, and the contour of the girth was depicted. The girth measurement performance of our proposed system was verified by the experiments of bust and waist measurement for woman and man. The results show that our system is efficient and reliable in the practical application of girth measurement. In our measurements, the measured girths have a maximum bust absolute error of 1.28cm for woman and 1.31cm for man, which are within the ± 2.0cm error limit of China national standard GB/T 2665-2017 and the ± 2.0cm error limit of GB/T 2664-2017. The measured girths also have a maximum waist absolute

error of 1.18cm for woman and 0.99cm for man, which are within the ± 1.5cm error limit of China textile industry standard FZ/T 81004-2012 and the ± 1.0cm error limit of China national standard GB/T 2666-2017. In particular, our system is passive and portable, suitable for quick and accurate girth measurement, and with low cost.

REFERENCES

[1] J. Zhang, K. Liu, M. Dong, H. Yuan, C. Zhu, and X. Zeng, "An intelligent garment recommendation system based on fuzzy techniques," *J. Textile Inst.*, vol. 111, pp. 1324–1330, 2020.

[2] K. Liu, X. Zeng, X. Tao, and P. Bruniaux, "Associate design of fashion sketch and pattern," *IEEE Access*, vol. 7, pp. 48830–48837, 2019.

[3] K. Liu, X. Zeng, P. Bruniaux, J. Wang, E. Kamalha, and X. Tao, "Fit evaluation of virtual garment try-on by learning from digital pressure data," *Knowl.-Based Syst.*, vol. 133, pp. 174–182, Oct. 2017.

[4] K. Liu, X. Zeng, J. Wang, X. Tao, J. Xu, X. Jiang, J. Ren, E. Kamalha, T.-K. Agrawal, and P. Bruniaux, "Parametric design of garment flat based on body dimension," *Int. J. Ind. Ergonom.*, vol. 65, pp. 46–59, May 2018.

[5] K. Liu, X. Zeng, P. Bruniaux, X. Tao, X. Yao, V. Li, and J. Wang, "3D interactive garment pattern-making technology," *Comput.-Aided Des.*, vol. 104, pp. 113–124, Nov. 2018.

[6] K. Liu, C. Zhu, X. Tao, P. Bruniaux, and X. Zeng, "Parametric design of garment pattern based on body dimensions," *Int. J. Ind. Ergonom.*, vol. 72, pp. 212–221, Jul. 2019.

[7] B. Gu, G. Liu, and B. Xu, "Individualizing women's suit patterns using body measurements from two-dimensional images," *Textile Res. J.*, vol. 87, no. 6, pp. 669–681, Apr. 2017.

[8] J. Su, Y. Ke, C. Kuang, B. Gu, and B. Xu, "Converting lower-body features from three-dimensional body images into rules for individualized pant patterns," *Textile Res. J.*, vol. 89, no. 11, pp. 2199–2208, Jun. 2019.

[9] *Anthropometric Definitions and Methods for Garment*, China National Standard Gb/t 16160-2017, AQSIQ, 2017.

[10] *Cyberware*. Accessed: Sep. 2, 2020. [Online]. Available: https://www.alibaba.com/product-detail/Cyberware-3D-Whole-Body-Scanner_112532870.html

[11] H. A. M. Daanen, S. E. Taylor, M. A. Brunsman, and J. H. Nurre, "Absolute accuracy of the cyberware WB4 whole body scanner," *Proc. SPIE*, vol. 3023, p. 34, Mar. 1997.

[12] *Vitus Smart LC3*. Accessed: Sep. 2, 2020. [Online]. Available: http://www.leatech.net/plus/view.php?aid=52

[13] *Boss-21*. Accessed: Sep. 2, 2020. [Online]. Available: http://www.vis.ca/product.asp?id=17

[14] P. Treleaven and J. Wells, "3D body scanning and healthcare applications," *Computer*, vol. 40, no. 7, pp. 28–34, Jul. 2007.

[15] *Ditus*. Accessed: Sep. 2, 2020. [Online]. Available: http://www.leatech.net/plus/view.php?aid=338

[16] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D full human bodies using kinects," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 4, pp. 643–650, Apr. 2012.

[17] P. Qiujuan, "Eye safety analysis for 400–1400 nm pulsed lasers systems," *Laser Infr.*, vol. 40, no. 8, pp. 821–824, 2010.

[18] S. Bragança, P. Arezes, M. Carvalho, S. P. Ashdown, I. Castellucci, and C. Leão, "A comparison of manual anthropometric measurements with kinect-based scanned measurements in terms of precision and reliability," *Work*, vol. 59, no. 3, pp. 325–339, Apr. 2018.

[19] S. Bragança, P. Arezes, M. Carvalho, S. P. Ashdown, B. Xu, and I. Castellucci, "Validation study of a kinect based body imaging system," *Work*, vol. 57, no. 1, pp. 9–21, 2017.

[20] 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. 2016.

[21] Y. Chen, Z.-Q. Cheng, C. Lai, R. R. Martin, and G. Dang, "Realtime reconstruction of an animating human body from a single depth camera," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 8, pp. 2000–2011, Aug. 2016.

[22] Q. Ran, K. Zhou, Y.-L. Yang, J. Kang, L. Zhu, Y. Tang, and J. Feng, "High-precision human body acquisition via multi-view binocular stereopsis," *Comput. Graph.*, vol. 87, pp. 43–61, Apr. 2020.

[23] C. BenAbdelkader and Y. Yacoob, "Statistical body height estimation from a single image," in *Proc. 8th IEEE Int. Conf. Automat. Face Gesture Recognit.*, Sep. 2008, pp. 1–7.

[24] B. Penders, R. Brecheisen, A. Gerver, G. van Zonneveld, and W.-J. Gerver, "Validating paediatric morphometrics: Body proportion measurement using photogrammetric anthropometry," *J. Pediatric Endocrinol. Metabolism*, vol. 28, nos. 11–12, pp. 1357–1362, Jan. 2015.

[25] B. Gu, G. Liu, and B. Xu, "Girth prediction of young female body using orthogonal silhouettes," *J. Textile Inst.*, vol. 108, no. 1, pp. 140–146, Jan. 2017.

[26] B. Gu, X. Li, Y. Yan, and J. Su, "Predicting human dimensions in body images for automatic generation of female pants patterns," *Textile Res. J.*, vol. 89, no. 18, pp. 3792–3801, Sep. 2019.

[27] Z. Li, W. Jia, Z.-H. Mao, J. Li, H.-C. Chen, W. Zuo, K. Wang, and M. Sun, "Anthropometric body measurements based on multi-view stereo image reconstruction," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 366–369.

[28] Y. Liu, A. Sowmya, and H. Khamis, "Single camera multi-view anthropometric measurement of human height and mid-upper arm circumference using linear regression," *PLoS ONE*, vol. 13, no. 4, Apr. 2018, Art. no. e0195600.

[29] J. Ma, H. Zhang, P. Yi, and Z. Wang, "SCSCN: A separated channel-spatial convolution net with attention for single-view reconstruction," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8649–8658, Oct. 2020.

[30] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[31] S. Kpodjedo, F. Ricca, P. Galinier, G. Antoniol, and Y.-G. Gueheneuc, "MADMatch: Many-to-many approximate diagram matching for design comparison," *IEEE Trans. Softw. Eng.*, vol. 39, no. 8, pp. 1090–1111, Aug. 2013.

[32] J. Vanne, E. Aho, T. D. Hamalainen, and K. Kuusilinna, "A high-performance sum of absolute difference implementation for motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 876–883, Jul. 2006.

[33] M. B. Hisham, S. N. Yaakob, R. A. A. Raof, A. B. A. Nazren, and N. M. Wafi, "Template matching using sum of squared difference and normalized cross correlation," in *Proc. IEEE Student Conf. Res. Develop. (SCOReD)*, Dec. 2015, pp. 100–104.

[34] J. Yang, Y. Shi, and G. Jia, "Finger-vein image matching based on adaptive curve transformation," *Pattern Recognit.*, vol. 66, pp. 34–43, Jun. 2017.

[35] C. Zhang, C. He, Z. Chen, W. Liu, M. Li, and J. Wu, "Edge-preserving stereo matching using minimum spanning tree," *IEEE Access*, vol. 7, pp. 177909–177921, 2019.

[36] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, "Robust L_2E estimation of transformation for non-rigid registration," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015.

[37] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[38] H. Bay, "Surf: Speeded up robust features," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 404–417, 2006.

[39] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.

[40] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, May 2019.

[41] J. Ma, X. Jiang, J. Jiang, J. Zhao, and X. Guo, "LMR: Learning a two-class classifier for mismatch removal," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4045–4059, Aug. 2019.

[42] B. Dasgupta and B. N. Chatterji, "Fourier-Mellin transform based image matching algorithm," *IETE J. Res.*, vol. 42, no. 1, pp. 3–9, Jan. 1996.

[43] G. Lazaridis and M. Petrou, "Image registration using the Walsh transform," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2343–2357, Aug. 2006.

[44] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.

[45] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, Dec. 2013.

[46] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, pp. 736–746, 2020.

[47] J. Zhang and J. Hu, "Image segmentation based on 2D Otsu method with histogram analysis," in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, 2008, pp. 105–108.

[48] N.-J. Kwak, D.-J. Kwon, Y.-G. Kim, and J.-H. Ahn, "Color image segmentation using edge and adaptive threshold value based on the image characteristics," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, 2004, pp. 555–558.

[49] A. K. C. Wong and P. K. Sahoo, "A gray-level threshold selection method based on maximum entropy principle," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 4, pp. 866–871, Jul./Aug. 1989.

[50] J. Q. Miao, "Based on canny operator with regional growth combined research on accurate image segmentation algorithm," *Adv. Mater. Res.*, vols. 989–994, pp. 2088–2092, Jul. 2014.

[51] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[52] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, Jul. 2001, pp. 105–112.

[53] S. Han, W. Tao, D. Wang, X.-C. Tai, and X. Wu, "Image segmentation based on GrabCut framework integrating multiscale nonlinear structure tensor," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2289–2302, Oct. 2009.

[54] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 512–519.

[55] C. Peng and J. Ma, "Semantic segmentation using stride spatial pyramid pooling and dual attention decoder," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107498.

[56] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[57] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1520–1528.

[58] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, "Semantic image segmentation via deep parsing network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1377–1385.

[59] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5168–5177.

[60] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1529–1537.

[61] G. Lin, C. Shen, A. van den Hengel, and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3194–3203.

[62] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[63] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[64] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 1, Sep. 1999, pp. 666–673.

[65] 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.

[66] *Look Into Person*. Accessed: Sep. 2, 2020. [Online]. Available: http://sysu-hcp.net/lip/overview.php

[67] X. Song, L. Yang, Y. Wu, and Z. Liu, "A new depth measuring method for stereo camera based on converted relative extrinsic parameters," in *Proc. Int. Symp. Photoelectron. Detection Imag., Imag. Sensors Appl.*, Aug. 2013, pp. 826–890.

[68] G. Zhang, *Vision Measurements*. Beijing, China: Beijing Science Press, 2008.

[69] G. Xuan, "Three dimensional moment invariants under rigid transformation," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 1993, pp. 518–522.

[70] A. S. Deane, E. P. Kremer, and D. R. Begun, "New approach to quantifying anatomical curvatures using high-resolution polynomial curve fitting (HR-PCF)," *Amer. J. Phys. Anthropol.*, vol. 128, no. 3, pp. 630–638, Nov. 2005.

[71] H. Li, Y. Luo, and J. Wu, "Collision-free path planning for intelligent vehicles based on Bézier curve," *IEEE Access*, vol. 7, pp. 123334–123340, 2019.

[72] J.-L. Hu, J.-B. Deng, and C. Hu, "An algorithm for circle curve fitting based on the constrained least square model represented by mosaic observation points," in *Proc. Wireless Commun., Netw. Mobile Comput.*, 2009, ppp. 1–4.

[73] *Women's Suits and Coats*, China National Standard Gb/t 2665-2017, AQSIQ, 2017.

[74] *Dress and Lady Suit*, China Textile Industry Standard Fz/t 81004-2012, AQSIQ, 2012.

[75] *Men's Suits and Coats*, China National Standard Gb/t 2664-2017s, AQSIQ, 2017.

[76] *Trousers*, China National Standard Gb/t 2666-2017, AQSIQ, 2017.

[77] H. Han, Y. Nam, and K. Choi, "Comparative analysis of 3D body scan measurements and manual measurements of size korea adult females," *Int. J. Ind. Ergonom.*, vol. 40, no. 5, pp. 530–540, Sep. 2010.

[78] J.-M. Lu and M.-J.-J. Wang, "The evaluation of scan-derived anthropometric measurements," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 8, pp. 2048–2054, Aug. 2010.

[79] L. D. Dekker, "3D human body modelling from range data," Ph.D. dissertation, Univ. London, London, U.K., 2000.

XIAOWEI SONG (Member, IEEE) received the Ph.D. degree in signal and information processing from Tianjin University, Tianjin, China, in 2007. From September 2012 to September 2013, he was a Visiting Researcher with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. He is currently a Professor with the School of Electronic and Information, Zhongyuan University of Technology, Zhengzhou, China. His research interests include 3D image and video processing, 3D display, and computer vision.

MENGLONG LI received the B.S. degree in electronic and information engineering from the Information Engineering University of PLA, Zhengzhou, China, in 2011, and the M.S. degree in signal and information processing from the Zhongyuan University of Technology, Zhengzhou, in 2019. His research interests include 3D image processing, stereo vision, and computer vision.

CHUNPING HOU received the M.Eng. and Ph.D. degrees in electronic engineering from Tianjin University, Tianjin, China, in 1986 and 1998, respectively.

Since 1986, she has been the Faculty of the School of Electronic and Information Engineering, Tianjin University, where she is currently a Full Professor and the Director of the Broadband Wireless Communications and 3D Imaging Institute. Her current research interests include 3D image processing, 3D display, wireless communication, and design and applications of communication systems.

LEI YANG (Member, IEEE) received the Ph.D. degree in communication and information system from Tianjin University, Tianjin, China, in 2007. From July 2017 to July 2018, she was a Visiting Researcher with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. She is currently an Associate Professor with the School of Electronic and Information, Zhongyuan University of Technology, Zhengzhou, China. Her research interests include 3D image processing, 3D display, and computer vision.

QIANQIAN HUANG received the B.S. degree in electronic and information engineering from the Zhongyuan University of Technology, Zhengzhou, China, in 2017, where she is currently pursuing the M.S. degree with the School of Electronics and Information. Her research interests include 3D image processing, stereo vision, and computer vision.
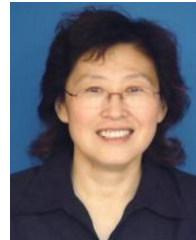
ZIXIANG XIONG (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, in 1996.

From 1995 to 1997, he was a Visiting Student with Princeton University, where he was a Research Associate. From 1997 to 1999, he was with the University of Hawaii. Since 1999, he has been with the Department of Electrical and Computer Engineering, Texas A&M University, where he is currently a Professor and the Associate Department Head. He spent his sabbatical leaves at Stanford University, in spring 2010, and Monash University, Australia, from 2017 to 2018. His main research interests include image/video processing, computer vision, virtual/augmented reality, big data, and communications. He was a co-recipient of the 2006 IEEE Signal Processing Magazine Best Paper Award, the top 10% paper awards at the 2011 and 2015 IEEE Multimedia Signal Processing Workshops, the IBM Best Student Paper Award at the 2016 IEEE International Conference on Pattern Recognition, and the Best Demo Paper Award at the 2018 IEEE International Conference on Multimedia and Expo. He received the NSF Career Award, in 1999, the ARO Young Investigator Award, in 2000, and the ONR Young Investigator Award, in 2001. He has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, from 1999 to 2005, the IEEE TRANSACTIONS ON IMAGE PROCESSING, from 2002 to 2005, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, from 2002 to 2006, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS (Part B), from 2005 to 2009, and the IEEE TRANSACTIONS ON COMMUNICATIONS, from 2008 to 2013. He is currently an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA.

● ● ●