



3D foot scanning using multiple RealSense cameras

Munan Yuan^{1,2} · Xiaofeng Li^{1,2} · Jinlin Xu^{1,2} · Chaochuan Jia³ · Xiru Li^{1,2}

Received: 19 December 2018 / Revised: 26 May 2020 / Accepted: 9 September 2020 /
Published online: 10 October 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

3D scanning of the foot is of great significance for footwear customization, intelligent shoe size recommendation and foot disease diagnosis. In this paper, we propose a 3D foot scanning system that consists of four Intel RealSense cameras and one host PC and scans both feet simultaneously. A novel calibration method that is based on a Tower-type block was proposed for calculating the extrinsic parameters of multiple RGB-Depth cameras. The Tower-type block was designed to realize the automatic execution of the multi-camera calibration process and reduce the operational complexity. This paper introduced the complete procedure of the system, including partial view scanning, point cloud filtering, registration, and non-visible area filling, reconstruction and foot measurement. The presented experimental results demonstrated that the proposed methods were efficient and versatile approaches for 3D foot scanning.

Keywords 3D scanning · Calibration · Reconstruction · RealSense

1 Introduction

In the traditional footwear industry, shoes are categorized according to length and width to fit the population. People generally select shoes only depend on their foot sizes even with some discomfort, which causes hidden dangers of foot disease [21, 33]. The customized shoe last with traditional manual methods have time-consuming and complex processes [15, 34]. With the development of optical measurement technology, binocular vision, phase measurement, digital holography and structured light methods develop as the major machine measurement for mass shoe customization. These optical methods measure foot parameters with a complex and expensive way [14, 19, 20]. Nowadays, 3D geometry models of real-world objects have

✉ Munan Yuan
mnyuan@hfcas.ac.cn

¹ Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China

² University of Science and Technology of China, Hefei 230026, China

³ Shandong University of Science and Technology, Qingdao 266590, China

been employed to collect anthropometric data for many application scenarios [23, 24]. There are several pioneering works, which focusing on interesting foot scene reconstruction from 3D scanner, have been studied extensively in computer graphics [9, 31]. Important researches have been obtained that are reliable to extract foot dimensions, compared with traditional methods [12, 26, 27]. 3D foot scanner has become a hot topic, which is of great significance for mass footwear customization, intelligent shoe size recommendation and foot disease diagnosis [22, 27, 28].

With the development of optical measurement technology, 3D foot scanners that use various technologies have been developed for obtaining the 3D shapes of feet [13, 18]. Some 3D scanners, using laser-based technology, have been able to obtain accurate 3D shapes of feet [9, 16]. However, these scanners are very expensive and the process usually takes a long time. In addition, people often have a psychological fear of lasers [31]. Therefore, laser-based 3D scanners are mainly used in research areas and difficult to widely promote. The multi-view method is fast; however, its scan accuracy is worse than that of the laser-based method. In addition, the multi-view method is computationally expensive and easily affected by the surrounding environment [8]. In recent years, devices of a new type, namely, RGB-depth cameras, such as PrimeSense, Kinect and RealSense, have received substantial attention. Compared with traditional 3D scanning equipment, these RGB-depth cameras are inexpensive, compact and easy to use. Many researchers have tried to build low-cost 3D scanning systems that are based on RGB-depth cameras.

Several methods [6, 7, 29] use a single RGB-depth camera to scan the object. In these methods, the RGB-depth camera must move around the target and merge all the single views correctly to obtain a closed 3D model [7]. Moving the camera increases the scan time and requires the target to maintain a static posture for a long time [32]. As a comparison, in a system that is based on multiple RGB-depth cameras, various parts of the subject can be scanned from multiple viewpoints simultaneously in a short time. For example, Tong et al. [25] proposed a scanning system that uses three Kinects to scan an object on a turntable. Liu et al. [11] used six RGB-depth cameras to reconstruct a real 3D moving object in the form of a polygonal mesh automatically. Alexiadis et al. [1, 2] proposed a reconstruction method that uses five Kinect sensors in a circular spatial arrangement that provides full coverage of the object. Although these multi-camera methods are available, they are mainly designed for body scanning [31].

In this paper, using four low-cost RealSense cameras, we design a 3D foot scanning system that can scan both feet simultaneously and complete the 3D scanning process in a few seconds. The detailed system setup is described in the second section. To merge the 3D point clouds that are captured by different cameras, we designed a Tower-type calibration block, based on which a novel calibration method for multiple RGB-depth cameras is proposed in the third section. The detailed implementation of scanning is described in the fourth section. The experimental procedure and results are described in section five. In the sixth part, we present the conclusions of the work of this paper.

2 System setup

2.1 Intel RealSense SR300 sensor

The proposed method is based on Intel RealSense SR300, which is the second generation of front-facing Intel RealSense cameras [5]. The device, as shown in Fig. 1, includes an FHD

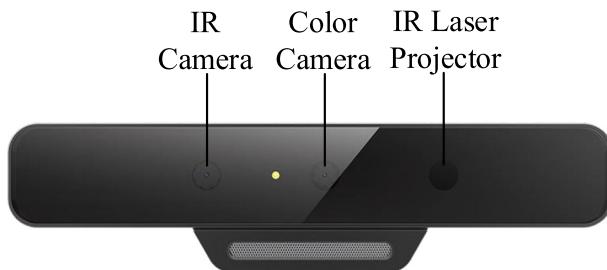


Fig. 1 Intel RealSense SR300

Color Camera, a fast VGA infrared (IR) camera and an IR laser projector. It can provide synchronized color, depth and IR video streaming. The device provides a depth camera resolution of 640×480 pixels, a depth frame rate of 60 fps and a depth field of view (FOV) of diagonal $88^\circ \times$ width $71.5^\circ \times$ height 55° . The effective range of depth is optimized from 0.2 m to 1.5 m and suggested to stay within 70 cm for use as a 3D scanner.

2.2 System layout

The proposed system, as shown in Fig. 2(a), consists of four SR300 cameras, which are connected to a single host computer. All cameras, which are labelled as #1, #2, #3 and #4, are horizontally fixed on four brackets, each of which is designed as shown in Fig. 2(b) and can adjust the position of the camera in the horizontal, vertical, front and back directions. Since the depth coordinate origin of SR300 is located at the center of the IR sensor, we align each SR300's IR sensor center with the diagonals of the square platform.

After configuration, the final layout of the system is shown in Fig. 3. Four cameras, with the IR sensor as the center, are positioned at the vertices of an imaginary square with a length of approximately 60 cm and point horizontally toward the center of the platform diagonally. Cameras #1 and #2 scan from the back of the object foot, and the other two cameras scan from the front. The height of cameras #1 and #2 is 10 cm, and the height of #3 and #4 is 12 cm. According to the camera's position, orientation and field of view, we define a square area with 36-cm sides as the effective area in the center of the scanning platform. When scanning, both feet are placed in the effective area to avoid incomplete capture.

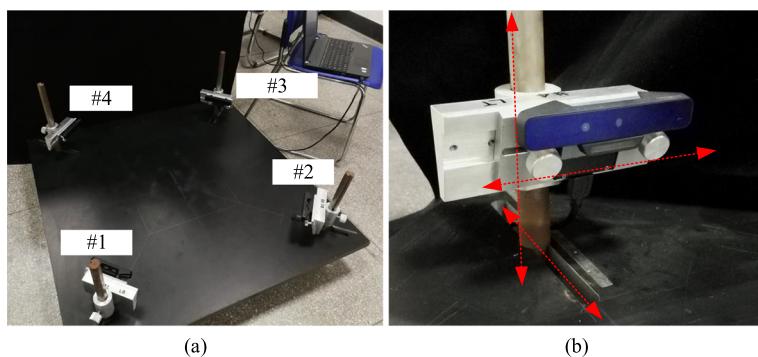


Fig. 2 System equipment: **a** Overview of the system; **b** the camera bracket

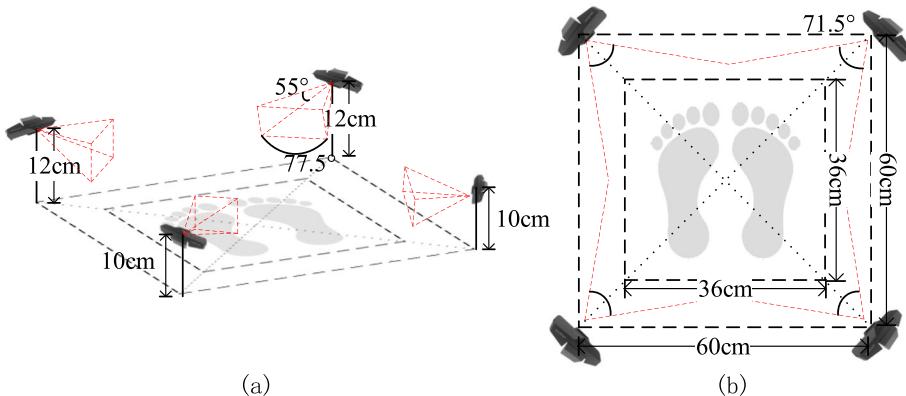


Fig. 3 System layout: **a** Overview and **b** top view

3 Calibration method for multiple cameras

The observed target is unable to be captured from different viewpoints simultaneously by multi cameras. 3D scanning systems based on RGB-D depth should to be calibrated prior to scanning, which promise measurement precision crossing a greater depth of field. To align the 3D point clouds that were captured by multiple cameras into unified 3D coordinates, external parameters need to be calculated to determine the relative position of each camera and the unified coordinate system. Many studies have been reported that checkerboard and extra objects methods obtain great progress in calibration [36]. These traditional calibration methods always need manual complicated operations, which attempt to be time consuming. In recent years, serval methods are proposed to improve calibration process. A T-shaped checkerboard introduced by Ge Wu makes contribution to multi-cameras calibration accuracy promotion and foot model alignment without manual intervention [31]. However, the accuracy of the calibration is limited due to noticeable depth errors.

3.1 Tower-type block

We propose an automatic calibration method for multiple depth cameras that uses a custom-made tower-type block. The tower-type block, as shown in Fig. 4, is white in color and has one base, three floors and four sides. The height of the base is 1.8 cm. Each floor is the same height of 5 cm. The widths of the floors from top to bottom are 12 cm, 16 cm and 20 cm. With a black hollow circle as the mark, each side of the bottom and middle floors has two marks and each side of the top floor has one mark.

3.2 Coordinate setting

In the proposed system, we define five left-handed 3D coordinate systems: four camera coordinates (XYZ_{ci} , $i = 1 \dots 4$) and the world coordinate system (XYZ_w). The origin of XYZ_{ci} is located at the center of the depth sensor of the camera. In XYZ_w , the origin is at the center of the platform, the Z-axis points forward in the same direction as the feet, the X-axis points to the right, and the Y-axis points upwards. As shown in Fig. 5(a), the tower-type block is placed at the center of the system platform so that the vertices of the base are aligned with the platform diagonally and each side of the tower is facing a camera.

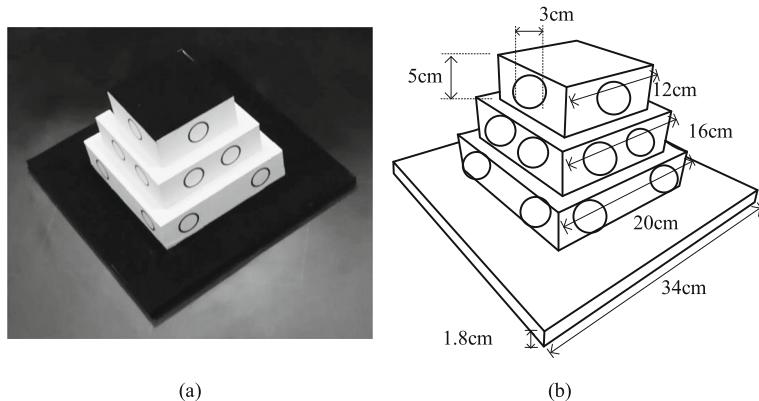


Fig. 4 Tower-type block: **a** Physical object and **b** structure and dimensions

With the center of the circle as the feature point, there are 5 feature points on each side of the tower-type block. According to the structure and dimensions of the block, we can measure the 3D coordinates of each feature point under XYZ_w , as shown in Fig. 5(b). Then, we obtain 4 groups of 3D points. Each point is numbered 1 to 5 from top to bottom and from left to right.

3.3 Calibration process

In the following sections, we describe the steps of our calibration method in detail.

Step 1. Identifying feature points from IR images

Each camera simultaneously captures a depth image and an IR image of one side of the tower-type block. We developed a program to identify feature points from the IR image using a circular Hough transform based method [3]. As shown in Fig. 6, the program has obtained the pixel coordinates of the feature points from each view. After repeated experiments, we found that the system has the best detection effect when the minimum detection radius is set to 12 and the maximum detection radius is set to 30.

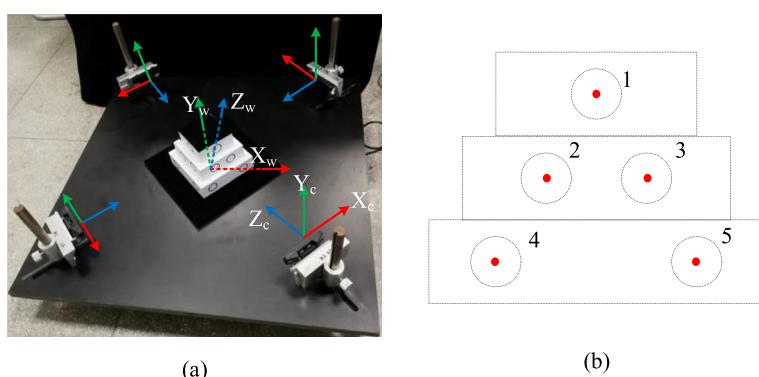


Fig. 5 Coordinate system: **a** Coordinates of world and cameras and **b** world coordinates of all target points



Fig. 6 Identifying feature points from IR images that were captured by different cameras

The order of the five feature points obtained on each view is random, and our program will sort them according to the order of the system coordinate setting. Since the depth image and IR image are aligned, the corresponding depth value of the feature points can be obtained according to the pixel coordinates of the feature points at the corresponding position of the depth image.

Step 2. Obtaining 3D coordinates of feature points under camera coordinates

Let point (u, v) be the 2D pixel coordinates of a feature point in the IR image, and let the point (x_c, y_c, z_c) be the corresponding 3D camera coordinates. Point (u, v) can be transformed to point (x_c, y_c, z_c) via Eq. (1):

$$\begin{cases} x_c = z_c \left(\frac{u}{u_0} - 1 \right) \tan\left(\frac{\alpha}{2}\right) \\ y_c = z_c \left(\frac{v}{v_0} - 1 \right) \tan\left(\frac{\beta}{2}\right) \\ z_c = \text{Depth}(u, v) \end{cases} \quad (1)$$

where (u_0, v_0) is the 2D pixel coordinates of the center point of the imaging plane of the depth camera, as shown in Fig. 7, where $u_0 = 420$ and $v_0 = 240$. α and β are FOV angles of the depth camera in the horizontal and vertical directions, respectively; $\alpha = 71.5^\circ$ and $\beta = 55^\circ$. When the IR image is aligned with the depth image, $\text{Depth}(u, v)$ is the depth value of point (u, v) in the depth image.

Step 3. Preparing 3D point pairs

As shown in Fig. 6, one camera captures a set of 3D points $P_c = \{\overrightarrow{pc_i}\}$ under the camera coordinates. Let $P_w = \{\overrightarrow{pw_i}\}$ be the corresponding 3D point set under the

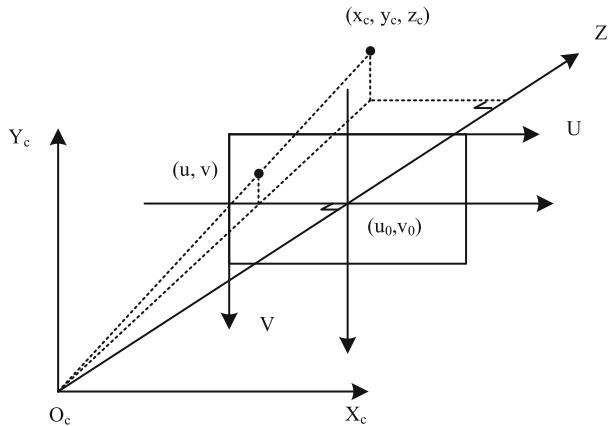


Fig. 7 The corresponding relation of camera coordinates and image coordinates

world coordinates. The number of 3D points in both a and b is n . P_c can be transformed to P_w via Eq. (2)

$$P_w = RP_c + T \quad (2)$$

where R and T are the rotation and translation matrices that need to be calculated.

Step 4. Calculating transformation matrices

As traditional 3D coordinate transformation models, such as Bursa-Wolf, are limited in their ability to solve big rotation angle transformations [10, 35], we use the unit quaternion method [4] to calculate matrices R and T . The unit quaternion is a vector \vec{q} that is given by Eq. (3).

$$\vec{q} = [q_0 \quad q_1 \quad q_2 \quad q_3]^T \quad (3)$$

where $q_0 \geq 0$ and $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. The mean value $\overrightarrow{m_c}$ of the point set P_c and the mean value $\overrightarrow{m_w}$ of the point set P_w are given by Eqs. (4) and (5).

$$\overrightarrow{m_c} = \frac{1}{n} \sum_{i=1}^n \overrightarrow{pc_i} \quad (4)$$

$$\overrightarrow{m_w} = \frac{1}{n} \sum_{i=1}^n \overrightarrow{pw_i} \quad (5)$$

The cross-covariance matrix of P_c and P_w is given by Eq. (6).

$$D_{cw} = \frac{1}{n} \sum_{i=1}^n \left[(\overrightarrow{pc_i} - \overrightarrow{m_c})(\overrightarrow{pw_i} - \overrightarrow{m_w})^T \right] \quad (6)$$

They are used to form a 4×4 matrix $Q(D_{cw})$, which is given by Eq. (7).

$$Q(D_{cw}) = \begin{bmatrix} tr(D_{cw}) & \Delta^T \\ \Delta & D_{cw} + D_{cw}^T - tr(D_{cw}^T)I \end{bmatrix} \quad (7)$$

where $\Delta = [A_{23} \ A_{31} \ A_{12}]$, $A_{ij} = (D_{cw} - D_{cw}^T)_{ij}$, $tr(D_{cw})$ is the trace of D_{cw} , and I is the 3×3 identity matrix. The unit quaternion $\vec{q} = [q_0 \ q_1 \ q_2 \ q_3]^T$ is the maximum eigenvalue of $Q(D_{cw})$. Then, the optimal rotation matrix R is given by Eq. (8).

$$R = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 + q_2^2 - q_1^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 - q_0 q_1) & q_0^2 + q_3^2 - q_1^2 - q_2^2 \end{bmatrix} \quad (8)$$

The optimal translation vector is given by Eq. (9).

$$T = \vec{m}_c - R \vec{m}_c \quad (9)$$

4 Scanning process

After calibration, our multi-camera 3D foot scanning system will be utilized to scan, reconstruct and measure the feet automatically. The proposed scanning process for this system is illustrated in Fig. 8. In the partial view scanning stage, four cameras capture the 3d point cloud of the target from different views. For the point cloud filtering process, the band-pass filtering algorithm and a new proposed filtering method for the reflection noise will be used to filter the original 3D point cloud. Then, in the process of the registration, we propose a two-step method of coarse and fine registration to align point clouds from different perspectives into a point cloud with uniform coordinates. In the non-visible area filling stage, we propose a method to fill the invisible areas at the top and bottom of the feet to obtain a watertight model. After filling the invisible area, we used the Poisson method to reconstruct the surface of the 3d point cloud and obtain a 3D mesh model of the feet. In the last stage, we will calculate the parameters of the feet scanned by the system by identifying the feature points.

4.1 Partial view scanning

Since the simultaneous operation of multiple IR laser projectors can cause interference problems, we developed a program that controls four cameras to turn on and off in sequence

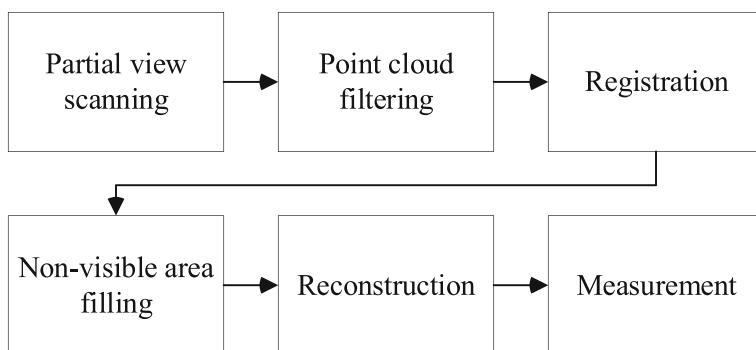


Fig. 8 The pipeline of the 3D foot scanning

to avoid multiple cameras working at the same time. Each SR300 camera scans the target on the center of the platform at a frame rate of 60 fps.

To obtain stable 3D point clouds, during the scanning process, each camera will continuously capture 60 frames of depth images and obtain an average depth image. Each frame depth image contains 640*480 pixels, and each pixel corresponds to a depth value. Before averaging the depth data of all frames in every pixel location, invalid pixels, which are pixels with values that are outside the valid range, will be excluded.

4.2 Point cloud filtering

Band-pass filter: We find that PCD_i , $i = 1\dots4$, that are obtained from partial views are affected by noise and outliers. To reduce the influence of noise on the point cloud registration and 3D foot reconstruction, we propose methods for filtering noise for various situations. In one case, the noise is mainly generated by the background because each camera will capture the opposite camera or bracket, as shown in Fig. 9. In addition, there are 307,200 (640×480) 3D points in one partial view, but many (more than a half) are invalid points, which have a depth value of 0. We use the band-pass filter $F(x, y, z)$ to deal with the noise and invalid 3D points in this case. It is expressed as Eq. (10).

$$F(x, y, z) = \begin{cases} 1, & z_0 \leq z \leq z_1 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

where (x, y, z) denotes one 3D point in PCD_i , and z_0 and z_1 denote the range of valid distance from the center of the depth sensor. In practice, $z_0 = 240$ mm and $z_1 = 750$ mm. Before registering the 3D point cloud from four perspectives to the unified world coordinate system, we do not filter points according to the range limits on the x and y axes.

Remove reflection noise: During the experiment, we found that the 3D point cloud that was captured by the camera has reflection noise under the contact surface between the foot and the platform, as shown in Fig. 10. The reflection noise and the point cloud of a real foot are closely connected, the area is large, and the shape is similar to that of the points of the real foot.

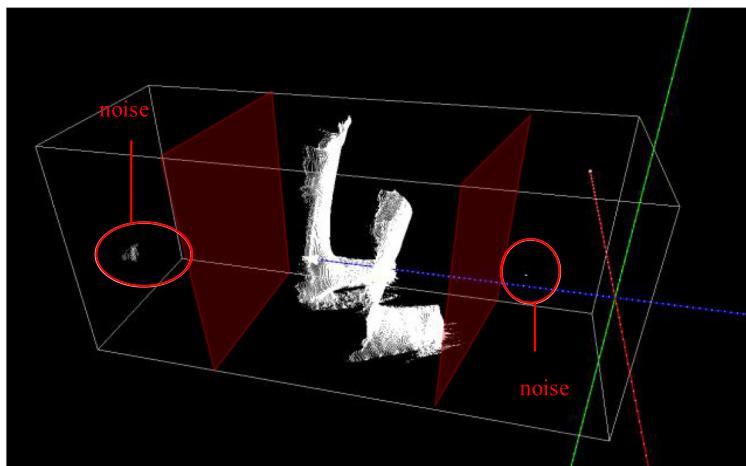


Fig. 9 Band-pass filter for a partial view 3D point cloud

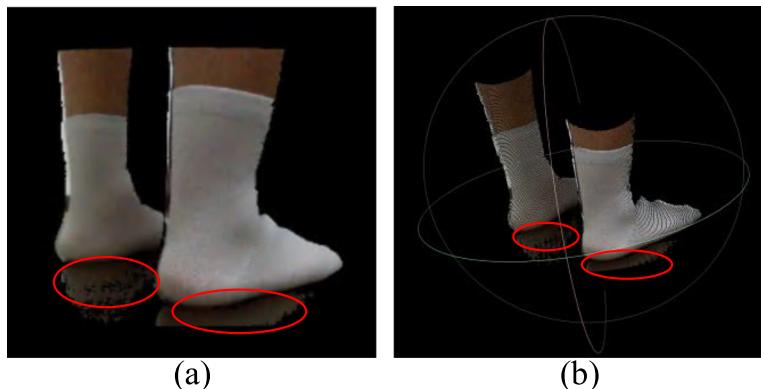


Fig. 10 Reflection noise: **a** View by 2D image and **b** view by 3D point cloud

Therefore, it is very difficult to remove the reflection points via the common point cloud filtering algorithm.

A new 3D point cloud filtering method that is based on gray image segmentation is proposed. a) We obtain the IR image (IR_i , $i = 1 \dots 4$) while obtaining the PCD_i from a partial view. Each pixel on IR_i corresponds to a 3D point in PCD_i . b) We calculate the threshold of the target and the background in IR_i using a threshold selection method [17] to minimize the intraclass variance of the black and white pixels and convert IR_i to the binary image BW_i based on the threshold. c) After using a mathematical morphological opening operation to eliminate the outliers of BW_i , we obtain a new binary image, which is denoted as BM_i . d) We use the binary image as an index map to filter out the points that correspond to pixels with a depth value of 0 and obtain the resulting 3D point cloud. The process of this method is shown in Fig. 11.

4.3 Registration

After preprocessing, the background and noise points of the point cloud that was captured by four cameras have been significantly reduced, as shown in Fig. 12(a). The point cloud registration method that we proposed includes two steps: coarse alignment and fine alignment. In the coarse alignment process, four sets of point clouds in camera coordinates XYZ_{ci} , $i = 1 \dots 4$, are transferred into the world coordinates XYZ_w using the rotation matrix R_i and the

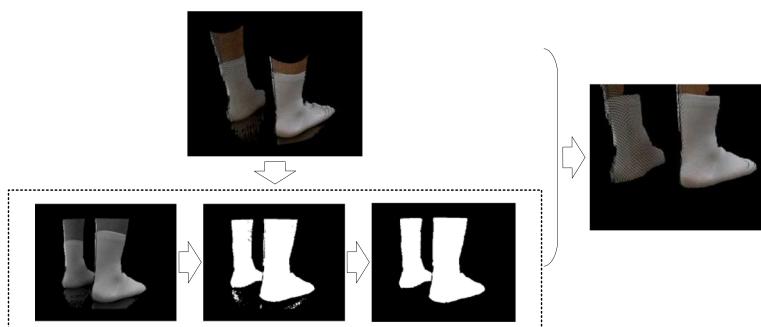


Fig. 11 Filtering process of the “inverted image” noise

translation matrix T_i , which are calculated by the calibration algorithm. Although the coarse alignment process has already aligned the point clouds that were captured by different cameras, it still needs to further resolve the gaps, overlaps and other problems through the fine alignment process. As shown in Fig. 12(b), we divide the transformed point cloud into eight pieces such that each foot consists of four pieces of point clouds from different cameras. The iterative closest point (ICP) algorithm is used to register each adjacent point cloud for each foot, and the moving least square algorithm is used to smooth the point cloud after registration. The point cloud registration result is shown in Fig. 12(c).

4.4 Non-visible area filling

Because there are non-visible areas of the scan, the registered foot point cloud has no data at the top or bottom. To obtain a watertight model for 3D reconstruction, we propose a point cloud hole filling method for closing the top and bottom surface holes. At the top surface of a foot, we fill a hole by constructing horizontal sections. First, we find the highest point, and its Y-axis is y_{\max} . Take all points in the 3-mm range from the highest point and project them onto the plane $y = y_{\max}$ to obtain a ring from the points, as shown in Fig. 13(a). The contour line of the point ring is obtained by using the convex packet algorithm, and the grid points are generated according to the outer rectangular scope of the contour line. The point of the intersecting range of the contour line and the grid, as shown in Fig. 13(b), is the point to be filled, and the filling result of the top surface is shown in Fig. 13(c).

To fully reconstruct the scanning details of the bottom contour, we fill the bottom hole by generating 3D curve lines. We select points (P_{bottom}) from the lowest point up to approximately 1 cm (no projection), as shown in Fig. 14(a). All points in P_{bottom} are scanned along the z-axis to obtain a cross-section of the vertical Z-axis at every 1 mm of distance. Then, we calculate the lowest point on both sides of the cross-section to validate the endpoints on both sides of the feet. According to the two endpoints, a parabola is generated on the cross section and a point is inserted at every 1 mm along the X-axis on the curve, as shown in Fig. 14(b).

4.5 Reconstruction

After filling the holes, we use the voxelized grid down sampling approach to reduce the number of points to reduce the computational cost of 3D reconstruction and make the newly added points and the original points more uniform. Poisson surface reconstruction was used to encapsulate the point cloud. In our implementation, we use two child threads to perform the

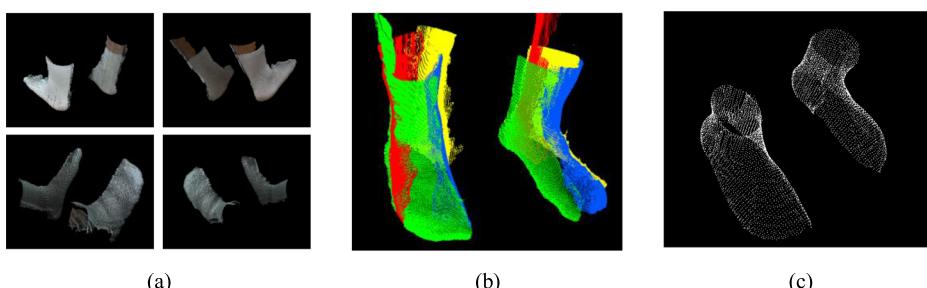


Fig. 12 Registration of the point cloud: **a** Point clouds after preprocessing and **b** point clouds after being transferred to the same coordinate system; **c** registration result

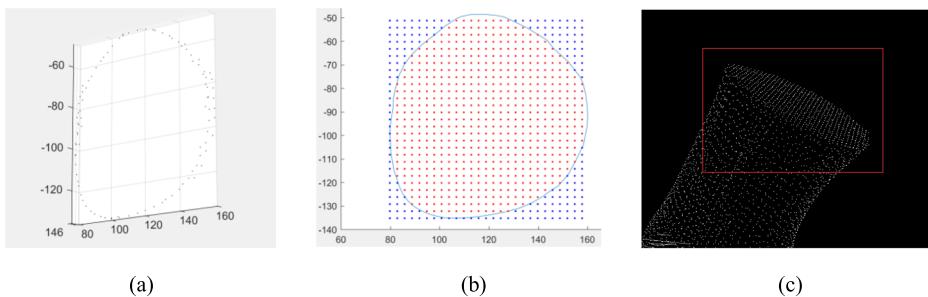


Fig. 13 Filling the hole in the top surface: **a** point projection; **b** contour and grid points; and **c** filling result

Poisson reconstruction algorithm for point clouds of the feet. In the main thread, we merge the results of the two child threads and generate a complete 3D mesh model that contains both feet. In practice, the maximum reconstruction depth was set to eight and the number of solver iterations was set to 4 to preserve the details of scans.

4.6 Foot measurement

According to the requirements of shoe making and previous studies [28, 30, 31], the scanning system in this paper selected foot length(L), foot width(W), ball girth(C_1), instep girth (C_2), short heel girth (C_3), the height of metatarsus(H_1), the height of tarsus(H_2) and the height of talus(H_3) as the feature parameters to be measured, as shown in Fig. 15(a). As shown in Fig. 15(b), P_1, P_2, \dots, P_7 represent several main feature points that need to be extracted related to the measurement.

As the 3D point cloud of the scanning result contains the data of two feet, the data should be divided into the left foot and the right foot according to the positive and negative X-axis coordinates before extracting the foot parameters. The basic steps of the measurement method are introduced as follows.

Step 1. Project the point whose Y-axis coordinate value is less than 20 mm in the foot point cloud onto the plane XOZ to get the outline points of the foot (F_{XZ}).

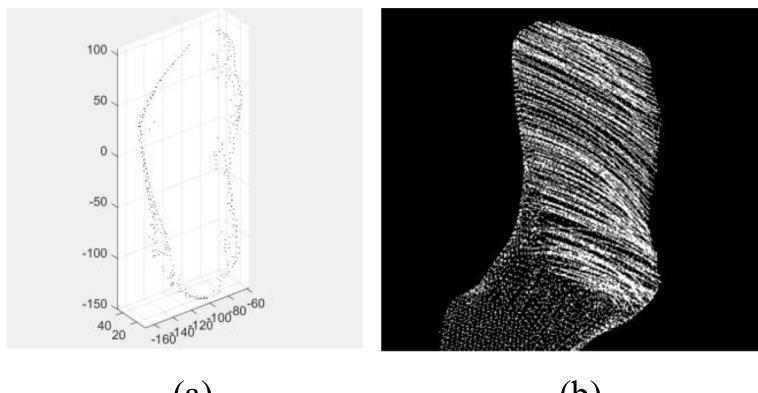


Fig. 14 Filling the hole in the bottom surface: **a** Points of the bottom contour and **b** filling result of one foot

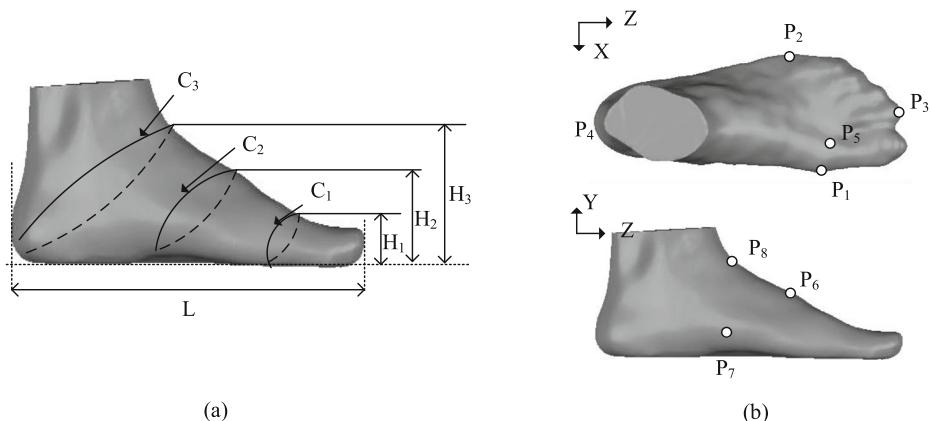


Fig. 15 Positions of measurement: **a** Feature parameters; **b** Landmarks

- Step 2. By solving the minimum enclosing rectangle for the contour points, we obtain four feature points: P_1 is the most medially prominent point on the first metatarsal head, P_2 is the most laterally prominent point on the fifth metatarsal head, P_3 is the front end of the second metatarsal and P_4 is the back heel point.
- Step 3. We use Frances's method [28] to obtain the base axis(F-axis) of foot measurement. In outline points F_{XZ} , this base axis is defined by the point P_3 and P_4 .
- Step 4. Calculate the angle (α) between axis and system Z axis on plane XOZ. Then, we adjusted the position and posture of the 3D point cloud of the foot according to α , so that the F-axis was parallel to Z-axis. Now, the Z-coordinate value of P_3 is z_{\max} , the Z-coordinate value of P_4 is z_{\min} , the X-coordinate value of from P_1 is x_1 , and the X-coordinate value of from P_2 is x_2 , so the foot length is given by Eq. (11) and the foot width is given by Eq. (12).

$$L = |z_{\max} - z_{\min}| \quad (11)$$

$$W = |x_2 - x_1| \quad (12)$$

- Step 5. According to Wu's method [31], we extracted the feature points P_5 , P_6 , P_7 and P_8 . The point $P_5(x_5, y_5, z_5)$ is top of the distal inter-phalangeal joint of the second toe, $P_6(x_6, y_6, z_6)$ is the instep point and $P_7(x_7, y_7, z_7)$ is the highest point in the arch region. Therefore, the Y-coordinate values of P_5 , P_6 and P_7 respectively represent the value of H_1 , H_2 and H_3 .
- Step 6. The plane defined by points P_1 , P_2 and P_5 intersects the foot contour to form an outline line whose length is $C1$.
- Step 7. Define a plane equation with two fixed points (P_6, P_7) and one variable point P_x . P_x moves along the outside of the bottom outline of the foot. Instep girth($C2$) is smallest girth formed by the intersection of the plane and the foot contour.

- Step 8. Define a plane that passes through the point P_4 and P_8 and is perpendicular to the YOZ plane. C3 is the length of the outline line formed by the intersection of the plane and the foot contour.

5 Experiment and results

All procedures of our methods were tested on a laptop computer with a Core i5 2.3 GHz CPU and 8 GB RAM. The four RealSense SR300 cameras were connected to a USB3.0 HUB via the USB3.0 cable, and the USB3.0 HUB was connected to the laptop computer. The hardware system environment is illustrated in Fig. 16.

5.1 Calibration evaluation

Compared with the checkerboard-based calibration method [28, 31, 32], the method in this paper has the following characteristics: (1) No manual intervention required: After placing the calibration block at the predetermined position of the scanning platform, the calibration program can automatically complete the calculation process of each camera without manual intervention. (2) Various planes: Feature points captured by camera would lie on various planes, which ensured system accuracy across a larger depth of field after calibration. (3) The calculation is simple and fast: Feature points are simple to identify, and can be directly identified in infrared images rather than in RGB images. (4) Large rotation angle: Using the unit quaternion method, it is more suitable for the calibration of large rotation Angle.

To evaluate the performance of our calibration method, which is based on a tower-type block, we capture the feature points on the calibration tower and compare the actual measured coordinates with the coordinate transformation results. As shown in Fig. 17(a), the red dots represent the actual measured coordinates of the feature points in the world coordinate system and the green squares represent the transformation results of the feature points that were captured by cameras. As shown in Fig. 17(b), the distances of 20 feature points are between 0.28 and 1.67 mm and the mean distance is 0.91 mm. The mean distance is used as the error evaluation value. According to repeated tests, the average error is 0.95 mm and the standard deviation is 0.019 mm.

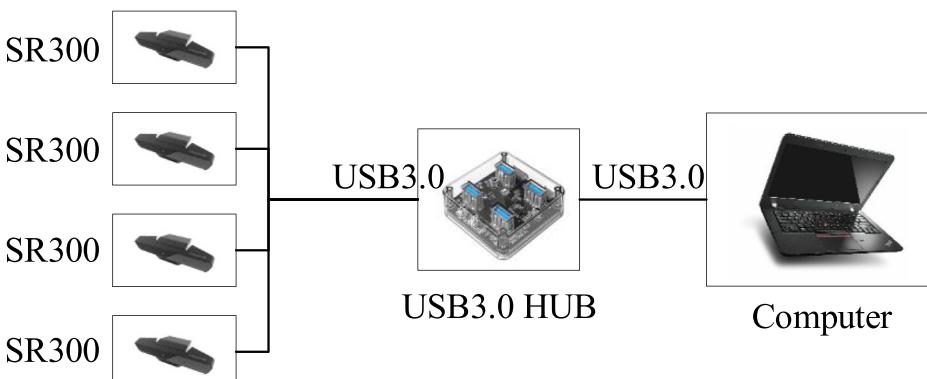


Fig. 16 Hardware system environment

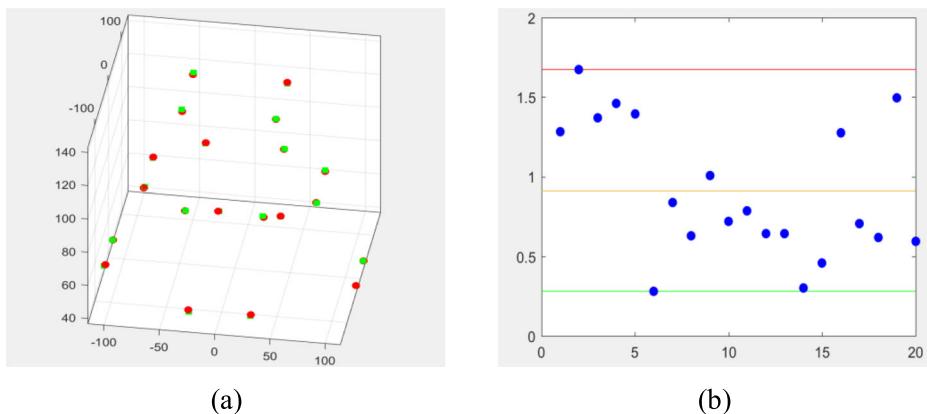


Fig. 17 Calibration experiment: **a** Coordinates of actual locations and transformation results and **b** distances of feature points

5.2 Scanning experiment

The scanning experiment is shown in Fig. 18(a) in which the subject stands with his or her feet in the center of the platform and the program controls the cameras to start capturing images. Each of the four cameras captured the 3D point cloud data from a single view, and a total of four sets of original 3D point cloud data were obtained, as shown in Fig. 18(b). The original point cloud data contain many noise points, which must be processed with bandpass filtering and removal of reflection noise points. The data after processing are shown in Fig. 18(c). According to the transformation matrix that was obtained by the calibration method, the four sets of processed 3D point cloud are transformed from the coordinates of the four cameras to a

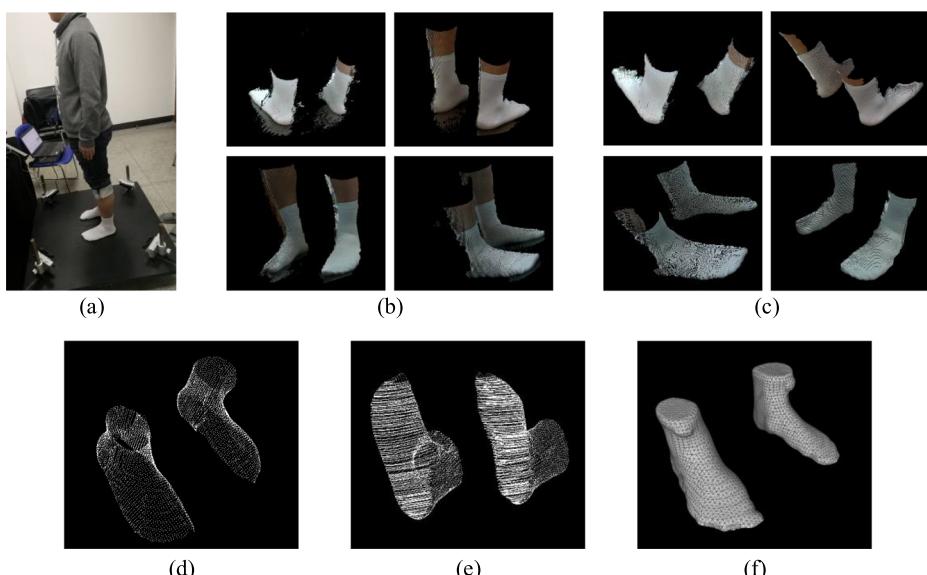


Fig. 18 Scanning results: **a** Scanning experiment; **b** original 3D point cloud data; **c** filtering result; **d** aligned 3D point cloud; **e** 3D point cloud after the holes have been filled; and **f** 3D reconstruction result

Table 1 Results of manual and automatic parameter measurements

Dimensions	Manual (mm)	Automatic (mm)			SD.
		Max.	Min.	Mean	
Foot length	259.3	260.1	259.1	259.7	0.37
foot width	100.5	101.1	100.0	100.6	0.29
ball girth	244.6	245.0	243.2	244.2	0.67
instep girth	242.5	243.0	241.2	242.1	0.66
short heel girth	338.0	340.0	338.0	339.1	0.76
height of metatarsus	33.1	34.9	33.2	34.0	0.62
height of tarsus	64.5	65.8	64.1	64.9	0.58
height of talus	87.5	87.9	86.1	87.1	0.58

unified world coordinate system. Then, the ICP algorithm is used to further align the 3D point cloud of each part to obtain a complete 3D point cloud model, as shown in Fig. 18(d). Next, the holes are filled, and the closed 3D point cloud model is shown in Fig. 18(e). Finally, using the Poisson distribution algorithm, the 3D reconstruction result of the feet is obtained, as is shown in Fig. 18(f).

5.3 Foot measurement result

In this experiment, we invited 20 participants (10 men and 10 women between the ages of 10 and 45) to test our scanning system. The system successfully completed the scanning work and obtained the measurement results of each participant. This shows that our system is suitable for different genders and covers a wide range of ages.

On the other hand, we randomly selected a participant to manually measure the foot 3 times, and used the average value of the measurement results as the comparison standard. Then, we use the scanning system proposed in this article to automatically scan and measure the participant's feet 10 times. According to the results shown in Table 1, the automatic measurement results are close to the manual results.

5.4 Performance evaluation

Based on repeated experiments, the average processing time that was required for each phase of our approach is shown in Table 2. The calibration method is time-consuming and requires approximately 8 s; however, it is a separate process and only needs to be performed once, during the initial setting of the system. The cameras take 4 s to capture the 3D point cloud data from all partial views. The data processing, including filtering, smoothing, and point cloud alignment, takes approximately 5 s. Finally, the reconstruction takes approximately 5 s to

Table 2 Average time cost of each phase of our approach

Process	Average time (s)
Calibration	8
Data capture	4
Data processing	5
Reconstruction & measurement	5

generate the 3D model and calculate the foot parameters. Without the calibration, our approach takes approximately 14 s to scan the feet and generate a 3D model. Although the entire scanning process is time-consuming, it does not make people feel uncomfortable, because people only need to stand on the scanning platform for about 4 s to leave, and the program can automatically perform subsequent steps after collecting the original data.

6 Conclusion and future work

In this paper, a novel and low-cost 3D foot scanning system was proposed. The complete procedure of using multiple RealSense cameras to capture the point cloud and reconstruct the 3D foot model is explained in detail. There are three main innovative approaches: (1) We presented a novel tower-type block-based calibration method that can automatically and accurately complete the calibration of multiple depth cameras. (2) A novel 3D point cloud filtering method that is based on gray image segmentation was proposed to solve the “inverted image” noise point problem. (3) To obtain a watertight model, we propose a point cloud hole filling method for closing the holes in the top and bottom surfaces of the non-visible areas. It can be seen from the calibration evaluation results that the calibration method in this paper has small errors and can quickly and automatically complete the calibration for multiple cameras of the system. In addition, the results of foot measurements show that the method proposed in this paper is accurate and stable. According to the scanning experimental results, with an ordinary consumer notebook computer, the system requires less than 8 s to capture a full 3D point cloud of a pair of feet and the complete reconstruction process takes no more than 40 s. It demonstrates that this system, which uses multiple low-cost depth cameras, can achieve stable, accurate and fast scanning of both feet under an open environment.

The approach proposed in this paper still suffers from several limitations: (1) In order to avoid infrared interference between multiple cameras during the scanning process, the system in this paper controls the cameras to start working one by one, which will increase the overall scanning time. (2) In this paper, the registration process mainly considers the rigid transformation of the point cloud, but in the actual scanning process, there is slight sway of the human body, which will cause the loss of depth data and increase registration error. In our future work, we intend to solve the infrared interference problem of multiple cameras by studying the depth image filtering algorithm so that multiple depth cameras can capture the 3D point cloud in parallel. Moreover, we plan to study the non-rigid shape characteristics of 3D models with time-varying shapes. Finally, we hope to study the real-time 3D foot reconstruction based on static foot scans, to analyze the gait characteristics of walkers in the 3D shape.

References

1. Alexiadis D, Zarpalas D (2015) Daras P fast and smooth 3D reconstruction using multiple RGB-depth sensors. Visual Communications and Image Processing Conference, In, pp 173–176
2. Alexiadis DS, Zarpalas D, Daras P (2013) Real-time, full 3-D reconstruction of moving foreground objects from multiple consumer depth cameras. Ieee T Multimedia 15(2):339–358. <https://doi.org/10.1109/Tmm.2012.2229264>
3. Atherton TJ, Kerbyson DJ (1999) Size invariant circle detection. Image Vis Comput 17(11):795–803. [https://doi.org/10.1016/S0262-8856\(98\)00160-7](https://doi.org/10.1016/S0262-8856(98)00160-7)

4. Besl PJ, Mckay ND (2002) A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Mach Intell* 14(2):239–256
5. Carfagni M, Furferi R, Governi L, Servi M, Uccheddu F, Volpe Y (2017) On the performance of the Intel SR300 depth camera: metrological and critical characterization. *IEEE Sensors J* 17(14):4508–4519. <https://doi.org/10.1109/jsen.2017.2703829>
6. Cui Y, Stricker D (2011) 3D body scanning with one kinect. In: International Conference on 3d Body Scanning Technologies, Lugano, Switzerland, 25–26 October, 2011. pp 121–129
7. Dou M, Taylor J, Fuchs H, Fitzgibbon A, Izadi S (2015) 3D scanning deformable objects with a single RGBD sensor. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p 493–501. <https://doi.org/10.1109/CVPR.2015.7298647>
8. Gao F, Wang Q, Geng W, Pan Y (2011) Acquisition of time-varying 3D foot shapes from video. *SCIENCE CHINA Inf Sci* 54(1):2256–2268. <https://doi.org/10.1007/s11432-011-4361-1>
9. Lee Y-C, Lin G, Wang M-JJ (2014) Comparing 3D foot scanning with conventional measurement methods. *Journal of Foot and Ankle Research* 7(1):44. <https://doi.org/10.1186/s13047-014-0044-7>
10. Liu C, Liu C (2017) Nonlinear gauss-Helmlert model and its solution for 3D coordinate transformation. *Science of Surveying and Mapping* 42(5):118–123
11. Liu ZB, Qin HL, Bu SH, Yan M, Huang JX, Tang XJ, Han JW (2015) 3D real human reconstruction via multiple low-cost depth cameras. *Signal Process* 112(C):162–179. <https://doi.org/10.1016/j.sigpro.2014.10.021>
12. Ma X, Luximon A (2014) 3D foot prediction method for low cost scanning. *Int J Ind Ergon* 44(6):866–873. <https://doi.org/10.1016/j.ergon.2014.08.006>
13. Maimone A, Fuchs H (2011) Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, Basel, vol 1. pp 137–146. <https://doi.org/10.1109/ISMAR.2011.6092379>
14. Maruyama M, Tabata S, Watanabe Y, Ishikawa M (2018) Multi-pattern embedded phase shifting using a high-speed projector for fast and accurate dynamic 3D measurement. In: 2018 IEEE winter conference on applications of computer vision (WACV), 12–15 March 2018. pp 921–929. <https://doi.org/10.1109/WACV.2018.00106>
15. Molfini R, Zoppi M (2005) Mass customized shoe production: a highly reconfigurable robotic device for handling limp material. *IEEE Robot Autom Mag* 12(2):66–76. <https://doi.org/10.1109/MRA.2005.1458327>
16. Novak B, Babnik A, Mozina J, Jezersek M (2014) Three-dimensional foot scanning system with a rotational laser-based measuring head. *Stroj Vestn-J Mech E* 60(11):685–693. <https://doi.org/10.5545/sv-jme.2014.1950>
17. Otsu N (2007) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9(1):62–66
18. Pesce M, Galantucci LM, Percoco G, Lavecchia F (2015) A low-cost multi camera 3D scanning system for quality measurement of non-static subjects. *Proc Cirp* 28:88–93. <https://doi.org/10.1016/j.procir.2015.04.015>
19. Pioaru I (2017) Visualizing virtual reality imagery through digital holography. In: 2017 International Conference on Cyberworlds (CW), 20–22 Sept. 2017. pp 241–244. <https://doi.org/10.1109/CW.2017.50>
20. Psota P, Lédl V, Kaván F, Mokré P, Matoušek O (2017) Surface profilometry by digital holography. In: 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 12–15 Sept. 2017. pp 1–5. <https://doi.org/10.1109/ETFA.2017.8247748>
21. Rout N, Zhang YF, Khandual A, Luximon A (2010) 3D foot scan to custom shoe last. Special Issue of IJCCT 2010(1):5
22. Seo SG, Lee DY, Kim J-B, Kim SH, Park HS, Yoo HJ, Kim SJ, Kim J, Lee KM, Chung CY, Choi IH (2014) Gender differences in segmental foot motions during gait using 3D multi-segment foot model. *J Foot Ankle Res* 7(1):A75. <https://doi.org/10.1186/1757-1146-7-S1-A75>
23. Shi Z, Yu L, A AE-LA, Niu X (2012) Skeleton modulated topological perception map for rapid viewpoint selection. *IEICE Trans Inf Syst* 95(10):2585–2588. <https://doi.org/10.1587/transinf.E95.D.2585>
24. Shi Z, Yu L, El-Latif AAA, Le D, Niu X (2012) A kinematics significance based skeleton map for rapid viewpoint selection. *Res J Appl Sci Eng Tech* 4(17):2887–2892
25. Tong J, Zhou J, Liu L, Pan Z, Yan H (2012) Scanning 3D full human bodies using Kinects. *IEEE Trans Vis Comput Graph* 18(4):643–650
26. Van den Herrewegen I, Cuppens K, Broeckx M, De Raeve E, Vertommen H, Mertens M, Peeraer L (2012) Foot inter-segment angles and rotation axes based on dynamic 3D surface point clouds. *J Biomech* 45:S197
27. Van den Herrewegen I, Cuppens K, Broeckx M, Barisch-Fritz B, Vander Sloten J, Leardini A, Peeraer L (2014) Dynamic 3D scanning as a markerless method to calculate multi-segment foot kinematics during stance phase: methodology and first application. *J Biomech* 47(11):2531–2539. <https://doi.org/10.1016/j.jbiomech.2014.06.010>

28. Wan FKW, Yick K-L, Yu WWM (2017) Validation of a 3D foot scanning system for evaluation of forefoot shape with elevated heels. Measurement 99:134–144. <https://doi.org/10.1016/j.measurement.2016.12.005>
29. Weiss A, Hirshberg D, Black MJ (2013) Home 3D body scans from a single Kinect. In: Fossati A, Gall J, Grabner H, Ren X, Konolige K (eds) Consumer depth cameras for computer vision: research topics and applications. Springer London, London, pp 99–117. https://doi.org/10.1007/978-1-4471-4640-7_6
30. Witana CP, Xiong S, Zhao J, Goonetilleke RS (2006) Foot measurements from three-dimensional scans: a comparison and evaluation of different methods. Int J Ind Ergon 36(9):789–807. <https://doi.org/10.1016/j.ergon.2006.06.004>
31. Wu G, Li D, Hu P, Zhong Y, Pan N (2016) Automatic foot scanning and measurement based on multiple RGB-depth cameras. Text Res J 88(2):1–15
32. Wu G, Li D, Zhong YQ, Hu PP (2017) A study on improving the calibration of body scanner built on multiple RGB-depth cameras. Int J Cloth Sci Tech 29(3):314–329. <https://doi.org/10.1108/Ijbst-05-2016-0051>
33. Xiong SP, Zhao JH, Jiang ZH, Dong M (2010) A computer-aided design system for foot-feature-based shoe last customization. Int J Adv Manuf Tech 46(1–4):11–19. <https://doi.org/10.1007/s00170-009-2087-7>
34. Xiong S, Li Y, Zhu Y, Qian J, Dong Y (2010) Foot measurements from 2D digital images. In: 2010 IEEE International Conference on Industrial Engineering and Engineering Management, 7–10 Dec. 2010. pp 497–501. <https://doi.org/10.1109/IEEM.2010.5674498>
35. Závoti J, Kalmár J (2016) A comparison of different solutions of the Bursa–wolf model and of the 3D, 7-parameter datum transformation. Acta Geodaetica Et Geophysica 51(2):245–256
36. Zhang C, Zhang Z (2011) Calibration between depth and color sensors for commodity depth cameras. In: 2011 IEEE International Conference on Multimedia and Expo, 11–15 July 2011. pp 1–6. <https://doi.org/10.1109/ICME.2011.6012191>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Munan Yuan received an M.S. degree in software engineering from Henan University, Kaifeng, China, in 2015. He is currently pursuing a Ph.D. degree in computer application technology through research from the University of Science and Technology of China, Hefei, China. His current research interests include computer software technology, 3D scanning and image processing.



Xiaofeng Li received a B.S. degree from Tianjin University, Tianjin, China, in 1987. He is currently working as a professor and doctoral supervisor at Hefei Institutes of Physical Science, Chinese Academy of Sciences. His research interests include Network Management and Auto-Control.



Jinlin Xu received an M.S. degree in circuits and systems from China University of Mining and Technology, Xuzhou, China, in 2010. He is currently pursuing a Ph.D. degree in computer application technology through research from the University of Science and Technology of China, Hefei, China. His current research interests include sensors, computer vision and bioelectrical impedance techniques.



Chaochuan Jia received an M.S. degree in Communication and Information Systems from Shandong University of Science and Technology, Qingdao, China, in 2015. He is currently pursuing a Ph.D. degree in Mechanical and Electrical Engineering through research from Shandong University of Science and Technology. His current research interests include mechanical electronics, 3D scanning and point cloud data processing.



Xiru Li received a B.S. degree in computer science and technology from Henan University, Kaifeng, China, in 2008, and a Ph.D. degree in computer application technology from the University of Science and Technology of China, Hefei, China, in 2017. She is currently with the Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei. Her current research interests include wearable devices and human physiological parameter monitoring techniques.