IOP Publishing Measurement Science and Technology
Meas. Sci. Technol. **31** (2020) 045901 (13pp) https://doi.org/10.1088/1361-6501/ab48b3

An extrinsic calibration method for multiple RGB-D cameras in a limited field of view

Jia Chaochuan[1], Yang Ting[2], Wang Chuanjiang[1], Fan Binghui[1] and He Fugui[2]

[1] College of Mechanical & Electronic Engineering, Shandong University of Science and Technology, Qingdao, People's Republic of China
[2] College of Electronics and Information Engineering, West Anhui University, Lu'an, People's Republic of China

E-mail: fanbh58@163.com and ccjia@hfcas.ac.cn

Received 24 June 2019, revised 25 September 2019
Accepted for publication 27 September 2019
Published 6 January 2020

Abstract

A 3D scanning system involving multiple RGB-D cameras has the potential to accelerate the reconstruction of an object and improve the measurement accuracy because it can capture an object in a comprehensive way. However, the extrinsic calibration of a multi-RGB-D-camera system is a fundamental and challenging problem, especially in a limited field of view. In this work, a system with four SR300 cameras on a platform with a size of approximately 0.64 m^2 and an extrinsic calibration method assisted by a tower calibration pattern with circular markers in a limited field of view are proposed. The Hough transform algorithm is used to identify the centres of the circular markers in a colour image, and then the 3D coordinates are extracted by employing the alignment relationship between the colour image and the depth image. An improved adaptive cuckoo search (IACS) algorithm and unit quaternion are proposed to optimize the extrinsic parameters. The effectiveness of the IACS is verified by a comparison with the singular value decomposition (SVD), standard cuckoo search (CS) and two CS variant algorithms. In addition, the accuracy of the proposed extrinsic calibration method is verified by reconstructing and measuring a human foot and a cube box. The mean errors from the standard values of the length and width of the reconstructed foot mode are 0.18 cm and 0.13 cm, and the diameter of the yellow sphere are 0.17 cm, respectively. All of the experimental results show that the proposed extrinsic calibration method has a high calibration accuracy and good potential application prospects.

Keywords: multi RGB-D cameras, tower calibration pattern, extrinsic calibration, improved adaptive cuckoo search algorithm, optimization

(Some figures may appear in colour only in the online journal)

1. Introduction

With the development of perception and optical technology, non-contact visual sensors such as the RGB-D camera, which can capture both a colour image and a depth image of the scene, have the advantages of low cost, small size, good stability and high precision and are widely used in many visual applications such as 3D reconstruction [1, 2], measurement [3, 4], recognition [5], augmented reality (AR) [6] and virtual reality (VR) [7]. In these applications, there are two alternative solutions for capturing the entire contour of the target object. One alternative is a single RGB-D camera scanning system that has a projection blind spot problem due to an occlusion, a limited field of view (FOV) or other reasons, resulting in the entire contour of the object not being captured by a single scan. However, the entire contour can be captured by rotating the camera several times and then matching the point cloud data after several scans. Although many point cloud registration methods have been proposed by researchers around the world and remarkable results have been achieved, there are still many limitations in a single-sensor system, such as the time-consuming, cumbersome operation and determination

of the camera position. Compared with the first scheme, the multi-RGB-D camera scanning system is very appealing. Nevertheless, to accurately collect all of the 3D data points of an object, this type of system must be calibrated. This calibration process is known as an extrinsic calibration, which is used to estimate the relative pose between the RGB-D cameras. The precision of the scanning system to some extent depends on the precision of the extrinsic calibration.

A large number of extrinsic calibration methods for multiple cameras have been proposed by researchers, and specific methods will be described in detail in the related work section.

The explicit requirements for the extrinsic calibration methods depend on the specific application. Generally, there are four main requirements for a calibration method [8–10]. The *accuracy* is the first and most important requirement. An extrinsic calibration method should always achieve the optimal relative pose estimation between the cameras. The second requirement is *repeatability*; since solving for the extrinsic parameters is an optimization problem, it must converge to the global optimal solution rather than a local suboptimal solution. This means that the deviation between the results of multiple calibrations must be small. Third, *simplicity* of operation is required. As the camera is replaced or moved, the calibration method needs to be used frequently. Finally, the calibration method should be *robust* to noise and outliers. These disturbances may be caused by environmental influences, sensor noise of the camera, or other factors. Thus, the development of an extrinsic calibration method to meet these requirements is an ongoing problem that needs to be solved.

The main contribution of this paper is a compact 3D scanning system consisting of multiple RGB-D cameras and an extrinsic calibration method that is assisted by a tower calibration pattern with circular markers in a limited FOV. The rest of the paper is organized as follows: following the introduction, the related works are reviewed in section 2. Section 3 elaborates the system setup of the RGB-D camera array and an overview of the proposed calibration method. Detailed methods are described in section 4. Experimental studies on the extrinsic calibration and foot scanning results utilizing the proposed prototype are presented and analysed in section 5. The conclusions are presented in section 6.

2. Related works

In general, the extrinsic calibration methods of multiple cameras can be divided into methods that use calibration objects and methods that do not use calibration objects. In recent decades, extrinsic calibration methods that do not use special calibration objects have been proposed. In [11], Miyata *et al* proposed a novel extrinsic calibration algorithm for multiple cameras in a broad area. The correspondences based on scene features between the two cameras were determined by the position of an omnidirectional camera, which acted as a reference point. Perez-Yus *et al* [12] and Ly *et al* [13] first extracted and matched the lines of the scene, and then geometric constraints were applied to estimate the relative pose between RGB-D cameras. Yang *et al* [14] proposed an extrinsic calibration method based on scene constraints to estimate the relative pose between a depth camera and an RGB camera. In [15], some large planes in the scene, such as the ceiling or floor, were used to calibrate the extrinsic parameters of multiple cameras. In [16], the contours of the human body were used to calibrate the extrinsic parameters of multiple cameras. Li *et al* [17] proposed an extrinsic calibration method based on a human's joint positions. However, the calibration precisions of these methods are low due to their strong dependence on the scene and inaccurate understanding of the unknown scene features, which will lead to many incorrect correspondences. Moreover, these methods are not suitable for multiple cameras with a limited FOV.

Meanwhile, methods that use special calibration objects have also been studied. For an extrinsic calibration of an RGB-D camera, the commonly used calibration objects are mainly 2D and 3D and include planar checkerboards, planar mirrors, circular patterns, spherical and other customized objects. For example, Liu *et al* [18] proposed a novel feature-plane-based method for estimating the relative poses between multiple RGB-D cameras with minimal overlapping fields. The authors matched 2D feature points with the corresponding depth values to construct the matched 3D feature points, which were used to calibrate the extrinsic parameters of multiple RGB-D cameras. In [19], the authors used a 2D planar calibration object with holes to calibrate the extrinsic parameters of multiple three-dimensional time-of-flight (3D-TOF) cameras. The four corners of a planar calibration object were used in [20], and the corners of a planar checkerboard in [21] were used to calibrate the extrinsic parameters of multiple RGB-D cameras. In [22, 23], the authors also used planarity constraints to find the correspondences between depth images. In [24], Shim *et al* directly looked for 3D correspondences instead of 2D feature correspondences to calibrate the extrinsic parameters of multiple RGB-D cameras. However, these extrinsic parameter calibration methods based on planar objects are usually time-consuming, especially when there are a large number of cameras. On the other hand, when the angle between the optical axes of the adjacent cameras is large, the planar image captured by the RGB-D camera may be distorted, thus resulting in many incorrect correspondences. In [25], a moving stick with one fixed end was used to calibrate an RGB-D camera. Two fixed objects connected by metal bars were used by Xia *et al* [9] to calibrate multiple cameras. The relative pose between the two objects, which was determined by a large-FOV camera, was used to estimate the extrinsic parameters of multiple cameras. This method is only suitable for a stereo camera calibration. In [26], the authors used a planar mirror to calibrate the extrinsic parameters of cameras. A virtual camera was built by the reflection of the planar mirror, and the relative pose between the virtual camera and the real camera could be estimated by markers positioned on the planar mirror. Kumar *et al* [27] also used a planar mirror to calibrate the extrinsic parameters of multiple cameras. Compared with the method in [26], the advantage of this method is that there is no need to position markers on the planar mirror. However, the reflection range of the planar

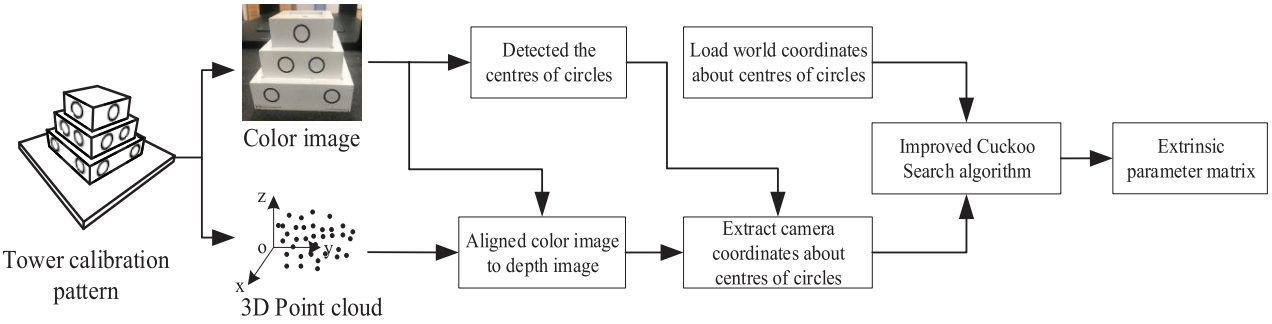

Figure 1. Flow chart of the proposed calibration method.

mirror is difficult to control, especially when the number of cameras is large and in a limited FOV.

In [28, 29], the authors used a 3D reconstruction method to estimate the extrinsic parameters of multiple cameras. First, the calibration plane with some coding markers was reconstructed by a 3D scanner system. Then, different cameras captured different parts of the calibration plane, and multiple cameras were connected by coding markers. Finally, the extrinsic parameters of the multiple cameras were calculated. However, these methods need to reconstruct the calibration plane in advance. Therefore, the accuracy of the calibration has a strong dependence on the accuracy of the reconstruction; if the calibration plane is very large, it is difficult to guarantee the reconstruction accuracy.

Shen *et al* [30] and Su *et al* [31] used a coloured sphere as a calibration object, and the correspondences of the different views were identified by fitting the coloured sphere. In [29], a spherical object was also used as a calibration object by Ruan *et al* and the position of the sphere centre, which was used to calculate the extrinsic matrix of multiple RGB-D cameras, was estimated by fitting the spherical object. In [32, 33], it was shown that the calibration results with spherical objects are better than those with the planar checkerboards in [23]. Nonetheless, these methods are also usually time-consuming due to the need to fit the spherical object. Furthermore, the cameras of the system require strict time synchronization. There are other extrinsic parameter calibration methods that employ auxiliary equipment. For instance, Meng *et al* [34] used an inertial measurement unit (IMU) as an auxiliary sensor to calibrate multiple RGB-D cameras, which were used for simultaneous localization and mapping (SLAM). In [35], images captured by mobile devices and the GPS positions of mobile devices were used to calibrate extrinsic parameters of a camera network.

The extrinsic calibration of multiple RGB-D cameras consists of two steps: determining the correspondences and calculating the extrinsic parameter transformation matrix. The calibration objects are only used to provide the correspondences, and the precision of the extrinsic calibration depends on the method used to calculate the extrinsic parameter transformation matrix. The problem of solving for the extrinsic parameters of multiple cameras is actually a nonlinear optimization problem. However, many of the above-mentioned methods approximate this nonlinear problem as a linear problem that can be solved. Commonly used methods include the least squares method (LSM) and singular value

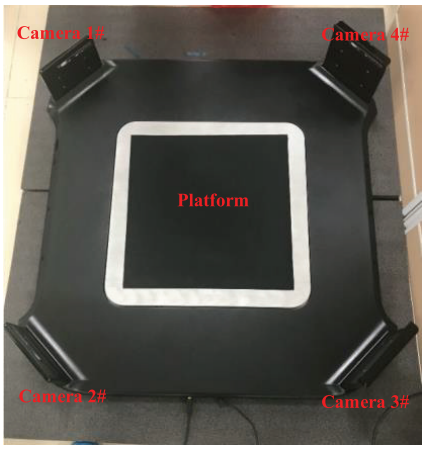

Figure 2. 3D Scanning system prototype with four SR300 cameras.

decomposition (SVD) [36]. Employing linear methods to solve a nonlinear problem will inevitably reduce the accuracy. It is known that meta-heuristic optimization algorithms are very good at solving nonlinear optimization problems. For example, Safaei *et al* [37] used metaheuristic optimization algorithms such as the genetic algorithm (GA), the colonial competitive algorithm (CCA), particle swarm optimization (PSO) and the shuffled frog leaping algorithm (SFLA) to estimate the extrinsic and intrinsic parameters and lens distortions of RGB-D cameras. In this paper, a tower with circular markers is used as the calibration object, and the improved adaptive cuckoo search (IACS) [38] algorithm is used to optimize the extrinsic transformation matrix.

3. Methods overview

A flow chart of the proposed extrinsic parameter calibration method for multiple RGB-D cameras is shown in figure 1. First, the tower calibration pattern with 5 circular marks is placed in the centre of the platform of the 3D scanning system, which is shown in figure 2, so that each face of the tower calibration pattern faces an RGB-D camera. The colour images, depth images, and 3D point clouds are captured by each RGB-D camera, and then the mapped image is generated by aligning the colour images with the depth images.

Meanwhile, the image coordinates of the circle centre on the tower calibration pattern are detected by applying the Hough transform algorithm to the colour images. The 3D camera coordinates of the circle centre are extracted according to the

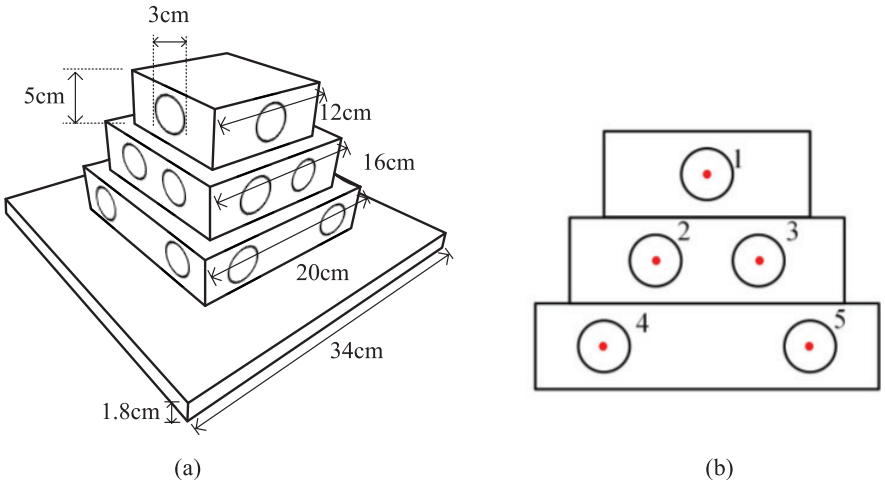

Figure 3. Schematic diagram of the tower calibration pattern. (a) Stereogram and (b) ichnography.

detected image coordinates of the circle centre and mapped image. Next, the world coordinates of the circle centre, which are known, are loaded from the file.

Afterwards, a unit quaternion and IACS algorithm are applied to optimize the extrinsic parameter matrix of each RGB-D camera by combining the 3D camera coordinates and the world coordinates of the circle centre on the tower. The last step is to estimate the relative pose between the two adjacent RGB-D cameras.

4. Detailed methods

4.1. Tower calibration pattern

Because of the symmetry of the circular marker, the centre detection algorithm of the circle is easy and has strong resistance to noise. In this paper, a tower calibration pattern, which is shown in figure 3(a), is designed for the 3D scanning system of multiple RGB-D cameras, which is shown in figure 2. Five circular markers distributed on three different steps are set at the designated position on each face of the tower calibration pattern; to ensure that the markers are not planar, they are sorted in the order shown in figure 3(b). Since the size of the tower is known, after a world coordinate system is established by selecting a point on the tower, the world coordinate of the centre of each circle mark with respect to the world coordinate system on the tower will be calculated. Thus, the world coordinates $(x_{wi}, y_{wi}, z_{wi})(i = 1, 2, ...5)$ of the centres of the circle marks on each face of the tower calibration pattern will be obtained.

4.2. Colour image and depth image alignment algorithm

An RGB-D camera generally contains two physical sensors, namely, an infrared sensor and a colour sensor, which are at different positions in the camera and have different coordinate systems. To establish the corresponding relationship between the 3D coordinate of a point in 3D space and the colour information, a pose transformation matrix of the coordinate systems of the two sensors must be obtained. In figure 4, P represents a point in 3D space, (u_1, v_1) and (x_1, y_1, z_1) denote the 2D pixel coordinate and the 3D position coordinate relative to the coordinate system of the infrared sensor, respectively, and (u_2, v_2) and (x_2, y_2, z_2) represent the 2D pixel coordinate and 3D position coordinate relative to the coordinate system of the colour sensor, respectively. The relationship between the 2D pixel coordinate and the 3D position coordinate can be formulated as equations (1) and (2).

$$Z_1 \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f_1}{\mathrm{d}x_1} & 0 & u_{01} \\ 0 & \frac{f_1}{\mathrm{d}y_1} & v_{01} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} \tag{1}$$

$$Z_2 \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f_2}{\mathrm{d}x_2} & 0 & u_{02} \\ 0 & \frac{f_2}{\mathrm{d}y_2} & v_{02} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix}. \tag{2}$$

Here, $f_1, \mathrm{d}x_1, \mathrm{d}y_1, u_{01}, v_{01}, f_2, \mathrm{d}x_2, \mathrm{d}y_2, u_{02}, v_{02}$ represent the intrinsic parameters of the infrared sensor and colour sensor. Assume that the matrix T that contains the extrinsic parameters denotes the pose between the infrared sensor and the colour sensor. Because the infrared sensor and the colour sensor are configured for binocular stereo vision mode, the intrinsic and extrinsic parameters of an RGB-D camera can be calibrated by using the checkerboard calibration method. In this paper, we assume that these parameters are known. Then, the relationship between (x_1, y_1, z_1) and (x_2, y_2, z_2) can be formulated as equation (3).

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{bmatrix} = T \cdot \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{r} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{bmatrix}. \tag{3}$$

Here, $\mathbf{r}$ and $\mathbf{t}$ denote the rotation matrix and the translation vector, respectively. From equations (1) to (3), the alignment between the colour image and depth image can be established.

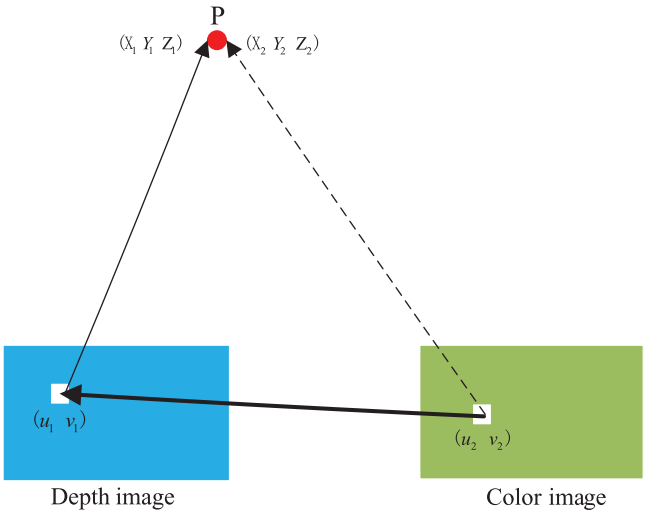

Figure 4. Alignment relationship diagram.

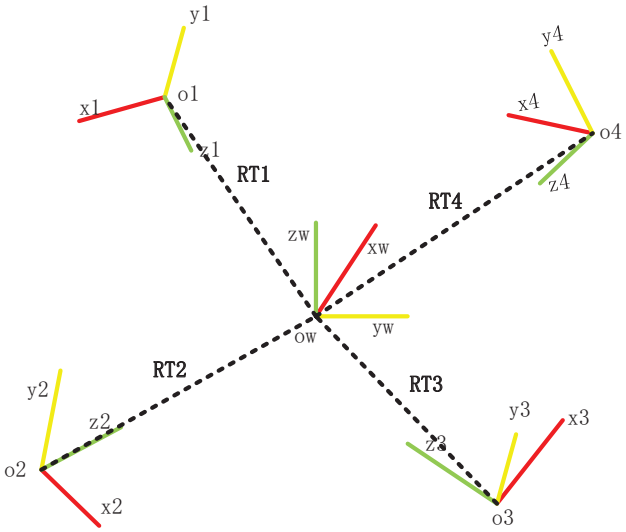

Figure 5. Diagram of the coordinate system.

Figure 6. The prototype of the tower calibration pattern.

4.3. Identification of the centre of a circle

Due to their regularity, symmetry, rotation invariance and ease of detection, circular marks are placed on the tower calibration pattern. First, the image coordinates $(u_i, v_i)(i = 1, 2, 3, 4, 5)$ of the centres of the circular marks are detected by using the circle Hough transformation, which is one of the popular circle detection algorithms in the computer vision field used for colour images. Afterwards, the image coordinates (u_i, v_i) are sorted according to the numbers shown in figure 3(b). Finally, the camera coordinates $(x_{ci}, y_{ci}, z_{ci})(i = 1, 2, 3, 4, 5)$ of the centres of the circular marks on each face of the tower calibration pattern are obtained according to the alignment relationship between the colour image and depth image that is described in section 4.2. Assume that $W_j = \left(x^j_{wi}, y^j_{wi}, z^j_{wi}\right)$ and $C_j = \left(x^j_{ci}, y^j_{ci}, z^j_{ci}\right)(i = 1, 2, 3, 4, 5, j = 1, 2, 3, 4)$ (the subscript i denotes the ith circle centre, and the superscript j denotes the jth face of the tower) represent the world coordinates and the camera coordinates of the centres of the circular marks on the tower pattern, respectively. Then, the corresponding point set $(W_j, C_j)(j = 1, 2, 3, 4)$ will be obtained.

4.4. Extrinsic calibration between the depth camera and the tower calibration pattern

A rotation matrix and a translation vector can be used to describe the relative pose of the camera in 3D space. A coordinate system diagram of the 3D foot scanning system proposed in this paper is shown in figure 5. The coordinate system $o_w - z_w y_w x_w$ indicates the world coordinate system located on the tower calibration pattern, the coordinate system $o_i - z_i y_i x_i (i = 1, 2, 3, 4)$ represents the depth camera coordinate system of the RGB-D camera, and $RT_j (j = 1, 2, 3, 4)$ denotes the pose of the depth camera with respect to the world coordinate system. For the corresponding point set $(W_j, C_j)(j = 1, 2, 3, 4)$, the transformation matrix RT_j is determined by a rotation matrix R_j and a translation vector t_j between the world coordinate system and the depth camera coordinate system:

$$w^i_j = \begin{bmatrix} x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \end{bmatrix}_j = RT_j \cdot c^i_j = RT_j \cdot \begin{bmatrix} x_{ci} \\ y_{ci} \\ z_{ci} \\ 1 \end{bmatrix}_j = \begin{bmatrix} R_j t_j \\ \mathbf{0}_{1\times 3} 1 \end{bmatrix} \cdot \begin{bmatrix} x_{ci} \\ y_{ci} \\ z_{ci} \\ 1 \end{bmatrix}_j \tag{4}$$

Here, $w^i_j = [x_{wi}, y_{wi}, z_{wi}, 1]^T_j$ and $c^i_j = [x_{ci}, y_{ci}, z_{ci}, 1]^T_j$ belong to W_j and C_j, respectively. R_j and t_j are unknown parameters, and we need at least 3 corresponding points to solve for these parameters. In other words, our goal is to find R_j and t_j that minimize the following objective function:

$$\min_{R_j, t_j} f_j(R_j, t_j) = \frac{1}{n} \sum_{i=1}^{n} \left\| w^i_j - R_j c^i_j - t_j \right\|^2 \tag{5}$$

where $R_j = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}_j$ and $t_j = [t_x, t_y, t_z]^T_j$. Generally, a least-squares-based method can be used to solve the above objective function in equation (5). However, the solution cannot satisfy the orthogonality of the rotation matrix R_j. Therefore, to guarantee the orthogonality, the unit quaternion

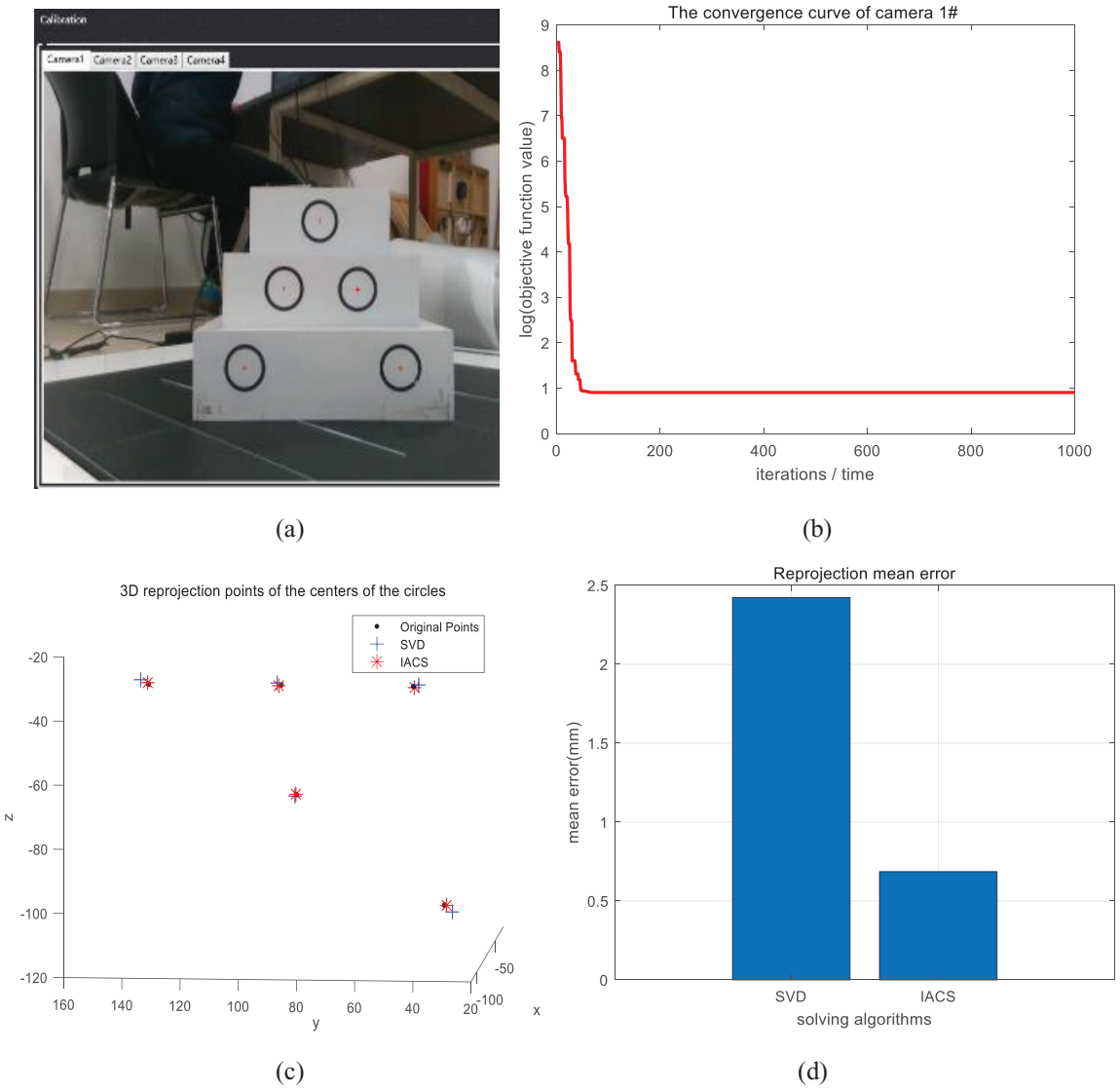

Figure 7. The extrinsic calibration results of depth camera 1#. (a) The circular centre recognition results for the 2D colour image. (b) The convergence curve of the IACS. (c) The 3D reprojection points of the centres of the circles. (d) Reprojection mean error.

method is applied to represent the rotation matrix. Assume that a unit quaternion $q_j = [q_{j0}, q_{j1}, q_{j2}, q_{j3}]^T$ denotes the rotation matrix R_j; then, the relationship between q_j and R_j is shown as follows:

$$\begin{cases} R_{11} = q_{j0}^2 + q_{j1}^2 - q_{j2}^2 - q_{j3}^2 \\ R_{12} = 2(q_{j1}q_{j2} - q_{j0}q_{j3}) \\ R_{13} = 2(q_{j1}q_{j3} + q_{j0}q_{j2}) \\ R_{21} = 2(q_{j1}q_{j2} + q_{j0}q_{j3}) \\ R_{22} = q_{j0}^2 - q_{j1}^2 + q_{j2}^2 - q_{j3}^2, \quad q_{j0}^2 + q_{j1}^2 + q_{j2}^2 + q_{j3}^2 = 1. \\ R_{23} = 2(q_{j2}q_{j3} - q_{j0}q_{j1}) \\ R_{31} = 2(q_1q_3 - q_0q_2) \\ R_{32} = 2(q_{j2}q_{j3} + q_{j0}q_{j1}) \\ R_{33} = q_{j0}^2 - q_{j1}^2 - q_{j2}^2 + q_{j3}^2 \end{cases} \tag{6}$$

The translation vector t_j between W_j and C_j can be calculated by the following equation:

$$t_j = \rho_{cj} - R_j \cdot \rho_{wj}. \tag{7}$$

Here, $\rho_{wj} = \frac{1}{n}\sum_{i=1}^{n} w_j^i$ and $\rho_{cj} = \frac{1}{n}\sum_{i=1}^{n} c_j^i$ represent the barycenter coordinates of W_j and C_j, respectively. Then, the problem of solving equation (5) can be transformed into an optimization problem with a constraint, and its objective function will be rewritten as the following equation:

$$\min_{q_j} f(q_j) = \frac{1}{n}\sum_{i=1}^{n} ||c_{ji} - R_j \cdot w_{ji} - t_j||^2 + ... \\ \lambda(q_{j0}^2 + q_{j1}^2 + q_{j2}^2 + q_{j3}^2 - 1), \lambda > 0. \tag{8}$$

Here, λ represents any positive value. Thus, some meta-heuristic optimization algorithms can be used to solve the above optimization problem with a constraint.

4.5. IACS algorithm

The cuckoo search (CS) algorithm was proposed by Yang and Deb [39] in 2009; the CS algorithm is a meta-heuristic evolutionary algorithm inspired by the brood parasitic behaviour of a type of cuckoo in combination with the Lévy flight

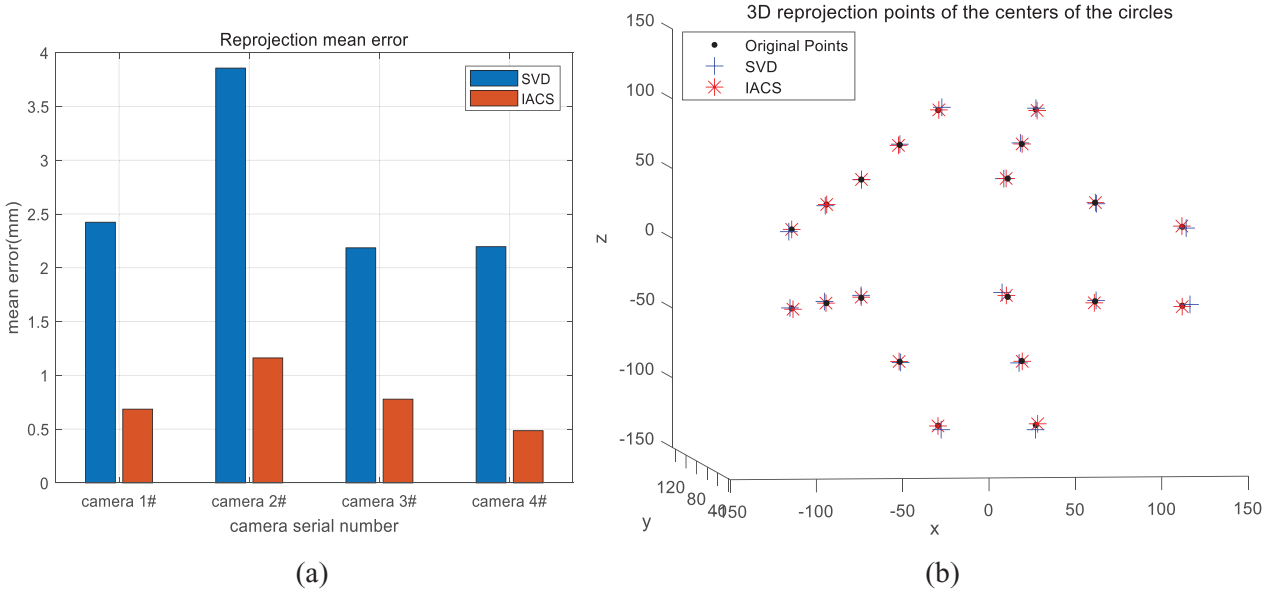

Figure 8. The extrinsic calibration results of all cameras. (a) Reprojection mean error. (b) 3D reprojection points of the centres of the circles.

Table 1. The results of the calibration after 100 experiments.

Method	Mean (*Err*) (mm)	Std. (*Err*)
SVD	2.6674	0.0036
IACS	0.8925	0.0029

behaviour of some animals. A Lévy flights random walk (LFRW) and biased/selective random walk (BSRW) are used to search for new solutions in the CS algorithm. After each random walk, a greedy strategy is used to select a better solution from the current solutions and newly generated solutions by comparing their fitness values. The optimal solutions obtained by the CS are far better than the best solutions obtained by the efficient PSO and GA [40].

However, the CS is the same as other population-based intelligent optimization algorithms in that it has some disadvantages, such as a weak local search ability, easily falling into local optima, and a slow convergence speed. In terms of the standard implementation of the CS, the parameters p_a, which is the probability of a cuckoo's egg being found, and α_0, which is the scaling factor of the CS, are beneficial for searching globally and locally improved solutions, respectively. These two parameters are very important for fine-tuning the solutions and can potentially be used to improve the convergence rate and performance of the CS algorithm. Thus, to improve the convergence speed and avoid falling into a local optimum, we propose an IACS algorithm based on population fitness information and iteration information feedback. Two adaptive strategies are integrated into the standard CS algorithm.

The first adaptive strategy is to adjust the parameters p_a and α_0 according to the iteration information in each iteration, which can be formulated as equations (9)–(11).

$$c = \frac{1}{G_{\max}} \ln\left(\frac{\alpha_{\min}}{\alpha_{\max}}\right) \tag{9}$$

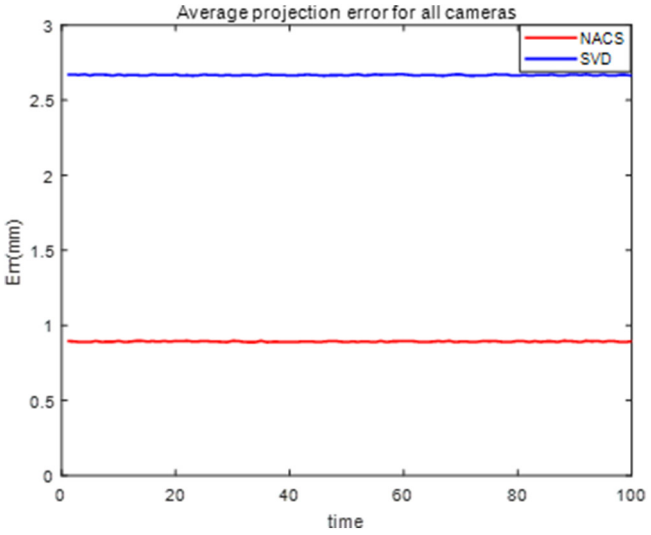

Figure 9. Average projection error for all cameras.

$$\alpha_0(G) = \alpha_{\max} \exp(c \cdot G) \tag{10}$$

$$p_a(G) = p_{a\max} - \frac{G}{G_{\max}} (p_{a\max} - p_{a\min}) \tag{11}$$

Here, G and $G_{\max}$ are the current iteration and the total iteration times, respectively. $\alpha_{\min}$ and $p_{a\min}$ represent the minimum values of α_0 and p_a, respectively, and $\alpha_{\max}$ and $p_{a\max}$ represent the maximum values. From equations (9)–(11), it can be seen that the parameters p_a and α_0 are large in the early iterations, and the diversity of the population is increased, which improves the global search ability of the CS; however, to obtain a better local adjustment solution, these parameters will be decreased in a later iteration.

The second strategy is to adjust the Lévy step size of each individual according to the category of the individual, where

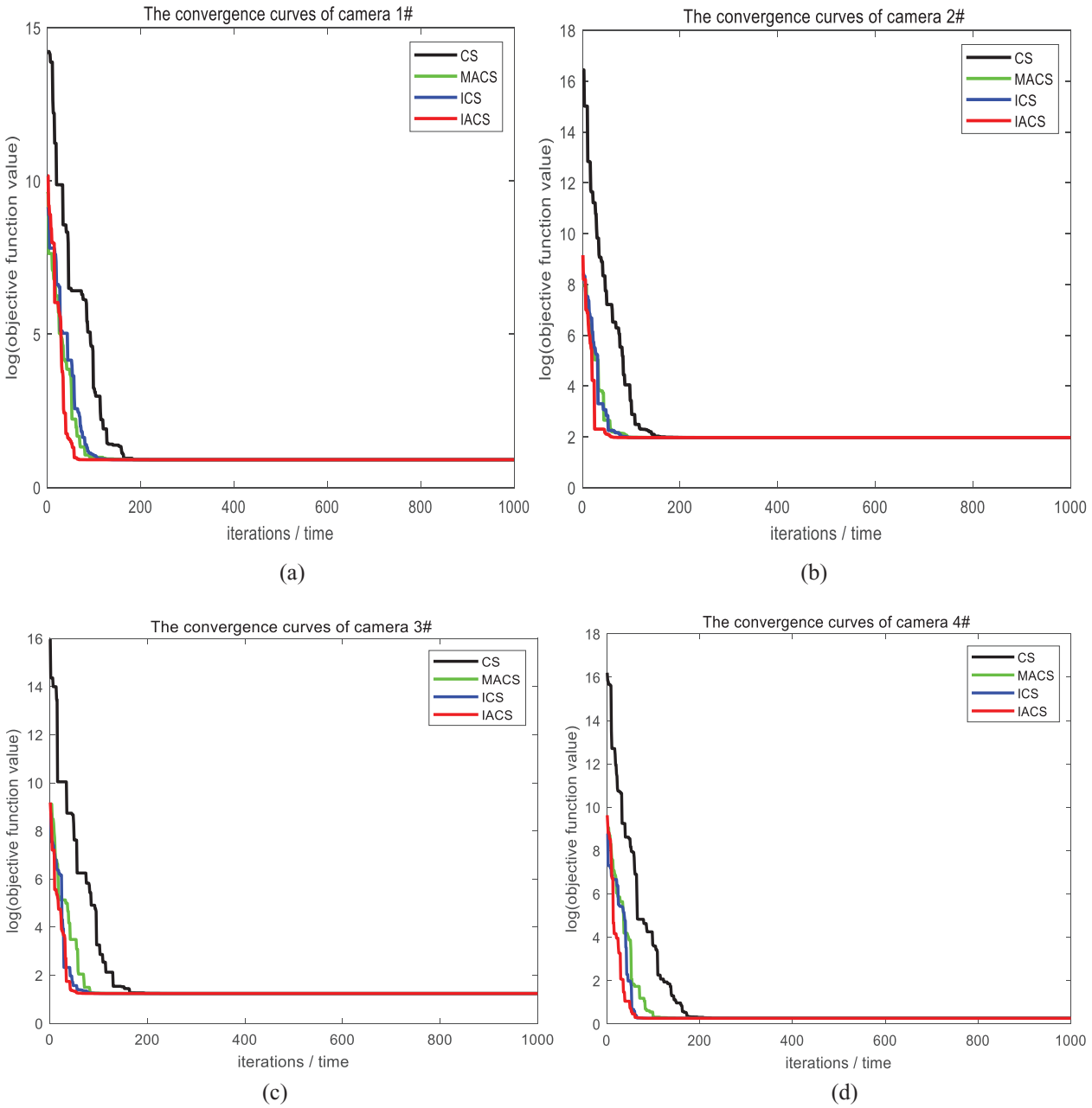

Figure 10. The convergence curves of camera 1# for the four optimization algorithms. (a) The convergence curves of camera 1#. (b) The convergence curves of camera 2#. (c) The convergence curves of camera 3#. (d) The convergence curves of camera 4#.

the categories include the individual fitness, best fitness, and mean population fitness in the current iteration, which can be formulated as equations (12)–(14).

$$\alpha_1 = \alpha_0(gn) \cdot \left| \frac{f(x_i) - f'}{f_m - f'} \right| \text{ if}, f_{\min} \leqslant f(x_i) \leqslant f' \quad (12)$$

$$\alpha_2 = \alpha_0(gn) \text{ if}, f' < f(x_i) < \bar{f} \quad (13)$$

$$\alpha_3 = 1.0 - \frac{1}{1 + \exp(-\Delta)} \text{ if}, f(x_i) > \bar{f}. \quad (14)$$

Here, $\bar{f} = \frac{1}{Np} \sum_{i=1}^{Np} f(x_i)$ and $f_{\min} = \min f(x_i)$ are the mean fitness and minimum fitness of the current generation, respectively. f' represents the average between $\bar{f}$ and $f_{\min}$. $\Delta = f' - f_{\min}$ is a degree factor. For a detailed description about the IACS, please refer to my previous work [38]. In this paper, the IACS is applied to optimize equation (8) to solve for the optimal q_j^*.

5. Experiments

The proposed calibration algorithm has no special requirement in terms of the type of depth sensing technology, which may be structured light, stereo vision or a time-of-flight RGB-D camera. For example, the Intel Realsense SR300 structured light RGB-D camera, which is inexpensive and small in size, is chosen to capture the depth and colour images in the experiments. The RGB-D camera setup of the 3D scanning

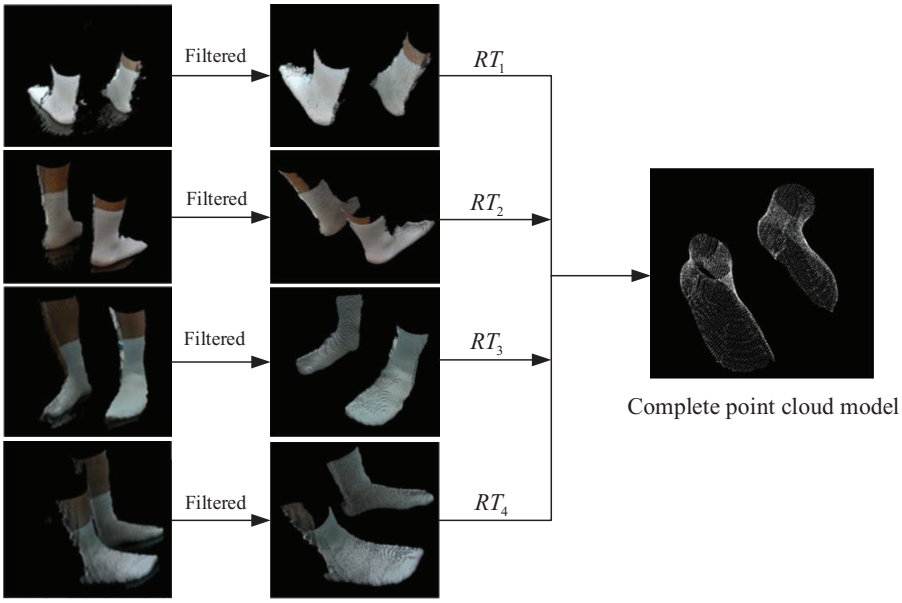

Figure 11. The foot reconstruction process.

system is shown in figure 2. The camera network consists of 4 SR300 cameras sparsely placed on a platform of 0.64 m^2, and all cameras point to the centre of the platform. The parameters α_0, p_a, $\alpha_{\min}$, $\alpha_{\max}$, $p_{a\min}$ and $p_{a\max}$ associated with the IACS algorithm are set to 0.01, 0.25, 0.05, 1.5, 0.005 and 0.95, respectively.

5.1. Single depth camera extrinsic calibration

The tower calibration pattern is produced as shown in figure 6, according to section 4.1. Assume that the world coordinate system $o_w - z_w y_w x_w$ is set at the centre of the tower bottom; then, the world coordinate of each centre of the circle will be calculated according to the size of the tower. The coordinate system $o_w - z_w y_w x_w$ can be regarded as a unified coordinate system. The proposed extrinsic calibration method can realize a sequential calibration for 4 depth cameras in a 3D scanning system. The calibration results of camera 1# are shown in figure 7, and the calibration results of the other cameras are the same as those shown in figure 7. The circular centre recognition results for the 2D colour image are shown in figure 7(a); the red asterisk indicates the correctly identified centres of the circles. The convergence curve of the IACS is shown in figure 7(b); it is easy to see that the proposed optimization method gradually converges after approximately 80 iterations. To evaluate the quality of the solution of the IACS algorithm, the solution obtained by SVD is used for a comparison. The 3D reprojection points of the centres of the circles are shown in figure 7(c); in the figure, the black dot represents the original point of the centre of the circle, the blue plus represents the reprojection results obtained by SVD, and the red asterisk represents the results obtained by the IACS. Notably, the red asterisk is closer to the original point than the blue plus, which shows that the calibration results obtained by the IACS are better than those obtained by SVD. The 3D reprojection error of the centres of the circles can be formulated as follows:

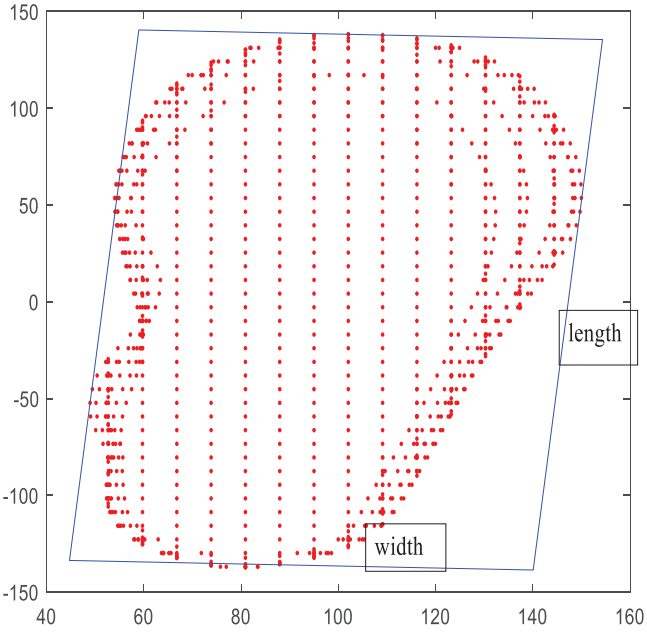

Figure 12. Diagram for the extraction of the right foot length and width parameters.

$$\mathrm{RMSE}_j = \frac{1}{n}\sum_{i=1}^{n} \left\| w_{ji} - R_j^* c_{ji} - t_j^* \right\|^2. \tag{15}$$

Here, w_{ji} represents the world coordinate of the centre of the i circular marker on the jth face of the calibration tower, n represents the number of circular markers, and R_j^* and t_j^* denote the optimal rotation and translation matrix, respectively, which are obtained by the IACS or SVD. RMSE_j denotes the root mean square error, which is obtained from projecting the 3D circle centre locations in each local camera coordinate system to the world coordinate system through the R_j^* and t_j^*. The RMSE_j values, which are obtained by the IACS and SVD, are shown in figure 7(d). The RMSE_j (0.6857 mm)

Table 2. The results of the parameters of the right foot (mm).

	Standard		SVD						IACS					
Volunteer	L	W	L	E	ME	W	E	ME	L	E	ME	W	E	ME
1	257	95	263	6	4.3	98	3	3.9	258	1	1.8	96	1	1.3
2	261	97	265	4		100	3		263	2		98	1	
3	245	90	249	4		93	3		247	2		91	1	
4	267	101	270	3		105	4		269	2		102	1	
5	263	100	268	5		103	3		265	2		101	1	
6	255	95	259	4		100	5		257	2		97	2	
7	251	92	256	5		97	5		252	1		94	2	
8	261	96	265	4		101	5		263	2		97	1	
9	249	91	253	4		95	4		251	2		92	1	
10	253	93	257	4		97	4		255	2		95	2	

Note: L: length, W: width, E: error, ME: mean error.

of the IACS is much smaller than the RMSE_j (2.4237 mm) of SVD. The extrinsic calibration results of all the cameras are shown in figure 8.

5.2. Pose evaluation between two depth cameras

As described in section 5.1, the pose transformation matrix $RT_j, (j = 1, 2, 3, 4)$ of each depth camera with respect to the world coordinate system on the tower can be calculated. However, when the tower calibration pattern is moved to another position, $RT_j, (j = 1, 2, 3, 4)$ will change accordingly. Nevertheless, the pose transformation matrix RT_i^j between the ith camera and the jth camera will not change since these cameras are in a static state. Without loss of generality, the camera coordinate system of camera 1# is set to the unified coordinate system; then, the pose transformation matrix of the remaining cameras with respect to camera 1# can be expressed as the following equation:

$$RT_1^j = (RT_1)^{-1} * RT_j, (j = 1, 2, 3, 4). \tag{16}$$

Therefore, according to the pose transformation matrix between camera 1# and the tower calibration pattern, the world coordinates $W_j, (j = 1, 2, 3, 4)$ of the centres of the circles on the tower can be converted to the camera coordinates PC_1^j relative to camera 1#. The conversion formula is as follows:

$$PC_1^j = (RT_1)^{-1} * W_j, (j = 1, 2, 3, 4). \tag{17}$$

Meanwhile, according to RT_i^j, $C_j, (j = 2, 3, 4)$ can be converted to the camera coordinates Pc_1^j relative to camera 1#. The conversion formula is as follows:

$$Pc_1^j = (RT_1)^{-1} * RT_j * C_j, (j = 2, 3, 4). \tag{18}$$

If there is no calibration error, the results of equation (17) are equal to those of equation (18). However, there is bound to be error. In this section, the calibration error is represented as follows:

$$\mathrm{Err} = \frac{1}{4}\sum_{j=1}^{4} \mathrm{projection\_err}_j. \tag{19}$$

Figure 13. Manual measurement.

Here,

$$\mathrm{projection\_err}_j = \left\| Pc_1^j - PC_1^j \right\|^2, (j = 1, 2, 3, 4). \tag{20}$$

To verify the effectiveness and stability of the proposed method and exclude contingency, the tower calibration pattern was moved to 100 different positions in a limited space, and the results of all positions were recorded. The mean errors and variances of the two methods are shown in table 1 and figure 9. As shown in table 1, the error variances of multiple calibrations of the two methods are very small and similar, indicating that both methods have good effectiveness and stability. However, the mean error of the IACS is smaller than that of the SVD method, indicating that the proposed method can achieve a high calibration accuracy.

5.3. Convergence evaluation of the IACS algorithm

The convergence of the IACS algorithm is verified in a comparative experiment. The ICAS is compared to the standard CS algorithm and two CS variants. The first CS variant is a memory-adaptive CS algorithm (MACS) proposed by Qin *et al* [41]. The authors adjusted the scaling factor α_0 adaptively according to population average fitness and individual fitness, and a memory strategy was added to the BSRW. The second CS variant is an improved CS algorithm (ICS) proposed by Ehsan *et al* [42]. In this variant, the authors adjusted

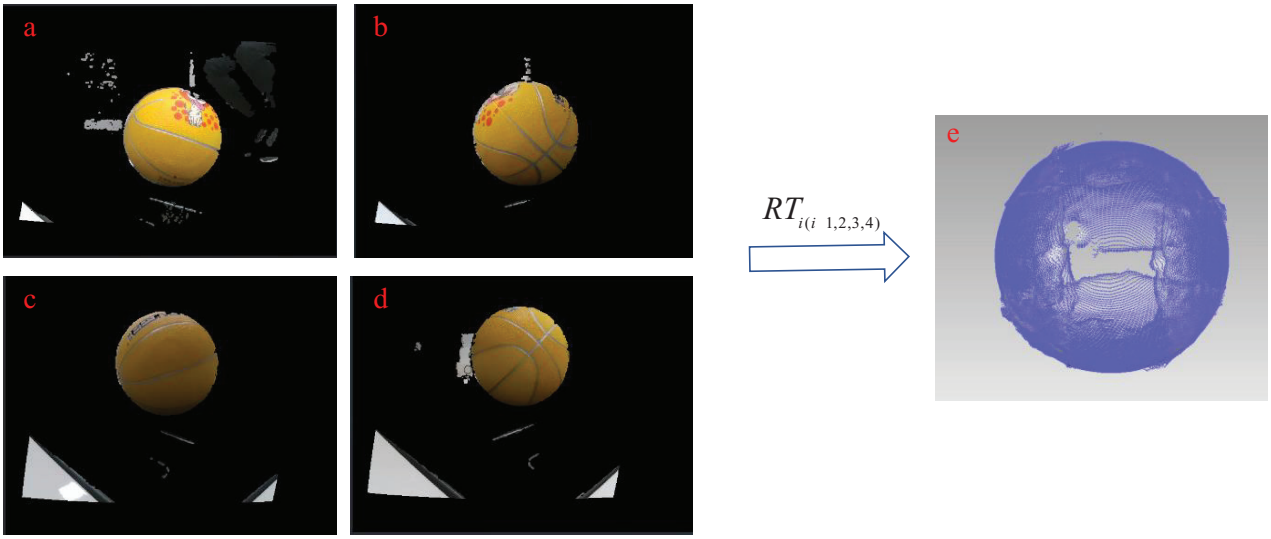

Figure 14. The scan results of the sphere. (a). Raw point cloud of camera 1#. (b) Raw point cloud of camera 2#. (c) Raw point cloud of camera 3#. (d) Raw point cloud of camera 4#. (e) complete point cloud of the sphere.

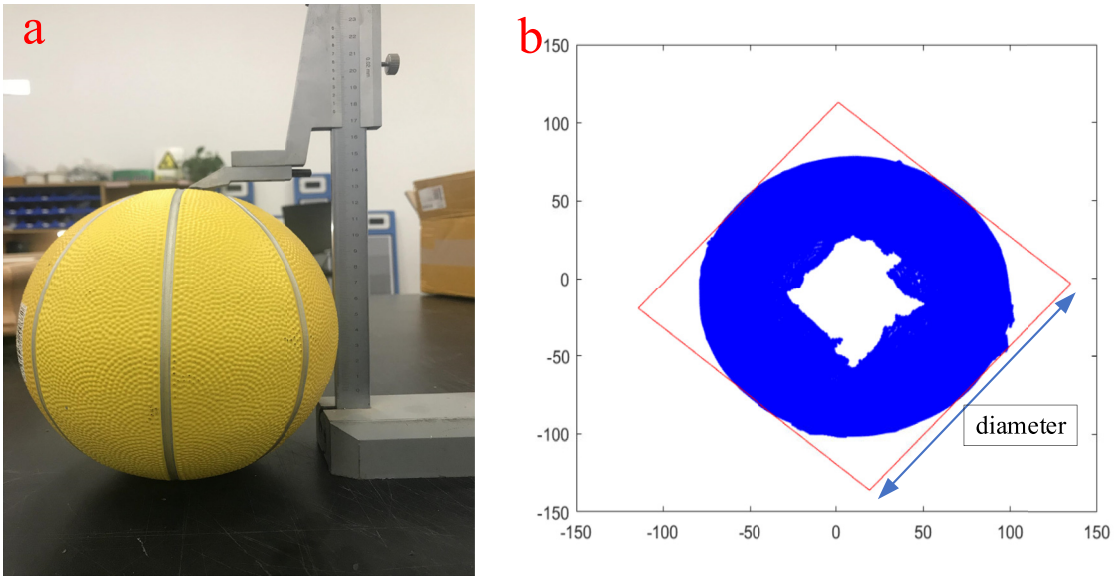

Figure 15. Diameter measurement methods. (a) Manually measurement. (b) Automatic measurement.

the parameters p_a and α_0 according to the iteration number. These CS algorithms are applied to optimize equation (8) for each camera. The parameters of the IACS are described above, and the parameters of the standard CS, MACS and ICS algorithms are described in the corresponding literature. The convergence curves of these CS algorithms for camera 1# are shown in figure 10(a). As shown in the figure, although these four optimization algorithms converge to the same optimal level, the IACS achieves the fastest convergence speed among these CS algorithms. The optimization results of the remaining cameras are the same as those shown in figures 10(b)–(d).

5.4. Qualitative evaluation

In this section, we evaluate the results of the proposed calibration method when scanning a human foot by a 3D scanning system with multiple RGB-D cameras. We evaluate the quality of the human foot reconstruction for different optimal calibration methods. The process of the foot reconstruction is shown in figure 11. Ten volunteers were invited to participate in the experiment. According to the established complete foot point cloud model, the length and width of each volunteer's foot were extracted. The length and width of a foot were the length and width of the smallest external rectangle of the foot projection point cloud, respectively, as shown in figure 12. The foot of each volunteer was measured 10 times, and then the means of the parameters were obtained. The results of the parameters of the right foot are shown in table 2. In table 2, the standard values are obtained by using a manual measurement, which is shown in figure 13. From table 2, in terms of the foot length, the maximum error from the standard value of SVD is 6 mm, that the maximum error of the IACS is 3 mm,

Table 3. The results of the diameter of the sphere (mm).

	Standard	SVD			IACS		
Time	D	D	E	ME	D	E	ME
1	175	178	3	3.8	177	2	1.7
2	175	179	4		177	2	
3	175	179	4		176	1	
4	175	180	5		178	3	
5	175	179	4		176	1	
6	175	178	3		176	1	
7	175	178	3		178	3	
8	175	180	5		176	1	
9	175	179	4		176	1	
10	175	178	3		177	2	

Note: D: diameter, E: error, ME: mean error.

and the mean errors from the standard values are 4.3 mm and 1.8 mm. In terms of the foot width, the maximum error from the standard value of SVD is 5 mm, the maximum error of the IACS is 2 mm, and the mean errors from the standard values are 3.9 mm and 1.3 mm. The same conclusions can be drawn from the parameters of the left foot, so the parameters of the left foot do not appear in table 2. In addition, a yellow sphere is also used to verify the accuracy of the proposed method, the scan results of the sphere are shown in figure 14. The diameter measurement methods are shown in figure 15. The results of measuring the diameter of the sphere 10 times are shown in table 3. From table 3, in terms of the diameter of the sphere, the maximum error from the standard value of SVD is 5 mm, that the maximum error of the IACS is 3 mm, and the mean errors from the standard values are 3.8 mm and 1.7 mm. In summary, compared with the SVD method, the proposed method in this paper achieves a higher accuracy in both the foot and cube box measurements.

6. Conclusions

In this paper, a compact multi-RGB-D-camera 3D scanning system and an extrinsic calibration method assisted by a tower calibration pattern with circular markers in a limited FOV are proposed. First, the Hough transform is used to identify the image coordinates of the centres of circles in a colour image, and then the 3D coordinates are extracted by employing the alignment relationship between the colour image and the depth image; therefore, the correspondences between the camera coordinate system and the world coordinate system are constructed. Second, IACS algorithm and a unit quaternion are used to optimize the extrinsic parameters. The optimization results of the proposed method are compared with those of the SVD method, and it is shown that the IACS achieves a higher calibration accuracy than the SVD method. Then, the IACS is compared with the standard CS algorithm and two CS variants. The experimental results indicate that the IACS achieves a faster convergence speed than the other CS methods. Finally, the proposed extrinsic calibration method is applied to a proposed 3D scanning system that is used to reconstruct a foot and a yellow sphere. The results indicate that the multi-perspective point cloud registration is very effective, which provides a good basis for subsequent processing. The accuracy and the effeciency of the proposed extrinsic calibration method are validated by these experiments, and it is explained that the proposed extrinsic calibration method has a high calibration accuracy and a higher speed and good potential application prospects compared with the existing methods.

Acknowledgment

This work supported by the National Young Natural Science Foundation (NO. 61702375, NO. 61803235) of China, the Key Research Programs of Shandong Province (NO. 2018GGX103011), the Key Science and Technology Innovation Programs in Shandong Province (NO. 2017CXGC0919).

ORCID iDs

Fan Binghui https://orcid.org/0000-0003-3393-7900

References

[1] Zhu C P *et al* 2019 Error estimation of 3D reconstruction in 3D digital image correlation *Meas. Sci. Technol.* **2** 1–10
[2] Meerits S *et al* 2018 FusionMLS: highly dynamic 3D reconstruction with consumer-grade RGB-D cameras *Comput. Vis. Media* **4** 287–303
[3] Chun S *et al* 2017 A foot-arch parameter measurement system using a RGB-D camera *Sensors* **17** 1796–821
[4] Filko D, Cupec R and Nyarko E K 2018 Wound measurement by RGB-D camera *Mach. Vis. Appl.* **29** 633–54
[5] Hoang D-C 2017 Sub-OBB based object recognition and localization algorithm using range images *Meas. Sci. Technol.* **2** 1–11
[6] Williem W *et al* 2018 Visual-Inertial RGB-D SLAM for mobile augmented reality *Pacific Rim Conf. on Multimedia* pp 928–38
[7] Su P C *et al* 2017 Real-time rendering of physical scene on virtual curved mirror with RGB-D camera networks *IEEE Int. Conf. on Multimedia & Expo Workshops* pp 79–84
[8] Lin H, Da F P and Gai S Y 2019 Research on multi-camera calibration and point cloud correction method based on three-dimensional calibration object *Opt. Lasers Eng.* **115** 32–41
[9] Xia R *et al* 2018 Global calibration of multi-cameras with non-overlapping field of views based on photogrammetry and reconfigurable target *Meas. Sci. Technol.* **29** 065005
[10] Zhang S 2018 High-speed 3D shape measurement with structured light methods: a review *Opt. Lasers Eng.* **106** 119–31
[11] Miyata S *et al* 2018 Extrinsic camera calibration without visible corresponding points using omnidirectional cameras *IEEE Trans. Circ. Syst. Video Technol.* **28** 2210–9
[12] Perez-Yus A *et al* 2018 Extrinsic calibration of multiple RGB-D cameras from line observations *IEEE Robot. Autom. Lett.* **3** 273–80
[13] Ly D S *et al* 2014 Extrinsic calibration of heterogeneous cameras by line images *Mach. Vis. Appl.* **25** 1601–14
[14] Yang J *et al* 2013 Single-shot extrinsic calibration of a generically configured RGB-D camera rig from scene constraints *IEEE Int. Symp. on Mixed and Augmented Reality* pp 181–8

[15] Fernández-Moral E *et al* 2014 Extrinsic calibration of a set of range cameras in 5 seconds without pattern *Int. Conf. on Intelligent Robots and Systems (IROS)* pp 429–35
[16] Sinha S N, Pollefeys M and McMillan L 2004 Camera network calibration from dynamic silhouettes *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* pp 195–202
[17] Li S, Pathirana P N and Caelli T 2014 Multi-kinect skeleton fusion for physical rehabilitation monitoring *36th Ann. Int. Conf IEEE Eng. in Med. Biol. Soc.* 5060–3
[18] Liu H *et al* 2019 A novel method for extrinsic calibration of multiple RGB-D cameras using descriptor-based patterns *Sensors* **19** 349–51
[19] Park J *et al* 2011 High quality depth map upsampling for 3D-TOF cameras *IEEE Int. Conf. on Computer Vision* pp 1623–30
[20] Daniel H C, Kannala J and Heikkilä J 2011 Accurate and practical calibration of a depth and color camera pair *Int. Conf. on Computer Analysis of Images and Patterns* pp 437–45
[21] Daniel H C, Kannala J and Heikkilä J 2012 Joint depth and color camera calibration with distortion correction *IEEE Trans. Pattern Anal. Mach. Intell.* **34** 2058–64
[22] Kim Y *et al* 2008 Design and calibration of a multi-view TOF sensor fusion system *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* pp 1–7
[23] Zhang C and Zhang Z 2011 Calibration between depth and color sensors for commodity depth cameras *IEEE Int. Conf. on Multimedia and Expo* pp 1–6
[24] Shim H *et al* 2012 Time-of-flight sensor and color camera calibration for multi-view acquisition *Vis. Comput.* **28** 1139–51
[25] Liu W *et al* 2012 A new method for calibrating depth and color camera pair based on kinect *Int. Conf. on Audio Language and Image Processing* pp 212–7
[26] Lébraly P *et al* 2010 Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: application to vision-based robotics *Int. Conf. on IEEE Intelligent Robots and Systems* pp 5640–7
[27] Kumar R K *et al* 2008 Simple calibration of non-overlapping cameras with a mirror *IEEE Conf. on Computer Vision and Pattern Recognition* pp 1–7
[28] Gong Z, Liu Z and Zhang G 2017 Flexible global calibration of multiple cameras with non-overlapping fields of view using circular targets *Appl. Opt.* **56** 3122–31
[29] Dong S *et al* 2016 Extrinsic calibration of a non-overlapping camera network based on close-range photogrammetry *Appl. Opt.* **55** 6363–70
[30] Shen J *et al* 2014 Extrinsic calibration for wide-baseline RGB-D camera network *IEEE Int. Workshop on Multimedia Signal Processing* pp 21–6
[31] Su P *et al* 2018 A fast and robust extrinsic calibration for RGB-D camera networks *Sensors* **18** 235–58
[32] Ruan M and Huber D 2014 Calibration of 3D sensors using a spherical target *Int. Conf. on 3D Vision* pp 187–93
[33] Staranowicz A N *et al* 2015 Practical and accurate calibration of RGB-D cameras using spheres *Comput. Vis. Image Underst.* **137** 102–14
[34] Meng X, Gao W and Hu Z 2018 Dense RGB-D SLAM with multiple cameras *Sensors* **18** 2118–29
[35] Kuo T *et al* 2014 Calibrating a wide-area camera network with non-overlapping views using mobile devices *ACM Trans. Sensor Netw.* **10** 1–26
[36] Xu G *et al* 2016 Global calibration method of a camera using the constraint of line features and 3D world points *Meas. Sci. Rev.* **16** 190–6
[37] Safaei A and Fazli S 2018 A novel solution in the simultaneous deep optimization of RGB-D camera calibration parameters using metaheuristic algorithms *Turkish J. Electr. Eng. Comput. Sci.* **26** 743–54
[38] Jia C C *et al* 2020 An improved adaptive cuckoo search algorithm based the population feature and iteration information *Int. J. Commun. Netw. Distrib. Syst.* in press (https://doi.org/10.1504/IJCNDS.2020.10018198)
[39] Yang X S and Deb S 2009 Cuckoo search via Lévy flights *IEEE World Congress on Nature & Biologically Inspired Computing* pp 210–4
[40] Jiang Y L *et al* 2018 TDOA passive location based on cuckoo search algorithm *J. Shanghai JiaoTong Univ.* **23** 368–75
[41] Qin L and Dai R 2017 An adaptive cuckoo search algorithm with memory *Microelectron. Comput.* **34** 15–9
[42] Ehsan V, Shahram M and Saeed T 2011 Improved cuckoo search algorithm for global optimization *Int. J. Commun. Inf. Technol.* **1** 31–44