

Research Article

A Novel Method for Automatic Extrinsic Parameter Calibration of RGB-D Cameras

Qin Shi , Huansheng Song , and Shijie Sun 

School of Information Engineering, Chang'an University, Xi'an 710064, China

Correspondence should be addressed to Huansheng Song; hshsong@chd.edu.cn

Received 8 July 2020; Accepted 6 July 2021; Published 15 July 2021

Academic Editor: Manuel De la Sen

Copyright © 2021 Qin Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Calibration of extrinsic parameters of the RGB-D camera can be applied in many fields, such as 3D scene reconstruction, robotics, and target detection. Many calibration methods employ a specific calibration object (i.e., a chessboard, cuboid, etc.) to calibrate the extrinsic parameters of the RGB-D color camera without using the depth map. As a result, it is difficult to simplify the calibration process, and the color sensor gets calibrated instead of the depth sensor. To this end, we propose a method that employs the depth map to perform extrinsic calibration automatically. In detail, the depth map is first transformed to a 3D point cloud in the camera coordinate system, and then the planes in the 3D point cloud are automatically detected using the Maximum Likelihood Estimation Sample Consensus (MLE-SAC) method. After that, according to the constraint relationship between the ground plane and the world coordinate system, all planes are traversed and screened until the ground plane is obtained. Finally, the extrinsic parameters are calculated using the spatial relationship between the ground plane and the camera coordinate system. The results show that the mean roll angle error of extrinsic parameter calibration was -1.14° . The mean pitch angle error was 4.57° , and the mean camera height error was 3.96 cm. The proposed method can accurately and automatically estimate the extrinsic parameters of a camera. Furthermore, after parallel optimization, it can achieve real-time performance for automatically estimating a robot's attitude.

1. Introduction

RGB-D cameras, such as Kinect [1–6], PrimeSense, and Asus Xtion Pro, are traditional RGB cameras with added infrared cameras. Figure 1 shows the structure of a Kinect camera. It can be seen that it includes a color camera, infrared camera, and an infrared illuminator. The color camera outputs a color image, the infrared camera outputs a depth image, and the infrared illuminator emits infrared light for the calculation of the depth image. With the emergence of low-cost RGB-D cameras such as Kinect, the camera and illuminator are increasingly used for tasks such as 3D scene reconstruction and navigation [8, 9], target recognition and tracking [10, 11], 3D measurement [12–15], and even social networks [16], for which the extrinsic parameters of the camera often must be calibrated.

For example, in target detection and recognition [11], the extrinsic parameters of the camera should be calibrated first,

and then pedestrians in the scene are tracked. If the extrinsic parameters of the camera can be obtained, the negative effects of perspective can be eliminated. In addition, the detection algorithm can be unified, and the recognition process can be simplified, thus accelerating recognition speed.

However, studies of the calibration of RGB-D cameras mainly focus on the extrinsic parameters of the color camera relative to the infrared camera [17, 18]. Calibrating the extrinsic parameters of an RGB-D camera is often similar to the calibration of the color camera, and the method of chessboard calibration is generally used to calibrate the color camera [19]. Munaro et al. [10] used a chessboard to calibrate the extrinsic parameters of multiple cameras and carried out pedestrian detection based on this. This calibration method did not make full use of the depth information provided by the RGB-D camera, and the calibration results were essentially the extrinsic parameters of the color

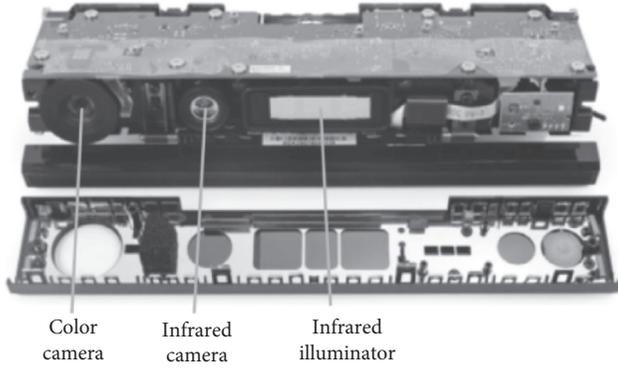


FIGURE 1: Sensor structure of the Kinect V2 [7].

camera. If it were directly used for 3D reconstruction of the depth map, then there would be a large error. Shibo and Qing [20] designed a calibration board for RGB-D infrared camera recognition, which had holes with regular intervals and calibrated the infrared camera by automatically identifying holes. The need to design special calibration objects increased the difficulty of calibration. Liao et al. [17] divided the calibration into three categories: calibration that required a calibrator, calibration that required human intervention, and fully automatic calibration. The proposed method belonged to the third category.

Liao et al. [17] divided the extrinsic parameter calibration methods for RGB-D cameras into three categories: (1) the first was to calibrate the extrinsic parameters of the color camera and then to use the extrinsic parameter $T_{C'}^D$ of the color camera and images collected by the infrared camera to obtain the extrinsic parameter T_D of the infrared camera through transformation $T_D = T_{C'}^D T_{C'}$. This method can directly use the color image calibration method, but it needs the extrinsic parameters of the color camera relative to the infrared camera and does not use the information provided by the depth map, so the process is complicated. (2) The second was to detect the features on the depth map provided by the infrared camera by designing a specific calibration object to obtain the extrinsic parameters of the camera based on the features (such as chessboard corners). (3) The third method used the depth map to calibrate the extrinsic parameters of the camera. The calibration was carried out by detecting the target on the depth map and using the relationship between the target and the world coordinate system. This method directly used the depth information, which greatly simplified the process and improved the efficiency of calibration. Our proposed method, which we will call the ground plane calibration method, is based on the automatic calibration of the extrinsic parameters of a ground plane detection camera and belongs to the third category. This method can directly calibrate the infrared camera of an RGB-D camera using its depth information.

2. Extrinsic Parameter Calibration

RGB-D cameras can be divided into color cameras and infrared cameras. For this study, the extrinsic parameters of an infrared camera were calibrated. Thus, extrinsic

parameter calibration refers to that of infrared cameras. The extrinsic parameter of the camera is

$$T_C^W = \begin{bmatrix} R_C^W & 0 \\ t_C^W & 1 \end{bmatrix}, \quad (1)$$

where $T_C^W \in \mathbb{R}^{4 \times 4}$ and $R_C^W \in \mathbb{R}^{3 \times 3}$ are the rotation matrices of the camera, $t_C^W \in \mathbb{R}^{1 \times 3}$ is the translation matrix of the camera, and $(\cdot)_C^W$ represents the transformation from the camera coordinate system to the world coordinate system.

Figure 2 shows the flowchart of extrinsic parameter calibration. After the establishment of a world coordinate system, the depth image is first obtained from the RGB-D camera. Then, the depth image is transformed to a 3D point cloud under the camera coordinate system using the internal parameters of the infrared camera. After that, the ground plane in the point cloud is solved by iteration. Subsequently, the extrinsic parameters of the camera are obtained by calculation.

2.1. Establishment of the Camera Coordinate System and World Coordinate System. The camera is in a 3D coordinate system, where the infrared camera is the origin, as shown in Figure 3. The infrared camera is the origin O_C of the camera coordinate system. The X_C axis is along the transverse direction of Kinect. The Z_C is perpendicular to Kinect and points in the shooting direction.

The world coordinates can generally be established at will. For facilitating the calculation of extrinsic parameters of the camera, as shown in Figure 4, the world coordinates should meet the following requirements.

- (1) The origin of the world coordinate system is the projection point of the origin of the camera coordinate system on the ground plane
- (2) The Y_W axis is the projection of the Z_C axis of the camera coordinate system on the ground
- (3) The Z_W axis is downwardly perpendicular to the ground plane
- (4) The coordinate system is a right-hand system

In this way, it is convenient to find the point group corresponding to camera coordinates and world coordinates and simplify the calibration process.

2.2. Calculation of Extrinsic Parameters of the Camera. The matrix of transformation from the camera coordinate system to the world coordinate system, that is, the extrinsic parameter T_C^W of the camera, is a 4×4 matrix with 12 unknown parameters, as shown in (1). First, the transformation matrix T_W^C from the world coordinate system to the camera coordinate system is calculated to obtain the extrinsic parameters of the camera:

$$T_C^W = T_W^C^{-1}. \quad (2)$$

A point in the camera coordinate system is $P_C^{(i)}$, and the corresponding point in the world coordinate system is $P_W^{(i)}$, as shown in Figure 5. Four special points are selected in the world coordinate system:

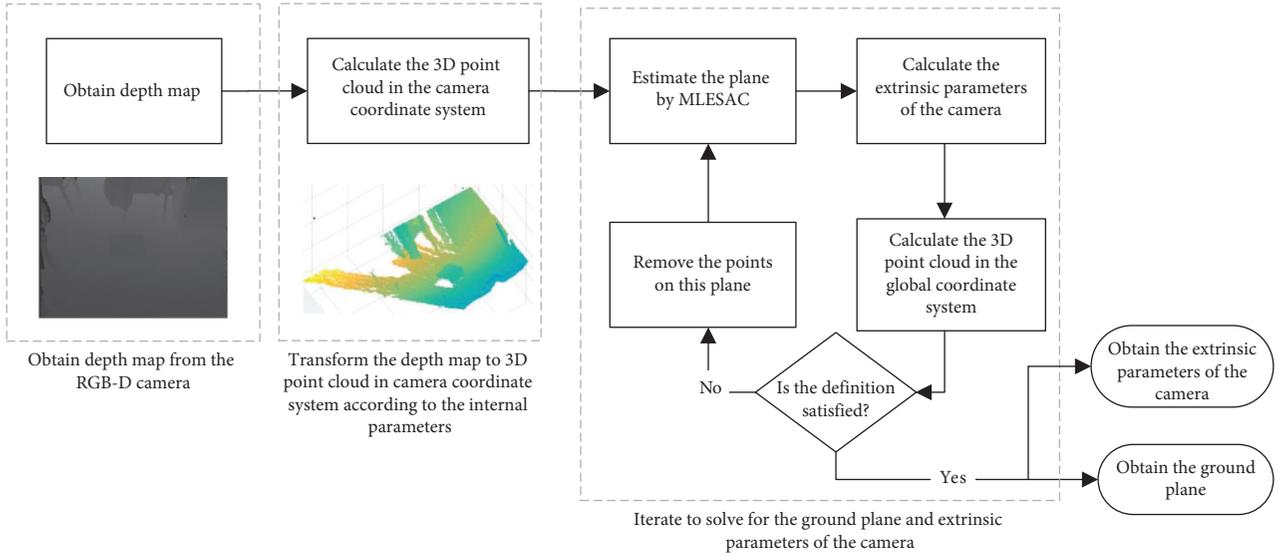


FIGURE 2: Process of extrinsic parameter calibration of the camera.



FIGURE 3: Camera coordinate.

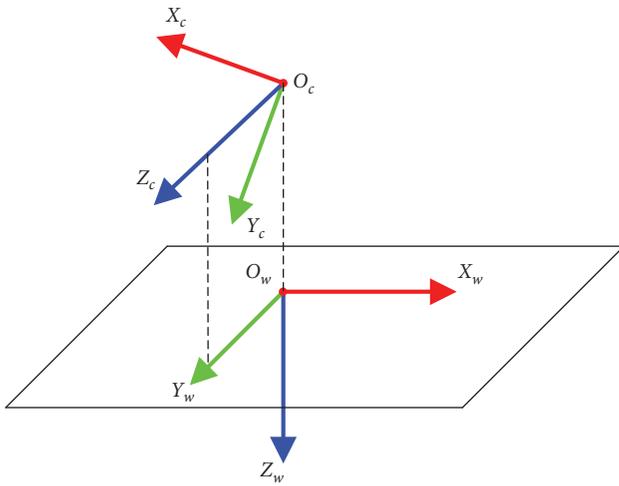


FIGURE 4: The position of world coordinate.

$$\begin{cases} P_w^{(0)} = [0 \ 0 \ 0], \\ P_w^{(1)} = [1 \ 0 \ 0], \\ P_w^{(2)} = [0 \ 1 \ 0], \\ P_w^{(3)} = [0 \ 0 \ 1]. \end{cases}$$

(3)

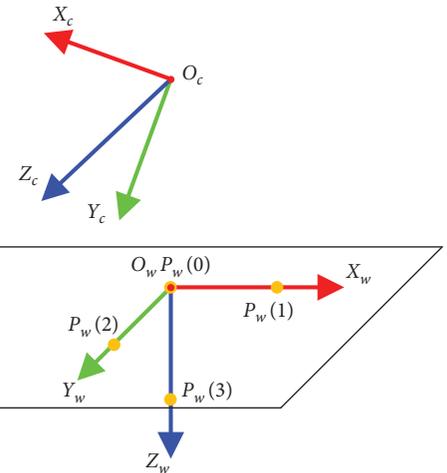


FIGURE 5: Selected points in world coordinate.

By calculating the corresponding points in the camera coordinate system, that is, $P_C^{(0)}$, $P_C^{(1)}$, $P_C^{(2)}$, and $P_C^{(3)}$, the value of ${}^C_W T$ can be obtained as

$${}^C_W T = \begin{bmatrix} \frac{(P_C^{(1)} - P_C^{(0)})^T}{\|P_C^{(1)} - P_C^{(0)}\|_2} & \frac{(P_C^{(2)} - P_C^{(0)})^T}{\|P_C^{(2)} - P_C^{(0)}\|_2} & \frac{(P_C^{(3)} - P_C^{(0)})^T}{\|P_C^{(3)} - P_C^{(0)}\|_2} & 0 \\ & \frac{P_C^{(0)}}{\|P_C^{(0)}\|_2} & & 1 \end{bmatrix}. \quad (4)$$

It is known that the plane in the camera coordinate system is $ax + by + cz + d = 0$. Its normal vector perpendicular to the ground is $\vec{n}_C = [a \ b \ c]$ and $\|\vec{n}_C\|_2 = 1$. As the origin of the world coordinate system is the projection of the origin of the camera coordinate system on the plane, the

coordinates corresponding to the origin $P_W^{(0)}$ of the world coordinate in the camera coordinate system are

$$P_C^{(0)} = \frac{|d|}{\sqrt{a^2 + b^2 + c^2}} \vec{n}_C. \quad (5)$$

If point $P_W^{(3)} = [0 \ 0 \ 1]$ is on the Z axis of the world coordinate system and $O_C, P_C^{(0)}$, and $P_C^{(3)}$ are collinear, then

$$P_C^{(3)} = P_C^{(0)} + \vec{n}_C. \quad (6)$$

Because the Y_W axis is the projection of the Z_C axis on the plane,

$$P_C^{(2)} = P_C^{(0)} + \vec{n}_y, \quad (7)$$

where \vec{n}_y is the unit vector of $\overrightarrow{P_C^{(0)}P_C^{(2)}}$,

$$\vec{n}_y = \frac{\overrightarrow{P_{Z_c} - P_C^{(0)}}}{\left\| \overrightarrow{P_{Z_c} - P_C^{(0)}} \right\|_2}. \quad (8)$$

P_{Z_c} is the projection of point $[0 \ 0 \ 1]$ in the camera coordinate system to the plane. By solving (9), the projection point can be obtained as

$$\begin{cases} ax + by + cz + d = 0, \\ \frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}, \end{cases} \quad (9)$$

where a, b, c , and d are the parameters of the plane equation and (x_0, y_0, z_0) is a random point on the plane.

After $P_C^{(0)}, P_C^{(2)}$, and $P_C^{(3)}$ are obtained, $P_C^{(1)}$ can be obtained by vector cross-multiplication as

$$P_C^{(1)} = \frac{\overrightarrow{P_C^{(2)} - P_C^{(0)}} \times \overrightarrow{P_C^{(3)} - P_C^{(0)}}}{\left\| \overrightarrow{P_C^{(2)} - P_C^{(0)}} \times \overrightarrow{P_C^{(3)} - P_C^{(0)}} \right\|_2}. \quad (10)$$

By solving (5)–(7) and (10), the coordinates of four points selected from the world coordinate system in the camera coordinate system are obtained. Then, the transformation matrix from the world coordinate point to the camera coordinate point is obtained by solving (4). Finally, the extrinsic parameter matrix of the infrared camera is obtained by solving (2).

3. Ground Plane Estimation

Toward facilitating plane detection, the depth map is first transformed to a 3D point cloud in the camera coordinate system using internal parameters. Then, the maximum likelihood estimation sample consensus (MLE-SAC) method [21] is used to extract the plane. The ergodic plane is iterated to obtain the extrinsic parameters. Next, the extrinsic parameters and point cloud are used to determine whether the plane is a ground plane, and we finally obtain the extrinsic parameters of the camera, as shown in Figure 6.

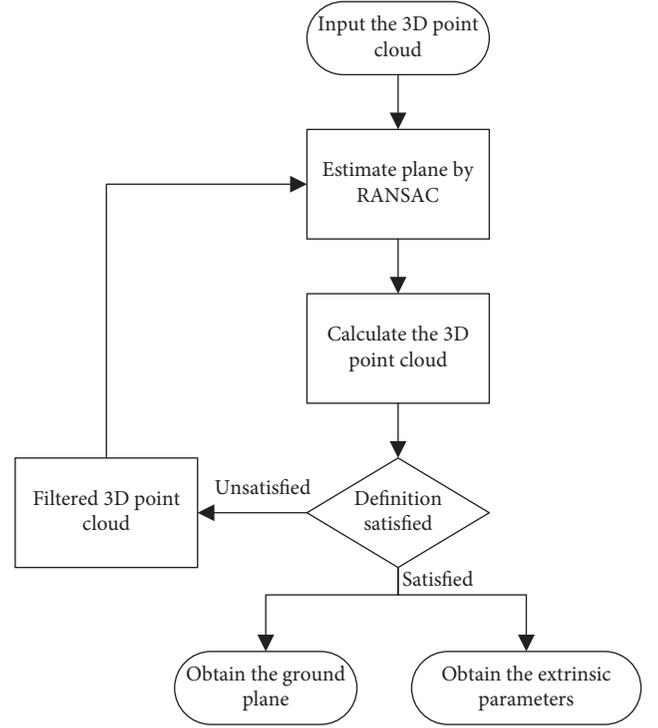


FIGURE 6: The process of ground plane estimation.

3.1. Transformation of Depth Map to 3D Point Cloud in the Camera Coordinate System. Consider the following:

$$\begin{cases} f_x = \frac{f}{d_x}, \\ f_y = \frac{f}{d_y}, \\ Z_C = d_p * s, \\ X_C = \frac{(u_I - c_x) \cdot Z_C}{f_x}, \\ Y_C = \frac{(v_I - c_y) \cdot Z_C}{f_y}. \end{cases} \quad (11)$$

Figure 7 shows the imaging model of the camera. The coordinate axes of the image plane are X_I and Y_I , respectively. The imaging point of point $P_C = (X_C, Y_C, Z_C)$ in the camera coordinate system on the image plane is $p_I = (u_I, v_I)$. The focal length of the camera is f . The coordinates of the intersection of the optical axis and the image plane are c_x and c_y . The length and width of the pixels are d_x and d_y , respectively. The value of the pixel is d_p . The proportion of Z coordinate values corresponding to the value of pixels in the camera coordinate system is s .

Then, (u, v, d_p) can be transformed to (X_C, Y_C, Z_C) , as shown in (11).

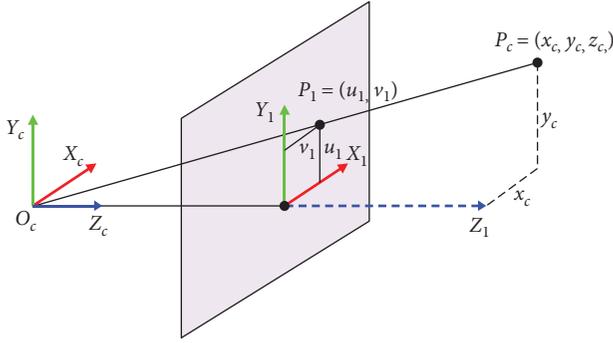


FIGURE 7: The model of the camera.

3.2. Ground Plane Estimation. Ground plane estimation is the basis of subsequent extrinsic parameter calculation of the camera. Its accuracy determines the accuracy of extrinsic camera parameters. In a scene, multiple planes will be fitted out. Then, whether the plane is a ground plane is determined according to the following conditions:

$$\begin{cases} \theta < \frac{\pi}{4}, \\ \text{median } P_{W_i}^Z > \min P_{C_o}^Z + \varepsilon, \end{cases} \quad (12)$$

where θ is the angle between the X_c axis of the camera and the plane; median is the operation of taking the median value; $P_{W_i}^Z$ is the set of interior point Z -values of the fitted plane model; $P_{C_o}^Z$ is the set of exterior point Z -values of the fitted plane model; and ε is the set tolerance value. Equation (12) represents two conditions: (1) the inclination angle of the camera relative to the plane does not exceed 45° and (2) after the exterior parameter is calculated according to the plane, the point cloud is transformed to the world coordinate system. The point set on the plane has the largest Z -value.

According to this definition, the planes were screened to calculate the qualified ground plane. The flowchart is shown in Figure 6.

First, the 3D point cloud in the camera coordinate system was input, and the plane was calculated using the MLESAC method. The set of interior points satisfying the plane was recorded. By using this plane and the method in Section 2.2, the extrinsic parameters of the camera were calculated. Combined with the internal parameters of the camera, the 3D point cloud in the camera coordinate system was transformed to the 3D point cloud in the world coordinate system, and it was judged whether the conditional equation (12) was met. If not, then the recorded set of interior points was removed from the point cloud, and fitting of the plane continued. Otherwise, the plane was the ground plane. No further operation was conducted, and the extrinsic parameters of the camera were output.

4. Experiment

4.1. Experiment Process. In this experiment, a PrimeSense camera was used to collect video data, and MATLAB was used for simulation to validate the proposed algorithm. To

facilitate the accuracy comparison, the chessboard calibration results of the color camera were used as the reference data. Because the matrix was not suitable for comparison, the extrinsic parameters were transformed to camera height (H), roll angle (θ), and pitch angle (ϕ). The calibration accuracy was measured by comparing the camera height, roll angle, and pitch angle. Following the world coordinates established in Section 2, the values can be obtained as

$$\begin{cases} H = Z_c, \\ \phi = \arccos(P_c^{(2)} \cdot P_w^{(2)}), \\ \theta = \arccos(P_c^{(0)} \cdot P_w^{(2)}), \end{cases} \quad (13)$$

where $P_c^{(2)}$ is the camera Z -axis in the world coordinate, $P_c^{(0)}$ is the camera X -axis in the world coordinate, and $P_w^{(2)}$ is the world Z -axis.

The experiment was carried out according to the following steps.

- (1) A PrimeSense camera was used to collect video data (each frame of the video had a clear chessboard), including color video and depth video.
- (2) N ($N > 20$) video frames were selected randomly from the color video as the input to the Zhang camera calibration method [19], and the internal camera parameters were estimated.
- (3) Each frame of the color video was traversed. The camera's extrinsic parameters were estimated using the internal parameters and the chessboard corner detected by the current frame, and the camera attitudes ($H_{\text{chessboard}}$, $\phi_{\text{chessboard}}$, and $\theta_{\text{chessboard}}$) of each frame of color video were obtained.
- (4) Each frame of depth video was traversed. The ground plane was detected using the proposed method, and the extrinsic parameters of the camera were estimated to obtain the camera attitudes (H_{plane} , ϕ_{plane} , and θ_{plane}) of each frame of depth video.

The camera height difference (ΔH), roll angle difference ($\Delta\theta$), and pitch angle difference ($\Delta\phi$) of each frame was calculated, using the following equations:

$$\begin{cases} \Delta H = H_{\text{chessboard}} - H_{\text{plane}}, \\ \Delta\phi = \phi_{\text{chessboard}} - \phi_{\text{plane}}, \\ \Delta\theta = \theta_{\text{chessboard}} - \theta_{\text{plane}}. \end{cases} \quad (14)$$

The video formats collected in Step 1 are shown in Table 1.

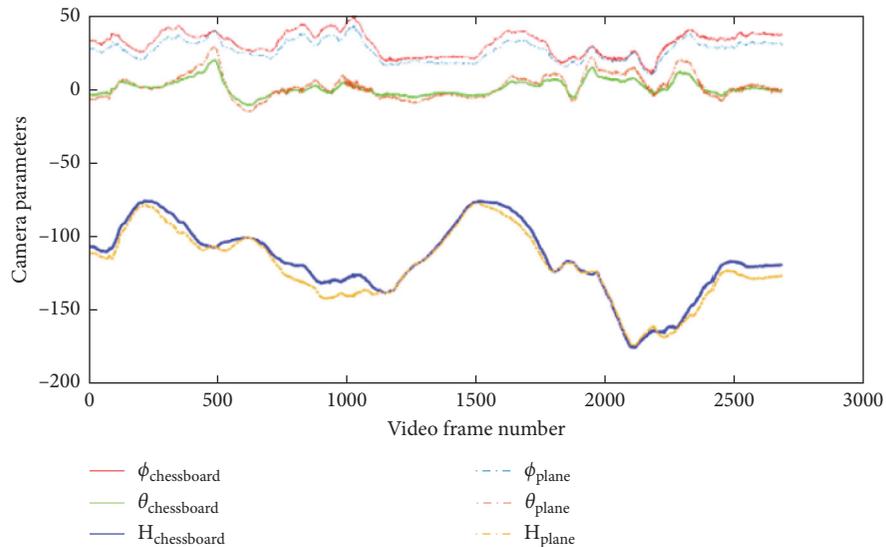
Each frame of the video contained a chessboard, as shown in Figure 8. In the experiment, the chessboard was used to estimate the camera's extrinsic parameters, which were taken as the reference data. Because different chessboards corresponded to different internal parameters when the camera's internal parameters were calculated using the Zhang calibration method [19], Steps 1 to 3 were executed 86 times. Each video frame could estimate 86 extrinsic parameters, with their medians as the reference data.

TABLE 1: Parameters of video.

	Color video	Depth video
Video size (pixel)	320 × 240	320 × 240
Frame rate (frame/s)	30	30
Duration (s)	89.4	89.4
Pixel byte (byte)	24	16
Video format	RGB24	Mono16



FIGURE 8: Color video frame comprising chessboard.

FIGURE 9: The result of the checkboard calibration method. The units of $\phi_{\text{chessboard}}$, $\theta_{\text{chessboard}}$, ϕ_{plane} , and θ_{plane} are degree, and the units of $H_{\text{chessboard}}$ and H_{plane} are centimeter.

The size of a chessboard was 7×9 , and both the length and width of each chessboard were 40 mm. The camera's extrinsic parameters, as obtained through chessboard calibration, are shown in Figure 8.

In Figure 9, $\theta_{\text{chessboard}}$, $\phi_{\text{chessboard}}$, and $H_{\text{chessboard}}$ are the medians of the camera pitch angle, roll angle, and height, respectively, as measured using the chessboard method and θ_{plane} , ϕ_{plane} , and H_{plane} are the medians of the camera pitch angle, roll angle, and height, respectively, as measured using the proposed method.

Figure 10 shows the experimental errors, and Table 2 shows the final measured results.

4.2. Error Analysis. Because there was no high-precision instrument to measure the extrinsic parameters of the camera in the experiment, these were measured using the chessboard calibration method and taken as the reference data to measure the accuracy of the proposed method. The factors affecting the accuracy of this experiment are as follows.

- (1) First is quantization error of corner detection. Because the video frame used was 240×320 pixels, the quantization error of corner detection was quite large when calculating the camera's internal and extrinsic parameters.

TABLE 2: Table of extrinsic parameters error of camera.

	Maximum	Minimum	Mean
θ	7.49	-0.75	4.47
ϕ	4.87	-9.62	-3.87
H	15.21	-1.15	3.96

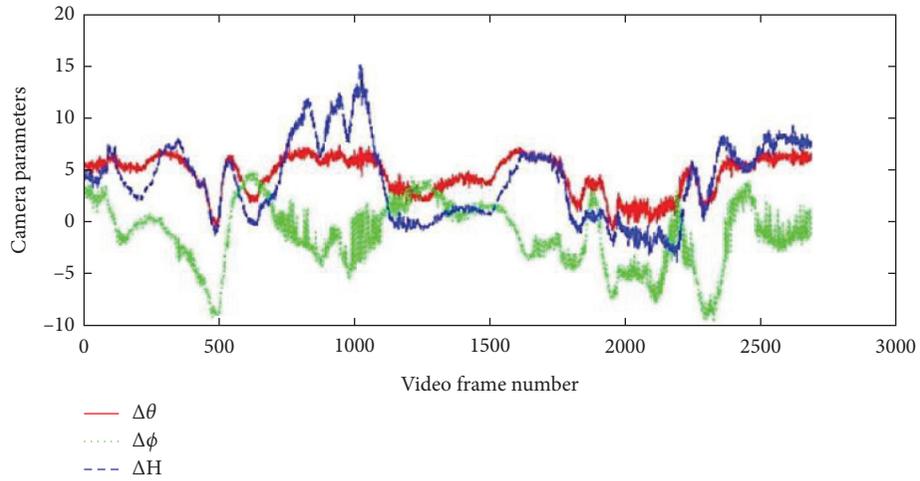


FIGURE 10: Test result. The camera height difference (ΔH), roll angle difference ($\Delta\theta$), and yaw angle difference ($\Delta\phi$) of each frame were calculated by (13). The units of $\Delta\theta$ and $\Delta\phi$ are degree, and the unit of ΔH is centimeter.

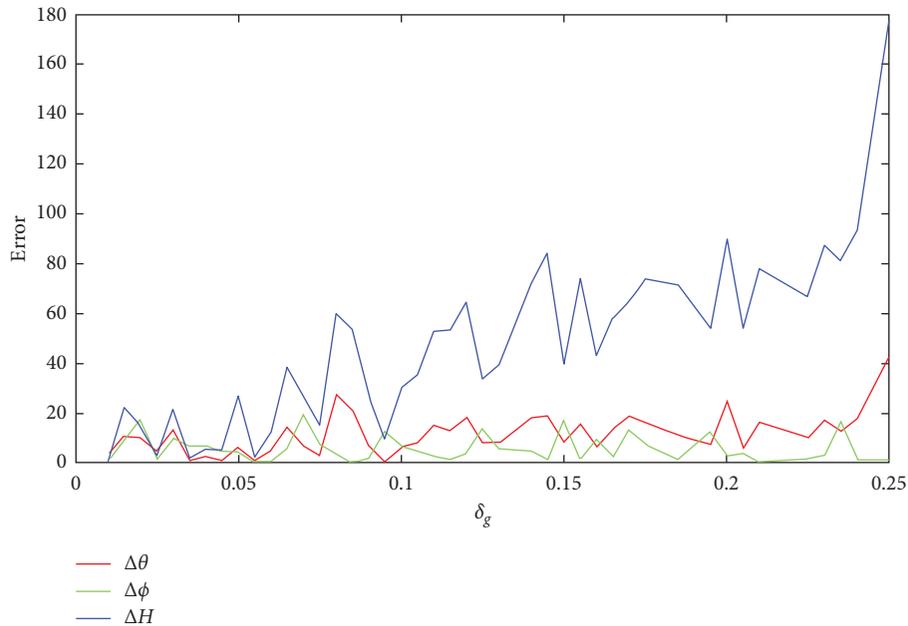


FIGURE 11: The effect of noise on the result. The error units of roll angle difference ($\Delta\theta$) and yaw angle difference ($\Delta\phi$) are degrees, and the error unit of camera height difference (ΔH) is centimeter.

- (2) The chessboard method and the proposed method were, respectively, used to measure the extrinsic parameters of the RGB-D camera's color sensor and infrared camera.
- (3) The noise of the RGB-D camera scanning scene data would affect the detection of the ground plane, thus affecting the accuracy of the camera's extrinsic parameters.

- (4) The parameters to stop MLESAC iteration in ground plane detection would also affect the results of this experiment.

4.3. Influence of Scene Noise on Ground Plane Detection. RGB-D cameras have many noise sources, such as temperature, incident angle and intensity of ambient light, and texture [22]. The MLESAC used in this study could deal with small amounts of noise, but not with scenes with too much noise or data loss.

- (1) Strong sunlight would cause too many ground plane noise points, resulting in inaccurate estimation of plane parameters and a slight influence on the camera height and roll angle.
- (2) If the reflectivity of the ground was too high (for example, a mirror was placed on the ground), the data of this area would be lost. If too much data were missing, then the ground plane could not be detected, and the extrinsic parameters of the camera could not be estimated.

To verify the influence of noise on camera attitude, Figure 11 shows the error change of camera attitude with the increase of Gaussian noise variance in the 929th frame. The mean value of Gaussian noise was 0, and the variance was δ_g . It can be seen that with the increase of noise variance, the height error increases, and the stability of pitch angle and roll angle decreases. In addition, when the variance is greater than 0.25, the plane cannot be correctly estimated.

5. Conclusion

During pedestrian detection based on RGB-D cameras, due to the impact of environmental vibration, an RGB-D camera will shift its original position, and the extrinsic parameters will change greatly, which will directly affect the detection accuracy. By using the automatic extrinsic parameter calibration method proposed in this study, the extrinsic parameters can be automatically corrected when there is no pedestrian, so as to solve the above problem.

The proposed method can be applied to the automatic adjustment of extrinsic parameters of 3D cameras (speckle, TOF, and binocular camera) and has high parallelism. After parallel implementation, it is instantaneous and can be used for the automatic calibration of extrinsic parameters of 3D cameras on mobile robots.

In this method, we extracted the plane from the 3D point cloud and used the position relationship between the plane and world coordinates to obtain a camera's extrinsic parameters, which were used to determine whether the current plane was a ground plane. If not, then the next plane would be used to calculate the extrinsic parameters until the ground plane was found, and the extrinsic parameters of the infrared camera were obtained. In this study, the conditions of the ground plane were given, which could guarantee the correctness of the established world coordinate system and creatively combine the plane detection with the extrinsic parameter estimation, so as to achieve the objective of

automatic extrinsic parameter calibration. This method does not require an additional calibration object, and it is aimed at the calibration of infrared cameras. The results are reliable.

The currently proposed method has the limitation that there needs to be a ground plane in the scene. If the ground plane cannot get detected, then the calibration cannot be carried out regularly. Two problems still must be solved: (1) to improve the accuracy of calibration and use targets such as pedestrians in scenes to carry out fine calibration and (2) to calibrate the extrinsic parameters of multiple cameras automatically according to the common area taken by multiple cameras. If no common area is taken by the two cameras, then a simple calibration object should be designed for automatic calibration to be carried out according to the geometric size of the calibration object.

The MATLAB resource code and the video data used to support the findings of this study are available from the corresponding author upon request.

Data Availability

Data are available on request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under Grant 61572083, in part by the Ministry of Education Joint Fund Project of China under Grant 6141A02022610, in part by the Fundamental Research Fund for the Central Universities of China under Grants 310824173601, 300102249304, and 300102248303, in part by the Fundamental Research Funds for the Central Universities Team Cultivation Project under Grant 300102248402, and in part by the Funds for Shaanxi Key R&D Program under Grant 2018ZDXMGY-047.

References

- [1] Microsoft. Kinect for X-Box 360[EB/OL]. 2017-01-01[2017-10-12], <https://support.xbox.com/en-US/browse/xbox-one/accessories/Kinect>.
- [2] Wikipedia. Kinect[EB/OL]. 2018-01-12 [2017-08-29], <https://en.wikipedia.org/wiki/Kinect>.
- [3] F. Barak, S. Alexander, M. Meir et al., "Depth mapping using projected patterns," US8493496B2[P], 2013.
- [4] R. A. El-laithy, J. Huang, and M. Yeh, "Study on the use of Microsoft Kinect for robotics applications," in *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, pp. 1280–1288, IEEE, Myrtle Beach, SC, USA, April 2012.
- [5] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, 2012.
- [6] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: a review," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1318–1334, 2013.
- [7] P. Fankhauser, M. Bloesch, D. Rodriguez et al., "Kinect v2 for mobile robot navigation: evaluation and modeling," in

- Proceedings of International Conference on Advanced Robotics (ICAR)*, IEEE, Istanbul, Turkey, July 2015.
- [8] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, “3-D mapping with an RGB-D camera,” *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 177–187, 2014.
- [9] M. Labbé and F. Michaud, “Online global loop closure detection for large-scale multi-session graph-based SLAM,” in *Proceedings of 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2661–2666, IEEE, Chicago, IL, USA, September 2014.
- [10] M. Munaro, F. Basso, and E. Menegatti, “OpenPTrack: open source multi-camera calibration and people tracking for RGB-D camera networks,” *Robotics and Autonomous Systems*, vol. 75, pp. 525–538, 2016.
- [11] M. Munaro and E. Menegatti, “Fast RGB-D people tracking for service robots,” *Autonomous Robots*, vol. 37, no. 3, pp. 227–242, 2014.
- [12] E. Lachat, H. Macher, T. Landes, and P. Grussenmeyer, “Assessment and calibration of a RGB-D camera (Kinect v2 sensor) towards a potential use for close-range 3D modeling,” *Remote Sensing*, vol. 7, no. 10, pp. 13070–13097, 2015.
- [13] C. Chen, B. Yang, S. Song et al., “Calibrate multiple consumer RGB-D cameras for low-cost and efficient 3D indoor mapping,” *Remote Sensing*, vol. 10, no. 2, 2018.
- [14] J. Chow, D. Lichti, J. Hol, G. Bellusci, and H. Luinge, “IMU and multiple RGB-D camera fusion for assisting indoor stop-and-go 3D terrestrial laser scanning,” *Robotics*, vol. 3, no. 3, pp. 247–280, 2014.
- [15] D. R. dos Santos, M. A. Basso, K. Khoshelham, E. de Oliveira, N. L. Pavan, and G. Vosselman, “Mapping indoor spaces by adaptive coarse-to-fine registration of RGB-D data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 2, pp. 262–266, 2016.
- [16] J. Li, T. Cai, K. Deng, X. Wang, T. Sellis, and F. Xia, “Community-diversified influence maximization in social networks,” *Information Systems*, vol. 92, Article ID 101522, 2020.
- [17] Q. H. Liao, M. Liu, L. Tai et al., “Extrinsic calibration of 3D range finder and camera without auxiliary object or human intervention,” 2017, <https://arxiv.org/abs/1703.04391>.
- [18] F. Basso, E. Menegatti, and A. Pretto, “Robust intrinsic and extrinsic calibration of RGB-D cameras,” 2017, <https://arxiv.org/abs/1701.05748>.
- [19] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [20] L. Shibo and Z. Qing, “A new approach to calibrate range image and color image from kinect,” in *Proceedings of International Conference on Intelligent Human-Machine Systems and Cybernetics*, pp. 252–255, Washington, DC, USA, August 2012.
- [21] P. H. S. Torr and A. Zisserman, “MLESAC: a new robust estimator with application to estimating image geometry,” *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.
- [22] A. Belhedi, A. Bartoli, S. Bourgeois, V. Gay-Bellile, K. Hamrouni, and P. Sayd, “Noise modelling in time-of-flight sensors with application to depth noise removal and uncertainty estimation in three-dimensional measurement,” *IET Computer Vision*, vol. 9, no. 6, pp. 967–977, 2015.